AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 9(9): 26195–26282.
DOI: 10.3934/math.20241280
Received: 07 June 2024
Revised: 09 August 2024
Accepted: 20 August 2024
Published: 11 September 2024

Research article

Limit theorems for nonparametric conditional U-statistics smoothed by asymmetric kernels

Salim Bouzebda*, **Amel Nezzal and Issam Elhattab**

Laboratory of Applied Mathematics of Compiègne (LMAC), Université de technologie de Compiègne, CS 60319 - 57 avenue de Landshut, Compiègne, France

* **Correspondence:** Email: salim.bouzebda@utc.fr.

Abstract: U-statistics represent a fundamental class of statistics used to model quantities derived from responses of multiple subjects. These statistics extend the concept of the empirical mean of a d-variate random variable X by considering sums over all distinct m-tuples of observations of X. Within this realm, W. Stute [134] introduced conditional U-statistics, a generalization of the Nadaraya-Watson estimators for regression functions, and demonstrated their strong point-wise consistency. This paper presented a first theoretical examination of the Dirichlet kernel estimator for conditional U-statistics on the dm-dimensional simplex. This estimator, being an extension of the univariate beta kernel estimator, effectively addressed boundary biases. Our analysis established its asymptotic normality and uniform strong consistency. Additionally, we introduced a beta kernel estimator specifically tailored for conditional U-statistics, demonstrating both weak and strong uniform convergence. Our investigation considered the expansion of compact sets and various sequences of smoothing parameters. For the first time, we examined conditional U-statistics based on mixed categorical and continuous regressors. We presented new findings on conditional U-statistics smoothed by multivariate Bernstein kernels, previously unexplored in the literature. These results are derived under sufficiently broad conditions on the underlying distributions. The main ingredients used in our proof were truncation methods and sharp exponential inequalities tailored to the U-statistics in connection with the empirical processes theory. Our theoretical advancements significantly contributed to the field of asymmetric kernel estimation, with potential applications in areas such as discrimination problems, ℓ-sample conditional U-statistics, and the Kendall rank correlation coefficient. Finally, we conducted some simulations to demonstrate the small sample performances of the estimators.

Keywords: asymmetric kernels; asymptotic normality; beta kernel; Bernstein kernels; boundary bias; conditional U-statistics; Dirichlet kernel; nonparametric regression estimation; rates of convergence; strong convergence
Mathematics Subject Classification: 60F05, 60G15, 60K15, 62F40

1. Introduction

Motivated by numerous applications, the theory of U-statistics (introduced in the seminal work by [80] and [71]) and U-processes has garnered considerable attention over the past decades. U-statistics are instrumental in solving complex statistical problems such as density estimation, nonparametric regression tests, and goodness-of-fit tests. Specifically, U-statistics are essential in analyzing estimators (including function estimators) with varying degrees of smoothness. For instance, Stute in [135] applied almost sure uniform bounds for $\mathbb{P}$-canonical U-processes to analyze the product limit estimator for truncated data. In [8], Arcones and Wang presented two novel tests for normality based on U-processes. Leveraging results from [68], Schick et al. in [126] introduced new normality tests using weighted L_1-distances between the standard normal density and local U-statistics based on standardized observations. In [87], Joly and Lugosi discussed estimating the mean of multivariate functions under possibly heavy-tailed distributions and introduced the median-of-means approach based on U-statistics. U-processes are crucial tools in various statistical applications, including testing for qualitative features of functions in nonparametric statistics [67], cross-validation for density estimation, and establishing the limiting distributions of M-estimators (see, e.g., [7]). Foundational asymptotic results for U-statistics under the assumption of independent and identically distributed (i.i.d.) random variables were provided by [71,80,146] and [59]. Under weak dependency assumptions, asymptotic results were demonstrated in [18,48]. For a comprehensive resource on U-statistics and U-processes, readers may refer to [7, 19, 92, 98, 100], and [46]. Recent advancements and references can be found in [30, 31]. U-statistics also appear naturally in other contexts, such as counting occurrences of certain subgraphs (e.g., triangles) in random graph theory [85]. In machine learning, U-statistics are utilized in a wide range of problems, including clustering, image recognition, ranking, and learning on graphs, where risk estimates often take the form of U-statistics. For instance, the ranking problem can be framed as pairwise classification, with the empirical ranking error being a U-statistic of order 2 [37]. For U-statistics with random kernels of diverging orders, see [60, 75, 131]. Infinite-order U-statistics are valuable for constructing simultaneous prediction intervals that quantify the uncertainty of ensemble methods such as sub-bagging and random forests [118]. The Mean nearest-neighbors approach for differential entropy estimation introduced by [58] is a specific application of U-statistics. Using U-statistics, [107] proposed a new test statistic for goodness-of-fit tests. In [45], Cybis et al. explored a model-free approach for clustering and classifying genetic data using U-statistics, offering alternative perspectives on these problems. In [103], Lim and Stojanovic proposed employing U-statistics for analyzing random compressed sensing matrices in the non-asymptotic regime. In [13], Bello et al. introduced a comprehensive framework for clustering within multiple groups using a U-statistics-based approach designed for high-dimensional datasets. In a related context in [91], Kim and Ramdas focused on dimension-agnostic inference, developing methods whose validity remains independent of assumptions regarding dimension versus sample size. Their approach utilized variational representations of the existing test statistics, incorporating sample splitting and self-normalization to yield a refined test statistic with a Gaussian limiting distribution. This involved modifying degenerate U-statistics by dropping diagonal blocks and retaining off-diagonal blocks. Further exploration by [39] involved U-statistics-based empirical risk minimization, while in [86], Janson examined asymmetric U-statistics based on a stationary sequence of m-dependent variables, with applications motivated by pattern matching in random strings and permutations. Additionally in [139], Sudheesh et al. developed innovative U-statistics considering left truncation and right censoring, proposing a straightforward nonparametric test for assessing independence between time to failure and cause of failure in competing risks under such censoring conditions. In [95], Le Minh investigated the quadruplet U-statistic with applications in network analysis statistical inference. In [102], Li et al. introduced a learning framework utilizing pairwise loss and minimizing empirical risk through U-processes and Rademacher complexity. In [66], Ghannadpour et al. demonstrated the high efficacy of using U-statistics to identify modification zones, accounting for the structural characteristics of the data and neighboring samples. In [44], Cintra et al. presented a model-free technique using U-statistics to construct control charts for effectively monitoring batch processes, considering multiple sources of variability. In [81], Huang et al. examined distributed inference for two-sample U-statistics in the context of large datasets, proposing blockwise linear two-sample U-statistics to reduce computational complexity. Finally in [122], Randles proposed a general method for obtaining asymptotic distribution theory for U-statistics with estimated parameters, extended recently by [50].

In this paper, we consider the conditional U-statistics introduced by [134]. These statistics generalize the Nadaraya-Watson estimators of a regression function by [112] and [150]. Specifically, let $\{(\mathbf{X}_i, \mathbf{Y}_i), i \in \mathbb{N}^*\}$ be a sequence of i.i.d. random elements with $\mathbf{X}_i \in \mathbb{R}^d$ and $\mathbf{Y}_i \in \mathbb{R}^q$, where $d, q \geq 1$. Let $\varphi : \mathbb{R}^{qm} \to \mathbb{R}$ be a measurable function. We focus on estimating the conditional expectation or regression function

$$r^{(m)}(\varphi, \mathbf{t}) = \mathbb{E}\left(\varphi(\mathbf{Y}_1, \ldots, \mathbf{Y}_m) \mid (\mathbf{X}_1, \ldots, \mathbf{X}_m) = \mathbf{t}\right), \quad \text{for } \mathbf{t} \in \mathbb{R}^{dm}, \tag{1.1}$$

whenever it exists, i.e., $\mathbb{E}(|\varphi(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)|) < \infty$. We introduce a kernel function $K : \mathbb{R}^d \to \mathbb{R}$ with support contained in $[-B, B]^d$, $B > 0$, satisfying

$$\sup_{\mathbf{x}\in\mathbb{R}^d} |K(\mathbf{x})| =: \kappa < \infty \quad \text{and} \quad \int K(\mathbf{x})\, d\mathbf{x} = 1. \tag{1.2}$$

In [134], Stute introduced a class of estimators for $r^{(m)}(\varphi, \mathbf{t})$, called conditional U-statistics, defined for each $\mathbf{t} \in \mathbb{R}^{dm}$ as

$$\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n) = \frac{\sum_{(i_1,\ldots,i_m)\in I(m,n)} \varphi(\mathbf{Y}_{i_1}, \ldots, \mathbf{Y}_{i_m}) K\left(\frac{\mathbf{t}_1-\mathbf{X}_{i_1}}{h_n}\right) \cdots K\left(\frac{\mathbf{t}_m-\mathbf{X}_{i_m}}{h_n}\right)}{\sum_{(i_1,\ldots,i_m)\in I(m,n)} K\left(\frac{\mathbf{x}_1-\mathbf{X}_{i_1}}{h_n}\right) \cdots K\left(\frac{\mathbf{x}_m-\mathbf{X}_{i_m}}{h_n}\right)}, \tag{1.3}$$

where

$$I(m, n) = \left\{\mathbf{i} = (i_1, \ldots, i_m) : 1 \leq i_j \leq n \text{ and } i_j \neq i_r \text{ if } j \neq r\right\}$$

is the set of all m-tuples of distinct integers between 1 and n, and $\{h_n\}_{n\geq 1}$ is a sequence of positive constants converging to zero at the rate $nh_n^{dm} \to \infty$. For $m = 1$, $r^{(m)}(\varphi, \mathbf{t})$ reduces to $r^{(1)}(\varphi, \mathbf{t}) = \mathbb{E}(\varphi(\mathbf{Y}) \mid \mathbf{X} = \mathbf{t})$, and Stute's estimator becomes the Nadaraya-Watson estimator of $r^{(1)}(\varphi, \mathbf{t})$ given by

$$\widehat{r}_n^{(1)}(\varphi, \mathbf{t}, h_n) = \frac{\sum_{i=1}^n \varphi(\mathbf{Y}_i) K\left(\frac{\mathbf{X}_i-\mathbf{t}}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{X}_i-\mathbf{t}}{h_n}\right)}.$$

In [128], Sen gave the rate of uniform convergence in $\mathbf{t}$ of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ to $r^{(m)}(\varphi, \mathbf{t})$. In [121], Rao and Sen discussed the limit distributions of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$ and compared them with those obtained by

Stute. In [74], Harel and Puri extended the results of [134] to weakly dependent data under appropriate mixing conditions and applied their findings to verify the Bayes risk consistency of the corresponding discrimination rules. In [138], Stute proposed symmetrized nearest-neighbors conditional U-statistics as alternatives to the usual kernel-type estimators. In [53], Dony and Mason established a much stronger form of consistency, namely, uniform in $\mathbf{t}$ and bandwidth consistency (i.e., $h_n \in [a_n, b_n]$ where $a_n < b_n \to 0$ at a specific rate) of $\widehat{r}_n^{(m)}(\varphi, \mathbf{t}; h_n)$. Additionally, they demonstrated uniform consistency over $\varphi \in \mathscr{F}$ for a suitably restricted class of functions $\mathscr{F}$, extended in [24, 25, 30, 31, 33, 35]. The main tool in their results was the use of the local conditional U-process investigated in [68]. [6], Arcones examined the distributional and almost sure point-wise Bahadur-Kiefer representation for U-quantiles, demonstrating that the order of this representation is contingent on the local variance of the empirical process of the U-statistic structure at the U-quantile. Additionally in [4], Arcones explored the order of the Kolmogorov-Smirnov distance for the bootstrap of U-quantiles. For further details on U-quantile estimation, readers are referred to the works of [47, 76, 145, 151, 156–158].

It is well-known that using standard symmetric kernels to estimate unknown curves on a bounded support, such as on a half-real line or a compact set, leads to boundary bias near the boundary (refer to [51, 55, 73, 110, 113, 119, 124, 140, 147, 152]). Despite various boundary correction methods proposed since the early works of [111] and [88] on boundary-adapted kernel estimation, smoothing with a nonstandard asymmetric kernel function has emerged as a viable alternative (see [62]). An asymmetric kernel, based on a probability density function (pdf) with the same support as the curves, inherently avoids boundary bias. Additionally, the kernel shape varies according to the position where smoothing is performed, allowing for adaptive smoothing. Over two decades have passed since the introduction of asymmetric kernels, and numerous studies have reported favorable evidence from their application to empirical models in economics and finance. For instance in [40], Chen considered kernel estimators using nonnegative kernels to estimate pdfs with compact supports. In [41], Chen examined regression estimation, comparing a beta smoother with a local linear smoother. In [20], Bouezmarni and Rolin studied the asymptotic properties of the beta kernel density estimator proposed by [40], deriving the exact asymptotic behavior of the expected L_1-loss and demonstrating uniform weak consistency for continuous underlying density functions. In [155], Zhang and Karunamuni showed that the performance of beta kernel estimator is very similar to that of the reflection estimator, which avoids the boundary problem only for densities with a shoulder at the support endpoints. In [16], Bertin and Klutchnikoff analyzed the beta kernel density estimator from an asymptotic minimax perspective, and this was later corrected and extended to the general setting of the Dirichlet kernel by [17]. When the underlying density has a fourth-order derivative, in [82], Igarashi improved the beta kernel estimator using bias correction techniques based on two beta kernel estimators with different smoothing parameters (see [77] for more details on asymmetric kernels and their applications). Various other statistical topics related to beta kernels are addressed in [38, 83, 94, 154]. Additionally, Bernstein polynomial density and distribution estimators have gained popularity and have been discussed in several works. In [144], Vitale first proposed the Bernstein estimator for estimating a density function for an i.i.d. sample $\{X_n\}_{n\geq 1}$. The asymptotic normality of the resulting density estimator, such as the uniform weak law of large numbers, the central limit theorem, and uniform strong consistency are studied in [133] and [64]. In [141], Tenbusch investigated the Bernstein estimation in the multivariate context, obtaining the bias, variance, and Mean Squared Error (MSE) for the Bernstein estimator, focusing on the two-dimensional unit simplex and the unit square $[0, 1]^2$. Several extensions have been provided by [2, 9, 12, 15, 89, 120, 142, 149]. In [96], Leblanc demonstrated that Bernstein estimators of distribution functions have excellent boundary properties, including the absence of boundary bias (see also [97]). The latest results were generalized in [116].

However, little is known about the uniform consistency and convergence rates of asymmetric nonlinear kernel estimators. This paper aims to fill this gap by considering a general framework for conditional U-statistics. This study examines the uniform convergence and the asymptotic normality of nonparametric estimators on a compact set smoothed using an asymmetric kernel. The compactness of the support frequently emerges due to the data's nature or as a theoretical concept in econometrics. Economic and financial variables described as shares or proportions are typically constrained within a specific range, with an upper bound for the former and a lower bound for the latter. Examples include the allocation of funds, the distribution of budget resources, the percentage of unemployed individuals, desired exchange rate ranges, and failure-to-repay and recovery rates. Compactness is required for certain aspects, such as the support of nonparametric copulas (see, e.g., [125]), the nonparametric part of partial linear regressions (see, e.g., [153]), and the covariates used for nearest-neighbor matching (see, e.g., [1]). Additionally, the fully nonparametric estimator for first-price auctions relies on the compactness of the supports of the distributions of private values and observed bids (see [70] for more information). In [3], Aitchison first defined the Dirichlet kernel estimator for density estimation and compared their performance empirically with an alternative approach, the logistic-normal kernel method, where the data on the simplex is first transformed to $\mathbb{R}^d$ via an additive log-ratio transformation, followed by multivariate Gaussian kernel smoothing. In [36], Brown and Chen first studied beta kernel ($d = 1$) theoretically in the context of smoothing for regression curves with equally spaced and fixed design points. The asymptotics of the point-wise bias, the integrated variance, and the mean integrated squared error (MISE) for the estimator of the regression function were found, with the optimal MISE shown to be $O(n^{-4/5})$. These results extended to the beta kernel estimator some findings from [133], who worked with the closely related Bernstein estimator. In [40], Chen first studied the beta kernel estimator theoretically in the context of density estimation. In [41], Chen generalized the results of [36] to arbitrary collections of fixed design points using a [63] type estimator. In [20], Bouezmarni and Rolin computed the asymptotics of the expected average absolute error for the beta kernel estimator, extending analogous results for traditional kernel estimators found in Theorem 2 of [52]. In [123], Renault and Scaillet first used a beta kernel to estimate recovery rate densities of defaulted bonds. In [21], Bouezmarni and Rombouts generalized the results of [40, 42] to the multidimensional setting, considering the products of one-dimensional asymmetric kernels. In [155], Zhang and Karunamuni showed that the performance of the beta kernel estimator is similar to that of the reflection estimator of [127], which avoids the boundary problem only for densities exhibiting a shoulder condition at the support endpoints. It should be noted that Bernstein density estimators, studied theoretically by authors such as [11, 61, 65, 69, 78, 84, 105, 108, 109, 148], share many of the same asymptotic properties with proper reparameterization. For an overview of the extensive literature on Bernstein estimators, see [115].

This paper presents a first exploration into establishing a rigorous theoretical foundation for asymmetric kernel-type estimators applied to conditional U-statistics, including Dirichlet kernels, multivariate beta kernels, and Bernstein polynomials. However, as we elucidate later, this challenge goes beyond merely combining ideas from existing literature. It necessitates intricate mathematical derivations to address the inherent nonlinearity characterizing conditional U-statistics. To the best of our knowledge, this broader contextualization remains unexplored within the scholarly domain, presenting a more formidable task compared to the Nadaraya-Watson estimator or the density estimator. Our analysis leverages Hoeffding's decomposition, a fundamental tool in the analysis of U-statistics. Specifically, we offer novel insights into the Dirichlet kernel regression estimator within the framework of Hoeffding's decomposition. We encounter a challenge concerning the applicability of the martingale structure to the nonlinear terms in Hoeffding's decomposition within our context. To address this, we employ exponential inequality principles drawn from the intricate theory of U-processes. By exploring the conditional U-statistics smoothed by Dirichlet kernels and Bernstein polynomial, we present several interesting new results regarding the Nadaraya-Watson estimators that are of independent interest. Furthermore, we broaden the scope of our investigation to beta kernel estimation by exploring the realm of mixed-type vectors.

The remainder of the paper is structured as follows. In Section 3, we introduce the conditional U-statistics estimator using Dirichlet kernel for the first time in the literature. We commence by establishing uniform convergence for the regression estimator in Section 3.1, specifically in Theorem 3.1, which holds independent significance. Extensions of this result are then presented in Section 3.2, as shown in Corollary 3.5. The asymptotic normality of the proposed estimators is explored in Section 3.3, wherein Theorem 3.6 and Corollary 3.7 are discussed. In Section 4, we consider the conditional U-statistics using Bernstein polynomials. Section 4.1 is devoted to the Nadaraya-Watason estimator (see Theorem 4.2 and Corollary 4.3) while Section 4.2 investigates the conditional U-statistics presenting the main results in Theorem 4.5 and Corollary 4.7. Moving forward, Section 5 unveils the estimator of conditional U-statistics through beta kernel smoothing. Section 5.1 outlines our findings regarding weak uniform convergence, detailed in Corollary 5.4. In Section 5.3, we establish the strong uniform convergence of the beta conditional U-statistics estimator, as presented in Corollary 5.8. Furthermore, in Section 5.4, we introduce the conditional U-statistics estimators for mixed categorical and continuous data for the first time. The main result regarding weak uniform convergence is stated in Corollary 5.9. Section 9 provides numerical experiments that illustrate the favorable small-sample properties of the proposed method. Finally, we provide concluding remarks and discuss potential future developments in Section 10. All proofs are consolidated in Section A to maintain the coherence of the presentation, while a selection of pertinent technical results is provided in the Appendix.

2. Preliminaries and estimation procedure

Let us consider a sequence of i.i.d. random vectors $\{(\mathbf{X}_i, \mathbf{Y}_i), i \in \mathbb{N}^*\}$, where $\mathbf{X}_i \in \mathcal{X} \subseteq [0,1]^d$, and the $\mathbf{Y}_i \in \mathcal{Y} := \mathbb{R}^q$. Let $\varphi : \mathcal{Y}^m \to \mathbb{R}$ be a measurable function. In this paper, we are primarily concerned with the estimation of the conditional expectation, or regression function of $\varphi(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)$ evaluated at $(\mathbf{X}_1, \ldots, \mathbf{X}_m) = \tilde{\mathbf{x}}$, given by

$$r^{(m)}(\varphi, \tilde{\mathbf{x}}) = \mathbb{E}\left(\varphi(\mathbf{Y}_1, \ldots, \mathbf{Y}_m) \mid (\mathbf{X}_1, \ldots, \mathbf{X}_m) = \tilde{\mathbf{x}}\right), \quad \text{for } \tilde{\mathbf{x}} \in \mathcal{X}^m, \tag{2.1}$$

whenever it exists, i.e., $\mathbb{E}\left(|\varphi(\mathbf{Y}_1,\ldots,\mathbf{Y}_m)|\right)<\infty$. [134] presented a class of estimators for $r^{(m)}(\varphi,\tilde{\mathbf{x}})$, called the conditional U-statistics, which is defined for each $\tilde{\mathbf{x}}\in\mathcal{X}^m$ and $\ell\in\{1,2,3\}$ to be:

$$\widehat{r}_{n,\ell}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}}))=\frac{\sum_{(i_1,\ldots,i_m)\in I(m,n)}\varphi(\mathbf{Y}_{i_1},\ldots,\mathbf{Y}_{i_m})K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_1)}(\mathbf{X}_{i_1})\ldots K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_m)}(\mathbf{X}_{i_m})}{\sum_{(i_1,\ldots,i_m)\in I(m,n)}K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_1)}(\mathbf{X}_{i_1})\ldots K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_m)}(\mathbf{X}_{i_m})},\tag{2.2}$$

where $\bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}})=(\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_1),\ldots,\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_m))$ will be specified later in the sections below. In the particular case $m=1$, the $r^{(m)}(\varphi,\tilde{\mathbf{x}})$ is reduced to $r^{(1)}(\varphi,\tilde{\mathbf{x}})=\mathbb{E}(\varphi(\mathbf{Y})|\mathbf{X}=\mathbf{x})$ and Stute's estimator becomes the Nadaraya-Watson estimator of $r^{(1)}(\varphi,\tilde{\mathbf{x}})$ is given by

$$\widehat{r}_{n,\ell}^{(1)}(\varphi,\mathbf{x}):=\frac{\sum_{i=1}^{n}\varphi(\mathbf{Y}_i)K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x})}(\mathbf{X}_i)}{\sum_{i=1}^{n}K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x})}(\mathbf{X}_i)}.\tag{2.3}$$

Throughout this paper, any multivariate point will be written in bold. To avoid confusion, we note $\mathbf{x}=(x_1,\ldots,x_d)$ for $\mathbf{x}\in\mathcal{X}$ and we denote $\tilde{\mathbf{x}}:=(\mathbf{x}_1,\ldots,\mathbf{x}_m)$ an m-tuple of multivariate points $\mathbf{x}_i\in\mathcal{X}$, $1\leq i\leq m$. Accordingly, we denote $\mathbf{1}=(1,\ldots,1)$ is a d-dimensional vector whose components are all equal to 1, and $\tilde{\mathbf{1}}=(\mathbf{1},\ldots,\mathbf{1})$ an m-tuple of points $\mathbf{1}$. From now on, we shall use the following notation:

$$\tilde{\mathbf{X}}:=(\mathbf{X}_1,\ldots,\mathbf{X}_m)\in\mathcal{X}^m,\ \text{ and }\ \tilde{\mathbf{X}}_{\mathbf{i}}:=(\mathbf{X}_{i_1},\ldots,\mathbf{X}_{i_m})\in\mathcal{X}^m,\quad \mathbf{i}\in I(m,n),$$

$$\tilde{\mathbf{Y}}:=(\mathbf{Y}_1,\ldots,\mathbf{Y}_m)\in\mathbb{R}^{qm},\ \text{ and }\ \tilde{\mathbf{Y}}_{\mathbf{i}}:=(\mathbf{Y}_{i_1},\ldots,\mathbf{Y}_{i_m})\in\mathbb{R}^{qm},\quad \mathbf{i}\in I(m,n).$$

We also define for all $\tilde{\mathbf{x}}=(\mathbf{x}_1,\ldots,\mathbf{x}_m)\in\mathcal{X}^m$ and $\ell\in\{1,2,3\}$

$$\mathcal{G}_{\varphi,\tilde{\mathbf{x}},\ell}(\tilde{\mathbf{t}},\tilde{\mathbf{y}})=\varphi(\tilde{\mathbf{y}})\tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}})}(\tilde{\mathbf{t}}),\quad (\tilde{\mathbf{t}},\tilde{\mathbf{y}})\in\mathcal{X}_d^m\times\mathbb{R}^{qm},$$

where $\tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}})}(\tilde{\mathbf{t}}):=\prod_{i=1}^{m}K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_i)}(\mathbf{t}_i)$. For $\ell\in\{1,2,3\}$, we now define

$$u_{n,\ell}(\varphi,\tilde{\mathbf{x}}):=u_{n,\ell}^{(m)}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}},\ell})=\frac{(n-m)!}{n!}\sum_{\mathbf{i}\in I(m,n)}\mathcal{G}_{\varphi,\tilde{\mathbf{x}},\ell}(\tilde{\mathbf{X}}_{\mathbf{i}},\tilde{\mathbf{Y}}_{\mathbf{i}}).$$

We can see that

$$\widehat{r}_{n,\ell}^{(m)}(\varphi,\tilde{\mathbf{x}},\bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}}))=\frac{u_{n,\ell}(\varphi,\tilde{\mathbf{x}})}{u_{n,\ell}(1,\tilde{\mathbf{x}})}.\tag{2.4}$$

In establishing the uniform consistency of $\widehat{r}_{n,\ell}^{(m)}(\varphi,\tilde{\mathbf{x}},\bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}}))$ with respect to $r^{(m)}(\varphi,\tilde{\mathbf{x}})$, an alternative and more suitable centering factor will be considered instead of the expectation $\mathbb{E}\left[\widehat{r}_{n,\ell}^{(m)}(\varphi,\tilde{\mathbf{x}},\bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}}))\right]$, which may either be non-existing or computationally challenging to determine. This alternative centering is defined as

$$\widehat{\mathbb{E}}\left[\widehat{r}_{n,\ell}^{(m)}(\varphi,\tilde{\mathbf{x}},\bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}}))\right]=\frac{\mathbb{E}\left[u_{n,\ell}(\varphi,\tilde{\mathbf{x}})\right]}{\mathbb{E}\left[u_{n,\ell}(1,\tilde{\mathbf{x}})\right]}.\tag{2.5}$$

The notation and facts presented below should be included in the continuation of this discussion. For a kernel L of $m \geq 1$ variables we define

$$U_{n,\ell}^{(m)}(L) = \frac{(n-m)!}{n!} \sum_{i \in I(m,n)} L\left(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_m}\right).$$

Suppose that L is a function of $k \geq 1$ variables, symmetric in its entries. Then, the Hoeffding projections (see [80] and [46]) with respect to $\mathbb{P}$, for $1 \leq m \leq k$, are defined as

$$\pi_{m,k} L\left(\mathbf{x}_1, \ldots, \mathbf{x}_m\right) = \left(\Delta_{\mathbf{x}_1} - \mathbb{P}\right) \times \cdots \times \left(\Delta_{\mathbf{x}_m} - \mathbb{P}\right) \times \mathbb{P}^{k-m}(L),$$

and

$$\pi_{0,k} L = \mathbb{E} L\left(\mathbf{X}_1, \ldots, \mathbf{X}_k\right).$$

For some measures $\mathbb{Q}_i$ on S, we denote

$$\mathbb{Q}_1 \cdots \mathbb{Q}_m L = \int_{S^m} L(\mathbf{x}_1, \ldots, \mathbf{x}_m) \mathrm{d}\mathbb{Q}_1(\mathbf{x}_1) \cdots \mathrm{d}\mathbb{Q}_m(\mathbf{x}_m),$$

and $\Delta_{\mathbf{x}}$ denotes the Dirac measure at point $\mathbf{x} \in \mathcal{X}$. Then, the decomposition of [80] gives

$$U_{n,\ell}^{(m)}(L) - \mathbb{E} L = \sum_{k=1}^{m} \binom{k}{m} U_{n,\ell}^{(k)}\left(\pi_{m,k} L\right),$$

which is easy to check. For $L \in L_2\left(\mathbb{P}^k\right)$ this denotes an orthogonal decomposition and $\mathbb{E}(\pi_m L \mid \mathbf{X}_2, \ldots, \mathbf{X}_m) = 0$ for $m \geq 1$; then, the kernels $\pi_{m,k} L$ are canonical for $\mathbb{P}$. Also, $\pi_{m,k}, m \geq 1$, are nested projections, that is, $\pi_{m,k} \circ \pi_{m',k} = \pi_{m,k}$ if $m \leq m'$, and

$$\mathbb{E}\left(\pi_{m,k} L\right)^2 \leq \mathbb{E}(L - \mathbb{E} L)^2 \leq \mathbb{E} L^2.$$

For example,

$$\pi_{1,k} h(x) = \mathbb{E}(h(\mathbf{X}_1, \ldots, \mathbf{X}_k) \mid X_1 = x) - \mathbb{E} h(\mathbf{X}_1, \ldots, \mathbf{X}_k).$$

Remark 2.1. *The functions $\mathcal{G}_{\varphi,\tilde{\mathbf{x}},\ell}$ are not necessarily symmetric; when we need to symmetrize them, we have*

$$\overline{\mathcal{G}}_{\varphi,\tilde{\mathbf{x}},\ell}(\tilde{\mathbf{t}}, \tilde{\mathbf{y}}) := \frac{1}{m!} \sum_{\sigma \in I_m^m} \mathcal{G}_{\varphi,\tilde{\mathbf{x}},\ell}(\tilde{\mathbf{t}}_\sigma, \tilde{\mathbf{y}}_\sigma) = \frac{1}{m!} \sum_{\sigma \in I_m^m} \varphi(\tilde{\mathbf{y}}_\sigma) \tilde{\mathcal{K}}_{\tilde{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}})}(\tilde{\mathbf{t}}_\sigma),$$

where $\tilde{\mathbf{t}}_\sigma = (\mathbf{t}_{\sigma_1}, \ldots, \mathbf{t}_{\sigma_m})$ and $\tilde{\mathbf{y}}_\sigma = (\mathbf{y}_{\sigma_1}, \ldots, \mathbf{y}_{\sigma_m})$. After symmetrization, the expectation

$$\mathbb{E}\left[\overline{\mathcal{G}}_{\varphi,\tilde{\mathbf{x}},\ell}(\tilde{\mathbf{t}}, \tilde{\mathbf{y}})\right] = \mathbb{E}\left[\mathcal{G}_{\varphi,\tilde{\mathbf{x}},\ell}(\tilde{\mathbf{t}}, \tilde{\mathbf{y}})\right]$$

and the U-statistic

$$u_{n,\ell}^{(m)}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}},\ell}) = u_{n,\ell}^{(m)}(\overline{\mathcal{G}}_{\varphi,\tilde{\mathbf{x}},\ell}) := u_{n,\ell}(\varphi, \tilde{\mathbf{x}})$$

do not change.

Before presenting the conditions and primary results, let's delve into the following notation. For $a > 0$,

$$\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t)\mathrm{dt}$$

is the gamma function; for $\mathbf{z} \in \mathbb{R}^d, \nabla\{h(\mathbf{z})\}$ signifies a d-column vector of the first-order partial derivatives of a function $h(\mathbf{z})$;

$$\tilde{f}(\tilde{\mathbf{x}}) := f(\mathbf{x}_1) \times \cdots \times f(\mathbf{x}_m), \quad \tilde{\mathbf{x}} \in \mathcal{X}^m,$$

where $f_{\mathbf{X}}(\cdot)$ is the marginal pdf of $\mathbf{X} \in \mathcal{X}$ and we denote

$$\mathcal{R}(\varphi, \tilde{\mathbf{x}}) := \tilde{f}(\tilde{\mathbf{x}}) r^{(m)}(\varphi, \tilde{\mathbf{x}}).$$

The expression "$X \overset{\mathcal{D}}{=} Y$" denotes that the random variable X has the same distribution as Y, while "a.s." stands for "almost surely". Moreover, $\|\mathbf{A}\|$ represents the Frobenius norm of the matrix $\mathbf{A}$, defined as $\|\mathbf{A}\| = \{\mathrm{tr}\,(\mathbf{A}^\top\mathbf{A})\}^{1/2}$.

2.1. Conditions and comments

(C.1) The function $\mathcal{R}(\varphi, \cdot)$ is Lipschitz continuous with respect to $\tilde{\mathbf{x}} \in \mathcal{X}^m$;

(C.2) The second-order partial derivatives of $\tilde{f}(\tilde{\mathbf{x}})$ and $\mathcal{R}(\varphi, \tilde{\mathbf{x}})$ are continuous on $\tilde{\mathbf{x}} \in (0, 1)^{dm}$;

(C.3) There are some constants $\gamma > 0$ and $C_1 \in [1, \infty)$ such that $\mathbb{E}|\varphi(\tilde{\mathbf{Y}})|^{2+\gamma} < \infty$ and

$$\sup_{\tilde{\mathbf{x}} \in (0,1)^{dm}} \mathbb{E}\left(|\varphi(\tilde{\mathbf{Y}})|^{2+\gamma} \mid \tilde{\mathbf{X}} = \tilde{\mathbf{x}}\right) \tilde{f}(\tilde{\mathbf{x}}) \leq C_1. \tag{2.6}$$

Comments:

Notice that **(C.2)** implies that there is some constant $C_0 \in [1, \infty)$ such that

$$\sup_{\tilde{\mathbf{x}} \in (0,1)^{dm}} \tilde{f}(\tilde{\mathbf{x}}) \leq C_0. \tag{2.7}$$

The uniform boundedness condition (2.6) in **(C.3)** implies that $\mathbb{E}\left(|\varphi(\tilde{\mathbf{Y}})|^{2+\gamma} \mid \tilde{\mathbf{X}} = \tilde{\mathbf{x}}\right)$ is allowed to diverge at boundaries but no faster than $\{f(\mathbf{x}_1) \cdots f(\mathbf{x}_m)\}^{-1}$. A similar condition can be found, for instance, in [72, Assumption 2] and [94, Assumption A3]. Notice that **(C.3)** is important to the truncation method. Note that **(C.3)** may be replaced by more general hypotheses upon moments of $\tilde{\mathbf{Y}}$ as in [14], that is

(C.3)″ We denote by $\{\mathcal{M}(x) : x \geq 0\}$ a nonnegative continuous function, increasing on $[0, \infty)$ such that, for some $s > 2$, ultimately as $x \uparrow \infty$,

$$(i)\ x^{-s}\mathcal{M}(x) \downarrow; (ii)\ \ x^{-1}\mathcal{M}(x) \uparrow. \tag{2.8}$$

For each $t \geq \mathcal{M}(0)$, we define $\mathcal{M}^{inv}(t) \geq 0$ by $\mathcal{M}(\mathcal{M}^{inv}(t)) = t$. We assume further that:

$$\mathbb{E}\left(\mathcal{M}\left(\left|\varphi(\tilde{\mathbf{Y}})\right|\right)\right) \tilde{f}(\tilde{\mathbf{x}}) < \infty.$$

The following choices of $\mathcal{M}(\cdot)$ are of particular interest:

(i) $\mathcal{M}(x) = x^p$ for some $p > 2$ or
(ii) $\mathcal{M}(x) = \exp(sx)$ for some $s > 0$.

3. Conditional U-statistics using Dirichlet kernels

In this section, we consider $\mathcal{X} = \mathbb{S}_{d,1}$, a d-dimensional simplex, defined by

$$\mathbb{S}_{d,1} := \left\{ \mathbf{x} \in [0,1]^d : \|\mathbf{x}\|_1 \leq 1 \right\},$$

and its interior

$$\operatorname{Int}(\mathbb{S}_{d,1}) := \left\{ \mathbf{x} \in (0,1)^d : \|\mathbf{x}\|_1 < 1 \right\},$$

where $\|\mathbf{x}\|_1 := \sum_{i=1}^{d} |x_i|$ and $d \in \mathbb{N}$. For $\alpha_1, \ldots, \alpha_d, \beta > 0$, the density of the Dirichlet($\boldsymbol{\alpha}$,β) distribution is

$$K_{\boldsymbol{\alpha},\beta}(\mathbf{x}) := \frac{\Gamma\left(\|\boldsymbol{\alpha}\|_1 + \beta\right)}{\Gamma(\beta) \prod_{i=1}^{d} \Gamma\left(\alpha_i\right)} \left(1 - \|\mathbf{x}\|_1\right)^{\beta-1} \prod_{i=1}^{d} x_i^{\alpha_i - 1}, \quad \mathbf{x} \in \mathbb{S}_{d,1}.$$

We refer to [93, Chapter 49] and [114]. In [3], Aitchison and Lauder introduced a significant aspect wherein the kernel $K_{\boldsymbol{\alpha},\beta}(\cdot)$ form alters with the position $\mathbf{x}$ within the simplex. This adaptation mitigates the boundary bias issue prevalent in conventional estimators, where the kernel remains constant across all points. Throughout this section, for each $j = 1, \ldots, m$, we set

$$\boldsymbol{\Lambda}_{n,1}(\mathbf{x}_j) = (\boldsymbol{\alpha}_j, \beta_j) := \left(\frac{\mathbf{x}_j}{\breve{b}} + \mathbf{1}, \frac{1 - \|\mathbf{x}_j\|_1}{\breve{b}} + 1 \right) \text{ for } \mathbf{x}_j \in \mathbb{S}_{d,1}, \breve{b} > 0. \tag{3.1}$$

The smoothing parameter $\breve{b}$, denoted as $\breve{b}(n)$, inherently depends on the sample size n. Now, we can introduce a new conditional U-statistic regression estimator using the Dirichlet kernel by replacing (3.1) in (2.2), that is,

$$\widehat{r}_{n,1}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\boldsymbol{\Lambda}}_{n,1}(\tilde{\mathbf{x}})) = \frac{\sum_{(i_1,\ldots,i_m) \in I(m,n)} \varphi(Y_{i_1}, \ldots, Y_{i_m}) \prod_{j=1}^{m} K_{(\boldsymbol{\alpha}_j,\beta_j)}\left(\mathbf{X}_{i_j}\right)}{\sum_{(i_1,\ldots,i_m) \in I(m,n)} \prod_{j=1}^{m} K_{(\boldsymbol{\alpha}_j,\beta_j)}\left(\mathbf{X}_{i_j}\right)}. \tag{3.2}$$

Below, we will present key findings regarding the regression function in the scenario where $m = 1$. These findings are essential for examining the estimator (3.2).

3.1. Nonparametric regression estimation

We consider the following quantities:

$$\widehat{g}_n(\varphi, \mathbf{x}, \boldsymbol{\Lambda}_{n,1}) := \frac{1}{n} \sum_{i=1}^{n} \varphi(\mathbf{Y}_i) K_{(\boldsymbol{\alpha},\beta)}\left(\mathbf{X}_i\right)$$

and

$$\widehat{f}_n(\mathbf{x}, \boldsymbol{\Lambda}_{n,1}) := \frac{1}{n} \sum_{i=1}^{n} K_{(\boldsymbol{\alpha},\beta)}\left(\mathbf{X}_i\right).$$

In this section, we establish uniform strong consistency of the estimator defined by

$$\widehat{r}^{(1)}_{n,1}(\varphi, \mathbf{x}) = \frac{\widehat{g}_n(\varphi, \mathbf{x}, \mathbf{\Lambda}_{n,1})}{\widehat{f}_n(\mathbf{x}, \mathbf{\Lambda}_{n,1})}. \tag{3.3}$$

Finally, we represent the expectation of $\widehat{g}_n(\varphi, \mathbf{x}, \mathbf{\Lambda}_{n,1})$ as

$$\mathbb{E}\left[\widehat{g}_n(\varphi, \mathbf{x}, \mathbf{\Lambda}_{n,1})\right] = \mathbb{E}\left[\varphi(\mathbf{Y})K_{(\alpha,\beta)}(\mathbf{X})\right] = \int_{\mathbb{S}_{d,1}} r^{(1)}(\varphi, \mathbf{u}) f(\mathbf{u}) K_{(\alpha,\beta)}(\mathbf{u})\, d\mathbf{u} = \mathbb{E}\left[\mathcal{R}(\varphi, \boldsymbol{\xi}_{\mathbf{x}})\right],$$

where $\boldsymbol{\xi}_{\mathbf{x}} \sim \text{Dirichlet}(\alpha, \beta)$ and $\mathcal{R}(\varphi, \mathbf{x}) = f(\mathbf{x}) r^{(1)}(\varphi, \mathbf{x})$. To derive uniform consistency results, as usual, we rewrite

$$\begin{aligned}
\widehat{r}^{(1)}_{n,1}(\varphi, \mathbf{x}) - r^{(1)}(\varphi, \mathbf{x}) \;=\; & \frac{1}{\widehat{f}_n(\mathbf{x}, \mathbf{\Lambda}_{n,1})}\left(\widehat{g}_n(\varphi, \mathbf{x}, \mathbf{\Lambda}_{n,1}) - \mathbb{E}[\widehat{g}_n(\varphi, \mathbf{x}, \mathbf{\Lambda}_{n,1})]\right) \\
& - \frac{\mathbb{E}[\widehat{g}_n(\varphi, \mathbf{x}, \mathbf{\Lambda}_{n,1})]}{\widehat{f}_n(\mathbf{x}, \mathbf{\Lambda}_{n,1})\mathbb{E}[\widehat{f}_n(\mathbf{x}, \mathbf{\Lambda}_{n,1})]}\left(\widehat{f}_n(\mathbf{x}, \mathbf{\Lambda}_{n,1}) - \mathbb{E}[\widehat{f}_n(\mathbf{x}, \mathbf{\Lambda}_{n,1})]\right) \\
& - \left(\mathbb{E}(\varphi(Y)|X = \mathbf{x}) - \frac{\mathbb{E}[\widehat{g}_n(\varphi, \mathbf{x}, \mathbf{\Lambda}_{n,1})]}{\mathbb{E}[\widehat{f}_n(\mathbf{x}, \mathbf{\Lambda}_{n,1})]}\right).
\end{aligned}$$

For $\delta > 0$, define

$$\mathbb{S}_{d,1}(\delta) := \left\{\mathbf{x} \in \mathbb{S}_{d,1} : 1 - \|\mathbf{x}\|_1 \geq \delta \text{ and } x_i \geq \delta, \ \forall i = 1, \ldots, d\right\}. \tag{3.4}$$

The Dirichlet Nadaraya-Watson kernel estimators have not yet been proven to yield the following result.

Theorem 3.1. *Assume that* ***(C.1)*** *and* ***(C.3)*** *hold. If, in addition, $\breve{b}^{-d} \leq n$ as $n \to \infty$, then, as $n \to \infty$,*

$$\sup_{\mathbf{x} \in \mathbb{S}_{d,1}(\breve{b}d)} \left|\widehat{r}^{(1)}_{n,1}(\varphi, \mathbf{x}) - r^{(1)}(\varphi, \mathbf{x})\right| = O\left(\breve{b}^{1/2}\right) + O\left(\frac{|\log \breve{b}|(\log n)^{3/2}}{\breve{b}^{d+1/2}\sqrt{n}}\right) \quad a.s. \tag{3.5}$$

In particular, if $|\log \breve{b}|^2 \breve{b}^{-2d-1} = o\left(n/(\log n)^3\right)$ as $n \to \infty$, then

$$\sup_{\mathbf{x} \in \mathbb{S}_{d,1}(\breve{b}d)} \left|\widehat{r}^{(1)}_{n,1}(\varphi, \mathbf{x}) - r^{(1)}(\varphi, \mathbf{x})\right| \to 0 \quad a.s.$$

3.2. Uniform convergence of conditional U-statistics

In this section, we unveil our principal findings regarding the uniform almost sure consistency for the conditional U-statistics. Below, we state the uniform consistency of conditional U-statistics when the function φ is not necessarily bounded.

Theorem 3.2. *If* ***(C.2.)*** *and* ***(C.3.)*** *hold, then as $n \to \infty$, we have*

$$\sup_{\tilde{\mathbf{x}} \in \mathbb{S}^m_{d,1}} \left|u_{n,1}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}\left[u_{n,1}(\varphi, \tilde{\mathbf{x}})\right]\right| = O\left(\frac{|\log \breve{b}|^m (\log n)^{3/2}}{\breve{b}^{m(d+1/2)}\sqrt{n}}\right) \quad a.s. \tag{3.6}$$

Theorem 3.3. *If (C.2) and (C.3.) hold, then as $n \to \infty$, we have*

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\left|\widehat{r}_{n,1}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,1}(\tilde{\mathbf{x}}))-\widehat{\mathbb{E}}\left[\widehat{r}_{n,1}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,1}(\tilde{\mathbf{x}}))\right]\right| = O\left(\frac{|\log\breve{b}|^m(\log n)^{3/2}}{\breve{b}^{m(d+1/2)}\sqrt{n}}\right) \quad a.s. \tag{3.7}$$

Theorem 3.4. *If (C.2) holds, then, we have*

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\left|r^{(m)}(\varphi,\tilde{\mathbf{x}})-\widehat{\mathbb{E}}\left[\widehat{r}_{n,1}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,1}(\tilde{\mathbf{x}}))\right]\right| = O\left(\breve{b}^{1/2}\right). \tag{3.8}$$

Corollary 3.5. *Under the assumptions of Theorems 3.3 and 3.4, we have, as $n \to \infty$,*

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\left|\widehat{r}_{n,1}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,1}(\tilde{\mathbf{x}}))-r^{(m)}(\varphi,\tilde{\mathbf{x}})\right| = O\left(\breve{b}^{1/2}\right)+O\left(\frac{|\log\breve{b}|^m(\log n)^{3/2}}{\breve{b}^{m(d+1/2)}\sqrt{n}}\right) \quad a.s. \tag{3.9}$$

3.3. Asymptotic normality

Within this section, we establish the asymptotic normality for the estimator defined in (3.2). To achieve this, we rely on the following set of assumptions:

(A.1) Let $\tilde{\mathbf{x}} = (\mathbf{x}_1,\ldots,\mathbf{x}_m)$ be a point of continuity for each

$$r_{jl}(\tilde{\mathbf{x}}) = \begin{cases} 0 & \text{if } \mathbf{x}_j \neq \mathbf{x}_l, \\ \mathbb{E}_{j,l}(\tilde{\mathbf{x}}) & \text{if } \mathbf{x}_j = \mathbf{x}_l, \end{cases}$$

where

$$\mathbb{E}_{j,l}(\tilde{\mathbf{x}}) = \mathbb{E}\Big[\varphi\left(\mathbf{Y}_1,\ldots,\mathbf{Y}_{j-1},\mathbf{Y},\mathbf{Y}_{j+1},\ldots,\mathbf{Y}_m\right)\varphi\left(\mathbf{Y}_{m+1},\ldots,\mathbf{Y}_{m+j-1},\mathbf{Y},\mathbf{Y}_{m+j+1},\ldots,\mathbf{Y}_{2m}\right) \\ \mid \mathbf{X}_i = \mathbf{x}_i \text{ for } i\neq j, \mathbf{X}_{m+r} = \mathbf{x}_r \text{ for } r\neq l \text{ and } \mathbf{X} = \mathbf{x}_j = \mathbf{x}_l\Big];$$

(A.2) The density function $f(\cdot)$ is continuous at each $\mathbf{x}_j$, $1 \leq j \leq m$, with $f(\mathbf{x}_j) > 0$;

(A.3) $r_{j,l,s}(\cdot,\cdot,\cdot)$ is bounded in a neighborhood of $(\tilde{\mathbf{x}},\tilde{\mathbf{x}},\tilde{\mathbf{x}}) \in \mathbb{S}_{d,1}^{3m}$, where, for all $1 \leq j,l,s \leq m$

$$\begin{aligned} r_{j,l,s}(\tilde{\mathbf{z}}_m,\tilde{\mathbf{z}}_{2m},\tilde{\mathbf{z}}_{3m}) \quad = \quad & \mathbb{E}\Big[\varphi\left(\mathbf{Y}_1,\ldots,\mathbf{Y}_{j-1},\mathbf{Y},\mathbf{Y}_{j+1}\ldots,\mathbf{Y}_m\right) \\ & \times\varphi\left(\mathbf{Y}_{m+1},\ldots,\mathbf{Y}_{m+j-1},\mathbf{Y},\mathbf{Y}_{m+j+1}\ldots,\mathbf{Y}_{2m}\right) \\ & \times\varphi\left(\mathbf{Y}_{2m+1},\ldots,\mathbf{Y}_{2m+j-1},\mathbf{Y},\mathbf{Y}_{2m+j+1}\ldots,\mathbf{Y}_{3m}\right) \\ & \mid \mathbf{X}_i = \mathbf{z}_i; 1\leq i\leq 3m, i\neq j, m+1, 2m+s, \mathbf{X} = \mathbf{z}\Big], \end{aligned}$$

and for $1 \leq s \leq 3$, $\tilde{\mathbf{z}}_{sm} = (\mathbf{z}_{(s-1)m+1},\ldots,\mathbf{z}_{(s-1)m+j-1},\mathbf{z},\mathbf{z}_{(s-1)m+j+1},\ldots,\mathbf{z}_{sm})$;

(A.4) $r_{1,2}^{(m)}(\cdot,\cdot)$ is bounded in a neighborhood of $(\tilde{\mathbf{x}},\tilde{\mathbf{x}})$, where

$$r_{1,2}^{(m)}(\tilde{\mathbf{x}}_1,\tilde{\mathbf{x}}_2) = \mathbb{E}\left[\varphi(\mathbf{Y}_{i_1},\ldots,\mathbf{Y}_{i_m})\varphi(\mathbf{Y}_{j_1},\ldots,\mathbf{Y}_{j_m}) \mid (\mathbf{X}_{i_1},\ldots,\mathbf{X}_{i_m}) = \tilde{\mathbf{x}}_1, (\mathbf{X}_{j_1},\ldots,\mathbf{X}_{j_m}) = \tilde{\mathbf{x}}_2\right];$$

(A.5) $r^{(m)}(\varphi,\cdot)$ admits an expansion

$$r^{(m)}(\varphi,\mathbf{t}+\boldsymbol{\Delta}) = r^{(m)}(\varphi,\mathbf{t}) + \left\{\frac{\partial}{\partial \mathbf{t}} r^{(m)}(\varphi,\mathbf{t})\right\}^{\top} \boldsymbol{\Delta} + \frac{1}{2}\boldsymbol{\Delta}^{\top}\left\{\frac{\partial^2}{\partial \mathbf{t}^2} r^{(m)}(\varphi,\mathbf{t})\right\}\Delta + o\left(\boldsymbol{\Delta}^{\top}\boldsymbol{\Delta}\right)$$

as $\Delta \to 0$, for all $\mathbf{t}$ in a neighborhood of $\tilde{\mathbf{x}}$.

Below, we write $Z \stackrel{\mathcal{D}}{=} \mathcal{N}(\mu,\sigma^2)$ whenever the random variable Z follows a normal law with expectation μ and variance σ^2, $\stackrel{\mathcal{D}}{\to}$ denotes convergence in distribution. We also denote

$$\mathbb{U}_{n,1}(\varphi,\tilde{\mathbf{x}}) = \frac{u_{n,1}(\varphi,\tilde{\mathbf{x}})}{N},$$

where

$$N = \prod_{j=1}^{m} \mathbb{E}\left[K_{(\alpha,\beta)}\left(\mathbf{X}_j\right)\right].$$

Our main result in this section is summarized as follows.

Theorem 3.6. *If (A.1)–(A.4) and (C.2.) hold and if* $r^{(m)}(\cdot,\cdot)$ *is continuous at* $\tilde{\mathbf{x}}$*, then*

$$\sqrt{n\breve{b}^{d/2}}\left(\widehat{r}_{n,1}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\boldsymbol{\Lambda}}_{n,1}(\tilde{\mathbf{x}}))) - \mathbb{E}[\mathbb{U}_{n,1}(\varphi,\tilde{\mathbf{x}})]\right) \stackrel{\mathcal{D}}{\longrightarrow} \mathcal{N}\left(0,\rho^2\right),$$

where

$$\rho^2 := \sum_{i=1}^{m}\sum_{j=1}^{m} \mathbf{1}_{\{\mathbf{x}_i=\mathbf{x}_j\}}\left[r_{ij}(\tilde{\mathbf{x}}) - (r^{(m)}(\tilde{\mathbf{x}}))^2\right]\int K^2_{\alpha,\beta}(\mathbf{u})\mathrm{d}\mathbf{u}/f(\mathbf{x}_i). \tag{3.10}$$

The proof of Theorem 3.6 is postponed to Appendix-1 A.

Corollary 3.7. *If, in addition to the assumptions of Theorem 3.6,* **(A.5)** *holds, then*

$$\begin{aligned}
&\breve{b}^{-d/2}\left(\mathbb{E}\left[\mathbb{U}_{n,1}(\varphi,\tilde{\mathbf{x}})\right] - r^{(m)}(\varphi,\tilde{\mathbf{x}})\right)\\
&= \left[\int \prod_{j=1}^{m} K_{\alpha,\beta}(\mathbf{t}_j)\left\{\mathcal{R}^{(m)'}(\varphi,\tilde{\mathbf{t}})\right\}^{\top}\mathbf{t}\cdot \mathrm{d}\mathbf{t}/\tilde{f}(\tilde{\mathbf{x}}) - \int \prod_{j=1}^{m} K_{\alpha,\beta}(\mathbf{t}_j)\mathbf{t}^{\top}\left\{\tilde{f}'(\tilde{\mathbf{x}})\right\}\mathbf{t}\cdot \mathrm{d}\mathbf{t}\frac{r^{(m)}(\varphi,\tilde{\mathbf{x}})}{\tilde{f}(\tilde{\mathbf{x}})}\right]\\
&\quad + \frac{\breve{b}^{d/2}}{2}\left[\int \prod_{j=1}^{m} K_{\alpha,\beta}(\mathbf{t}_j)\mathbf{t}^{\top}\left\{\mathcal{R}^{(m)''}(\varphi,\tilde{\mathbf{t}})\right\}\mathbf{t}\cdot \mathrm{d}\mathbf{t}/\tilde{f}(\tilde{\mathbf{x}})\right.\\
&\quad \left. - \int \prod_{j=1}^{m} K_{\alpha,\beta}(\mathbf{t}_j)\mathbf{t}^{\top}\left\{\tilde{f}''(\tilde{\mathbf{x}})\right\}\mathbf{t}\cdot \mathrm{d}\mathbf{t}\frac{r^{(m)}(\varphi,\tilde{\mathbf{x}})}{\tilde{f}(\tilde{\mathbf{x}})}\right] + o(1).
\end{aligned}$$

In addition, we have

$$\left(n\breve{b}^{d/2}\right)^{1/2}\left[\widehat{r}_{n,1}^{(m)}\left(\varphi,\tilde{\mathbf{x}};\bar{\boldsymbol{\Lambda}}_{n,1}(\tilde{\mathbf{x}})\right) - r^{(m)}(\varphi,\tilde{\mathbf{x}})\right] \to \mathscr{N}\left(0,\rho^2\right),$$

provided that $n\breve{b}^{(d+2)/4} \to 0$.

Remark 3.8. *[117] According to Theorem 3.1.15 in [119], the convergence rate for the conventional d-dimensional kernel density estimator for i.i.d. data, using bandwidth h, is* $O\left(n^{-1/2}h^{-d/2}\right)$*. In contrast, the estimator* $\widehat{f_n}(\mathbf{x}, \mathbf{\Lambda}_{n,1})$ *achieves a convergence rate of* $O\left(n^{-1/2}\breve{b}^{-d/4}\right)$*. Consequently, the relationship between the bandwidths of* $\widehat{f_n}(\mathbf{x}, \mathbf{\Lambda}_{n,1})$ *and the traditional multivariate kernel density estimator is expressed as* $\breve{b} \approx h^2$*.*

Remark 3.9. *In their work, [117] demonstrated that for all* $\mathbf{x} \in Int(\mathbb{S}_{d,1})$ *and as n tends to infinity, the MSE of the estimator* $\widehat{f_n}(\mathbf{x}, \mathbf{\Lambda}_{n,1})$ *with respect to the density function* $f(\cdot)$ *can be expressed as:*

$$\begin{aligned}\mathrm{MSE}\left[\widehat{f_n}(\mathbf{x}, \mathbf{\Lambda}_{n,1})\right] &:= \mathbb{E}\left[\left|\widehat{f_n}(\mathbf{x}, \mathbf{\Lambda}_{n,1}) - f(\mathbf{x})\right|^2\right] \\ &= \mathrm{Var}\left[\widehat{f_n}(\mathbf{x}, \mathbf{\Lambda}_{n,1})\right] + \left\{\mathrm{Bias}\left[\widehat{f_n}(\mathbf{x}, \mathbf{\Lambda}_{n,1})\right]\right\}^2 \\ &= n^{-1}\breve{b}^{-d/2}\psi(\mathbf{x})f(\mathbf{x}) + \breve{b}^2 g^2(\mathbf{x}) + O_{\mathbf{x}}\left(n^{-1}\breve{b}^{-d/2+1/2}\right) + o\left(\breve{b}^2\right),\end{aligned}$$

where $\psi(\cdot)$ *is defined in Eq* (B.1) *in Lemma B.4, and*

$$g(\mathbf{x}) := \sum_{i=1}^{d} (1 - (d+1)x_i) \frac{\partial}{\partial x_i} f(\mathbf{x}) + \frac{1}{2} \sum_{i,j=1}^{d} x_i \left(\mathbf{1}_{\{i=j\}} - x_j\right) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}).$$

In particular, if $f(\mathbf{x}) \cdot g(\mathbf{x}) \neq 0$*, the asymptotically optimal choice of* $\breve{b}$*, concerning MSE, is given by:*

$$\breve{b}_{\mathrm{opt}}(\mathbf{x}) = n^{-2/(d+4)} \left[\frac{d}{4} \cdot \frac{\psi(\mathbf{x})f(\mathbf{x})}{g^2(\mathbf{x})}\right]^{2/(d+4)},$$

with

$$\mathrm{MSE}\left[\widehat{f_n}(\mathbf{x}, \mathbf{\Lambda}_{n,1}); \breve{b}_{\mathrm{opt}}\right] = n^{-4/(d+4)} \left[\frac{1 + \frac{d}{4}}{\left(\frac{d}{4}\right)^{\frac{d}{d+4}}}\right] \frac{(\psi(\mathbf{x})f(\mathbf{x}))^{4/(d+4)}}{(g^2(\mathbf{x}))^{-d/(d+4)}} + o_{\mathbf{x}}\left(n^{-4/(d+4)}\right).$$

Furthermore, if $n^{2/(d+4)}\breve{b}$ *tends to* λ *for some* $\lambda > 0$ *as n approaches infinity, then*

$$\mathrm{MSE}\left[\widehat{f_n}(\mathbf{x}, \mathbf{\Lambda}_{n,1})\right] = n^{-4/(d+4)}\left[\lambda^{-d/2}\psi(\mathbf{x})f(\mathbf{x}) + \lambda^2 g^2(\mathbf{x})\right] + o_{\mathbf{x}}\left(n^{-4/(d+4)}\right).$$

4. Conditional *U*-statistics using Bernstein polynomials

This section delves into the asymptotic properties of the conditional U-statistics using Bernstein polynomials. Let $F(\cdot)$ represent any joint cumulative distribution function on $\mathbb{S}_{d,1}$, where values outside $\mathbb{S}_{d,1}$ are either 0 or 1. Following [115, 116], we define the Bernstein polynomial of order ϑ for $F(\cdot)$ as follows

$$F_{\vartheta}^{\star}(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap \vartheta \mathbb{S}_{d,1}} F(\mathbf{k}/\vartheta) P_{\mathbf{k},\vartheta}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{S}_{d,1}, \vartheta \in \mathbb{N},$$

where the weights are probabilities from the multinomial$(\vartheta, \mathbf{x})$ distribution

$$P_{\mathbf{k},\vartheta}(\mathbf{x}) = \frac{\vartheta!}{(\vartheta - \|\mathbf{k}\|_1)! \prod_{i=1}^{d} k_i!} (1 - \|\mathbf{x}\|_1)^{\vartheta - \|k\|_1} \prod_{i=1}^{d} x_i^{k_i}, \quad \mathbf{k} \in \mathbb{N}_0^d \cap \vartheta \mathbb{S}_{d,1}.$$

The Bernstein estimator of $F(\cdot)$, denoted by $F^{\star}_{n,\vartheta}(\cdot)$, is the Bernstein polynomial of order ϑ for the empirical cumulative distribution function

$$F_n(\mathbf{x}) := n^{-1}\sum_{i=1}^{n}\mathbf{1}_{(-\infty,\mathbf{x}]}(\mathbf{X}_i),$$

where $\mathbf{X}_1,\ldots,\mathbf{X}_n$ are i.i.d according to $F(\cdot)$ and $\mathbf{1}\{A\}$ denotes as usual the indicator function of the set A. Precisely, we define

$$F^{\star}_{n,\vartheta}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{N}_0^d\cap\vartheta\mathbb{S}_{d,1}} F_n(\mathbf{k}/\vartheta)P_{\mathbf{k},\vartheta}(\mathbf{x}),\quad \mathbf{x}\in\mathbb{S}_{d,1},\vartheta,n\in\mathbb{N}.$$

For a density $f(\cdot)$ supported on $\mathbb{S}_{d,1}$, we define the Bernstein density estimator of $f(\cdot)$ as

$$\hat{f}_{n,\vartheta}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}} \frac{(\vartheta-1+d)!}{(\vartheta-1)!}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k}+\mathbf{1}}{\vartheta}\right]}(\mathbf{x}_i)\right\}P_{\mathbf{k},\vartheta-1}(\mathbf{x}),\quad \mathbf{x}\in\mathbb{S}_{d,1},\vartheta,n\in\mathbb{N}. \tag{4.1}$$

In this expression, we let

$$K_{\mathbf{x},\vartheta}(\mathbf{X}_i) = \sum_{\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}} \frac{(\vartheta-1+d)!}{(\vartheta-1)!}\left\{\mathbf{1}_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k}+\mathbf{1}}{\vartheta}\right]}(\mathbf{X}_i)\right\}P_{\mathbf{k},\vartheta-1}(\mathbf{x}),$$

and $(\vartheta-1+d)!/(\vartheta-1)!$ serves as a scaling factor proportional to the inverse of the volume of the hypercube $(\mathbf{k}/\vartheta,(\mathbf{k}+\mathbf{1})/\vartheta]$. It's worth noting that replacing this scaling factor by ϑ^d in (4.1) maintains $\hat{f}_{n,\vartheta}(\cdot)$ as an asymptotic density function. The asymptotic results presented in this paper remain largely unchanged under both definitions of the density estimator.

Remark 4.1. *[115] A different expression for the Bernstein density estimator (4.1) can be formulated as a specific finite mixture of Dirichlet densities, that is,*

$$\hat{f}_{n,\vartheta}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}} \left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k}+\mathbf{1}}{\vartheta}\right]}(\mathbf{x}_i)\right\}K_{(\mathbf{k}+\mathbf{1},\vartheta-\|\mathbf{k}\|_1)}(\mathbf{x}),$$

where the density value of the Dirichlet(α,β) *distribution at* $\mathbf{x}\in\mathbb{S}_{d,1}$ *is given by*

$$K_{(\alpha,\beta)}(\mathbf{x}) = \frac{(\beta+\|\boldsymbol{\alpha}\|_1-1)!}{(\beta-1)!\prod_{i=1}^{d}(\alpha_i-1)!}(1-\|\mathbf{x}\|_1)^{\beta-1}\prod_{i=1}^{d}x_i^{\alpha_i-1},\quad \alpha_i,\beta>0.$$

For further insights, refer to [115].

The conditional U-statistics smoothed by the Bernstein polynomials are defined as, for each $\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}$

$$\widehat{r}^{(m)}_{n,2}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,2}(\tilde{\mathbf{x}})) = \frac{\sum_{(i_1,\ldots,i_m)\in I(m,n)}\varphi(Y_{i_1},\ldots,Y_{i_m})K_{\mathbf{x}_1,\vartheta}(\mathbf{X}_{i_1})\ldots K_{\mathbf{x}_m,\vartheta}(\mathbf{X}_{i_m})}{\sum_{(i_1,\ldots,i_m)\in I(m,n)}K_{\mathbf{x}_1,\vartheta}(\mathbf{X}_{i_1})\ldots K_{\mathbf{x}_m,\vartheta}(\mathbf{X}_{i_m})}. \tag{4.2}$$

In the particular case $m = 1$, the $r^{(m)}(\varphi, \tilde{\mathbf{x}})$ is reduced to $r^{(1)}(\varphi, \mathbf{x}) = \mathbb{E}(\varphi(\mathbf{Y})|\mathbf{X} = \mathbf{x})$ and Stute's estimator becomes the Nadaraya-Watson estimator of $r^{(1)}(\varphi, \mathbf{x})$

$$\widehat{r}_{n,2}^{(1)}(\varphi, \mathbf{x}) := \frac{\sum_{i=1}^{n} \varphi(\mathbf{Y}_i) K_{\mathbf{x},\vartheta}(\mathbf{X}_i)}{\sum_{i=1}^{n} K_{\mathbf{x},\vartheta}(\mathbf{X}_i)} = \frac{\widehat{g}_{n,\vartheta}(\varphi, \mathbf{x})}{\widehat{f}_{n,\vartheta}(\mathbf{x})}.$$

4.1. Nonparametric regression estimation

In this section, we prove the uniform strong consistency of the regression estimator for $m = 1$.

Theorem 4.2. *Assume that **(C.1)–(C.3)** hold. If $2 \leq \vartheta \leq \dfrac{n}{\log n}$, as $n \longrightarrow \infty$, then*

$$\sup_{\mathbf{x} \in \mathbb{S}_{d,1}} \left| \mathbb{E}[\widehat{r}_{n,2}^{(1)}(\varphi, \mathbf{x})] - r^{(1)}(\varphi, \mathbf{x}) \right| = O(\vartheta^{-1/2}) \tag{4.3}$$

and

$$\sup_{\mathbf{x} \in \mathbb{S}_{d,1}} \left| \widehat{r}_{n,2}^{(1)}(\varphi, \mathbf{x}) - \mathbb{E}[\widehat{r}_{n,2}^{(1)}(\varphi, \mathbf{x})] \right| = O(\vartheta^{d-1/2}(n^{-1} \log n)^{1/2}) \ \ a.s. \tag{4.4}$$

Corollary 4.3. *Under the assumptions of Theorem 4.2, we have as $n \longrightarrow \infty$*

$$\sup_{\mathbf{x} \in \mathbb{S}_{d,1}} \left| \widehat{r}_{n,2}^{(1)}(\varphi, \mathbf{x}) - r^{(1)}(\varphi, \mathbf{x}) \right| = O(\vartheta^{d-1/2}(n^{-1} \log n)^{1/2}) + O(\vartheta^{-1/2}) \ \ a.s. \tag{4.5}$$

In addition, if $\vartheta^{2d-1} = o(n/\log n)$, then

$$\sup_{\mathbf{x} \in \mathbb{S}_{d,1}} \left| \widehat{r}_{n,2}^{(1)}(\varphi, \mathbf{x}) - r^{(1)}(\varphi, \mathbf{x}) \right| \longrightarrow 0 \ \ a.s. \tag{4.6}$$

4.2. Conditional U-statistics

In this section, we study the uniform strong consistency of the conditional U-statistics using Bernstein polynomials.

Theorem 4.4. *Assume that **(C.2)** and **(C.3)** hold. If $2 \leq \vartheta \leq \dfrac{n}{\log n}$, as $n \longrightarrow \infty$, then*

$$\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{d,1}^m} \left| u_{n,2}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}\left[u_{n,2}(\varphi, \tilde{\mathbf{x}})\right] \right| = O(\vartheta^{m(d-1/2)}(n^{-1} \log n)^{1/2}) \ \ a.s. \tag{4.7}$$

Theorem 4.5. *Assume that **(C.2)** and **(C.3)** hold. If $2 \leq \vartheta \leq \dfrac{n}{\log n}$, as $n \longrightarrow \infty$, then*

$$\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{d,1}^m} \left| \widehat{r}_{n,2}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,2}(\tilde{\mathbf{x}})) - \widehat{\mathbb{E}}\left[\widehat{r}_{n,2}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,2}(\tilde{\mathbf{x}}))\right] \right| = O(\vartheta^{m(d-1/2)}(n^{-1} \log n)^{1/2}) \ \ a.s. \tag{4.8}$$

Theorem 4.6. *Assume that **(C.1)** and **(C.2)** hold. If $2 \leq \vartheta \leq \dfrac{n}{\log n}$, as $n \longrightarrow \infty$, then*

$$\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{d,1}^m} \left| r^{(m)}(\varphi, \tilde{\mathbf{x}}) - \widehat{\mathbb{E}}\left[\widehat{r}_{n,2}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,2}(\tilde{\mathbf{x}}))\right] \right| = O\left(\vartheta^{-m/2}\right). \tag{4.9}$$

Corollary 4.7. *Under the assumptions of Theorems 4.5 and 4.6, as $n \longrightarrow \infty$, we have*

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}} \left|\widehat{r}^{(m)}_{n,2}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,2}(\tilde{\mathbf{x}})) - r^{(m)}(\varphi, \tilde{\mathbf{x}})\right| = O\left(\vartheta^{-m/2}\right) + O(\vartheta^{m(d-1/2)}(n^{-1}\log n)^{1/2}) \ \ a.s. \tag{4.10}$$

In addition, if $\vartheta^{2d-1} = o(n/\log n)$, then

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}} \left|\widehat{r}^{(m)}_{n,2}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,2}(\tilde{\mathbf{x}})) - r^{(m)}(\varphi, \tilde{\mathbf{x}})\right| \longrightarrow 0 \ \ a.s. \tag{4.11}$$

Remark 4.8. *[115] It is worth mentioning that, similar to Remark 3.8, the convergence rate for the conventional d-dimensional kernel density estimator for i.i.d. data, using bandwidth h, is $O\left(n^{-1/2}h^{-d/2}\right)$. However, the estimator $\widehat{f_n}(\mathbf{x}, \vartheta)$ achieves a convergence rate of $O\left(n^{-1/2}\vartheta^{d/4}\right)$. Consequently, the relationship between the bandwidths of $\widehat{f_n}(\mathbf{x}, \vartheta)$ and the traditional multivariate kernel density estimator is expressed as $\vartheta \approx h^{-2}$.*

Remark 4.9. *[115] demonstrated that the Mean Squared Error (MSE) of the density estimator $\widehat{f_n}(\mathbf{x}, \vartheta)$ satisfies for all $\mathbf{x} \in Int(\mathbb{S}_{d,1})$ and as n tends to infinity:*

$$\mathrm{MSE}\left(\widehat{f_n}(\mathbf{x}, \vartheta)\right) = n^{-1}\vartheta^{d/2}\psi(\mathbf{x})f(\mathbf{x}) + \vartheta^{-2}b^2(\mathbf{x}) + o_{\mathbf{x}}\left(n^{-1}\vartheta^{d/2}\right) + \mathrm{o}\left(\vartheta^{-2}\right),$$

where

$$b(\mathbf{x}) := \frac{d(d-1)}{2}f(\mathbf{x}) + \sum_{i=1}^{d}\left(\frac{1}{2} - x_i\right)\frac{\partial}{\partial x_i}f(\mathbf{x}) + \frac{1}{2}\sum_{i,j=1}^{d}\left(x_i\mathbf{1}_{\{i=j\}} - x_i x_j\right)\frac{\partial^2}{\partial x_i \partial x_j}f(\mathbf{x}),$$

and

$$\psi(\mathbf{x}) := \left[(4\pi)^d\,(1 - \|\mathbf{x}\|_1)\prod_{i=1}^{d} x_i\right]^{-1/2}.$$

In particular, when $f(\mathbf{x}) \cdot b(\mathbf{x}) \neq 0$, the asymptotically optimal choice for ϑ, minimizing MSE, is:

$$\vartheta_{opt}\,(\mathbf{x}) = n^{2/(d+4)}\left[\frac{4}{d}\cdot\frac{b^2(\mathbf{x})}{\psi(\mathbf{x})f(\mathbf{x})}\right]^{2/(d+4)},$$

with corresponding MSE:

$$\mathrm{MSE}\left[\widehat{f_n}(\mathbf{x}, \vartheta); \vartheta_{opt}\right] = n^{-4/(d+4)}\left[\frac{\frac{4}{d}+1}{\left(\frac{4}{d}\right)^{\frac{4}{d+4}}}\right]\frac{(\psi(\mathbf{x})f(\mathbf{x}))^{4/(d+4)}}{(b^2(\mathbf{x}))^{-d/(d+4)}} + \mathrm{o}_{\mathbf{x}}\left(n^{-4/(d+4)}\right).$$

Moreover, in the more general case where $n^{2/(d+4)}\vartheta^{-1} \to \lambda > 0$, as $n \to \infty$, the MSE becomes:

$$\mathrm{MSE}\left[\widehat{f_n}(\mathbf{x}, \vartheta)\right] = n^{-4/(d+4)}\left[\lambda^{-d/2}\psi(\mathbf{x})f(\mathbf{x}) + \lambda^2 b^2(\mathbf{x})\right] + o_{\mathbf{x}}\left(n^{-4/(d+4)}\right).$$

5. Conditional *U*-statistics estimators using beta kernel

Throughout this section, it is assumed, as in [79], without loss of generality, that the compact set is a d-dimensional unit hypercube $[0,1]^d$. Among all asymmetric kernels, our particular focus is on the beta kernel by [40]. The kernel takes the form

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

where

$$\breve{\alpha} := \frac{x}{b}+1 \text{ and } \breve{\beta} := \frac{1-x}{b}+1,\ x\in[0,1], b>0,$$

and $B(\breve{\alpha},\breve{\beta}) = \int_0^1 y^{\breve{\alpha}-1}(1-y)^{\breve{\beta}-1}dy$ for $\breve{\alpha},\breve{\beta}>0$ is the beta function. To cope with multivariate problems, we construct a product kernel for $\breve{\boldsymbol{\alpha}} = (\breve{\alpha}_1,\ldots,\breve{\alpha}_d)$ and $\breve{\boldsymbol{\beta}} = (\breve{\beta}_1,\ldots,\breve{\beta}_d)$

$$K_{\breve{\boldsymbol{\alpha}},\breve{\boldsymbol{\beta}}}(\mathbf{u}) = \prod_{i=1}^{d} K_{\breve{\alpha}_i,\breve{\beta}_i}(u_i) = \prod_{i=1}^{d} \frac{u_i^{x_i/b_i}(1-u_i)^{(1-x_i)/b_i}}{B\{x_i/b_i+1,(1-x_i)/b_i+1\}}\mathbf{1}\{u_i\in[0,1]\},$$

where $\mathbf{u} := (u_1,\ldots,u_d) \in [0,1]^d, \mathbf{x} := (x_1,\ldots,x_d) \in [0,1]^d$, and $\mathbf{b} := (b_1,\ldots,b_d) \in \mathbb{R}_+^d$ are d-dimensional vectors of data points, design points, and the smoothing parameter. We consider

$$\widehat{r}_{n,3}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\boldsymbol{\Lambda}}_{n,3}(\tilde{\mathbf{x}})) = \frac{\sum_{(i_1,\ldots,i_m)\in I(m,n)} \varphi(\mathbf{Y}_{i_1},\ldots,\mathbf{Y}_{i_m})K_{\breve{\boldsymbol{\alpha}}_1,\breve{\boldsymbol{\beta}}_1}(\mathbf{X}_{i_1})\ldots K_{\breve{\boldsymbol{\alpha}}_m,\breve{\boldsymbol{\beta}}_m}(\mathbf{X}_{i_m})}{\sum_{(i_1,\ldots,i_m)\in I(m,n)} \mathbb{K}_{B(\mathbf{x}_1,\mathbf{b})}(\mathbf{X}_{i_1})\ldots K_{\breve{\boldsymbol{\alpha}}_m,\breve{\boldsymbol{\beta}}_m}(\mathbf{X}_{i_m})}, \tag{5.1}$$

where

$$\breve{\boldsymbol{\alpha}}_j := \frac{\mathbf{x}_j}{\breve{\mathbf{b}}_j}+\mathbf{1} \text{ and } \breve{\boldsymbol{\beta}}_j := \frac{\mathbf{1}-\mathbf{x}_j}{\breve{\mathbf{b}}_j}+\mathbf{1}.$$

In the particular case $m=1$, the Nadaraya-Watson estimator of $r^{(1)}(\varphi,\tilde{\mathbf{x}})$ of [79] is given by

$$\widehat{r}_n^{(1)}(\varphi,\mathbf{x}) := \frac{\sum_{i=1}^{n}\varphi(\mathbf{Y}_i)K_{\breve{\alpha},\breve{\beta}}(\mathbf{X}_i)}{\sum_{i=1}^{n}K_{\breve{\alpha},\breve{\beta}}(\mathbf{X}_i)}.$$

5.1. Conditions and comments

Our analysis starts from demonstrating weak uniform consistency with rates of the sample average estimator (5.1) for (2.1) on a dm-hyper-rectangle

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

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

(C.4) $\mathbf{b}_j := \mathbf{b}_j(n) = \left(b_{j_1}, \ldots, b_{j_d}\right) > 0$ and $\boldsymbol{\eta}_j := \boldsymbol{\eta}_j(n) = \left(\eta_{j_1}, \ldots, \eta_{j_d}\right) > 0$, $j = 1, \ldots, m$, satisfy for $i = 1, \ldots, d$, $b_{j_i}, \eta_{j_i} \to 0$, $\dfrac{b_{j_i}}{\eta_{j_i}} \to 0$, and

$$\frac{\log n}{n \prod_{j=1}^{m} \left(\prod_{i=1}^{d} b_{j_i} \eta_{j_i} \right)} \longrightarrow 0 \ \text{ as } \ n \to \infty.$$

The conditions on η_{j_i} in **(C.4)** are intended for the case of an expanding set. In particular, the condition $b_{j_i}/\eta_{j_i} \to 0$ means that the boundary parameter η_{j_i} must shrink to zero at a slower rate than b_{j_i}, this is crucial for Stirling's approximation to the gamma function. This condition was used [79] for the novel proof of the convergence results that we have extended to our setting.

5.2. Weak uniform convergence of conditional U-statistics

In the following theorem, we state that the weak uniform convergence of conditional U-statistics, in the particular case of $m = 1$, reduces to the results obtained in [79].

Theorem 5.1. *If (C.2)–(C.4) hold, then, as $n \to \infty$,*

$$\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{\mathbf{X}}^m} \left| u_{n,3}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}\left[u_{n,3}(\varphi, \tilde{\mathbf{x}})\right] \right| = O_{\mathbb{P}}\left(\sqrt{\frac{(\log n/n)}{\prod_{j=1}^{m} \left(\prod_{i=1}^{d} b_{j_i} \eta_{j_i} \right)}} \right). \tag{5.2}$$

Theorem 5.2. *If (C.2)–(C.4) hold, then, as $n \to \infty$,*

$$\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{\mathbf{X}}^m} \left| \widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})) - \widehat{\mathbb{E}}\left[\widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})) \right] \right| = O_{\mathbb{P}}\left(\sqrt{\frac{(\log n/n)}{\prod_{j=1}^{m} \left(\prod_{i=1}^{d} b_{j_i} \eta_{j_i} \right)}} \right). \tag{5.3}$$

Theorem 5.3. *If (C.2) holds, then, as $n \to \infty$,*

$$\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{\mathbf{X}}^m} \left| r^{(m)}(\varphi, \tilde{\mathbf{x}}) - \widehat{\mathbb{E}}\left[\widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})) \right] \right| = O\left(\sum_{j=1}^{m} \sum_{i=1}^{d} b_{j_i} \right). \tag{5.4}$$

Corollary 5.4. *Under the assumptions of Theorem 5.2 and Theorem 5.3, as $n \to \infty$,*

$$\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{\mathbf{X}}^m} \left| \widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})) - r^{(m)}(\varphi, \tilde{\mathbf{x}}) \right| = O_{\mathbb{P}}\left(\sum_{j=1}^{m} \sum_{i=1}^{d} b_{j_i} + \sqrt{\frac{(\log n/n)}{\prod_{j=1}^{m} \left(\prod_{i=1}^{d} b_{j_i} \eta_{j_i} \right)}} \right). \tag{5.5}$$

5.3. Strong uniform convergence of conditional U-statistics

In this section, we establish strong uniform consistency with rates of $\widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}}))$. However, to do so, the assumption on smoothing parameters must be suitably strengthened.

(C.4') $\mathbf{b}_j := \mathbf{b}_j(n) = \left(b_{j_1}, \ldots, b_{j_d}\right) > 0$ and $\boldsymbol{\eta}_j := \boldsymbol{\eta}_j(n) = \left(\eta_{j_1}, \ldots, \eta_{j_d}\right) > 0$, $j = 1, \ldots, m$, satisfy for $i = 1, \ldots, d$, $b_{j_i}, \eta_{j_i} \to 0$, $\dfrac{b_{j_i}}{\eta_{j_i}} \to 0$ and

$$\frac{\log n}{n \prod_{j=1}^{m}\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)} \left(\sum_{j=1}^{m}\sum_{i=1}^{d}\frac{1}{b_{j_i}^2}\right)^{1-\kappa} = O(1), \tag{5.6}$$

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

The condition (5.6) is stronger than $\log n / \left(n\sqrt{\prod_{j=1}^{m}\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)}\right) \longrightarrow 0$ in **(C.4)** in that the former implies the latter. Under this condition, the statement in Corollary 5.4 can be strengthened to almost sure convergence. The following theorems generalize the results of [79] that are given for $m = 1$.

Theorem 5.5. *If (C.2)–(C.3) and (C.4') hold, then, as n → ∞,*

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\left|u_{n,3}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}\left[u_{n,3}(\varphi, \tilde{\mathbf{x}})\right]\right| = O\left(\sqrt{\frac{(\log n/n)}{\prod_{j=1}^{m}\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)}}\right) \quad a.s. \tag{5.7}$$

Theorem 5.6. *If (C.2)–(C.3) and (C.4') hold, then, as n → ∞,*

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\left|\widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})) - \widehat{\mathbb{E}}\left[\widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}}))\right]\right| = O\left(\sqrt{\frac{(\log n/n)}{\prod_{j=1}^{m}\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)}}\right) \quad a.s. \tag{5.8}$$

Theorem 5.7. *If (C.2) holds, then*

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\left|r^{(m)}(\varphi, \tilde{\mathbf{x}}) - \widehat{\mathbb{E}}\left[\widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}}))\right]\right| = O\left(\sum_{j=1}^{m}\sum_{i=1}^{d} b_{j_i}\right). \tag{5.9}$$

Corollary 5.8. *Under the assumptions of Theorems 5.6 and 5.9, as n → ∞,*

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\left|\widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})) - r^{(m)}(\varphi, \tilde{\mathbf{x}})\right| = O\left(\sum_{j=1}^{m}\sum_{i=1}^{d} b_{j_i} + \sqrt{\frac{(\log n/n)}{\prod_{j=1}^{m}\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)}}\right) \quad a.s. \tag{5.10}$$

5.4. *Conditional U-statistics estimators for mixed categorical and continuous data*

Let's delve into the methodology for handling a discrete random variable Z, which can assume c distinct values, $\{0, 1, \ldots, c-1\}$, where $c \geq 2$; we refer to [43, 101, 104, 106, 132]. We categorize this variable as either unordered or ordered, as the kernels utilized for these two types differ slightly. For an unordered variable, the univariate discrete kernel takes the form

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

Here, v represents the data point, z denotes the design point, and $\lambda \in (0, 1)$ denotes the bandwidth. Conversely, the univariate discrete kernel for an ordered variable is given by

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

Moving on to the product discrete kernel, when $q_1(\leq q)$ out of q discrete variables are unordered, it becomes

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

Here, $\mathbf{v} := \left(v_1, \ldots, v_q\right)$, $\mathbf{z} := \left(z_1, \ldots, z_q\right)$, and $\lambda := \left(\lambda_1, \ldots, \lambda_q\right)$. Combining this with the product beta kernel $K_{\breve{\alpha},\breve{\beta}}(\mathbf{u})$ yields the product kernel for mixed categorical and continuous data

$$\mathbb{W}(\mathbf{u}, \mathbf{v}; \mathbf{x}, \mathbf{z}, \mathbf{b}, \lambda) = K_{\breve{\alpha},\breve{\beta}}(\mathbf{u})\mathbb{L}(\mathbf{v}; \mathbf{z}, \lambda).$$

Given this kernel and n i.i.d. observations $\{(Y_i, \mathbf{X}_i, \mathbf{Z}_i)\}_{i=1}^n \in \mathbb{R} \times [0, 1]^d \times \mathbb{S}_{\mathbf{Z}}$, where $\mathbb{S}_{\mathbf{Z}} := \prod_{k=1}^q \{0, 1, \ldots, c_k - 1\}$, we turn to a regression estimator of the conditional mean

$$r^{(m)}(\varphi, \tilde{\mathbf{x}}, \tilde{\mathbf{z}}) = \mathbb{E}\left(\varphi(\mathbf{Y}_1, \ldots, \mathbf{Y}_m) \mid (\mathbf{X}_1, \ldots, \mathbf{X}_m) = \tilde{\mathbf{x}}, (\mathbf{Z}_1, \ldots, \mathbf{Z}_m) = \tilde{\mathbf{z}}\right).$$

This estimator, denoted as $\widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}}))$, is expressed as

$$\widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})) = \frac{\sum\limits_{(i_1,\ldots,i_m)\in I(m,n)} \varphi(Y_{i_1}, \ldots, Y_{i_m})\mathbb{W}\left(\mathbf{X}_{i_1}, \mathbf{Z}_{i_1}; \mathbf{x}_1, \mathbf{z}_1, \mathbf{b}, \lambda\right)\cdots\mathbb{W}\left(\mathbf{X}_{i_m}, \mathbf{Z}_{i_m}; \mathbf{x}_m, \mathbf{z}_m, \mathbf{b}, \lambda\right)}{\sum\limits_{(i_1,\ldots,i_m)\in I(m,n)} \mathbb{W}\left(\mathbf{X}_{i_1}, \mathbf{Z}_{i_1}; \mathbf{x}_1, \mathbf{z}_1, \mathbf{b}, \lambda\right)\cdots\mathbb{W}\left(\mathbf{X}_{i_m}, \mathbf{Z}_{i_m}; \mathbf{x}_m, \mathbf{z}_m, \mathbf{b}, \lambda\right)} \quad (5.11)$$

5.4.1. Weak uniform convergence

Before we state the uniform convergence results of the estimator, we assume

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

(C.2') Let $\tilde{f}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ be the joint pdf of $(\tilde{X}, \tilde{Y})$. Then, the second-order derivatives of $\tilde{f}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ and $\tilde{g}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) := r^{(m)}(\varphi, \tilde{\mathbf{x}}, \tilde{\mathbf{z}})\tilde{f}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$, with respect to $\tilde{\mathbf{x}}$, are continuous on $\tilde{\mathbf{x}} \in (0, 1)^{dm}$;

(C.3') There are some constants $\gamma > 0$ and $C_1 \in [1, \infty)$ such that $\mathbb{E}|\varphi(\mathbf{Y})|^{2+\gamma} < \infty$ and

$$\sup_{\tilde{\mathbf{x}} \in (0,1)^{dm} \times \mathbb{S}^m_{\mathbf{Z}}} \mathbb{E}\left(|\varphi(\mathbf{Y})|^{2+\gamma} \mid \tilde{\mathbf{X}} = \tilde{\mathbf{x}}, \tilde{\mathbf{Z}} = \tilde{\mathbf{z}}\right) \tilde{f}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \leq C_1; \tag{5.12}$$

(C.4'') $\mathbf{b}_j := \mathbf{b}_j(n) = \left(b_{j_1}, \ldots, b_{j_d}\right) > 0$, $\boldsymbol{\eta}_j := \boldsymbol{\eta}_j(n) = \left(\eta_{j_1}, \ldots, \eta_{j_d}\right) > 0$, $j = 1, \ldots, m$, and $\lambda_k := \lambda_k \in (0, 1)$, $k = 1, \ldots, q$, satisfy for $i = 1, \ldots, d$, $b_{j_i}, \eta_{j_i} \to 0$, $\dfrac{b_{j_i}}{\eta_{j_i}} \to 0$, $\lambda_k \to 0$, and

$$\frac{\log n}{n \prod_{j=1}^{m} \left(\prod_{i=1}^{d} b_{j_i} \eta_{j_i} \right)} \longrightarrow 0 \text{ as } n \to \infty;$$

(C.4''') $\mathbf{b}_j := \mathbf{b}_j(n) = \left(b_{j_1}, \ldots, b_{j_d}\right) > 0$, $\boldsymbol{\eta}_j := \boldsymbol{\eta}_j(n) = \left(\eta_{j_1}, \ldots, \eta_{j_d}\right) > 0$, $j = 1, \ldots, m$, and $\lambda_k := \lambda_k \in (0, 1)$, $k = 1, \ldots, q$ satisfy, for $i = 1, \ldots, d$, $b_{j_i}, \eta_{j_i} \to 0$, $\dfrac{b_{j_i}}{\eta_{j_i}} \to 0$, $\lambda_k \to 0$, and

$$\frac{\log n}{n \prod_{j=1}^{m} \left(\prod_{i=1}^{d} b_{j_i} \eta_{j_i} \right)} \left(\sum_{j=1}^{m} \sum_{i=1}^{d} \frac{1}{b^2_{j_i}} \right)^{1-\kappa} = O(1) \tag{5.13}$$

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

(C.5) Let $f^m_n := \inf_{\tilde{\mathbf{x}} \in \mathbb{S}^m_X} \tilde{f}(\tilde{\mathbf{x}}) > 0$, f^m_n tends to zero as $n \longrightarrow \infty$, and

$$f^m_n \left(\sum_{j=1}^{m} \sum_{i=1}^{d} b_{j_i} + \sum_{k=1}^{q} \lambda_k + \sqrt{\frac{(\log n/n)}{\prod_{j=1}^{m} \left(\prod_{i=1}^{d} b_{j_i} \eta_{j_i} \right)}} \right) \to 0. \tag{5.14}$$

Corollary 5.9. *If (C.1')–(C.3'), (C.4'') and (C.5) hold, then, as n → ∞,*

$$\sup_{(\tilde{\mathbf{x}}, \tilde{\mathbf{z}}) \in \mathbb{S}^m_{\mathbf{X}} \times \mathbb{S}^m_{\mathbf{Z}}} \left| \widehat{r}^{(m)}_{n,3}(\varphi, \tilde{\mathbf{x}}, \tilde{\mathbf{z}}, \mathbf{b}_n) - r^{(m)}(\varphi, \tilde{\mathbf{x}}, \tilde{\mathbf{z}}) \right| = O_{\mathbb{P}} \left\{ f^m_n \left(\sum_{j=1}^{m} \sum_{i=1}^{d} b_{j_i} + \sum_{k=1}^{q} \lambda_k + \sqrt{\frac{(\log n/n)}{\prod_{j=1}^{m} \left(\prod_{i=1}^{d} b_{j_i} \eta_{j_i} \right)}} \right) \right\}. \tag{5.15}$$

6. Applications

6.1. Discrimination problems

Now, we apply the results of the problem of discrimination described in Section 3 of [137], referring also to [136]. We will use a similar notation and setting. Let $\varphi(\cdot)$ be any function taking at most finitely many values, say, $1, \ldots, M$. The sets

$$A_j = \{(\mathbf{y}_1, \ldots, \mathbf{y}_k) : \varphi(\mathbf{y}_1, \ldots, \mathbf{y}_k) = j\}, \quad 1 \le j \le M,$$

then yield a partition of the feature space. Predicting the value of $\varphi(\mathbf{Y}_1, \ldots, \mathbf{Y}_k)$ is tantamount to predicting the set in the partition to which $(\mathbf{Y}_1, \ldots, \mathbf{Y}_k)$ belongs. For any discrimination rule g, we have

$$\mathbb{P}(g(\mathbf{X}) = \varphi(\mathbf{Y})) \le \sum_{j=1}^{M} \int_{\{\tilde{\mathbf{x}}:g(\tilde{\mathbf{x}})=j\}} \max_{1\le j\le M} \mathfrak{M}^j(\tilde{\mathbf{x}}) d\mathbb{P}(\tilde{\mathbf{x}}),$$

where

$$\mathfrak{M}^j(\tilde{\mathbf{x}}) = \mathbb{P}(\varphi(\mathbf{Y}) = j \mid \mathbf{X} = \tilde{\mathbf{x}}), \quad \tilde{\mathbf{x}} \in \mathbb{R}^d.$$

The above inequality becomes an equality if

$$g_0(\tilde{\mathbf{x}}) = \arg \max_{1\le j\le M} \mathfrak{M}^j(\tilde{\mathbf{x}}).$$

$g_0(\cdot)$ is called the Bayes rule, and the pertaining probability of error

$$\mathbf{L}^* = 1 - \mathbb{P}(g_0(\mathbf{X}) = \varphi(\mathbf{Y})) = 1 - \mathbb{E}\left\{\max_{1\le j\le M} \mathfrak{M}^j(\tilde{\mathbf{x}})\right\}$$

is called the Bayes risk. Each of the above unknown function $\mathfrak{M}^j$'s can be consistently estimated by one of the methods discussed in the preceding sections. Let, for $1 \le j \le M$ and $\ell \in \{1, 2, 3\}$,

$$\mathfrak{M}^j_{n,\ell}(\tilde{\mathbf{x}}) = \frac{\sum_{(i_1,\ldots,i_k)\in I(k,n)} \mathbf{1}\{\varphi(Y_{i_1}, \ldots, Y_{i_k}) = j\} K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_1)}(\mathbf{X}_{i_1}) \ldots K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_m)}(\mathbf{X}_{i_m})}{\sum_{(i_1,\ldots,i_k)\in I(k,n)} K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_1)}(\mathbf{X}_{i_1}) \ldots K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_m)}(\mathbf{X}_{i_m})}. \tag{6.1}$$

Set

$$g_{0,n,\ell}(\tilde{\mathbf{x}}) = \arg \max_{1\le j\le M} \mathfrak{M}^j_{n,\ell}(\tilde{\mathbf{x}}).$$

Let us introduce

$$\mathbf{L}^*_{n,\ell} = \mathbb{P}(g_{0,n,\ell}(\mathbf{X}) \ne \varphi(\mathbf{Y})).$$

The discrimination rule $g_{0,n,\ell}(\cdot)$ is asymptotically Bayes' risk consistent

$$\mathbf{L}^*_{n,\ell} \to \mathbf{L}^*.$$

This follows from Corollaries 3.5, 4.7, or 5.8 and the obvious relation

$$\left|\mathbf{L}^* - \mathbf{L}^*_{n,\ell}\right| \le 2\mathbb{E}\left[\max_{1\le j\le M} \left|\mathfrak{M}^j_{n,\ell}(\mathbf{x}) - \mathfrak{M}^j(\mathbf{x})\right|\right].$$

6.2. Generalized U-statistics

The extension to the case of several samples is straightforward. Consider $\tilde{\ell}$ independent collections of independent observations

$$\left\{\left(\mathbf{X}_1^{(1)}, \mathbf{Y}_1^{(1)}\right), \left(\mathbf{X}_2^{(1)}, \mathbf{Y}_2^{(1)}\right), \ldots\right\}, \ldots, \left\{\left(\mathbf{X}_1^{(\tilde{\ell})}, \mathbf{Y}_1^{(\tilde{\ell})}\right), \left(\mathbf{X}_2^{(\tilde{\ell})}, \mathbf{Y}_1^{(\tilde{\ell})}\right), \ldots\right\}.$$

Let, for $\mathbf{t} \in \mathbb{R}^{d(k_1+\cdots+k_\ell)}$,

$$\begin{aligned} r^{(k,\tilde{\ell})}(\varphi, \mathbf{t}) =& r^{(k,\tilde{\ell})}(\varphi, \mathbf{t}_1, \ldots, \mathbf{t}_\ell) \\ =& \mathbb{E}\left(\varphi\left(\mathbf{Y}_1^{(1)}, \ldots, \mathbf{Y}_{k_1}^{(1)}; \ldots; \mathbf{Y}_1^{(\tilde{\ell})}, \ldots, Y_{k_{\tilde{\ell}}}^{(\tilde{\ell})}\right) \mid \left(\mathbf{X}_1^{(j)}, \ldots, \mathbf{X}_{m_j}^{(j)}\right) = \mathbf{t}_j, j = 1, \ldots, \tilde{\ell}\right), \end{aligned}$$

where φ is assumed, without loss of generality, to be symmetric within each of its $\tilde{\ell}$ blocks of arguments. Corresponding to the "kernel" φ and assuming $n_1 \geq k_1, \ldots, n_k \geq k_{\tilde{\ell}}$, the conditional U-statistic for estimation of $r^{(k,\tilde{\ell})}(\varphi, \mathbf{t})$ is defined, for $\ell \in \{1, 2, 3\}$, by

$$\widehat{r}_{n,\ell}^{(k,\tilde{\ell})}(\varphi, \mathbf{t}) = \frac{\sum_c \varphi\left(\mathbf{Y}_{i_{11}}^{(1)} \ldots, \mathbf{Y}_{i_{1k_1}}^{(1)}; \ldots; \mathbf{Y}_{i_{k_1}}^{(\tilde{\ell})}, \ldots, \mathbf{Y}_{i_{kk_{\tilde{\ell}}}}^{(\tilde{\ell})}\right) \mathbf{K}_\ell\left(\mathbf{X}_{i_{11}}^{(1)} \ldots, \mathbf{X}_{i_{1k_1}}^{(1)}; \ldots; \mathbf{X}_{i_{k_1}}^{(\tilde{\ell})}, \ldots, \mathbf{X}_{i_{\tilde{\ell}k_{\tilde{\ell}}}}^{(\tilde{\ell})}\right)}{\sum_c \mathbf{K}_\ell\left(\mathbf{X}_{i_{11}}^{(1)} \ldots, \mathbf{X}_{i_{1k_1}}^{(1)}; \ldots; \mathbf{X}_{i_{k_1}}^{(\tilde{\ell})}, \ldots, \mathbf{X}_{i_{\tilde{\ell}k_{\tilde{\ell}}}}^{(\tilde{\ell})}\right)},$$

where

$$\mathbf{K}_\ell\left(\mathbf{X}_{i_{11}}^{(1)} \ldots, \mathbf{X}_{i_{1k_1}}^{(1)}; \ldots; \mathbf{X}_{i_{k_1}}^{(\ell)}, \ldots, \mathbf{X}_{i_{\ell k_\ell}}^{(\ell)}\right) = \prod_{j=1}^{\tilde{\ell}} K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{t}_1)}(\mathbf{X}_{i_1}^{(j)}) \ldots K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{t}_m)}(\mathbf{X}_{i_m}^{(j)}).$$

Here $\left\{i_{j1}, \ldots, i_{jm_j}\right\}$ denotes a set of k_j distinct elements of the set $\left\{1, 2, \ldots, n_j\right\}$ $1 \leq j \leq \tilde{\ell}$, and $\sum_c$ denotes summation over all such combinations. The extension of [80] treatment of one-sample U-statistics to the $\tilde{\ell}$ sample case is due to [99] and [54]. One can use Corollaries 3.5, 4.7, or 5.8 to infer that

$$\left|\widehat{r}_{n,\ell}^{(k,\tilde{\ell})}(\varphi, \mathbf{t}) - r^{(k,\tilde{\ell})}(\varphi, \mathbf{t})\right| \longrightarrow 0 \quad a.s. \tag{6.2}$$

6.3. Kendall rank correlation coefficient

To test the independence of one-dimensional random variables Y_1 and Y_2, [90] proposed a method based on the U-statistic K_n with the kernel function:

$$\varphi\left(\left(s_1, t_1\right), \left(s_2, t_2\right)\right) = \mathbf{1}_{\{(s_2-s_1)(t_2-t_1)>0\}} - \mathbf{1}_{\{(s_2-s_1)(t_2-t_1)\leqslant 0\}}. \tag{6.3}$$

Its rejection region is of the form $\left\{\sqrt{n}K_n > \gamma\right\}$. In this example, we consider a multivariate case. To test the conditional independence of $\boldsymbol{\xi}, \boldsymbol{\eta} : Y = (\boldsymbol{\xi}, \boldsymbol{\eta})$ given X, we propose a method based on the conditional U-statistic, for $\ell \in \{1, 2, 3\}$,

$$\widehat{r}_{n,\ell}^{(2)}(\varphi, \mathbf{t}) = \frac{\sum_{i\neq j}^{n} \varphi\left(Y_i, Y_j\right) K_{\mathbf{\Lambda}_{n,\ell}(t_1)}(\mathbf{X}_i) K_{\mathbf{\Lambda}_{n,\ell}(t_2)}(\mathbf{X}_j)}{\sum_{i\neq j}^{n} K_{\mathbf{\Lambda}_{n,\ell}(t_1)}(\mathbf{X}_i) K_{\mathbf{\Lambda}_{n,\ell}(t_2)}(\mathbf{X}_j)},$$

where $\mathbf{t} = (t_1, t_2) \in \mathbb{I} \subset \mathbb{R}^2$ and $\varphi(\cdot)$ is Kendall's kernel (6.3). Suppose that $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are d_1 and d_2-dimensional random vectors, respectively, and $d_1 + d_2 = d$. Furthermore, suppose that $\mathrm{Y}_1, \ldots, \mathrm{Y}_n$ are observations of $(\boldsymbol{\xi}, \boldsymbol{\eta})$, and we are interested in testing :

$$\mathrm{H}_0 : \boldsymbol{\xi} \text{ and } \boldsymbol{\eta} \text{ are conditionally independent given } X. \text{ vs } \mathrm{H}_a : \mathrm{H}_0 \text{ is not true.} \tag{6.4}$$

Let $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2) \in \mathbb{R}^d$ such as $\|\mathbf{a}\| = 1$ and $\mathbf{a}_1 \in \mathbb{R}^{d_1}, \mathbf{a}_2 \in \mathbb{R}^{d_2}$, and $\mathrm{F}(\cdot), \mathrm{G}(\cdot)$ be the distribution functions of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ respectively. Suppose $F^{a_1}(\cdot)$ and $G^{a_2}(\cdot)$ to be continuous for any unit vector $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2)$ where $\mathrm{F}^{\mathbf{a}_1}(t) = \mathbb{P}(\mathbf{a}_1\boldsymbol{\xi} < t)$ and $\mathrm{G}^{\mathbf{a}_2}(t) = \mathbb{P}(\mathbf{a}_2\boldsymbol{\eta} < t)$ and $\mathbf{a}_1^\top$ means the transpose of the vector $\mathbf{a}_i$, $1 \leqslant i \leqslant 2$. For $n = 2$, let $Y^{(1)} = \left(\boldsymbol{\xi}^{(1)}, \boldsymbol{\eta}^{(1)}\right)$ and $Y^{(2)} = \left(\boldsymbol{\xi}^{(2)}, \boldsymbol{\eta}^{(2)}\right)$ such as $\boldsymbol{\xi}^{(i)} \in \mathbb{R}^{d_1}$ and $\boldsymbol{\eta}^{(i)} \in \mathbb{R}^{d_2}$ for $i = 1, 2$, and:

$$\varphi^a\left(\mathrm{Y}^{(1)}, \mathrm{Y}^{(2)}\right) = \varphi\left(\left(\mathbf{a}_1\boldsymbol{\xi}^{(1)}, \mathbf{a}_2\boldsymbol{\eta}^{(1)}\right), \left(\mathbf{a}_1\boldsymbol{\xi}^{(2)}, \mathbf{a}_2\boldsymbol{\eta}^{(2)}\right)\right).$$

An application of Corollaries 3.5, 4.7, or 5.8 gives

$$\left|\widehat{r}_{n,\ell}^{(2)}(\varphi^a, \mathbf{x}; m_n) - r^{(2)}(\varphi^a, \mathbf{x})\right| \longrightarrow 0 \quad a.s. \tag{6.5}$$

7. Examples

Generally speaking, we may take for h any function that has been found interesting in the unconditional setup; cf. [129]. As mentioned before, the case $m = 1$ leads to the Nadaraya-Watson estimator if we set $\varphi(\cdot) = \mathrm{id}; \varphi(\cdot) = \mathbf{1}_{(-\infty,x]}(\cdot)$ yields the empirical conditional d.f. evaluated at x. We now discuss several examples for $m = 2$.

Example 7.1. *[134] For*

$$h(y_1, y_2) = \frac{1}{2}(y_1 - y_2)^2,$$

we obtain

$$m(x_1, x_1) = \mathrm{Var}(Y_1 \mid X_1 = x_1).$$

In this case,

$$\rho^2 = \left\{\mathbb{E}\left[(Y - Y_2)^2 (Y - Y_3)^2 \mid X = X_2 = X_3 = x_1\right] - 4r^{(2)^2}(x_1, x_1)\right\} \int K^2(u)\mathrm{d}u / f(x_1).$$

Compare ρ^2 with ζ_1 in [129], page 182.

Example 7.2. *[134] Assume $Y_i = (Y_{i1}, Y_{i2})^\top$, and define h by*

$$h\left[\begin{bmatrix} y_{11} \\ y_{12} \end{bmatrix}, \begin{bmatrix} y_{21} \\ y_{22} \end{bmatrix}\right] = \frac{1}{2}(y_{11}y_{12} + y_{21}y_{22} - y_{11}y_{22} - y_{12}y_{21}),$$

that is, $m = 2$, and

$$\begin{aligned} r^{(2)}(x_1, x_2) = \frac{1}{2} [&\mathbb{E}(Y_{11}Y_{12} \mid X_1 = x_1) + \mathbb{E}(Y_{21}Y_{22} \mid X_2 = x_2) \\ &-\mathbb{E}(Y_{11}Y_{22} \mid X_1 = x_1, X_2 = x_2) - \mathbb{E}(Y_{12}Y_{21} \mid X_1 = x_1, X_2 = x_2)]. \end{aligned}$$

In particular,

$$r^{(2)}(x_1, x_1) = \mathbb{E}(Y_{11}Y_{12} \mid X_1 = x_1) - \mathbb{E}(Y_{11} \mid X_1 = x_1)\mathbb{E}(Y_{12} \mid X_1 = x_1),$$

the conditional covariance of Y_1 given $X_1 = x_1$.

8. The bandwidth selection criterion

Following [53] and [30], the leave-one-out cross validation procedure allows us to define, for any fixed $\mathbf{i} = (i_1, \ldots, i_m) \in I(m,n)$ and $\ell \in \{1,2,3\}$,

$$\widehat{r}_{n,\ell,\mathbf{i}}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}})) = \frac{\sum_{(j_1,\ldots,j_m)\in I(m,n)\setminus\{\mathbf{i}\}} \varphi(\mathbf{Y}_{j_1}, \ldots, \mathbf{Y}_{j_m}) K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_1)}(\mathbf{X}_{j_1}) \ldots K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_m)}(\mathbf{X}_{j_m})}{\sum_{(i_1,\ldots,i_m)\in I(m,n)} K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_1)}(\mathbf{X}_{j_1}) \ldots K_{\mathbf{\Lambda}_{n,\ell}(\mathbf{x}_m)}(\mathbf{X}_{j_m})}. \tag{8.1}$$

To minimize the quadratic loss function, we introduce the following criterion, we have for some (known) nonnegative weight function $\mathcal{W}(\cdot)$, for $\ell \in \{1,2,3\}$,

$$CV_\ell\left(\varphi, h\right) := \frac{(n-m)!}{n!} \sum_{\mathbf{i}\in I(m,n)} \left(\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}}) - \widehat{r}_{n,\ell,\mathbf{i}}^{(m)}(\varphi, \tilde{\mathbf{X}}_{\mathbf{i}}; h)\right)^2 \widetilde{\mathcal{W}}(\tilde{\mathbf{X}}_{\mathbf{i}}), \tag{8.2}$$

where

$$\widetilde{\mathcal{W}}\left(\tilde{\mathbf{x}}\right) := \prod_{i=1}^{m} \mathcal{W}(\mathbf{x}_i).$$

A natural way for choosing the bandwidth is to minimize the preceding criterion, so let's choose $\widehat{h}_{n,\ell}$ minimizing

$$CV_\ell\left(\varphi, h\right).$$

One can replace (8.2) by

$$CV_\ell\left(\varphi, \widehat{h}_{n,\ell}\right) := \frac{(n-m)!}{n!} \sum_{\mathbf{i}\in I(m,n)} \left(\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}}) - \widehat{r}_{n,\ell,\mathbf{i}}^{(m)}(\varphi, \tilde{\mathbf{X}}_{\mathbf{i}}; \widehat{h}_{n,\ell})\right)^2 \widehat{\mathcal{W}}(\tilde{\mathbf{X}}_{\mathbf{i}}, \tilde{\mathbf{x}}), \tag{8.3}$$

where

$$\widehat{\mathcal{W}}\left(\tilde{\mathbf{s}}, \tilde{\mathbf{x}}\right) := \prod_{i=1}^{m} \widehat{W}(\mathbf{s}_i, \mathbf{x}_i).$$

In practice, one takes for $\mathbf{i} \in I(m,n)$, the uniform global weights $\widetilde{\mathcal{W}}(\tilde{\mathbf{X}}_{\mathbf{i}}) = 1$, and the local weights

$$\widehat{W}(\tilde{\mathbf{X}}_{\mathbf{i}}, \tilde{\mathbf{t}}) = \begin{cases} 1 & \text{if} \quad \|\tilde{\mathbf{X}}_{\mathbf{i}} - \tilde{\mathbf{x}}\| \leq h, \\ 0 & \text{otherwise.} \end{cases}$$

For the sake of brevity, we have just considered the most popular method, that is, the cross-validated selected bandwidth. This may be extended to any other bandwidth selector, such as the bandwidth based on Bayesian ideas [32, 33].

9. Monte Carlo experiments

The program codes are implemented in R. Our simulation setup closely follows that of [49] and [23, 32]. Specifically, we consider drawing i.i.d. random samples $(X_i, Y_{i,1}, Y_{i,2})$ for $i = 1, \ldots, n$, with univariate explanatory variables. We assume that the underlying distribution is continuous. Note that the conditional Kendall-τ, as described in [90], can be defined as follows:

$$\begin{aligned}\tau_{1,2|X=u} &= 4\mathbb{P}(Y_{1,1} > Y_{2,1}, Y_{1,2} > Y_{2,2} \big| X_1 = X_2 = u) - 1 \\ &= 1 - 4\mathbb{P}(Y_{1,1} > Y_{2,1}, Y_{1,2} < Y_{2,2} \big| X_1 = X_2 = u) \\ &= \mathbb{P}((Y_{1,1} - Y_{2,1})(Y_{1,2} - Y_{2,2}) > 0 \big| X_1 = X_2 = u) \\ &\quad -\mathbb{P}((Y_{1,1} - Y_{2,1})(Y_{1,2} - Y_{2,2}) < 0 \big| X_1 = X_2 = u).\end{aligned}$$

Motivated by these expressions, [49] introduced several kernel-based estimators of $\tau_{1,2|X=u}$:

$$\begin{aligned}\hat{\tau}^{(1)}_{1,2|X=u} &:= 4\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}(u)w_{j,n}(u)\mathbf{1}\{Y_{i,1} < Y_{j,1}, Y_{i,2} < Y_{j,2}\} - 1, \\ \hat{\tau}^{(2)}_{1,2|X=u} &:= \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}(u)w_{j,n}(u)\Big(\mathbf{1}\{(Y_{i,1} - Y_{j,1}) \cdot (Y_{i,2} - Y_{j,2}) > 0\} \\ &\quad -\mathbf{1}\{(Y_{i,1} - Y_{j,1}) \cdot (Y_{i,2} - Y_{j,2}) < 0\}\Big), \\ \hat{\tau}^{(3)}_{1,2|X=u} &:= 1 - 4\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}(u)w_{j,n}(u)\mathbf{1}\{Y_{i,1} < Y_{j,1}, Y_{i,2} > Y_{j,2}\},\end{aligned}$$

where $w_{i,n}(\cdot)$ is a sequence of Nadaraya-Watson weights defined by

$$w_{i,n}(u) = \frac{K_h(X_i - u)}{\sum_{j=1}^{n} K_h(X_j - u)}, \tag{9.1}$$

with the notation $K_h(\cdot) := h^{-1}K(\cdot/h)$ for some kernel $K(\cdot)$ on $\mathbb{R}$, and $h = h(n)$ denotes a usual bandwidth sequence that tends to zero as $n \to \infty$. In the simulation study, we consider the following simple case of bounded explanatory variables, as follows:

- The variable X is uniformly on $(0, 1)$. Conditionally on $X = u$, $Y_1|X = u$ and $Y_2|X = u$ both follow a Gaussian distribution $N(u, 1)$. Their associated conditional copula is Gaussian, and their conditional Kendall-τ is given by

$$\tau_{1,2|X=u} = 2u - 1.$$

We make use of the following kernels:

- The tricube kernel:

$$K(x) = \frac{70}{81}\left(1 - |x|^3\right)^3 \mathbf{1}_{|x|\le 1},$$

- the Gaussian kernel:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

- the Epanechnikov kernel [57]:

$$K(x) = \frac{3}{4}(1 - x^2)\mathbf{1}\{|x| \le 1\},$$

- the beta kernel:

$$K_{\alpha_1,\alpha_2}(x) = \frac{1}{B(\alpha_1,\alpha_2)} x^{\alpha_1-1}(1-x)^{\alpha_2-1}\mathbf{1}\{0 \le x \le 1\},$$

with the weight

$$w_{i,n}(u) = \frac{K_{\alpha_1,\alpha_2}(X_i)}{\sum_{j=1}^{n} K_{\alpha_1,\alpha_2}(X_j)}, \tag{9.2}$$

where $\alpha_1 = u/h + 1$ and $\alpha_2 = (1-u)/h + 1$,

- the Bernstein polynomials: For $x \in [0,1], \vartheta \in \mathbb{N}, k = 0,\ldots,\vartheta$, set $b_{k,\vartheta}(x) := \binom{\vartheta}{k} x^k\,(1-x)^{\vartheta-k}$, and

$$K_{u,\vartheta}(x) = \vartheta \sum_{k=0}^{\vartheta-1} \mathbf{1}\left(\frac{k}{\vartheta} < x \le \frac{k+1}{\vartheta}\right) b_{k,\vartheta-1}(u).$$

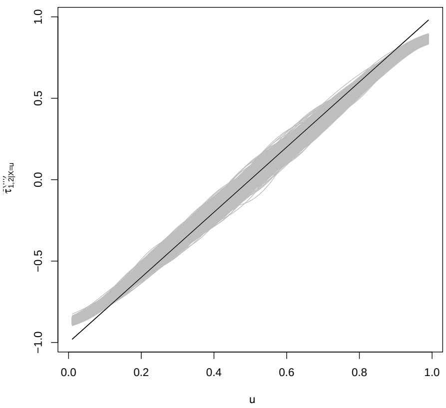
For the estimators based on the tricube kernel, the Gaussian kernel, and the Epanechnikov kernel, we use the "normal scale rule" or the rule-of-thumb method; for instance, see [130], to select the bandwidth. Specifically, we set h to be $\alpha_h \hat{\sigma}(X_1,\ldots,X_n)n^{-1/5}$, where α_h is some positive constant and $\hat{\sigma}(X_1,\ldots,X_n)$ is the empirical standard deviation of X. The parameter α_h is calculated by minimizing the L_2 distance between the kernel density estimator and the theoretical density. The choice of α_h is not the optimal one, since we are choosing this in order to minimize the distance between the densities rather than between the regression functions. This choice is sufficient for our needs. The flexibility of this choice is due to the rule-of-thumb method. For the sake of effective calculations of these measures, the theoretical density can be replaced by the empirical counterparts based, for example, on 10000 simulations. Corollary 4.7 clearly shows that the choice of ϑ is critical for accurately estimating the Kendall-τ. According to [10], it is noted that for the distribution function, ϑ can be as large as $n/\log n$, whereas for density estimation, it is preferable to have $\vartheta = \mathrm{o}(n/\log n)$. In our simulations, we choose $\vartheta = \lceil n/\log n \rceil$, where $\lceil x \rceil$ denotes the ceiling of x. This framework allows us to examine the small sample performances of the estimators $\widehat{\tau}^{(\ell)}_{1,2|X=u}$ for $\ell = 1,2,3$. Consequently, we compute our estimators for each of the kernels mentioned above and for each $n \in \{500, 1000, 2000\}$. We consider three local measures of goodness-of-fit: for a given u and for any Kendall-τ estimate (say $\widehat{\tau}_{1,2|X=u}$), let

- the (local) bias: $\mathrm{Bias}(u) := \mathbb{E}[\widehat{\tau}_{1,2|X=u}] - \widehat{\tau}_{1,2|X=u}$,
- the (local) standard deviation: $\mathrm{Sd}(u) := \mathbb{E}\Big[\big(\widehat{\tau}_{1,2|X=u} - \mathbb{E}[\widehat{\tau}_{1,2|X=u}]\big)^2\Big]^{1/2}$,
- the (local) MSE: $\mathrm{MSE}(u) := \mathbb{E}\Big[\big(\widehat{\tau}_{1,2|X=u} - \tau_{1,2|X=u}\big)^2\Big]$.

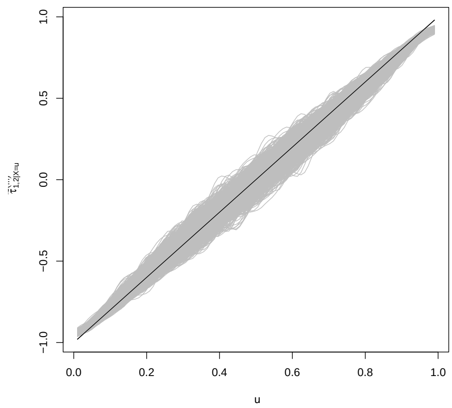
We also consider the integrated versions of these measures with respect to the usual Lebesgue measure over the entire support of t, denoted as $IBias$, ISd, and $IMSE$. For effective computation, we replace the theoretical expectations with their empirical counterparts based on 1000 simulations.

Table 1. Bold values indicate optimal choices for the chosen performance measure. These results are integrated performance measures over the (0, 1) interval.

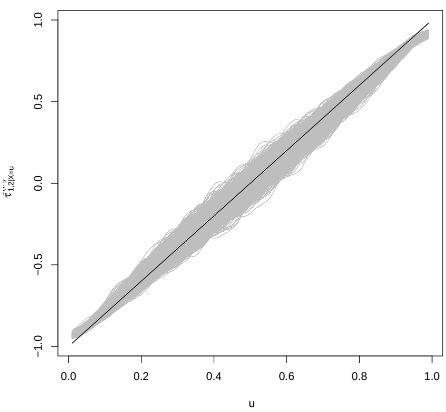
		n = 250			n = 500			n = 1000			n = 2000		
		IBias	ISd	IMSE	IBias	ISd	IMSE	IBias	ISd	IMSE	IBias	ISd	IMSE
Gaussian	$\hat{\tau}^{(1)}_{1,2\|X=\cdot}$	0,0034	-0,0074	0,0091	-0,0026	0,0018	0,0054	-0,0006	0,001	0,0033	0,0006	0,0005	0,0021
	$\hat{\tau}^{(2)}_{1,2\|X=\cdot}$	0,0034	0,0061	0,0091	0,0049	0,0018	0,0054	0,0036	0,001	0,0033	0,003	0,0005	0,0021
	$\hat{\tau}^{(3)}_{1,2\|X=\cdot}$	0,0034	0,0196	0,0094	0,0125	0,0018	0,0055	0,0079	0,001	0,0034	0,0054	0,0005	0,0021
Epanechnikov	$\hat{\tau}^{(1)}_{1,2\|X=\cdot}$	0,0067	-0,0258	0,0091	-0,0143	0,0038	0,0049	-0,0082	0,002	0,0027	-0,0045	0,0011	0,0015
	$\hat{\tau}^{(2)}_{1,2\|X=\cdot}$	0,0067	0,0011	0,0084	0,0008	0,0038	0,0047	0,0004	0,002	0,0026	0,0003	0,0011	0,0015
	$\hat{\tau}^{(3)}_{1,2\|X=\cdot}$	0,0067	0,0281	0,0092	0,016	0,0038	0,0049	0,009	0,002	0,0027	0,0052	0,0011	0,0015
Tricube	$\hat{\tau}^{(1)}_{1,2\|X=\cdot}$	0,008	-0,0309	0,0105	-0,0173	0,0044	0,0055	-0,01	0,0024	0,003	-0,0056	0,0013	0,0016
	$\hat{\tau}^{(2)}_{1,2\|X=\cdot}$	0,0079	0,0007	0,0095	0,0004	0,0044	0,0052	0,0001	0,0024	0,0029	**0,0001**	0,0013	0,0016
	$\hat{\tau}^{(3)}_{1,2\|X=\cdot}$	0,0079	0,0324	0,0106	0,0183	0,0044	0,0056	0,0102	0,0024	0,003	0,0059	0,0013	0,0017
Beta	$\hat{\tau}^{(1)}_{1,2\|X=\cdot}$	0,0028	-0,0004	0,0193	0,005	0,0012	0,0042	0,0062	0,0006	0,0028	0,0068	**0,0003**	0,0019
	$\hat{\tau}^{(2)}_{1,2\|X=\cdot}$	0,0028	0,01	0,0194	0,0115	0,0012	0,0043	0,0096	0,0006	0,0028	0,0087	**0,0003**	0,002
	$\hat{\tau}^{(3)}_{1,2\|X=\cdot}$	0,0028	0,0205	0,0197	0,0179	0,0012	0,0045	0,013	0,0006	0,0029	0,0105	**0,0003**	0,002
Bernstein	$\hat{\tau}^{(1)}_{1,2\|X=\cdot}$	0,0028	0,0452	0,0111	0,0016	0,0298	0,0048	0,001	0,0169	0,0023	0,0006	0,0106	**0,0011**
	$\hat{\tau}^{(2)}_{1,2\|X=\cdot}$	0,0027	0,0589	0,0117	0,0016	0,0387	0,0051	0,001	0,0228	0,0024	0,0006	0,0143	**0,0012**
	$\hat{\tau}^{(3)}_{1,2\|X=\cdot}$	0,0027	0,0726	0,0129	0,0016	0,0476	0,0056	0,001	0,0287	0,0026	0,0006	0,0181	**0,0013**



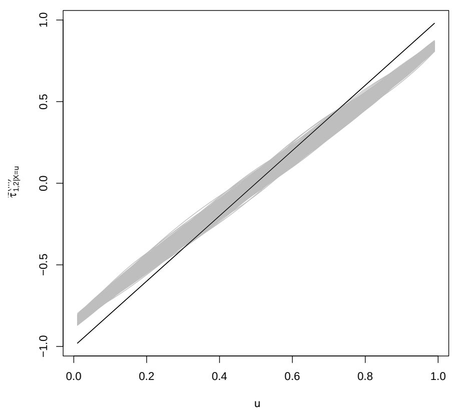
(a) Normal kernel.



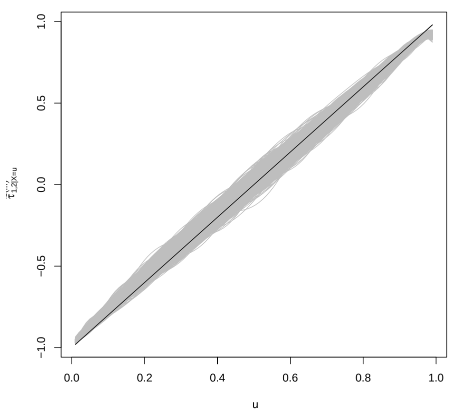
(b) Tricube kernel.



(c) Epanechnikov kernel.

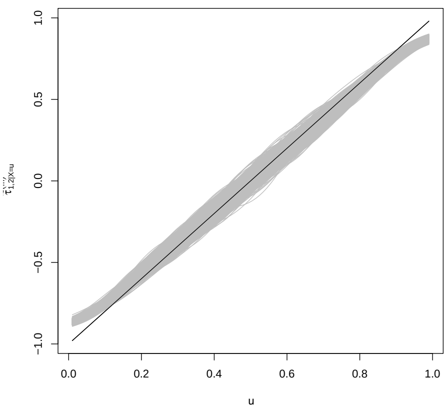


(d) Beta kernel.

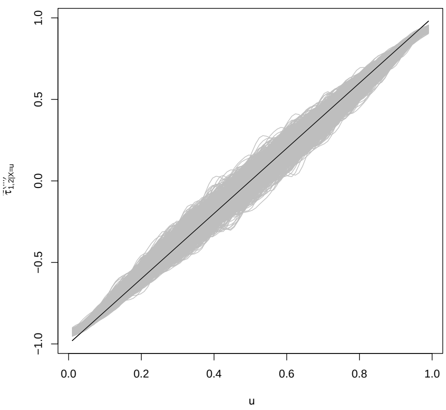


(e) Bernstein kernel.

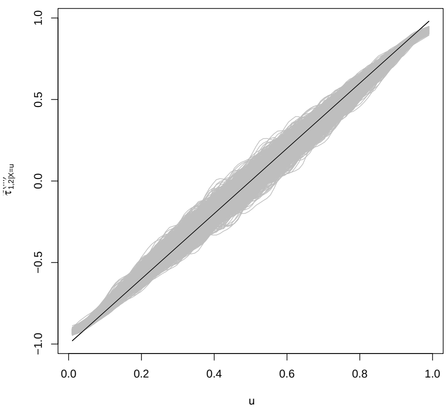
Figure 1. Fitting of $\widehat{\tau}^{(1)}_{1,2|X=\cdot}$ curves based on 1000 simulations for $n = 2000$.



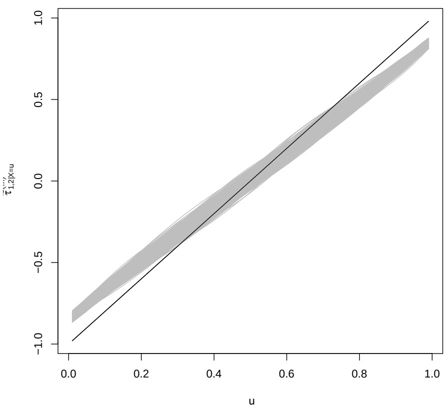
(a) Normal kernel.



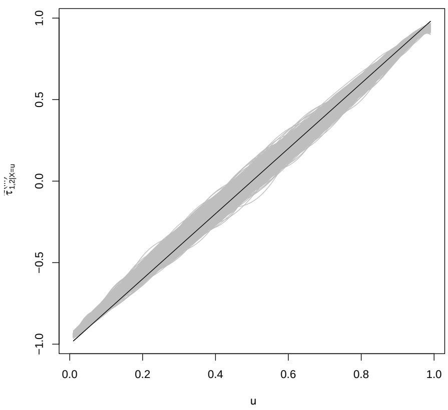
(b) Tricube kernel.



(c) Epanechnikov kernel.

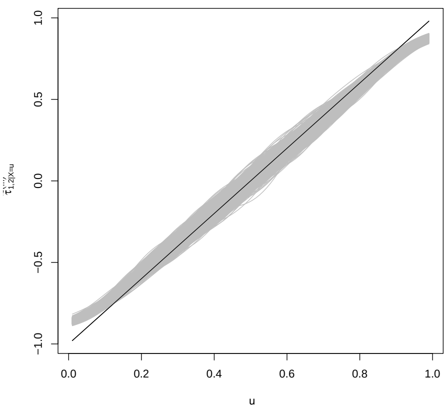


(d) Beta kernel.

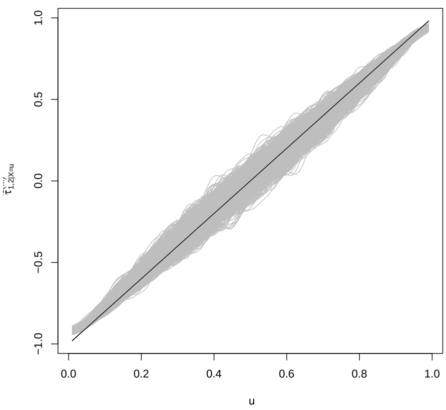


(e) Bernstein kernel.

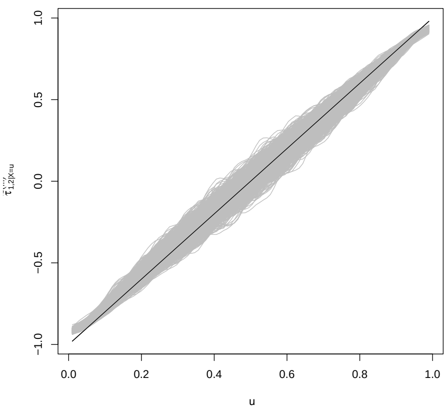
Figure 2. Fitting of $\widehat{\tau}^{(2)}_{1,2|X=\cdot}$ curves based on 1000 simulations for $n = 2000$.



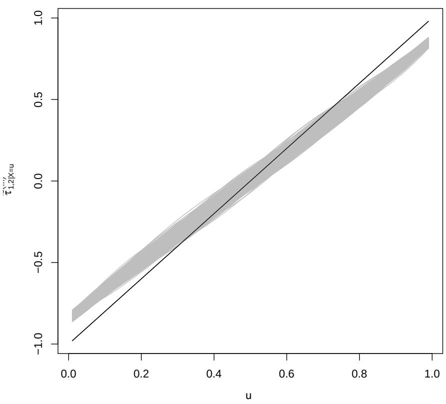
(a) Normal kernel.



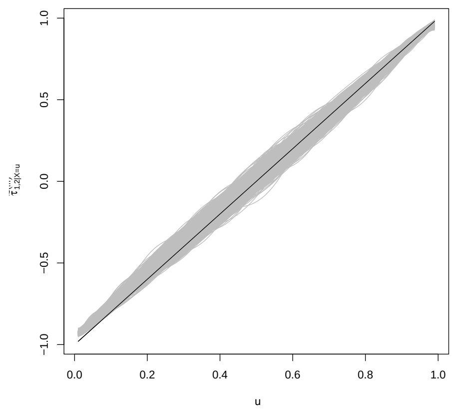
(b) Tricube kernel.



(c) Epanechnikov kernel.

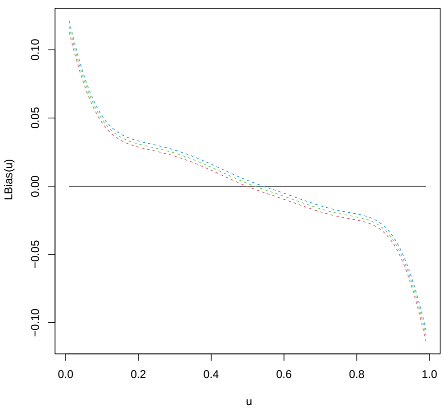


(d) Beta kernel.

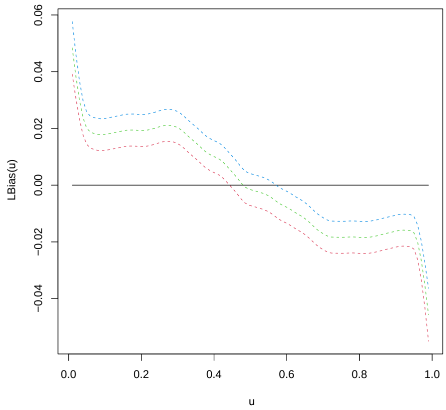


(e) Bernstein kernel.

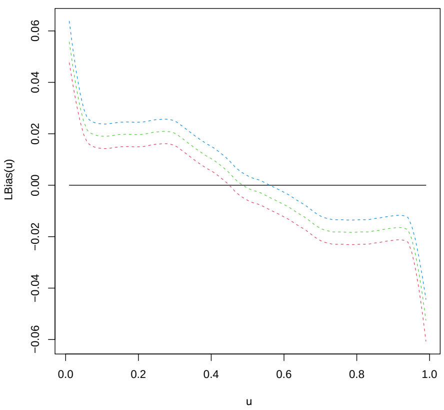
Figure 3. Fitting of $\widehat{\tau}^{(3)}_{1,2|X=\cdot}$ curves based on 1000 simulations for $n = 2000$.



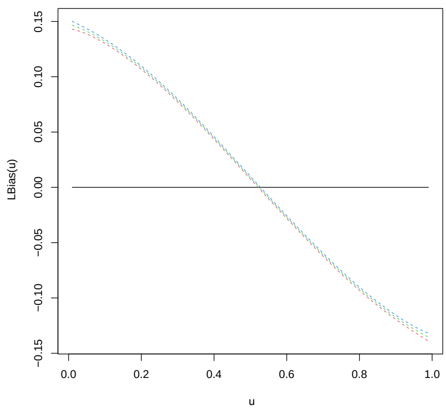
(a) Normal kernel.



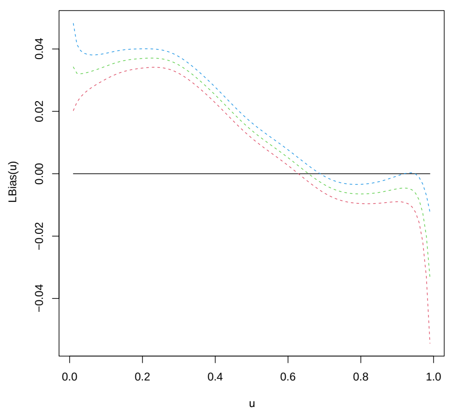
(b) Tricube kernel.



(c) Epanechnikov kernel.

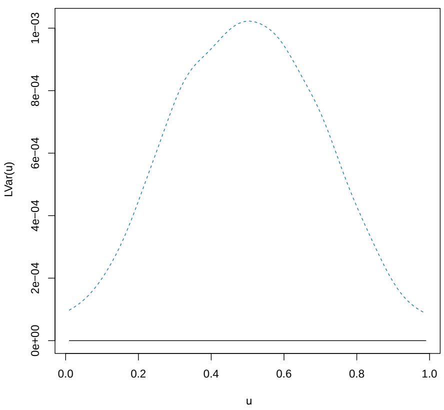


(d) Beta kernel.

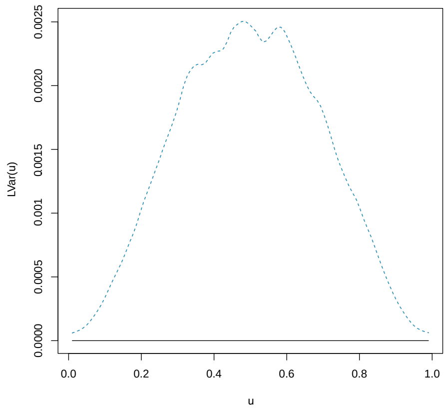


(e) Bernstein kernel.

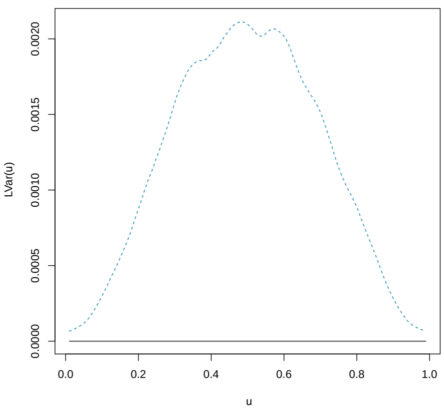
Figure 4. The local bias of $\widehat{\tau}^{(i)}_{1,2|X=\cdot}$, $i = 1, 2, 3$ estimator for $n = 2000$. The used kernel and α_h are: the Gaussian kernel with $\alpha_h = 1.25$ in (a), the Epanechnikov kernel with $\alpha_h = 2.5$ in (b), and the beta kernel with $\alpha_h = 1.5$ in (c).



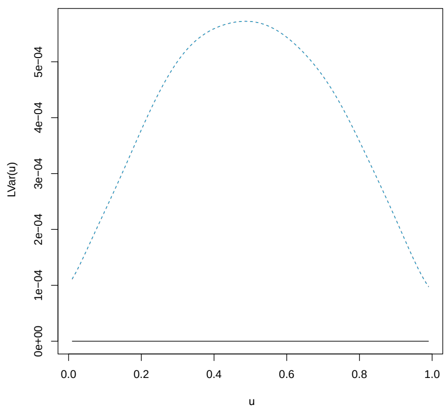
(a) Normal kernel.



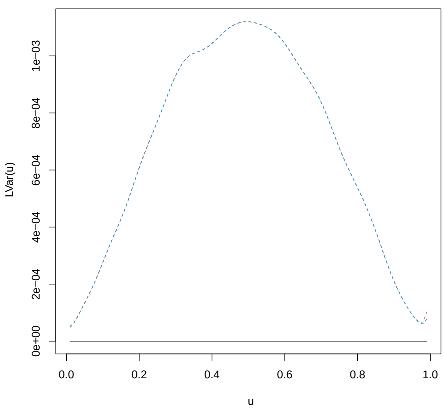
(b) Tricube kernel.



(c) Epanechnikov kernel.

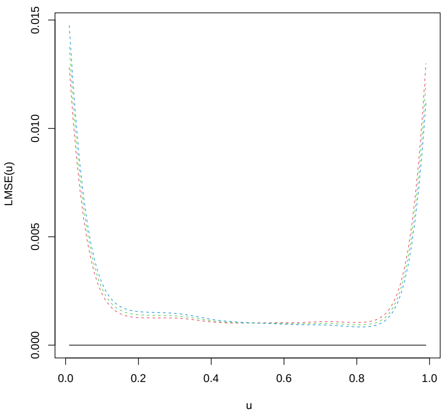


(d) Beta kernel.

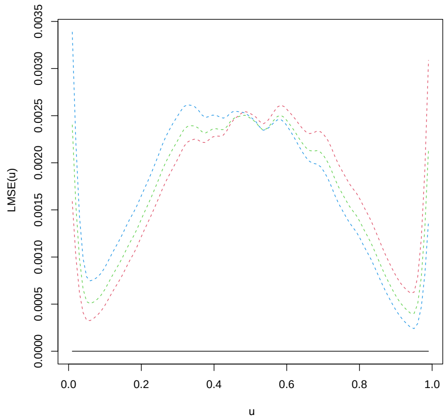


(e) Bernstein kernel.

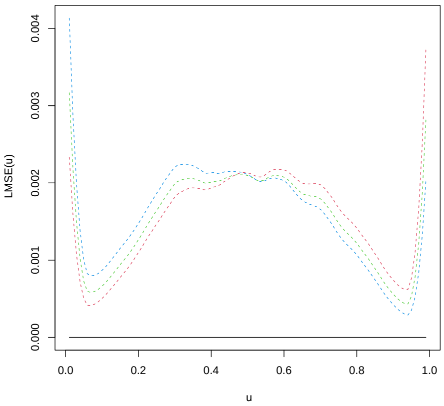
Figure 5. The local standard deviation of $\widehat{\tau}^{(2)}_{1,2|X=\cdot}$ estimator for $n = 2000$ in Setting 1. The used kernel and α_h are: the Gaussian kernel with $\alpha_h = 1.25$ in (a), the Epanechnikov kernel with $\alpha_h = 2.5$ in (b), and the beta kernel with $\alpha_h = 1.5$ in (c).



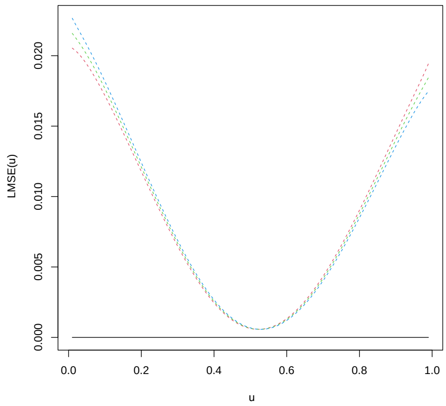
(a) Normal kernel.



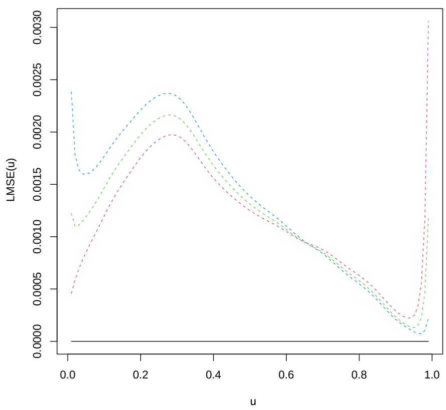
(b) Tricube kernel.



(c) Epanechnikov kernel.



(d) Beta kernel.



(e) Bernstein kernel.

Figure 6. The local MSE $\widehat{\tau}^{(2)}_{1,2|X=\cdot}$ estimator for $n = 2000$ in Setting 1. The used kernel and α_h are: the Gaussian kernel with $\alpha_h = 1.25$ in (a), the Epanechnikov kernel with $\alpha_h = 2.5$ in (b), and the beta kernel with $\alpha_h = 1.5$ in (c).

As is common in inferential contexts, larger sample sizes generally yield better performance, see Figures 1–6. Simple inspection of the results in Table 1 shows that larger sample sizes n lead to smaller $IBias$, ISd, and $IMSE$. The parameters of the beta kernel and Bernstein polynomial can be adjusted such that the mode, median, or mean aligns with the point x. This variable smoothing enables these asymmetric kernel estimators to outperform traditional kernel estimators near the boundary of the support by reducing bias. Additionally, because variable smoothing is directly integrated into the parameterization of the kernel function, asymmetric kernel estimators are generally simpler to implement compared to boundary kernel methods. It is important to note that for $m = 2$, the tendency for larger variance when $x_1 = x_2$ occurs because only data from the single neighborhood of x_1 is used. In contrast, when $x_1 \neq x_2$, data from two potentially disjoint sets are incorporated. Specifically, for Kendall's tau, the case $x_1 = x_2 = u$ is not ideal for minimizing variance. To provide methodological recommendations for using the proposed estimators, it would be beneficial to conduct extensive Monte Carlo experiments comparing our procedures with other alternatives in the literature. However, this is beyond the scope of the present paper.

10. Concluding remarks

Our study initiated exploration into the theoretical aspects of the Dirichlet kernel estimator, as initially proposed by [3], when applied to conditional U-statistics within the dm-dimensional simplex. By extending the unidimensional beta kernel estimator introduced by [40], the Dirichlet kernel estimator effectively mitigates boundary bias. Our comprehensive analysis established both its asymptotic normality and uniform strong consistency. Furthermore, we have considered the estimation of conditional U-statistics using Bernstein polynomials. In deriving our results for these estimators, we also presented novel insights into the Nadaraya-Watson estimators employing Bernstein polynomials, which are of independent interest. Additionally, we introduce a beta kernel estimator specifically designed for conditional U-statistics, providing an extensive suite of uniform consistency results along with associated rates. Our rigorous analysis demonstrated both weak and strong uniform convergence, leveraging the expansion of compact sets and general sequences of smoothing parameters as the analytical foundation. Also, we conducted some simulations to illustrate the small sample performances of the estimators. One aspect that remains unexplored in this paper is the optimal selection of smoothing parameters, a topic of significant importance warranting dedicated research effort, which we defer to a forthcoming investigation.

Several avenues exist for further development of our approach. Extending our results to encompass k-nearest-neighbors estimators is of notable interest, although achieving this goal necessitates the development of new technical arguments, as it currently lies beyond reasonable expectations. Exploring the realm of k-nearest-neighbors estimators would broaden the scope of our research and yield valuable insights into their performance and properties (see [31, 56]). The literature concerning asymmetric kernels for dependent data remains underdeveloped. Extending our results to account for dependence poses a more formidable challenge compared to previous extensions, as it necessitates the formulation of new probabilistic results, given that those employed in our present analysis are tailored specifically to i.i.d. samples (refer to [26, 34]). Change-point detection has become an essential tool for identifying points in a data sequence where a stochastic system undergoes sudden external influences. This can significantly enhance our understanding of the underlying processes.

While change-point analysis has been widely applied to various stochastic processes across numerous scientific fields (refer to [27–29]), its application to conditional U-statistics constitutes an unexplored and challenging research topic that warrants further investigation. Missing data are prevalent in modern statistics, presenting significant challenges across various applications. For example, missing data can occur when information is gathered from sources that measure different variables, such as in healthcare, where patient data may vary between clinics or hospitals. Additional causes of missing data include sensor failure, data censoring, and privacy concerns, among many others. It would be interesting to extend the results of the present paper to the missing data or censored data framework as in [22]. Spatial data, gathered from measurement sites across various disciplines, frequently appears in research fields such as econometrics, epidemiology, environmental science, image analysis, oceanography, meteorology, geostatistics, and many others. Extending this paper to effectively handle spatial data is a challenging task. Our results will be expanded in future investigations by examining the weak convergence of the conditional U-processes. This will require additional effort and advanced techniques to establish tightness.

Author contributions

Salim Bouzebda, Amel Nezzal and Issam Elhattab: Conceptualization, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft, writing – review & editing. All authors have read and approved the final version of the manuscript for publication

Acknowledgements

The authors extend their sincere gratitude to the Editor-in-Chief, the Associate Editor, and the five referees for their invaluable feedback. Their insightful comments have greatly refined and focused the original work, resulting in a markedly improved presentation.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Salim Bouzebda is the(a) Guest Editor of special issue "Advances in Statistical Inference and Stochastic Processes: Theory and Applications" for AIMS Mathematics. Prof. Salim Bouzebda was not involved in the editorial review and the decision to publish this article.

References

1. A. Abadie, G. W. Imbens, Large sample properties of matching estimators for average treatment effects, *Econometrica*, **74** (2006), 235–267. https://doi.org/10.1111/j.1468-0262.2006.00655.x
2. S. Abrams, P. Janssen, J. Swanepoel, N. Veraverbeke, Nonparametric estimation of risk ratios for bivariate data, *J. Nonparametr. Stat.*, **34** (2022), 940–963. https://doi.org/10.1080/10485252.2022.2085265

3. J. Aitchison, I. J. Lauder, Kernel density estimation for compositional data, *J. Roy. Statist. Soc. Ser. C*, **34** (1985), 129–137. https://doi.org/10.2307/2347365

4. M. A. Arcones, The asymptotic accuracy of the bootstrap of U-quantiles, *Ann. Statist.*, **23** (1995), 1802–1822. https://doi.org/10.1214/aos/1176324324

5. M. A. Arcones, A Bernstein-type inequality for U-statistics and U-processes, *Statist. Probab. Lett.*, **22** (1995), 239–247. https://doi.org/10.1016/0167-7152(94)00072-G

6. M. A. Arcones, The Bahadur-Kiefer representation for U-quantiles, *Ann. Statist.*, **24** (1996), 1400–1422. https://doi.org/10.1214/aos/1032526976

7. M. A. Arcones, E. Giné, Limit theorems for U-processes, *Ann. Probab.*, **21** (1993), 1494–1542. https://doi.org/10.1214/aop/1176989128

8. M. A. Arcones, Y. Wang, Some new tests for normality based on U-processes, *Statist. Probab. Lett.*, **76** (2006), 69–82. https://doi.org/10.1016/j.spl.2005.07.003

9. G. J. Babu, Y. P. Chaubey, Smooth estimation of a distribution and density function on a hypercube using Bernstein polynomials for dependent random vectors, *Statist. Probab. Lett.*, **76** (2006), 959–969. https://doi.org/10.1016/j.spl.2005.10.031

10. G. J. Babu, A. J. Canty, Y. P. Chaubey, Application of Bernstein polynomials for smooth estimation of a distribution and density function, *J. Statist. Plann. Inference*, **105** (2002), 377–392. https://doi.org/10.1016/S0378-3758(01)00265-8

11. M. Belalia, On the asymptotic properties of the Bernstein estimator of the multivariate distribution function, *Statist. Probab. Lett.*, **110** (2016), 249–256. https://doi.org/10.1016/j.spl.2015.10.004

12. M. Belalia, T. Bouezmarni, F. C. Lemyre, A. Taamouti, Testing independence based on Bernstein empirical copula and copula density, *J. Nonparametr. Stat.*, **29** (2017), 346–380. https://doi.org/10.1080/10485252.2017.1303063

13. D. Z. Bello, M. Valk, G. B. Cybis, Towards U-statistics clustering inference for multiple groups, *J. Stat. Comput. Simul.*, **94** (2024), 204–222. https://doi.org/10.1080/00949655.2023.2239978

14. N. Berrahou, S. Bouzebda, L. Douge, Functional Uniform-in-Bandwidth moderate deviation principle for the local empirical processes involving functional data, *Math. Meth. Statist.*, **33** (2024), 26–69. https://doi.org/10.3103/S1066530724700030

15. N. Berrahou, S. Bouzebda, L. Douge, A nonparametric distribution-free test of independence among continuous random vectors based on L_1-norm, *arXiv:2105.02164v3*, 2024.

16. K. Bertin, N. Klutchnikoff, Minimax properties of beta kernel estimators, *J. Statist. Plann. Inference*, **141** (2011), 2287–2297. https://doi.org/10.1016/j.jspi.2011.01.009

17. K. Bertin, C. Genest, N. Klutchnikoff, F. Ouimet, Minimax properties of Dirichlet kernel density estimators, *J. Multivariate Anal.*, **195** (2023), 105158. https://doi.org/10.1016/j.jmva.2023.105158

18. S. Borovkova, R. Burton, H. Dehling, Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation, *Trans. Amer. Math. Soc.*, **353** (2001), 4261–4318.

19. Y. V. Borovskikh, *U-statistics in Banach spaces*, Boston: De Gruyter, 1996. https://doi.org/10.1515/9783112313954

20. T. Bouezmarni, J.-M. Rolin, Consistency of the beta kernel density function estimator, *Canad. J. Statist.*, **31** (2003), 89–98. https://doi.org/10.2307/3315905

21. T. Bouezmarni, J. V. K. Rombouts, Nonparametric density estimation for multivariate bounded data, *J. Statist. Plann. Inference*, **140** (2010), 139–152. https://doi.org/10.1016/j.jspi.2009.07.013

22. T. Bouezmarni, F. C. Lemyre, A. El Ghouch, Estimation of a bivariate conditional copula when a variable is subject to random right censoring, *Electron. J. Stat.*, **13** (2019), 5044–5087. https://doi.org/10.1214/19-EJS1645

23. S. Bouzebda, General tests of conditional independence based on empirical processes indexed by functions, *Jpn. J. Stat. Data Sci.*, **6** (2023), 115–177. https://doi.org/10.1007/s42081-023-00193-3

24. S. Bouzebda, Limit theorems in the nonparametric conditional single-index U-processes for locally stationary functional random fields under stochastic sampling design, *Mathematics*, **12** (2024), 1996. https://doi.org/10.3390/math12131996

25. S. Bouzebda, Weak convergence of the conditional single index U-statistics for locally stationary functional time series, *AIMS Mathematics*, **9** (2024), 14807–14898. https://doi.org/10.3934/math.2024720

26. S. Bouzebda, S. Didi, Additive regression model for stationary and ergodic continuous time processes, *Comm. Statist. Theory Methods*, **46** (2017), 2454–2493. https://doi.org/10.1080/03610926.2015.1048882

27. S. Bouzebda, A. A. Ferfache, Asymptotic properties of M-estimators based on estimating equations and censored data in semi-parametric models with multiple change points, *J. Math. Anal. Appl.*, **497** (2021), 124883. https://doi.org/10.1016/j.jmaa.2020.124883

28. S. Bouzebda, A. Keziou, A new test procedure of independence in copula models via χ^2-divergence, *Comm. Statist. Theory Methods*, **39** (2010), 1–20. https://doi.org/10.1080/03610920802645379

29. S. Bouzebda, A. Keziou, A semiparametric maximum likelihood ratio test for the change point in copula models, *Stat. Methodol.*, **14** (2013), 39–61. https://doi.org/10.1016/j.stamet.2013.02.003

30. S. Bouzebda, A. Nezzal, Asymptotic properties of conditional U-statistics using delta sequences, *Comm. Statist. Theory Methods*, **53** (2024), 4602–4657. https://doi.org/10.1080/03610926.2023.2179887

31. S. Bouzebda, A. Nezzal, Uniform in number of neighbors consistency and weak convergence of kNN empirical conditional processes and kNN conditional U-processes involving functional mixing data, *AIMS Mathematics*, **9** (2024), 4427–4550. https://doi.org/10.3934/math.2024218

32. S. Bouzebda, N. Taachouche, On the variable bandwidth kernel estimation of conditional U-statistics at optimal rates in sup-norm, *Phys. A*, **625** (2023), 129000. https://doi.org/10.1016/j.physa.2023.129000

33. S. Bouzebda, N. Taachouche, Rates of the strong uniform consistency with rates for conditional U-statistics estimators with general kernels on manifolds, *Math. Meth. Stat.*, **33** (2024), 95 –153. https://doi.org/10.3103/S1066530724700066

34. S. Bouzebda, S. Didi, L. El Hajj, Multivariate wavelet density and regression estimators for stationary and ergodic continuous time processes: Asymptotic results, *Math. Meth. Stat.*, **24** (2015), 163–199. https://doi.org/10.3103/S1066530715030011

35. S. Bouzebda, A. Nezzal, T. Zari, Uniform consistency for functional conditional U-statistics using delta-sequences, *Mathematics*, **11** (2023), 161. https://doi.org/10.3390/math11010161

36. B. M. Brown, S. X. Chen, Beta-Bernstein smoothing for regression curves with compact support, *Scand. J. Statist.*, **26** (1999), 47–59. https://doi.org/10.1111/1467-9469.00136

37. Q. Cao, Z. C. Guo, Y. Ying, Generalization bounds for metric and similarity learning, *Mach. Learn.*, **102** (2016), 115–132. https://doi.org/10.1007/s10994-015-5499-7

38. A. Charpentier, A. Oulidi, Beta kernel quantile estimators of heavy-tailed loss distributions, *Stat. Comput.*, **20** (2010), 35–55. https://doi.org/10.1007/s11222-009-9114-2

39. L. Chen, A. T. K. Wan, S. Zhang, Y. Zhou, Distributed algorithms for U-statistics-based empirical risk minimization, *J. Mach. Learn. Res.*, **24** (2023), 1–43.

40. S. X. Chen, Beta kernel estimators for density functions, *Comput. Statist. Data Anal.*, **31** (1999), 131–145. https://doi.org/10.1016/S0167-9473(99)00010-9

41. S. X. Chen, Beta kernel smoothers for regression curves, *Statist. Sinica*, **10** (2000), 73–91.

42. S. X. Chen, Probability density function estimation using gamma kernels, *Ann. Inst. Statist. Math.*, **52** (2000), 471–480. https://doi.org/10.1023/A:1004165218295

43. T. C. Cheng, A. Biswas, Maximum trimmed likelihood estimator for multivariate mixed continuous and categorical data, *Comput. Statist. Data Anal.*, **52** (2008), 2042–2065. https://doi.org/10.1016/j.csda.2007.06.026

44. R. F. Cintra, M. Valk, D. M. Filho, A model-free-based control chart for batch process using U-statistics, *J. Process Contr.*, **132** (2023), 103097. https://doi.org/10.1016/j.jprocont.2023.103097

45. G. B. Cybis, M. Valk, S. R. C. Lopes, Clustering and classification problems in genetics through U-statistics, *J. Stat. Comput. Simul.*, **88** (2018), 1882–1902. https://doi.org/10.1080/00949655.2017.1374387

46. V. H. de la Peña, E. Giné, Decoupling, In: *Probability and its applications*, New York: Springer, 1999. https://doi.org/10.1007/978-1-4612-0537-1

47. H. Dehling, R. Fried, Asymptotic distribution of two-sample empirical U-quantiles with applications to robust tests for shifts in location, *J. Multivariate Anal.*, **105** (2012), 124–140. https://doi.org/10.1016/j.jmva.2011.08.014

48. M. Denker, G. Keller, On U-statistics and v.mise'statistics for weakly dependent processes, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete.*, **64** (1983), 505–522. https://doi.org/10.1007/BF00534953

49. A. Derumigny, J. D. Fermanian, On kernel-based estimation of conditional Kendall's tau: Finite-distance bounds and asymptotic behavior, *Depend. Model.*, **7** (2019), 292–321. https://doi.org/10.1515/demo-2019-0016

50. A. Desgagné, C. Genest, F. Ouimet, Asymptotics for non-degenerate multivariate U-statistics with estimated nuisance parameters under the null and local alternative hypotheses, *arXiv:2401.11272*, 2024. https://doi.org/10.48550/arXiv.2401.11272

51. L. Devroye, *A course in density estimation*, Birkhauser Boston Inc., 1978.
52. L. Devroye, C. S. Penrod, Distribution-free lower bounds in density estimation, *Ann. Statist.*, **12** (1984), 1250–1262. https://doi.org/10.1214/aos/1176346790
53. J. Dony, D. M. Mason, Uniform in bandwidth consistency of conditional U-statistics, *Bernoulli*, **14** (2008), 1108–1133. https://doi.org/10.3150/08-BEJ136
54. M. Dwass, The large-sample power of rank order tests in the two-sample problem, *Ann. Math. Statist.*, **27** (1956), 352–374.
55. P. P. B. Eggermont, V. N. LaRiccia, Maximum penalized likelihood estimation, In: *Springer series in statistics*, New York: Springer, 2001. https://doi.org/10.1007/978-1-0716-1244-6
56. Z. C. Elmezouar, F. Alshahrani, I. M. Almanjahie, S. Bouzebda, Z. Kaid, A. Laksaci, Strong consistency rate in functional single index expectile model for spatial data, *AIMS Mathematics*, **9** (2024), 5550–5581. https://doi.org/10.3934/math.2024269
57. V. A. Epanechnikov, Non-parametric estimation of a multidimensional probability density, *Theory Probab. Appl.*, **14** (1969), 153–158.
58. L. Faivishevsky, J. Goldberger, Ica based on a smooth estimation of the differential entropy, In: *Proceedings of the 21st international conference on neural information processing systems*, Curran Associates Inc., 2008, 433–440.
59. A. A. Filippova, Mises theorem on the limit behaviour of functionals derived from empirical distribution functions, *Dokl. Akad. Nauk SSSR*, **129** (1959), 44–47.
60. E. W. Frees, Infinite order U-statistics, *Scand. J. Statist.*, 1989, 29–45.
61. B. Funke, M. Hirukawa, Bias correction for local linear regression estimation using asymmetric kernels via the skewing method, *Econom. Stat.*, **20** (2021), 109–130. https://doi.org/10.1016/j.ecosta.2020.01.004
62. B. Funke, M. Hirukawa, Density derivative estimation using asymmetric kernels, *J. Nonparametr. Stat.*, 2023, 1–24. https://doi.org/10.1080/10485252.2023.2291430
63. T. Gasser, H. G. Müller, Kernel estimation of regression functions, In: *Lecture notes in mathematics*, Berlin, Heidelberg: Springer, **757** (2006), 23–68. https://doi.org/10.1007/BFb0098489
64. W. Gawronski, Strong laws for density estimators of Bernstein type, *Period. Math. Hung.*, **16** (1985), 23–43. https://doi.org/10.1007/BF01855801
65. W. Gawronski, U. Stadtmüller, On density estimation by means of Poisson's distribution, *Scand. J. Statist.*, **7** (1980), 90–94.
66. S. S. Ghannadpour, S. E. Kalkhoran, H. Jalili, M. Behifar, Delineation of mineral potential zone using u-statistic method in processing satellite remote sensing images, *Int. J. Mining Geo-Eng.*, **57** (2023), 445–453. https://doi.org/10.22059/IJMGE.2023.364690.595097
67. S. Ghosal, A. Sen, A. W. van der Vaart, Testing monotonicity of regression, *Ann. Statist.*, **28** (2000), 1054–1082. https://doi.org/10.1214/aos/1015956707
68. E. Giné, D. M. Mason, Laws of the iterated logarithm for the local U-statistic process, *J. Theor. Probab.*, **20** (2007), 457–485. https://doi.org/10.1007/s10959-007-0067-0

69. Z. Guan, Efficient and robust density estimation using Bernstein type polynomials, *J. Nonparametr. Stat.*, **28** (2016), 250–271. https://doi.org/10.1080/10485252.2016.1163349
70. E. Guerre, I. Perrigne, Q. Vuong, Optimal nonparametric estimation of first-price auctions, *Econometrica*, **68** (2000), 525–574. https://doi.org/10.1111/1468-0262.00123
71. P. R. Halmos, The theory of unbiased estimation, *Ann. Math. Statist.*, **17** (1946), 34–43. https://doi.org/10.1214/aoms/1177731020
72. B. E. Hansen, Uniform convergence rates for kernel estimation with dependent data, *Econometric Theory*, **24** (2008), 726–748.
73. W. Härdle, *Applied nonparametric regression*, Cambridge University Press, 1990. https://doi.org/10.1017/CCOL0521382483
74. M. Harel, M. L. Puri, Conditional U-statistics for dependent random variables, In: *Probability theory and extreme value theory*, New York: De Gruyter Mouton, **2** (2003), 533–549. https://doi.org/10.1515/9783110917826.533
75. C. Heilig, D. Nolan, Limit theorems for the infinite-degree U-process, *Statist. Sinica*, **11** (2001), 289–302.
76. R. Helmers, M. Hušková, Bootstrapping multivariate U-quantiles and related statistics, *J. Multivariate Anal.*, **49** (1994), 97–109. https://doi.org/10.1006/jmva.1994.1016
77. M. Hirukawa, *Asymmetric kernel smoothing: Theory and applications in economics and finance*, Singapore: Springer, 2018. https://doi.org/10.1007/978-981-10-5466-2
78. M. Hirukawa, M. Sakudo, Another bias correction for asymmetric kernel density estimation with a parametric start, *Statist. Probab. Lett.*, **145** (2019), 158–165. https://doi.org/10.1016/j.spl.2018.09.002
79. M. Hirukawa, I. Murtazashvili, A. Prokhorov, Uniform convergence rates for nonparametric estimators smoothed by the beta kernel, *Scand. J. Stat.*, **49** (2022), 1353–1382. https://doi.org/10.1111/sjos.12573
80. W. Hoeffding, A class of statistics with asymptotically normal distribution, *Ann. Math. Statist.*, **19** (1948), 293–325. https://doi.org/10.1214/aoms/1177730196
81. B. Huang, Y. Liu, L. Peng, Distributed inference for two-sample U-statistics in massive data analysis, *Scand. J. Stat.*, **50** (2023), 1090–1115. https://doi.org/10.1111/sjos.12620
82. G. Igarashi, Bias reductions for beta kernel estimation, *J. Nonparametr. Stat.*, **28** (2016), 1–30. https://doi.org/10.1080/10485252.2015.1112011
83. G. Igarashi, Y. Kakizawa, Limiting bias-reduced Amoroso kernel density estimators for non-negative data, *Comm. Statist. Theory Methods*, **47** (2018), 4905–4937. https://doi.org/10.1080/03610926.2017.1380832
84. G. Igarashi, Y. Kakizawa, Multiplicative bias correction for asymmetric kernel density estimators revisited, *Comput. Statist. Data Anal.*, **141** (2020), 40–61. https://doi.org/10.1016/j.csda.2019.06.010
85. S. Janson, A functional limit theorem for random graphs with applications to subgraph count statistics, *Random Structures Algorithms*, **1** (1990), 15–37. https://doi.org/10.1002/rsa.3240010103

86. S. Janson, Asymptotic normality for *m*-dependent and constrained *U*-statistics, with applications to pattern matching in random strings and permutations, *Adv. Appl. Probab.*, **55** (2023), 841–894. https://doi.org/10.1017/apr.2022.51

87. E. Joly, G. Lugosi, Robust estimation of *U*-statistics, *Stochastic Process. Appl.*, **126** (2016), 3760–3773. https://doi.org/10.1016/j.spa.2016.04.021

88. M. C. Jones, Corrigendum: "Variable kernel density estimates and variable kernel density estimates" [Austral. J. Statist. **32** (1991). 3, 361–371], *Austral. J. Statist.*, **33** (1991), 119. https://doi.org/10.1111/j.1467-842X.1991.tb00418.x

89. Y. Kakizawa, Bernstein polynomial probability density estimation, *J. Nonparametr. Stat.*, **16** (2004), 709–729. https://doi.org/10.1080/1048525042000191486

90. M. G. Kendall, A new measure of rank correlation, *Biometrika*, **30** (1938), 81–93.

91. I. Kim, A. Ramdas, Dimension-agnostic inference using cross U-statistics, *Bernoulli*, **30** (2024), 683–711. https://doi.org/10.3150/23-BEJ1613

92. V. S. Koroljuk, Y. V. Borovskich, Theory of *U*-statistics, In: *Mathematics and its applications*, Dordrecht: Springer, **273** (2013). https://doi.org/10.1007/978-94-017-3515-5

93. S. Kotz, N. Balakrishnan, N. L. Johnson, *Continuous multivariate distributions: Models and applications*, John Wiley & Sons, Inc., 2000.

94. D. Kristensen, Uniform convergence rates of kernel estimators with heterogeneous dependent data, *Econometric Theory*, **25** (2009), 1433–1445. https://doi.org/10.1017/S0266466609090744

95. T. Le Minh, *U*-statistics on bipartite exchangeable networks, *ESAIM Probab. Stat.*, **27** (2023), 576–620. https://doi.org/10.1051/ps/2023010

96. A. Leblanc, On estimating distribution functions using Bernstein polynomials, *Ann. Inst. Stat. Math.*, **64** (2012), 919–943. https://doi.org/10.1007/s10463-011-0339-4

97. A. Leblanc, On the boundary properties of Bernstein polynomial estimators of density and distribution functions, *J. Statist. Plann. Inference*, **142** (2012), 2762–2778. https://doi.org/10.1016/j.jspi.2012.03.016

98. A. J. Lee, *U-statistics: Theory and practice*, New York: Routledge, 1990. https://doi.org/10.1201/9780203734520

99. E. L. Lehmann, A general concept of unbiasedness, *Ann. Math. Statist.*, **22** (1951), 587–592. https://doi.org/10.1214/aoms/1177729549

100. E. L. Lehmann, Elements of large-sample theory, In: *Springer texts in statistics*, New York: Springer, 1999. https://doi.org/10.1007/b98855

101. C. Y. Leung, The effect of across-location heteroscedasticity on the classification of mixed categorical and continuous data, *J. Multivariate Anal.*, **84** (2003), 369–386. https://doi.org/10.1016/S0047-259X(02)00057-X

102. H. Li, C. Ren, L. Li, *U*-processes and preference learning, *Neural Comput.*, **26** (2014), 2896–2924. https://doi.org/10.1162/NECO_a_00674

103. F. Lim, V. M. Stojanovic, On U-statistics and compressed sensing I: Non-asymptotic average-case analysis, *IEEE Trans. Signal Process.*, **61** (2013), 2473–2485. https://doi.org/10.1109/TSP.2013.2247598

104. R. J. A. Little, M. D. Schluchter, Maximum likelihood estimation for mixed continuous and categorical data with missing values, *Biometrika*, **72** (1985), 497–512. https://doi.org/10.1093/biomet/72.3.497

105. B. Liu, S. K. Ghosh, On empirical estimation of mode based on weakly dependent samples, *Comput. Statist. Data Anal.*, **152** (2020), 107046. https://doi.org/10.1016/j.csda.2020.107046

106. C. Liu, D. B. Rubin, Ellipsoidally symmetric extensions of the general location model for mixed categorical and continuous data, *Biometrika*, **85** (1998), 673–688.

107. Q. Liu, J. Lee, M. Jordan, A kernelized stein discrepancy for goodness-of-fit tests, In:*Proceedings of the 33rd international conference on international conference on machine learning*, **48** (2016), 276–284.

108. D. Lu, L. Wang, J. Yang, The stochastic convergence of Bernstein polynomial estimators in a triangular array, *J. Nonparametr. Stat.*, **34** (2022), 987–1014. https://doi.org/10.1080/10485252.2022.2107643

109. L. Lu, On the uniform consistency of the Bernstein density estimator, *Statist. Probab. Lett.*, **107** (2015), 52–61. https://doi.org/10.1016/j.spl.2015.08.004

110. H. G. Müller, Nonparametric regression analysis of longitudinal data, In: *Lecture notes in statistics*, New York: Springer, **46** (1988). https://doi.org/10.1007/978-1-4612-3926-0

111. H. G. Müller, Smooth optimum kernel estimators near endpoints, *Biometrika*, **78** (1991), 521–530. https://doi.org/10.1093/biomet/78.3.521

112. E. A. Nadaraja, On a regression estimate, *Teor. Verojatnost. Primenen.*, **9** (1964), 157–159.

113. E. A. Nadaraya, Nonparametric estimation of probability densities and regression curves, In: *Mathematics and its applications*, Dordrecht: Springer, **20** (1989). https://doi.org/10.1007/978-94-009-2583-0

114. K. W. Ng, G. L. Tian, M. L. Tang, *Dirichlet and related distributions: Theory, methods and applications*, John Wiley & Sons, Ltd., 2011. Chichester. Theory, methods and applications. (2011)

115. F. Ouimet, Asymptotic properties of Bernstein estimators on the simplex, *J. Multivariate Anal.*, **185** (2021), 104784. https://doi.org/10.1016/j.jmva.2021.104784

116. F. Ouimet, On the boundary properties of Bernstein estimators on the simplex, *Open Statist.*, **3** (2022), 48–62. https://doi.org/10.1515/stat-2022-0111

117. F. Ouimet, R. Tolosana-Delgado, Asymptotic properties of Dirichlet kernel density estimators, *J. Multivariate Anal.*, **187** (2022), 104832. https://doi.org/10.1016/j.jmva.2021.104832

118. W. Peng, T. Coleman, L. Mentch, Rates of convergence for random forests via generalized U-statistics, *Electron. J. Statist.*, **16** (2022), 232–292. https://doi.org/10.1214/21-EJS1958

119. B. L. S. P. Rao, *Nonparametric functional estimation*, Academic Press, 1983. https://doi.org/10.1016/C2013-0-11326-8

120. B. L. S. P. Rao, Estimation of distribution and density functions by generalized Bernstein polynomials, *Indian J. Pure Appl. Math.*, **36** (2005), 63–88.

121. B. L. S. P. Rao, A. Sen, Limit distributions of conditional U-statistics, *J. Theor. Probab.*, **8** (1995), 261–301. https://doi.org/10.1007/BF02212880

122. R. H. Randles, On the asymptotic normality of statistics with estimated parameters, *Ann. Statist.*, **10** (1982), 462–474. https://doi.org/10.1214/aos/1176345787

123. O. Renault, O. Scaillet, On the way to recovery: A nonparametric bias free estimation of recovery rate densities, *J. Bank. Financ.*, **28** (2004), 2915–2931. https://doi.org/10.1016/j.jbankfin.2003.10.018

124. G. G. Roussas, Estimation of transition distribution function and its quantiles in Markov processes: strong consistency and asymptotic normality, In: *Roussas, G. (eds) Nonparametric functional estimation and related topics. NATO ASI Series*, Dordrecht: Springer, **335** (1991). https://doi.org/10.1007/978-94-011-3222-0_34

125. A. Sancetta, S. Satchell, The Bernstein copula and its applications to modeling and approximations of multivariate distributions, *Econometric Theory*, **20** (2004), 535–562.

126. A. Schick, Y. Wang, W. Wefelmeyer, Tests for normality based on density estimators of convolutions, *Statist. Probab. Lett.*, **81** (2011), 337–343. https://doi.org/10.1016/j.spl.2010.10.022

127. E. F. Schuster, Incorporating support constraints into nonparametric estimators of densities, *Comm. Statist. Theory Methods*, **14** (1985), 1123–1136. https://doi.org/10.1080/03610928508828965

128. A. Sen, Uniform strong consistency rates for conditional U-statistics, *Sankhyā*, **56** (1994), 179–194.

129. R. J. Serfling, *Approximation theorems of mathematical statistics*, New York: John Wiley & Sons, Inc., 1980.

130. B. W. Silverman, *Density estimation for statistics and data analysis*, New York: Routledge, 1998. https://doi.org/10.1201/9781315140919

131. Y. Song, X. Chen, K. Kato, Approximating high-dimensional infinite-order U-statistics: Statistical and computational guarantees, *Electron. J. Statist.*, **13** (2019), 4794–4848. https://doi.org/10.1214/19-EJS1643

132. M. Spiess, Estimation of a two-equation panel model with mixed continuous and ordered categorical outcomes and missing data, *J. Roy. Statist. Soc. Ser. C*, **55** (2006), 525–538. https://doi.org/10.1111/j.1467-9876.2006.00551.x

133. U. Stadmüller, Asymptotic distributions of smoothed histograms, *Metrika*, **30** (1983), 145–158. https://doi.org/10.1007/BF02056918

134. W. Stute, Conditional U-statistics, *Ann. Probab.*, **19** (1991), 812–825. https://doi.org/10.1214/aop/1176990452

135. W. Stute, Almost sure representations of the product-limit estimator for truncated data, *Ann. Statist.*, **21** (1993), 146–156. https://doi.org/10.1214/aos/1176349019

136. W. Stute, L^p-convergence of conditional U-statistics, *J. Multivariate Anal.*, **51** (1994), 71–82. https://doi.org/10.1006/jmva.1994.1050

137. W. Stute, Universally consistent conditional U-statistics, *Ann. Statist.*, **22** (1994), 460–473. https://doi.org/10.1214/aos/1176325378

138. W. Stute, Symmetrized NN-conditional U-statistics, In: *Research developments in probability and statistics*, 1996, 231–237.

139.K. K. Sudheesh, S. Anjana, M. Xie, U-statistics for left truncated and right censored data, *Statistics*, **57** (2023), 900–917. https://doi.org/10.1080/02331888.2023.2217314

140.R. A. Tapia, J. R. Thompson, Nonparametric probability density estimation, In: *Johns Hopkins series in the mathematical sciences*, Routledge: Johns Hopkins University Press, 1978.

141.A. Tenbusch, Two-dimensional Bernstein polynomial density estimators, *Metrika*, **41** (1994), 233–253. https://doi.org/10.1007/BF01895321

142.A. Tenbusch, Nonparametric curve estimation with Bernstein estimates, *Metrika*, **45** (1997), 1–30. https://doi.org/10.1007/BF02717090

143.A. W. van der Vaart, J. A. Wellner, Weak convergence and empirical processes—with applications to statistics, In: *Springer series in statistics*, Springer Cham, 2023. https://doi.org/10.1007/978-3-031-29040-4

144.R. A. Vitale, Bernstein polynomial approach to density function estimation, In: *Statistical inference and related topics*, Academic Press, 1975, 87–99. https://doi.org/10.1016/B978-0-12-568002-8.50011-2

145.D. Vogel, M. Wendler, Studentized U-quantile processes under dependence with applications to change-point analysis, *Bernoulli*, **23** (2017), 3114–3144. https://doi.org/10.3150/16-BEJ838

146.R. V. Mises, On the asymptotic distribution of differentiable statistical functions, *Ann. Math. Statist.*, **18** (1947), 309–348. https://doi.org/10.1214/aoms/1177730385

147.M. P. Wand, M. C. Jones, *Kernel smoothing*, New York: Chapman and Hall/CRC, 1994. https://doi.org/10.1201/b14876

148.L. Wang, D. Lu, On the rates of asymptotic normality for Bernstein density estimators in a triangular array, *J. Math. Anal. Appl.*, **511** (2022), 126063. https://doi.org/10.1016/j.jmaa.2022.126063

149.L. Wang, D. Lu, Application of Bernstein polynomials on estimating a distribution and density function in a triangular array, *Methodol. Comput. Appl. Probab.*, **25** (2023), 56. https://doi.org/10.1007/s11009-023-10032-3

150.G. S.Watson, Smooth regression analysis, *Sankhyā*, **26** (1964), 359–372.

151.M. Wendler, U-processes, U-quantile processes and generalized linear statistics of dependent data, *Stochastic Process. Appl.*, **122** (2012), 787–807. https://doi.org/10.1016/j.spa.2011.11.010

152.W. W. Göttingen, *Statistical density estimation: A survey*, Vandenhoeck & Ruprecht, 1978.

153.A. Yatchew, An elementary estimator of the partial linear model, *Econom. Lett.*, **57** (1997), 135–143. https://doi.org/10.1016/S0165-1765(97)00218-8

154.X. F. Yin, Z. F. Hao, Adaptive kernel density estimation using beta kernel, In: *2007 International conference on machine learning and cybernetics*, IEEE, 2007. https://doi.org/10.1109/ICMLC.2007.4370716

155.S. Zhang, R. J. Karunamuni, Boundary performance of the beta kernel estimators, *J. Nonparametr. Stat.*, **22** (2010), 81–104. https://doi.org/10.1080/10485250903124984

156.W. Zhou, *Generalized spatial U-quantiles: Theory and applications*, The University of Texas at Dallas, 2005.

157. W. Zhou, R. Serfling, Generalized multivariate rank type test statistics via spatial U-quantiles, *Statist. Probab. Lett.*, **78** (2008), 376–383. https://doi.org/10.1016/j.spl.2007.07.010

158. W. Zhou, R. Serfling, Multivariate spatial U-quantiles: A Bahadur-Kiefer representation, a Theil-Sen estimator for multiple regression, and a robust dispersion estimator, *J. Statist. Plann. Inference*, **138** (2008), 1660–1678. https://doi.org/10.1016/j.jspi.2007.05.043

A. Appendix-1

This section is dedicated to proving our results. We will continue to use the previously established notation. A crucial element in our proofs involves the truncation of the U-statistics. Specifically, we represent the U-statistics $u_{n,\ell}(\varphi,\tilde{\mathbf{x}})$ for $\ell\in\{1,2,3\}$ as follows:

$$\begin{aligned} u_{n,\ell}(\varphi,\tilde{\mathbf{x}}) &= u_{n,\ell}\left(\mathcal{G}^{(T)}_{\varphi,\tilde{\mathbf{x}},\ell}\right)+u_{n,\ell}\left(\mathcal{G}^{(R)}_{\varphi,\tilde{\mathbf{x}},\ell}\right)\\ &=: u^{(T)}_{n,\ell}(\varphi,\tilde{\mathbf{x}})+u^{(R)}_{n,\ell}(\varphi,\tilde{\mathbf{x}}), \end{aligned} \tag{A.1}$$

where for $\ell\in\{1,2,3\}$ and some $\omega_{n,\ell}$ (to be specified later in the proof of each section), we have:

$$\begin{aligned} \mathcal{G}_{\varphi,\tilde{\mathbf{x}},\ell}(\mathbf{x},\mathbf{y}) &= \mathcal{G}^{(T)}_{\varphi,\tilde{\mathbf{x}},\ell}(\mathbf{x},\mathbf{y})+\mathcal{G}^{(R)}_{\varphi,\tilde{\mathbf{x}},\ell}(\mathbf{x},\mathbf{y})\\ &= \mathcal{G}_{\varphi,\tilde{\mathbf{x}},\ell}(\mathbf{x},\mathbf{y})\mathbf{1}_{\{|\varphi(\mathbf{y})|\le\omega_{n,\ell}\}}+\mathcal{G}_{\varphi,\tilde{\mathbf{x}},\ell}(\mathbf{x},\mathbf{y})\mathbf{1}_{\{|\varphi(\mathbf{y})|>\omega_{n,\ell}\}}. \end{aligned}$$

Here, $u^{(T)}_{n,\ell}(\varphi,\tilde{\mathbf{x}})$ is the truncated part, and $u^{(R)}_{n,\ell}(\varphi,\tilde{\mathbf{x}})$ is the remainder part. We establish the uniform convergence rates of $u_{n,\ell}(\varphi,\tilde{\mathbf{x}})$ to $\mathbb{E}[u_{n,\ell}(\varphi,\tilde{\mathbf{x}})]$ based on the convergence rates of $u_{n,\ell}\left(\mathcal{G}^{(T)}_{\varphi,\tilde{\mathbf{x}},\ell}\right)$ to $\mathbb{E}[u_{n,\ell}\left(\mathcal{G}^{(T)}_{\varphi,\tilde{\mathbf{x}},\ell}\right)]$, while demonstrating that the remainder part is asymptotically negligible. Next, we can use these results to deduce the convergence rates of the stochastic part of the estimators $\widehat{r}^{(m)}_{n,\ell}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}}))$. Indeed, we can clearly see that based on the classical decomposition, for $\ell\in\{1,2,3\}$:

$$\begin{aligned} &\left|\widehat{r}^{(m)}_{n,\ell}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}}))-\widehat{\mathbb{E}}\left(\widehat{r}^{(m)}_{n,\ell}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}}))\right)\right|\\ \le\ & \frac{\left|u_{n,\ell}(\varphi,\tilde{\mathbf{x}})-\mathbb{E}\left(u_{n,\ell}(\varphi,\tilde{\mathbf{x}})\right)\right|}{\left|u_{n,\ell}(1,\tilde{\mathbf{x}})\right|}+\frac{\left|\mathbb{E}\left(u_{n,\ell}(\varphi,\tilde{\mathbf{x}})\right)\right|\cdot\left|u_{n,\ell}(1,\tilde{\mathbf{x}})-\mathbb{E}\left(u_{n,\ell}(1,\tilde{\mathbf{x}})\right)\right|}{\left|u_{n,\ell}(1,\tilde{\mathbf{x}})\right|\cdot\left|\mathbb{E}\left(u_{n,\ell}(1,\tilde{\mathbf{x}})\right)\right|}\\ =:\ & \mathscr{I}_{\ell,1}+\mathscr{I}_{\ell,2}. \end{aligned} \tag{A.2}$$

Later, based on the imposed regularity conditions in each section, we can easily control the terms $\left|u_{n,\ell}(\varphi,\tilde{\mathbf{x}})\right|$ and $\left|\mathbb{E}\left[u_{n,\ell}(\varphi,\tilde{\mathbf{x}})\right]\right|$ (including the particular case when $\varphi\equiv 1$), uniformly in $\tilde{\mathbf{x}}$ to obtain the desired rates of convergence. Lastly, we need to study the bias of each estimator. It is worth noting that the proof of the bias term for the three estimators proposed in this paper is based on the following decomposition for $\ell\in\{1,2,3\}$. We have

$$\left|\widehat{\mathbb{E}}\left[\widehat{r}^{(m)}_{n,\ell}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,\ell}(\tilde{\mathbf{x}}))\right]-r^{(m)}(\varphi,\tilde{\mathbf{x}})\right| = \frac{\left|\mathbb{E}\left(u_{n,\ell}(\varphi,\tilde{\mathbf{x}})\right)-r^{(m)}(\varphi,\tilde{\mathbf{x}})\mathbb{E}\left(u_{n,\ell}(1,\tilde{\mathbf{x}})\right)\right|}{\left|\mathbb{E}\left(u_{n,\ell}(1,\tilde{\mathbf{x}})\right)\right|}. \tag{A.3}$$

As a matter of fact, (A.3) implies that it suffices to control the term $\left|\mathbb{E}\left(u_{n,\ell}(\varphi,\tilde{\mathbf{x}})\right)-\mathcal{R}(\varphi,\tilde{\mathbf{x}})\right|$ uniformly in $\tilde{\mathbf{x}}$, to establish the desired results, as we can see in the sequel.

A.1. Proofs of Section 3: Dirichlet kernels

A.1.1. Proofs of Section 3.1

The regression proof for the case where $m = 1$ closely resembles the one given in [117]. We include it here in full detail for the reader's convenience and to ensure it is self-contained. However, the result concerning the regression function smoothed by the Dirichlet kernel has not been addressed in the literature, providing the primary motivation for presenting it in this paper.

Proof of Theorem 3.1. Observe that

$$\begin{aligned}\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\widehat{g}_n(\varphi,\mathbf{x},\mathbf{\Lambda}_{n,1})-\mathcal{R}(\varphi,\mathbf{x})\right| \quad\leq\quad & \sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\widehat{g}_n(\varphi,\mathbf{x},\mathbf{\Lambda}_{n,1})-\mathbb{E}\left[\widehat{g}_n(\varphi,\mathbf{x},\mathbf{\Lambda}_{n,1})\right]\right| \\ & +\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\mathbb{E}\left[\widehat{g}_n(\varphi^{(T)},\mathbf{x},\mathbf{\Lambda}_{n,1})\right]-\mathcal{R}(\varphi,\mathbf{x})\right|.\end{aligned}$$

Keep in mind the definition of the set $\mathbb{S}_{d,1}(\delta)$ given in (3.4). To begin, we need to prove the following result:

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}(\breve{b}d)}\left|\widehat{g}_n(\varphi,\mathbf{x},\mathbf{\Lambda}_{n,1})-\mathbb{E}\left[\widehat{g}_n(\varphi,\mathbf{x},\mathbf{\Lambda}_{n,1})\right]\right| = O\left(\frac{|\log\breve{b}|(\log n)^{3/2}}{\breve{b}^{d+1/2}\sqrt{n}}\right)\quad\text{a.s.}\tag{A.4}$$

The proof of (A.4) follows the same analogy as in [117] while applying the necessary changes to fit our context. We have

$$\begin{aligned}\widehat{g}_n(\varphi,\mathbf{x},\mathbf{\Lambda}_{n,1})-\mathbb{E}\left[\widehat{g}_n(\varphi,\mathbf{x},\mathbf{\Lambda}_{n,1})\right] &= \frac{1}{n}\sum_{i=1}^{n}\left\{\varphi(\mathbf{Y}_i)K_{(\alpha,\beta)}(\mathbf{X}_i)-\mathbb{E}[\varphi(\mathbf{Y}_i)K_{(\alpha,\beta)}(\mathbf{X}_i)]\right\}\\ &= \frac{1}{n}\sum_{i=1}^{n}Z_{i,b}(\mathbf{x}),\end{aligned}$$

where, for $i = 1,\ldots,n$,

$$Z_{i,b}(\mathbf{x}) := \varphi(\mathbf{Y}_i)K_{(\alpha,\beta)}(\mathbf{X}_i)-\mathbb{E}[\varphi(\mathbf{Y}_i)K_{(\alpha,\beta)}(\mathbf{X}_i)].$$

For some sequence $\omega_{n,1}$ tending to infinity, we also consider the following notation:

$$\begin{aligned}\varphi^{(T)}(\mathbf{y}) :&= \varphi(\mathbf{y})\mathbf{1}_{\{|\varphi(\mathbf{y})|\leq\omega_{n,1}\}},\\ \varphi^{(R)}(\mathbf{y}) :&= \varphi(\mathbf{y})\mathbf{1}_{\{|\varphi(\mathbf{y})|>\omega_{n,1}\}}.\end{aligned}$$

This allows us to write

$$\widehat{g}_n(\varphi,\mathbf{x},\mathbf{\Lambda}_{n,1})-\mathbb{E}\left[\widehat{g}_n(\varphi,\mathbf{x},\mathbf{\Lambda}_{n,1})\right] = \frac{1}{n}\sum_{i=1}^{n}\left\{Z_{i,b}^{(T)}(\mathbf{x})+Z_{i,b}^{(R)}(\mathbf{x})\right\},$$

where

$$Z_{i,b}^{(T)}(\mathbf{x}) := \varphi^{(T)}(\mathbf{Y}_i)K_{(\alpha,\beta)}(\mathbf{X}_i)-\mathbb{E}[\varphi^{(T)}(\mathbf{Y}_i)K_{(\alpha,\beta)}(\mathbf{X}_i)],\tag{A.5}$$

$$Z_{i,b}^{(R)}(\mathbf{x}) := \varphi^{(R)}(\mathbf{Y}_i)K_{(\alpha,\beta)}(\mathbf{X}_i)-\mathbb{E}[\varphi^{(R)}(\mathbf{Y}_i)K_{(\alpha,\beta)}(\mathbf{X}_i)].\tag{A.6}$$

We also denote

$$W_{i,b}(\mathbf{x}) := K_{(\alpha,\beta)}(\mathbf{X}_i)-\mathbb{E}[K_{(\alpha,\beta)}(\mathbf{X}_i)].\tag{A.7}$$

The following proposition, that is close to Proposition 1 of [117], will play an instrumental role in the sequel.

Proposition A.1. *Let* $\mathbf{x} \in \mathbb{S}_{d,1}(\breve{b}(d+1))$, $n \geq 1$, $0 < \breve{b} < \left(e^{-16\sqrt{2}} \wedge d^{-1}\right)$, $0 < a \leq e^{-1}\|f\|_\infty |\log \breve{b}|/\breve{b}^{d+1/2}$, *and take the unique*

$$\delta \in \left(0, e^{-1}\right] \text{ that satisfies } \delta|\log\delta| = \frac{\breve{b}^{d+1/2}a}{\|f\|_\infty |\log \breve{b}|}. \tag{A.8}$$

Then, for all $h \in \mathbb{R}$,

$$\begin{aligned} &\mathbb{P}\left(\sup_{\mathbf{x}'\in\mathbf{x}+[-\breve{b},\breve{b}]^d}\left|\frac{1}{n}\sum_{i=1}^n Z_{i,b}^{(T)}(\mathbf{x}')\right| \geq h + 2a\omega_{n,1}, \left|\frac{1}{n}\sum_{i=1}^n Z_{i,b}^{(T)}(\mathbf{x})\right| \leq h\right) \qquad (A.9)\\ \leq &C_{\varphi,d}\exp\left(-\frac{1}{100^2 d^4 \|f\|_\infty^2}\cdot\left(\frac{n^{1/2}\breve{b}^{d+1/2}a}{|\log\delta||\log\breve{b}|}\right)^2\right), \end{aligned}$$

where $C_{\varphi,d} > 0$ *is a constant that depends only on the function* $\varphi(\cdot)$ *and the dimension* d.

Proof of Proposition A.1. Following a similar approach to the proof in [117], we apply a union bound to show that the probability in (A.9) can be bounded as follows:

$$\begin{aligned} \leq\;& \mathbb{P}\left(\left\{\sup_{\mathbf{x}'\in\mathbf{x}+[-\breve{b},\breve{b}]^d}\left|\frac{1}{n}\sum_{i=1}^n \left(Z_{i,b}^{(T)}(\mathbf{x}') - Z_{i,b}^{(T)}(\mathbf{x})\right)\mathbf{1}_{\{\mathbf{x}_i\in\mathbb{S}_{d,1}\setminus\mathbb{S}_{d,1}(\delta)\}}\right| \geq a\omega_{n,1}\right\}\right.\\ &\left.\cap\left\{\sum_{i=1}^n \mathbf{1}_{\{\mathbf{x}_i\in\mathbb{S}_{d,1}\setminus\mathbb{S}_{d,1}(\delta)\}} \leq n\cdot 4\|f\|_\infty\delta\right\}\right) \qquad (A.10)\\ &+\mathbb{P}\left(\sum_{i=1}^n \mathbf{1}_{\{\mathbf{x}_i\in\mathbb{S}_{d,1}\setminus\mathbb{S}_{d,1}(\delta)\}} \geq n\cdot 4\|f\|_\infty\delta\right) \qquad (A.11)\\ &+\mathbb{P}\left(\sup_{\mathbf{x}'\in\mathbf{x}+[-\breve{b},\breve{b}]^d}\left|\frac{1}{n}\sum_{i=1}^n \left(Z_{i,b}^{(T)}(\mathbf{x}') - Z_{i,b}^{(T)}(\mathbf{x})\right)\mathbf{1}_{\{\mathbf{x}_i\in\mathbb{S}_{d,1}(\delta)\}}\right| \geq a\omega_{n,1}\right) \qquad (A.12)\\ =:\;& (A) + (B) + (C). \qquad (A.13) \end{aligned}$$

To clarify our notation, for any subset $\mathfrak{A} \subset \mathbb{R}^d$ and any point $\mathbf{x} \in \mathbb{R}^d$, we define $\mathbf{x} + \mathfrak{A} = \{\mathbf{x} + \mathbf{y} : \mathbf{y} \in \mathfrak{A}\}$. To bound term (A.10), consider the assumption $\mathbf{x} \in \mathbb{S}_{d,1}(\breve{b}(d+1))$ and $\mathbf{x}' = \mathbf{x} + [-\breve{b}, \breve{b}]^d$. This implies that both $\mathbf{x}$ and $\mathbf{x}'$ are in $\mathbb{S}_{d,1}(\breve{b})$, leading to the following relations:

$$\alpha_1 = \frac{x_1}{\breve{b}} + 1, \ldots, \alpha_d = \frac{x_d}{\breve{b}} + 1, \beta = \frac{1 - \|\mathbf{x}\|_1}{\breve{b}} + 1 \geq 2,$$

and for $\mathbf{x}'$,

$$\alpha_1' = \frac{x_1'}{\breve{b}} + 1, \ldots, \alpha_d' = \frac{x_d'}{\breve{b}} + 1, \beta' = \frac{1 - \|\mathbf{x}'\|_1}{\breve{b}} + 1 \geq 2.$$

Consequently, we have:

$$\sqrt{\frac{\|\boldsymbol{\alpha}\|_1 + \beta - 1}{(\beta - 1)\prod_{i=1}^d(\alpha_i - 1)}} \leq \sqrt{\|\boldsymbol{\alpha}\|_1 + \beta - 1} = \sqrt{\breve{b}^{-1} + d}, \tag{A.14}$$

$$\sqrt{\frac{\|\boldsymbol{\alpha}'\|_1 + \beta' - 1}{(\beta' - 1)\prod_{i=1}^d(\alpha_i' - 1)}} \leq \sqrt{\|\boldsymbol{\alpha}'\|_1 + \beta' - 1} = \sqrt{\breve{b}^{-1} + d}. \tag{A.15}$$

Combining these results with our assumption in (A.8) and the upper bound on the Dirichlet density from Lemma 2 in [117], we obtain:

$$\left\{\sum_{i=1}^{n} \mathbf{1}_{\{X_i \in \mathbb{S}_{d,1} \setminus \mathbb{S}_{d,1}(\delta)\}} \leq 4n\|f\|_\infty \delta\right\},$$

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n}\left(Z_{i,b}^{(T)}(\mathbf{x}') - Z_{i,b}^{(T)}(\mathbf{x})\right)\mathbf{1}_{\{X_i \in \mathbb{S}_{d,1} \setminus \mathbb{S}_{d,1}(\delta)\}}\right| &\leq 4 \cdot 4\omega_{n,1}\|f\|_\infty \delta \cdot \breve{b}^{-d}\sqrt{\breve{b}^{-1} + d} \\ &\leq \frac{16\sqrt{1+bd}}{|\log \delta||\log \breve{b}|} a\omega_{n,1}. \qquad (A.16)\end{aligned}$$

Given the assumptions $0 < \delta \leq e^{-1}$ and $0 < \breve{b} < \left(e^{-16\sqrt{2}} \wedge d^{-1}\right)$, it follows that:

$$(A) = 0. \qquad (A.17)$$

Term (A.11) represents the probability of encountering "too many bad observations", meaning too many $\mathbf{x}_i$s near the boundary of the simplex where the partial derivatives of the Dirichlet density with respect to $\alpha_1, \ldots, \alpha_d$, and β diverge. We can control this term using a concentration bound. To begin, note that the volume of $\mathbb{S}_{d,1} \setminus \mathbb{S}_{d,1}(\delta)$ is at most $2d\delta/d!$. Specifically, $\mathbb{S}_{d,1}(\delta)$ forms a simplex of side-length $1 - 2\delta$ within $\mathbb{S}_{d,1}$, so:

$$d! \cdot \text{Volume}\left(\mathbb{S}_{d,1} \setminus \mathbb{S}_{d,1}(\delta)\right) = 1 - (1-2\delta)^d \leq 1 - (1 + d \cdot (-2\delta)) = 2d\delta, \qquad (A.18)$$

where we used the inequality $(1+x)^n \geq 1 + nx$, which holds for all $n \in \mathbb{N}$ and $x \geq -1$. From (A.18) and knowing that $\|f\|_\infty$ is finite (since $f(\cdot)$ is continuous by assumption and $\mathbb{S}_{d,1}$ is compact), we get:

$$\mathbb{E}\left[\mathbf{1}_{\{X_i \in \mathbb{S}_{d,1} \setminus \mathbb{S}_{d,1}(\delta)\}}\right] \leq \frac{2\|f\|_\infty}{(d-1)!}\delta.$$

Applying Hoeffding's inequality and condition (A.8), we obtain:

$$(B) \leq \exp\left(-2n\left((2(d-1)!-1) \cdot \frac{2\|f\|_\infty}{(d-1)!}\delta\right)^2\right) \leq \exp\left(-2\left(\frac{n^{1/2}\breve{b}^{d+1/2}a}{|\log \delta||\log \breve{b}|}\right)^2\right). \qquad (A.19)$$

To bound the third probability in (A.13), the main idea is to decompose the supremum using a chaining argument and apply concentration bounds on the increments at each level of the d-dimensional tree. With the notation $\mathcal{H}_k := 2^{-k} \cdot \breve{b}\mathbb{Z}^d$, we have the following embedded sequence of lattice points:

$$\mathcal{H}_0 \subseteq \mathcal{H}_1 \subseteq \cdots \subseteq \mathcal{H}_k \subseteq \cdots \subseteq \mathbb{R}^d.$$

Hence, for $\mathbf{x} \in \mathbb{S}_{d,1}(\breve{b}(d+1))$ fixed, and for any $\mathbf{x}' \in \mathbf{x} + [-\breve{b}, \breve{b}]^d$, let $(\mathbf{x}_k)_{k \in \mathbb{N}_0}$ be a sequence that satisfies:

$$\mathbf{x}_0 = \mathbf{x}, \quad \mathbf{x}_k - \mathbf{x} \in \mathcal{H}_k \cap [-\breve{b}, \breve{b}]^d, \quad \lim_{k\to\infty}\|\mathbf{x}_k - \mathbf{x}'\|_\infty = 0,$$

and

$$(\mathbf{x}_{k+1})_i = (\mathbf{x}_k)_i \pm 2^{-k-1}b, \quad \text{for all } i = 1, \ldots, d.$$

Since the map $\mathbf{x} \mapsto \frac{1}{n}\sum_{i=1}^{n} Z_{i,b}^{(T)}(\mathbf{x})$ is almost surely continuous,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left(Z_{i,b}^{(T)}(\mathbf{x}') - Z_{i,b}^{(T)}(\mathbf{x})\right)\mathbf{1}_{\{\mathbf{x}_i\in\mathbb{S}_{d,1}(\delta)\}}\right| \leq \sum_{k=0}^{\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\left(Z_{i,b}^{(T)}(\mathbf{x}_{k+1}) - Z_{i,b}^{(T)}(\mathbf{x}_k)\right)\mathbf{1}_{\{\mathbf{x}_i\in\mathbb{S}_{d,1}(\delta)\}}\right|,$$

and since $\sum_{k=0}^{\infty}\frac{1}{2(k+1)^2} \leq 1$, we have the inclusion of events,

$$\begin{aligned}&\left\{\sup_{\mathbf{x}'\in\mathbf{x}+[-\breve{b},\breve{b}]^d}\left|\frac{1}{n}\sum_{i=1}^{n}\left(Z_{i,b}^{(T)}(\mathbf{x}') - Z_{i,b}^{(T)}(\mathbf{x})\right)\mathbf{1}_{\{\mathbf{x}_i\in\mathbb{S}_{d,1}(\delta)\}}\right| \geq a\omega_{n,1}\right\}\\ &\quad\subseteq \bigcup_{k=0}^{\infty}\bigcup_{\substack{\mathbf{x}_k\in\mathbf{x}+\mathcal{H}_k\cap[-\breve{b},\breve{b}]^d\\ (\mathbf{x}_{k+1})_i=(\mathbf{x}_k)_i\pm 2^{-k-1}b,\forall i\in[d]}}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\left(Z_{i,b}^{(T)}(\mathbf{x}_{k+1}) - Z_{i,b}^{(T)}(\mathbf{x}_k)\right)\mathbf{1}_{\{\mathbf{x}_i\in\mathbb{S}_{d,1}(\delta)\}}\right| \geq \frac{a\omega_{n,1}}{2(k+1)^2}\right\}.\end{aligned}$$

By a union bound and the fact that $\left|\mathcal{H}_k \cap [-\breve{b},\breve{b}]^d\right| \leq 2^{(k+2)d}$,

$$\begin{aligned}(C) &\leq \sum_{k=0}^{\infty} 2^{(k+2)d}2^d \sup_{\substack{\mathbf{x}_k\in\mathbf{x}+\mathcal{H}_k\cap[-\breve{b},\breve{b}]^d\\ (\mathbf{x}_{k+1})_i=(\mathbf{x}_k)_i\pm 2^{-k-1}b,\forall i\in[d]}} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(Z_{i,b}^{(T)}(\mathbf{x}_{k+1}) - Z_{i,b}^{(T)}(\mathbf{x}_k)\right)\mathbf{1}_{\{\mathbf{x}_i\in\mathbb{S}_{d,1}(\delta)\}}\right| \geq \frac{a\omega_{n,1}}{2(k+1)^2}\right)\\ &\leq \sum_{k=0}^{\infty} 2^{(k+2)d}2^d \sup_{\substack{\mathbf{x}_k\in\mathbf{x}+\mathcal{H}_k\cap[-\breve{b},\breve{b}]^d\\ (\mathbf{x}_{k+1})_i=(\mathbf{x}_k)_i\pm 2^{-k-1}b,\forall i\in[d]}} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(W_{i,b}(\mathbf{x}_{k+1}) - W_{i,b}(\mathbf{x}_k)\right)\mathbf{1}_{\{\mathbf{x}_i\in\mathbb{S}_{d,1}(\delta)\}}\right| \geq \frac{a}{2(k+1)^2}\right).\end{aligned} \tag{A.20}$$

Using Azuma's inequality and Lemma 4 [117] (note that $\mathbf{x} \in \mathbb{S}_{d,1}(\breve{b}(d+1))$ and $\mathbf{x}' \in \mathbf{x}+[-\breve{b},\breve{b}]^d$ implies $\mathbf{x}_k \in \mathbb{S}_{d,1}(\breve{b})$ for all $k \in \mathbb{N}_0$, so that $\alpha_1 = (\mathbf{x}_k)_1/\breve{b}+1,\ldots,\alpha_d = (\mathbf{x}_k)_d/\breve{b}+1, \beta = (1-\|\mathbf{x}_k\|_1)/\breve{b}+1 \geq 2$ for all $k \in \mathbb{N}_0$) and (A.7), the above is

$$\begin{aligned}&\leq \sum_{k=0}^{\infty} 2^{(k+3)d}\cdot 2\exp\left(-\frac{na^2}{8(k+1)^4}\cdot\left(25d^2\|f\|_\infty\frac{|\log\delta||\log\breve{b}|}{\breve{b}^{d+1/2}2^{k+1}}\right)^{-2}\right)\\ &\leq \sum_{k=0}^{\infty} 2^{(k+3)d}\cdot 2\exp\left(-\frac{2^{2k-1}}{25^2d^4\|f\|_\infty^2(k+1)^4}\cdot\left(\frac{n^{1/2}\breve{b}^{d+1/2}a}{|\log\delta||\log\breve{b}|}\right)^2\right).\end{aligned}$$

The minimum of $k \mapsto 0.99\cdot 2^{2k-1}(k+1)^{-4}$ on $\mathbb{N}_0$ is larger than, say, $1/16$, so we deduce

$$(C) \leq C_{f,d}\exp\left(-\frac{1}{100^2d^4\|f\|_\infty^2}\cdot\left(\frac{n^{1/2}\breve{b}^{d+1/2}a}{|\log\delta||\log\breve{b}|}\right)^2\right), \tag{A.21}$$

for some large constant $C_{f,d} > 0$. Putting (A.17), (A.19), and (A.21) together in (A.13) concludes the proof of Proposition A.1. □

The following corollary is very similar to Corollary 2 of [117].

Corollary A.2 (Large deviation estimates)**.** *Recall* $Z^{(T)}_{i,b}(\mathbf{x})$ *defined in* (A.5)*. Let* $\mathbf{x} \in \mathbb{S}_{d,1}(\breve{b}(d+1)), n \geq 100^6 d^6, n^{-1/d} \leq \breve{b} \leq \left(e^{-16\sqrt{2}} \wedge d^{-1}\right) 0 < a \leq e^{-1}\|f\|_\infty |\log \breve{b}|/\breve{b}^{d+1/2}$*, and take the unique*

$$\delta \in \left(0, e^{-1}\right] \quad \textit{that satisfies} \quad \delta|\log\delta| = \frac{\breve{b}^{d+1/2} a}{\|f\|_\infty |\log \breve{b}|}.$$

Then, we have

$$\mathbb{P}\left(\sup_{\mathbf{x}'\in\mathbf{x}+[-\breve{b},\breve{b}]^d}\left|\frac{1}{n}\sum_{i=1}^{n} Z^{(T)}_{i,b}(\mathbf{x}')\right| \geq 3a\omega_{n,1}\right) \leq C_{f,d}\exp\left(-\frac{1}{100^2 d^4 \|f\|_\infty^2}\cdot\left(\frac{n^{1/2}\breve{b}^{d+1/2}a}{|\log\delta||\log\breve{b}|}\right)^2\right), \tag{A.22}$$

where $C_{f,d} > 0$ *is a constant that depends only on the density* $f(\cdot)$ *and the dimension* d*.*

Proof of Corollary A.2. By applying a union bound, we find that the probability in Eq (A.22) can be bounded as follows:

$$\leq \mathbb{P}\left(\sup_{\mathbf{x}'\in\mathbf{x}+[-\breve{b},\breve{b}]^d}\left|\frac{1}{n}\sum_{i=1}^{n} Z^{(T)}_{i,b}(\mathbf{x}')\right| \geq 3a\omega_{n,1}, \left|\frac{1}{n}\sum_{i=1}^{n} Z^{(T)}_{i,b}(\mathbf{x})\right| \leq a\omega_{n,1}\right) + \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z^{(T)}_{i,b}(\mathbf{x})\right| \geq a\omega_{n,1}\right).$$

The first probability can be bounded using Proposition A.1, and the second probability can be similarly bounded by applying Azuma's inequality and Lemma 4 from [117], as was done in Eq (A.20). Now, we turn to the proof of (A.4). We start by noting:

$$\begin{aligned}
\widehat{g}_n(\varphi, \mathbf{x}, \mathbf{\Lambda}_{n,1}) &= \frac{1}{n}\sum_{i=1}^{n}\varphi(\mathbf{Y}_i)K_{(\alpha,\beta)}(\mathbf{X}_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}\varphi^{(T)}(\mathbf{Y}_i)K_{(\alpha,\beta)}(\mathbf{X}_i) + \frac{1}{n}\sum_{i=1}^{n}\varphi^{(R)}(\mathbf{Y}_i)K_{(\alpha,\beta)}(\mathbf{X}_i) \\
&= \widehat{g}_n(\varphi^{(T)}, \mathbf{x}, \mathbf{\Lambda}_{n,1}) + \widehat{g}_n(\varphi^{(R)}, \mathbf{x}, \mathbf{\Lambda}_{n,1}).
\end{aligned}$$

To prove (A.4), we need to show that the remainder term is asymptotically negligible, i.e.,

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}(\breve{b}d)}\left|\widehat{g}_n(\varphi^{(R)}, \mathbf{x}, \mathbf{\Lambda}_{n,1}) - \mathbb{E}\left[\widehat{g}_n(\varphi^{(R)}, \mathbf{x}, \mathbf{\Lambda}_{n,1})\right]\right| = o(1) \quad \text{a.s..}$$

This follows directly from the proof of the remainder term for the U-statistics developed subsequently. Additionally, we need to prove:

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}(\breve{b}d)}\left|\widehat{g}_n(\varphi^{(T)}, \mathbf{x}, \mathbf{\Lambda}_{n,1}) - \mathbb{E}\left[\widehat{g}_n(\varphi^{(T)}, \mathbf{x}, \mathbf{\Lambda}_{n,1})\right]\right| = O\left(\frac{|\log\breve{b}|(\log n)^{3/2}}{\breve{b}^{d+1/2}\sqrt{n}}\right) \quad \text{a.s..} \tag{A.23}$$

This equation is obtained by a union bound over the suprema on hypercubes of width $2\breve{b}$ centered at each $\mathbf{x} \in 2\breve{b}\mathbb{Z}^d \cap \mathbb{S}_{d,1}(\breve{b}(d+1))$, using the large deviation estimates in Corollary A.2, and choosing

$$a = 100d^2\frac{(\log n)^{3/2}}{\sqrt{n}}\cdot\frac{\|f\|_\infty|\log\breve{b}|}{\breve{b}^{d+1/2}}. \tag{A.24}$$

The upper bound condition on a is satisfied as long as $100d^2(\log n)^{3/2}/(\sqrt{n}) \leq e^{-1}$, which is valid if $n \geq 100^6 d^6$. For the unique $\delta \in (0, e^{-1}]$ that satisfies

$$\delta|\log\delta| = \frac{\breve{b}^{d+1/2}a}{\|f\|_\infty|\log\breve{b}|} \overset{(A.24)}{=} 100d^2\frac{(\log n)^{3/2}}{\sqrt{n}}, \tag{A.25}$$

we obtain:

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\mathbf{x}\in\mathbb{S}_{d,1}(\breve{b}d)}\left|\widehat{g}_n(\varphi^{(T)},\mathbf{x},\boldsymbol{\Lambda}_{n,1}) - \mathbb{E}\left[\widehat{g}_n(\varphi^{(T)},\mathbf{x},\boldsymbol{\Lambda}_{n,1})\right]\right| \geq 3a\omega_{n,1}\right)\\
&\leq \sum_{\mathbf{x}\in 2\breve{b}\mathbb{Z}^d\cap\mathbb{S}_{d,1}(\breve{b}(d+1))}\mathbb{P}\left(\sup_{\mathbf{x}'\in\mathbf{x}+[-\breve{b},\breve{b}]^d}\left|\frac{1}{n}\sum_{i=1}^n Z_{i,b}^{(T)}(\mathbf{x}')\right| \geq 3a\omega_{n,1}\right)\\
&\leq \breve{b}^{-d}\cdot C_{f,d}\exp\left(-\frac{1}{100^2d^4\|f\|_\infty^2}\left(\frac{n^{1/2}\breve{b}^{d+1/2}a}{|\log\delta||\log\breve{b}|}\right)^2\right)\\
&\leq \breve{b}^{-d}\cdot C_{f,d}\exp\left(-\frac{(\log n)^3}{|\log\delta|^2}\right).
\end{aligned}$$

The condition on δ in (A.25) implies:

$$n^{-1/2} \leq \delta \leq e^{-1}, \quad (\text{thus } |\log\delta| \leq \frac{1}{2}\log n), \tag{A.26}$$

since the function $x \mapsto x|\log x|$ is increasing on $(0, e^{-1}]$. Using (A.26) in (A.25), we get:

$$\mathbb{P}\left(\sup_{\mathbf{x}\in\mathbb{S}_{d,1}(\breve{b}d)}\left|\widehat{g}_n(\varphi^{(T)},\mathbf{x},\boldsymbol{\Lambda}_{n,1}) - \mathbb{E}\left[\widehat{g}_n(\varphi^{(T)},\mathbf{x},\boldsymbol{\Lambda}_{n,1})\right]\right| \geq 3a\omega_{n,1}\right) \leq C_{f,d}\exp(d|\log\breve{b}| - 4\log n).$$

Since we assumed that $\breve{b} \geq n^{-1/d}$, the above is $\leq C_{f,d}n^{-3}$, which is summable. By our choice of a in (A.24) and the Borel-Cantelli lemma, we obtain

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}(\breve{b}d)}\left|\widehat{g}_n(\varphi^{(T)},\mathbf{x},\boldsymbol{\Lambda}_{n,1}) - \mathbb{E}\left[\widehat{g}_n(\varphi^{(T)},\mathbf{x},\boldsymbol{\Lambda}_{n,1})\right]\right| = O\left(\frac{|\log\breve{b}|(\log n)^{3/2}}{\breve{b}^{d+1/2}\sqrt{n}}\right) \quad \text{a.s.}$$

Now, we only need to study the bias term,

$$\left|\mathbb{E}\left[\widehat{g}_n(\varphi^{(T)},\mathbf{x},\boldsymbol{\Lambda}_{n,1})\right] - \mathcal{R}(\varphi,\mathbf{x})\right| = O(\breve{b}^{1/2}). \tag{A.27}$$

Using the same reasoning as [79], we have

$$\begin{aligned}
\mathbb{E}\left[\widehat{g}_n(\varphi^{(T)},\mathbf{x},\boldsymbol{\Lambda}_{n,1})\right] &= \int_{\mathbb{S}_{d,1}} r^{(1)}(\varphi,\mathbf{u})f(\mathbf{u})K_{\alpha,\beta}(\mathbf{u})\mathbf{du}\\
&= \mathbb{E}\left[\mathcal{R}(\varphi,\zeta_{\mathbf{x}})\right],
\end{aligned}$$

where $\zeta_{\mathbf{x}} = (\zeta_{x_1},\ldots,\zeta_{x_d}) \sim \text{Dirichlet}(\boldsymbol{\alpha},\boldsymbol{\beta})$. By a second-order Taylor expansion around $\zeta_{\mathbf{x}} = \mathbf{x}$, we have

$$\mathbb{E}\left[\mathcal{R}(\varphi,\zeta_{\mathbf{x}})\right] = \mathcal{R}(\varphi,\mathbf{x}) + \sum_{j=1}^d\frac{\partial\mathcal{R}(\varphi,\mathbf{x})}{\partial x_j}\mathbb{E}\left(\zeta_{x_j} - x_j\right) + \frac{1}{2}\sum_{j=1}^d\frac{\partial^2\mathcal{R}(\varphi,\overline{\mathbf{x}})}{\partial x_j^2}\mathbb{E}\left(\zeta_{x_j} - x_j\right)^2$$

$$+\sum_{j=1}^{d}\sum_{k=1,k\neq j}^{d}\frac{\partial^2\mathcal{R}(\varphi,\overline{\mathbf{x}})}{\partial x_j\partial x_k}\mathbb{E}\left\{\left(\zeta_{x_j}-x_j\right)(\zeta_{x_k}-x_k)\right\},$$

for some $\overline{\mathbf{x}}$ joining $\zeta_{\mathbf{x}}$ and $\mathbf{x}$. In addition, for all $j,k\in\{1,\ldots,d\}$, straightforward calculations yield (for instance, see [117]):

$$\begin{aligned}
\mathbb{E}\left[\zeta_j\right] &= \frac{\frac{x_j}{\breve{b}}+1}{\frac{1}{\breve{b}}+d+1}=\frac{x_j+\breve{b}}{1+\breve{b}(d+1)}=x_j+\breve{b}\left(1-(d+1)x_j\right)+O\left(\breve{b}^2\right),\\
\operatorname{Cov}\left(\zeta_j,\zeta_k\right) &= \frac{\left(\frac{x_j}{\breve{b}}+1\right)\left(\left(\frac{1}{\breve{b}}+d+1\right)\mathbf{1}_{\{j=k\}}-\left(\frac{x_k}{\breve{b}}+1\right)\right)}{\left(\frac{1}{\breve{b}}+d+1\right)^2\left(\frac{1}{\breve{b}}+d+2\right)}\\
&= \frac{b\left(x_j+\breve{b}\right)\left(\mathbf{1}_{\{j=k\}}-x_k+\breve{b}(d+1)\mathbf{1}_{\{j=k\}}-\breve{b}\right)}{(1+\breve{b}(d+1))^2(1+\breve{b}(d+2))}\\
&= \breve{b}x_j\left(\mathbf{1}_{\{j=k\}}-x_k\right)+O\left(\breve{b}^2\right), \qquad (A.28)\\
\mathbb{E}\left[\left(\zeta_j-x_j\right)(\zeta_k-x_k)\right] &= \operatorname{Cov}\left(\zeta_j,\zeta_k\right)+\left(\mathbb{E}\left[\zeta_j\right]-x_j\right)(\mathbb{E}\left[\zeta_k\right]-x_k)\\
&= \breve{b}x_j\left(\mathbf{1}_{\{j=k\}}-x_k\right)+O\left(\breve{b}^2\right). \qquad (A.29)
\end{aligned}$$

Then, the Cauchy-Schwartz inequality, (A.28), and (A.29) yields:

$$\begin{aligned}
&\left|\mathbb{E}\left[\mathcal{R}(\varphi,\zeta_{\mathbf{x}})\right]-\mathcal{R}(\varphi,\mathbf{x})\right|\\
&=\sum_{j=1}^{d}O\left(\mathbb{E}\left(\zeta_{x_j}-x_j\right)\right)+\frac{1}{2}\sum_{j=1}^{d}O\left(\mathbb{E}\left(\zeta_{x_j}-x_j\right)^2\right)+\sum_{j=1}^{d}\sum_{k=1,k\neq j}^{d}O\left(\mathbb{E}\left[\left(\zeta_{x_j}-x_j\right)(\zeta_{x_k}-x_k)\right]\right)\\
&\leq\sum_{j=1}^{d}O\left(\sqrt{\mathbb{E}\left(\left|\zeta_{x_j}-x_j\right|^2\right)}\right)+O(\breve{b})+O(\breve{b}^2)\\
&\leq O(\breve{b}^{1/2})+O(\breve{b})+O(\breve{b}^2)\leq O(\breve{b}^{1/2})(1+o(1)).
\end{aligned}$$

Finally, we obtain

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\frac{\widehat{g}_n(\varphi,\mathbf{x},\mathbf{\Lambda}_{n,1})}{f(\mathbf{x})}-r^{(1)}(\varphi,\mathbf{x})\right|\leq\frac{\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\widehat{g}_n(\varphi,\mathbf{x},\mathbf{\Lambda}_{n,1})-\mathcal{R}(\varphi,\mathbf{x})\right|}{\inf_{\mathbf{x}\in\mathbb{S}_{d,1}}f(\mathbf{x})}, \qquad (A.30)$$

and

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\frac{\widehat{f}_n(\mathbf{x},\mathbf{\Lambda}_{n,1})}{f(\mathbf{x})}-1\right|\leq\frac{\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\hat{f}_n(\mathbf{x},b)-f(\mathbf{x})\right|}{\inf_{\mathbf{x}\in\mathbb{S}_{d,1}}f(\mathbf{x})}. \qquad (A.31)$$

By integrating the acquired findings with the given information,

$$r^{(1)}(\varphi,\mathbf{x})=\frac{\widehat{g}_n(\varphi,\mathbf{x},\mathbf{\Lambda}_{n,1})}{f(\mathbf{x})}\cdot\frac{f(\mathbf{x})}{\widehat{f}_n(\mathbf{x},\mathbf{\Lambda}_{n,1})},$$

gives us the desired result. Therefore, the proof is conclusive. □

□

A.1.2. Proofs of Section 3.2

Proof of Theorem 3.2. Let $\mathbf{x} \in \mathbb{S}_{d,1}(\breve{b}(d+1))$, $n \geq 1$, $0 < \breve{b} < \left(e^{-16\sqrt{2}} \wedge d^{-1}\right)$, $0 < a \leq e^{-1}\|f\|_\infty |\log \breve{b}|/\breve{b}^{d+1/2}$, and take the unique

$$\delta \in \left(0, e^{-1}\right] \text{ that satisfies } \delta|\log\delta| = \frac{\breve{b}^{d+1/2}a}{2\|f\|_\infty|\log\breve{b}|}.$$

Define $\tilde{\mathbf{x}}' = (\mathbf{x}'_1, \ldots, \mathbf{x}'_m)$ such that $\tilde{\mathbf{x}}' \in \tilde{\mathbf{x}} + [-\breve{\mathbf{b}}, \breve{\mathbf{b}}]^m$, where $\tilde{\mathbf{x}} = (\mathbf{x}_1, \ldots, \mathbf{x}_m) \in \mathbb{S}^m_{d,1}(\breve{b}(d+1))$, and $\breve{\mathbf{b}} := (b, \ldots, b)$ is a d-dimensional vector, then we have:

$$\begin{aligned}\left|u_{n,1}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}[u_{n,1}(\varphi, \tilde{\mathbf{x}})]\right| \leq & \left|u_{n,1}(\varphi, \tilde{\mathbf{x}}) - u_{n,1}(\varphi, \tilde{\mathbf{x}}')\right| + \left|\mathbb{E}[u_{n,1}(\varphi, \tilde{\mathbf{x}}')] - \mathbb{E}[u_{n,1}(\varphi, \tilde{\mathbf{x}})]\right| \\ & + \left|u_{n,1}(\varphi, \tilde{\mathbf{x}}') - \mathbb{E}[u_{n,1}(\varphi, \tilde{\mathbf{x}}')]\right|.\end{aligned}$$

As we explained before, to establish uniform convergence rates, we will be studying the convergence of the truncated part and the remainder part of $u_{n,1}(\varphi, \tilde{\mathbf{x}})$ respectively.

Truncated Part:

Notice that

$$\left|u^{(T)}_{n,1}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}\left(u^{(T)}_{n,1}(\varphi, \tilde{\mathbf{x}})\right)\right| = \frac{(n-m)!}{n!}\left|\sum_{\mathbf{i}\in I(m,n)} \mathcal{W}^{(T)}(\tilde{\mathbf{X}}_{\mathbf{i}}, \tilde{\mathbf{Y}}_{\mathbf{i}})\right|,$$

where

$$\mathcal{W}^{(T)}(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) := \mathcal{G}^{(T)}_{\varphi,\tilde{\mathbf{x}},1}(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) - \mathbb{E}\left[\mathcal{G}^{(T)}_{\varphi,\tilde{\mathbf{x}},1}(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})\right].$$

Mirroring the approach used in the proof of Theorem 3.1, we begin by establishing continuity estimates for the random fields $\tilde{\mathbf{z}} \mapsto \mathcal{W}^{(T)}(\tilde{\mathbf{z}})$ so that we get to control the probability that $\mathcal{W}^{(T)}(\tilde{\mathbf{z}})$ and $\mathcal{W}^{(T)}(\tilde{\mathbf{z}}')$ are too far apart when $\tilde{\mathbf{z}} = (\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ and $\tilde{\mathbf{z}}' = (\tilde{\mathbf{x}}', \tilde{\mathbf{y}})$ are close. Building on the framework established in Proposition 1 of [117], we present the following proposition to determine the behavior of $\left(u^{(m)}_{n,1}(\mathcal{W}^{(T)}(\tilde{\mathbf{x}}', \tilde{\mathbf{y}})) - u^{(m)}_{n,1}(\mathcal{W}^{(T)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}))\right)$.

Proposition A.3. *Let* $\mathbf{x} \in \mathbb{S}_{d,1}(\breve{b}(d+1))$, $n \geq 1$, $0 < \breve{b} < \left(e^{-16\sqrt{2}} \wedge d^{-1}\right)$, $0 < a \leq e^{-1}\|f\|_\infty|\log\breve{b}|/\breve{b}^{d+1/2}$, *and take the unique*

$$\delta \in \left(0, e^{-1}\right] \textit{ that satisfies } \delta|\log\delta| = \frac{\breve{b}^{d+1/2}a}{2\|f\|_\infty|\log\breve{b}|}. \tag{A.32}$$

Then, for all $h \in \mathbb{R}$, *we have*

$$\begin{aligned}&\mathbb{P}\left(\sup_{\tilde{\mathbf{x}}'\in\tilde{\mathbf{x}}+[-\breve{\mathbf{b}},\breve{\mathbf{b}}]^m}\left|u^{(m)}_{n,1}(\mathcal{W}^{(T)}(\tilde{\mathbf{x}}', \tilde{\mathbf{y}}))\right| \geq h + 2a^m, \left|u^{(m)}_{n,1}(\mathcal{W}^{(T)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}))\right| \leq h\right) \\ \leq\ & C_{\varphi,dm}\exp\left(-\frac{1}{100^2 d^4\|f\|^2_\infty}\cdot\left(\frac{n^{1/2}\breve{b}^{d+1/2}a^m}{|\log\delta||\log\breve{b}|}\right)^2\right),\end{aligned} \tag{A.33}$$

where $C_{\varphi,dm} > 0$ *is a constant that depends only on the function* $\varphi(\cdot)$*, the dimension* d*, and the degree* m*.*

Proof of Proposition A.3. Similar to the proof of Proposition A.1, by a union bound the probability in (A.33) is

$$\leq \mathbb{P}\left(\left\{\sup_{\tilde{\mathbf{x}}'\in\tilde{\mathbf{x}}+[-\breve{\mathbf{b}},\breve{\mathbf{b}}]^m}\left|\left(u_{n,1}^{(m)}(\mathcal{W}^{(T)}(\tilde{\mathbf{x}}',\tilde{\mathbf{y}}))-u_{n,1}^{(m)}(\mathcal{W}^{(T)}(\tilde{\mathbf{x}},\tilde{\mathbf{y}}))\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\right|\geq a^m\right\}\right.$$
$$\left.\cap\left\{\sum_{i\in I(m,n)}\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\leq 2\cdot 2^m C_n^m\|f\|_\infty^m\delta^m\right\}\right) \tag{A.34}$$
$$+\mathbb{P}\left(\sum_{i\in I(m,n)}\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\geq 2\cdot 2^m C_n^m\|f\|_\infty^m\delta^m\right) \tag{A.35}$$
$$+\mathbb{P}\left(\sup_{\tilde{\mathbf{x}}'\in\tilde{\mathbf{x}}+[-\breve{\mathbf{b}},\breve{\mathbf{b}}]^m}\left|\left(u_{n,1}^{(m)}(\mathcal{W}^{(T)}(\tilde{\mathbf{x}}',\tilde{\mathbf{y}}))-u_{n,1}^{(m)}(\mathcal{W}^{(T)}(\tilde{\mathbf{x}},\tilde{\mathbf{y}}))\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m(\delta)\right\}}\right|\geq a^m\right), \tag{A.36}$$

where C_m^n as usual is equal to $n!/\{m!(n-m)!\}$. Let us begin with (A.34). Following the same reasoning as Proposition A.1, on the event

$$\left\{(C_n^m)^{-1}\sum_{i\in I(m,n)}\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus S_{d,1}^m(\delta)\right\}}\leq 2\cdot 2^m\|f\|_\infty^m\delta^m\right\},$$

we have

$$\begin{aligned}
&\left|\left(u_{n,1}^{(m)}(\mathcal{W}^{(T)}(\tilde{\mathbf{x}}',\tilde{\mathbf{y}}))-u_{n,1}^{(m)}(\mathcal{W}^{(T)}(\tilde{\mathbf{x}},\tilde{\mathbf{y}}))\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus S_{d,1}^m(\delta)\right\}}\right|\\
&\leq \sup_{(\tilde{\mathbf{x}},\tilde{\mathbf{x}}')\in\mathbb{S}_{d,1}^{2m}(b)}\left|\mathcal{W}^{(T)}(\tilde{\mathbf{x}},\tilde{\mathbf{y}})-\mathcal{W}^{(T)}(\tilde{\mathbf{x}}',\tilde{\mathbf{y}})\right|(C_n^m)^{-1}\sum_{i\in I(m,n)}\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\\
&\leq 4\cdot 2\cdot 2^m\omega_{n,1}\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\tilde{\mathcal{K}}_{\mathbf{\Lambda}_{n,1}(\tilde{\mathbf{x}})}(\tilde{\mathbf{x}},\tilde{\mathbf{X}})\cdot\|f\|_\infty^m\delta^m\\
&\leq 4\cdot 2\cdot 2^m\omega_{n,1}\cdot\breve{b}^{-dm}(\breve{b}^{-1}+d)^{m/2}\|f\|_\infty^m\delta^m.
\end{aligned}$$

The latter equation is obtained by (A.14), (A.15), and Lemma 2 ([117]). Therefore, we have

$$\left|\left(u_{n,1}^{(m)}(\mathcal{W}^{(T)}(\tilde{\mathbf{x}}',\tilde{\mathbf{y}}))-u_{n,1}^{(m)}(\mathcal{W}^{(T)}(\tilde{\mathbf{x}},\tilde{\mathbf{y}}))\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\right|\leq\frac{8\omega_{n,1}(1+\breve{b}d)^{m/2}}{|\log(\delta)|^m|\log(\breve{b})|^m}a^m<a^m. \tag{A.37}$$

Note that $0<\delta\leq e^{-1}$ and $0<\breve{b}<\left(e^{-8\sqrt{2}}\wedge d^{-1}\right)$ by assumption. This implies that the probability in (A.34) equals zero. Next, we apply Hoeffding's inequality to control the probability in (A.35). Since $0\leq\mathbf{1}_{\left\{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\leq 1$ and

$$\mu=\mathbb{E}\left[\mathbf{1}_{\left\{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\right]\leq\prod_{j=1}^m\mathbb{E}\left[\mathbf{1}_{\left\{\mathbf{x}_j\in\mathbb{S}_{d,1}\setminus\mathbb{S}_{d,1}(\delta)\right\}}\right]\leq\frac{2^m\|f\|_\infty^m\delta^m}{((d-1)!)^m},$$

we have, for $t=2\cdot 2^m\|f\|_\infty^m\delta^m-\mu$, by considering (A.32)

$$\mathbb{P}\left(\sum_{i\in I(m,n)}\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}-\mu\geq t\right)\leq\exp\left\{-2[n/m]\left((2((d-1)!)^m-1)\cdot\frac{2^m\|f\|_\infty^m}{((d-1)!)^m}\delta^m\right)^2\right\}$$

$$\leq \exp\left\{-2[n/m]\left(\frac{\breve{b}^{d+1/2}a}{|\log\delta||\log\breve{b}|}\right)^{2m}\right\}.$$

Moving on to (A.36), then

$$\begin{aligned}
& u_{n,1}^{(m)}\left(\mathcal{W}^{(T)}(\tilde{\mathbf{x}}',\tilde{\mathbf{y}}) - \mathcal{W}^{(T)}(\tilde{\mathbf{x}},\tilde{\mathbf{y}})\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\right) \\
= \; & m u_{n,1}^{(1)}\left(\pi_{1,m}\left[\left(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}',1}^{(T)} - \mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(T)}\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\right]\right) \\
& + \sum_{q=2}^{m}\frac{m!}{(m-q)!}u_{n,1}^{(q)}\left(\pi_{q,m}\left[\left(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}',1}^{(T)} - \mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(T)}\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\right]\right), \qquad \text{(A.38)}
\end{aligned}$$

where the linear term

$$\begin{aligned}
& m u_{n,1}^{(1)}\left(\pi_{1,m}\left[\left(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}',1}^{(T)} - \mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(T)}\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\right]\right) \\
= \; & \frac{m}{n}\sum_{i=1}^{n}\pi_{1,m}\left[\left(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}',1}^{(T)} - \mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(T)}\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\right](\tilde{\mathbf{X}}_i,\tilde{\mathbf{Y}}_i) \qquad \text{(A.39)}
\end{aligned}$$

can be treated similarly to the proof of Proposition A.1. Now, for the nonlinear term, let us first introduce the following class of functions:

$$\mathcal{F} := \left\{\mathcal{G}(\varphi,\tilde{\mathbf{x}}') - \mathcal{G}(\varphi,\tilde{\mathbf{x}}) : \tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m(\breve{b}(d+1) \text{ and } \tilde{\mathbf{x}}'\in\tilde{\mathbf{x}} + [-\breve{\mathbf{b}},\breve{\mathbf{b}}]^m\right\},$$

then we have for all $\varepsilon > 0$

$$\begin{aligned}
& \mathbb{P}\left(\sup_{\tilde{\mathbf{x}}'\in\tilde{\mathbf{x}}+[-\breve{\mathbf{b}},\breve{\mathbf{b}}]^m}\left|\sum_{q=2}^{m}\frac{m!}{(m-q)!}u_{n,1}^{(q)}\left(\pi_{q,m}\left[\left(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}',1}^{(T)} - \mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(T)}\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\right]\right)\right| \geq \varepsilon\right) \\
\equiv \; & \mathbb{P}\left(\left\|\sum_{q=2}^{m}\frac{m!}{(m-q)!}u_{n,1}^{(q)}\left(\pi_{q,m}\left[\left(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}',1}^{(T)} - \mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(T)}\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\right]\right)\right\|_{\mathcal{F}} \geq \varepsilon\right).
\end{aligned}$$

We have

$$\begin{aligned}
& \mathbb{E}\left[\left\|n^{1-m}\sum_{I_m^n}\varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)}\left[\left(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}',1}^{(T)} - \mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(T)}\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i\in\mathbb{S}_{d,1}^m\setminus\mathbb{S}_{d,1}^m(\delta)\right\}}\right]\right\|_{\mathcal{F}}\right] \\
\leq \; & 2C\mathbb{E}\left[\left|n^{1-m}\sum_{I_m^n}\varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)}\left[\varphi(\mathbf{Y}_{i_1},\ldots,\mathbf{Y}_{i_m})\right]\right|\right].
\end{aligned}$$

Using the same reasoning as in [5], one can find a positive constant $c_0 > 0$ such that

$$\mathbb{E}\left[\left|n^{1-m}\sum_{I_m^n}\varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)}\left[\varphi(\mathbf{Y}_{i_1},\ldots,\mathbf{Y}_{i_m})\right]\right|\right] < c_0.$$

Now, making use of (A.32) and applying Proposition 4 of [5] gives us, for $\varepsilon = a^m n^{-1/2}$,

$$\begin{aligned}
&\mathbb{P}\left(\left\|\sum_{q=2}^{m} \frac{m!}{(m-q)!} u_{n,1}^{(q)}\left(\pi_{q,m}\left[\left(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}',1}^{(T)} - \mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(T)}\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i \in \mathbb{S}_{d,1}^m \setminus \mathbb{S}_{d,1}^m(\delta)\right\}}\right]\right)\right\|_{\mathcal{F}} \geqslant a^m n^{-1/2}\right) \\
\leq\ & 2\exp\left(-\frac{a^m n^{1/2}}{2^{m+5} m^{m+1} 2\omega_{n,1} \breve{b}^{-dm}(\breve{b}^{-1}+d)^{m/2} c_0}\right) \\
\leq\ & 2\exp\left(-\frac{(2\delta\left|\log\delta\right| \|f\|_\infty \left|\log\breve{b}\right|)^m n^{1/2}}{2^{m+5} m^{m+1} 2\omega_{n,1} \breve{b}^{m(d+1/2)} \breve{b}^{-dm}(\breve{b}^{-1}+d)^{m/2} c_0}\right) \\
\leq\ & 2\exp\left(-\frac{(\delta\left|\log\delta\right| \|f\|_\infty \left|\log\breve{b}\right|)^m n^{1/2}}{2^{6} m^{m+1} \omega_{n,1} (1+\breve{b}d)^{m/2} c_0}\right).
\end{aligned}$$

We can find a constant $\mathbb{C}_1 > 0$, such that

$$\frac{(\delta\left|\log\delta\right| \|f\|_\infty)^m}{2^6 m^{m+1}(1+\breve{b}d)^{m/2} c_0} \geq \mathbb{C}_1,$$

which implies

$$\begin{aligned}
\exp\left(-\frac{(\delta\left|\log\delta\right| \|f\|_\infty \left|\log\breve{b}\right|)^m n^{1/2}}{2^{6} m^{m+1} \omega_{n,1} (1+\breve{b}d)^{m/2} c_0}\right) &\leq \exp\left(-\mathbb{C}_1 \left|\log\breve{b}\right|^m n^{1/2-1/p}\right) \\
&\leq \exp\left(-\mathbb{C}_1 m \left|\log\breve{b}\right| n^{1/2-1/p}\right).
\end{aligned}$$

Therefore, we readily infer that

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left\|\sum_{q=2}^{m} \frac{m!}{(m-q)!} u_{n,1}^{(q)}\left(\pi_{q,m}\left[\left(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}',1}^{(T)} - \mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(T)}\right)\mathbf{1}_{\left\{\tilde{\mathbf{x}}_i \in \mathbb{S}_{d,1}^m \setminus \mathbb{S}_{d,1}^m(\delta)\right\}}\right]\right)\right\|_{\mathcal{F}} \geqslant a^m n^{-1/2}\right) < \infty.$$

Hence, the proof of the proposition is complete by an application of the Borel-Cantelli lemma. □

Remainder Part:

We now consider the remaining part. Recall that the U-statistic $u_{n,1}^{(R)}(\varphi, \tilde{\mathbf{x}})$ is related to the unbounded kernel given by

$$\mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(R)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})\mathbf{1}_{\{\varphi(\mathbf{y}) > \lambda\xi_n^{1/(1+\gamma)}\}}.$$

We have to establish that it is negligible, meaning that

$$\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{d,1}^m} \frac{\sqrt{n}\breve{b}^{m(d+1/2)}\left|u_{n,1}^{(m)}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(R)}) - \mathbb{E}\left(u_{n,1}^{(m)}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(R)})\right)\right|}{\left|\log\breve{b}\right|^m (\log n)^{3/2}} = o_{a.s}(1). \tag{A.40}$$

For $\tilde{\mathbf{x}}, \tilde{\mathbf{y}} \in \mathbb{S}_{d,1}^m$, observe that

$$\left|\mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})\right| \leq \breve{b}^{-dm}(\breve{b}^{-1}+d)^{m/2}\left|\varphi(\tilde{\mathbf{y}})\right| =: \widetilde{F}(\tilde{\mathbf{y}}).$$

Taking into account that $\widetilde{F}$ is symmetric, we have

$$\left|u_{n,1}^{(m)}\left(\mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(R)}\right)\right| \leq u_{n,1}^{(m)}\left(\widetilde{F}\mathbf{1}_{\{\widetilde{F}>\lambda\xi_n^{1/(1+\gamma)}\}}\right),$$

where $u_{n,1}^{(m)}\left(\widetilde{F}(\mathbf{y})\mathbf{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right)$ is a U-statistic based on the U- kernel $\widetilde{F}\mathbf{1}_{\{\varphi>\lambda\xi_n^{1/(1+\gamma)}\}}$,

$$\begin{aligned}\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\frac{\sqrt{n}\breve{b}^{m(d+1/2)}\left|u_{n,1}^{(m)}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(R)})\right|}{\left|\log\breve{b}\right|^m(\log n)^{3/2}} &\leq \frac{\sqrt{n}(1+\breve{b}d)^{m/2}}{\left|\log\breve{b}\right|^m(\log n)^{3/2}}u_{n,1}^{(m)}\left(\widetilde{F}\mathbf{1}_{\{\widetilde{F}>\lambda\xi_n^{1/(1+\gamma)}\}}\right)\\ &\leq C_7\xi_n u_{n,1}^{(m)}\left(\widetilde{F}\mathbf{1}_{\{\widetilde{F}>\lambda\xi_n^{1/(1+\gamma)}\}}\right) \qquad (A.41)\end{aligned}$$

and

$$\begin{aligned}\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\frac{\sqrt{n}\breve{b}^{m(d+1/2)}\left|u_{n,1}^{(m)}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(R)})\right|}{\left|\log\breve{b}\right|^m(\log n)^{3/2}} &\leq C_7\xi_n\mathbb{E}\left(u_{n,1}^{(m)}\left(\widetilde{F}\mathbf{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right)\right)\\ &\leq C_7\mathbb{E}\left(\widetilde{F}^{2+\gamma}\mathbf{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right).\end{aligned}$$

Therefore, as $n\longrightarrow\infty$, we have, almost surely,

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\frac{\sqrt{n}\breve{b}^{m(d+1/2)}\left|u_{n,1}^{(m)}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}^{(R)})\right|}{\left|\log\breve{b}\right|^m(\log n)^{3/2}} = o(1). \qquad (A.42)$$

Hence, to achieve the proof, it remains to establish that

$$u_{n,1}^{(m)}\left(\widetilde{F}\mathbf{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right) = o_{a.s}\left(\left(s_m^{-1}\xi_n\right)^{-1/2}\right). \qquad (A.43)$$

An application of the Chebyshev's inequality, for any $\eta>0$, gives

$$\begin{aligned}&\mathbb{P}\left\{\left|u_{n,1}^{(m)}\left(\widetilde{F}\mathbf{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right)-\mathbb{E}\left(u_{n,1}^{(m)}\left(\widetilde{F}\mathbf{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right)\right)\right|\geq\eta(s_m^{-1}\xi_n)^{-1/2}\right\}\\ \leq\ &\eta^{-2}(s_m^{-1}\xi_n)Var\left(u_{n,1}^{(m)}\left(\widetilde{F}\mathbf{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right)\right)\leq m\eta^{-2}\xi_n\mathbb{E}\left(\widetilde{F}^2\mathbf{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right)\\ \leq\ &\frac{m}{n^2}\eta^{-2}(\xi_n)^{1+\gamma}\mathbb{E}\left(\widetilde{F}^2\mathbf{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right)\leq\eta'\mathbb{E}\left(\widetilde{F}^3\mathbf{1}_{\{\varphi(\mathbf{Y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right)\frac{1}{n^2}=o(1),\end{aligned}$$

so by using the fact that

$$\eta'\mathbb{E}\left(\widetilde{F}^3\mathbf{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right)\sum_{n\geq1}\frac{1}{n^2}<\infty,$$

we deduce that

$$\sum_{n\geq1}\mathbb{P}\left\{\left|u_{n,1}^{(m)}\left(\widetilde{F}\mathbf{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right)-\mathbb{E}\left(u_{n,1}^{(m)}\left(\widetilde{F}\mathbf{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right)\right)\right|\geq\eta(s_m^{-1}\xi_n)^{-1/2}\right\}<\infty.$$

Finally, note that (A.41) implies

$$\mathbb{E}\left(u_{n,1}^{(m)}\left(\widetilde{F}\mathbf{1}_{\{\varphi(\mathbf{y})>\lambda\xi_n^{1/(1+\gamma)}\}}\right)\right)=o\left(s_m^{-1}\xi_n^{-1/2}\right).$$

The preceding results of the arbitrary choice of $\lambda > 0$ show that (A.43) holds, which, by combining with (A.42) and (A.41), completes the proof of (A.40). We finally obtain

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}} \left|u_{n,1}(\varphi,\tilde{\mathbf{x}}) - \mathbb{E}[u_{n,1}(\varphi,\tilde{\mathbf{x}})]\right| = O\left(\frac{|\log \breve{b}|^m(\log n)^{3/2}}{\breve{b}^{m(d+1/2)}\sqrt{n}}\right) \quad \text{a.s.}$$

Hence, the proof is complete. □

Proof of Theorem 3.3. Recall (A.2), and we have:

$$\left|\widehat{r}^{(m)}_{n,1}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,1}(\tilde{\mathbf{x}})) - \widehat{\mathbb{E}}\left(\widehat{r}^{(m)}_{n,1}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,1}(\tilde{\mathbf{x}}))\right)\right| \leq \mathscr{I}_{1,1} + \mathscr{I}_{1,2}, \tag{A.44}$$

where

$$\begin{aligned}\mathscr{I}_{1,1} &= \frac{\left|u_{n,1}(\varphi,\tilde{\mathbf{x}}) - \mathbb{E}\left(u_{n,1}(\varphi,\tilde{\mathbf{x}})\right)\right|}{\left|u_{n,1}(1,\tilde{\mathbf{x}})\right|},\\ \mathscr{I}_{1,2} &= \frac{\left|\mathbb{E}\left(u_{n,1}(\varphi,\tilde{\mathbf{x}})\right)\right| \cdot \left|u_{n,\ell}(1,\tilde{\mathbf{x}}) - \mathbb{E}\left(u_{n,1}(1,\tilde{\mathbf{x}})\right)\right|}{\left|u_{n,1}(1,\tilde{\mathbf{x}})\right| \cdot \left|\mathbb{E}\left(u_{n,1}(1,\tilde{\mathbf{x}})\right)\right|}.\end{aligned}$$

Notice that, given the imposed hypothesis and previously obtained results, and for some positive constants $c_1, c_2 > 0$, we readily infer

$$\begin{aligned}\sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}} \left|u_{n,1}(1,\tilde{\mathbf{x}})\right| &= c_1 \qquad \text{a.s.},\\ \sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}} \left|\mathbb{E}\left(u_{n,1}(1,\tilde{\mathbf{x}})\right)\right| &= c_2,\\ \sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}} \left|\mathbb{E}\left(u_{n,1}(\varphi,\tilde{\mathbf{x}})\right)\right| &= O(1).\end{aligned}$$

Hence by Theorem 3.2, for some $c'' > 0$, we get with probability 1:

$$\begin{aligned}&\sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}} \frac{\breve{b}^{m(d+1/2)}\sqrt{n}\left|\widehat{r}^{(m)}_{n,1}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,1}(\tilde{\mathbf{x}})) - \widehat{\mathbb{E}}\left(\widehat{r}^{(m)}_{n,1}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,1}(\tilde{\mathbf{x}}))\right)\right|}{|\log \breve{b}|^m(\log n)^{3/2}}\\ \leq\ &\sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}} \frac{\breve{b}^{m(d+1/2)}\sqrt{n}\,(\mathscr{I}_{1,1})}{|\log \breve{b}|^m(\log n)^{3/2}} + \sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}} \frac{\breve{b}^{m(d+1/2)}\sqrt{n}\,(\mathscr{I}_{1,2})}{|\log \breve{b}|^m(\log n)^{3/2}} \leq c''.\end{aligned}$$

Hence, the proof is complete. □

Proof of Theorem 3.4. By (A.3), it suffices to establish

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}} \left|\mathbb{E}\left(u_{n,1}(\varphi,\tilde{\mathbf{x}})\right) - \mathcal{R}(\varphi,\tilde{\mathbf{x}})\right| = O\left(\breve{b}^{1/2}\right), \tag{A.45}$$

and

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}} \left|\mathbb{E}\left(u_{n,1}(1,\tilde{\mathbf{x}})\right) - \tilde{f}(\tilde{\mathbf{x}})\right| = O\left(\breve{b}^{1/2}\right). \tag{A.46}$$

Let us start with (A.45). We have

$$\begin{aligned}
\mathbb{E}\left[u_{n,1}(\varphi,\tilde{\mathbf{x}})\right] &= \frac{(n-m)!}{n!}\sum_{\mathbf{i}\,\in I(m,n)}\mathbb{E}\left[\mathcal{G}_{\varphi,\tilde{\mathbf{x}},1}(\tilde{\mathbf{X}}_{\mathbf{i}},\tilde{\mathbf{Y}}_{\mathbf{i}})\right]\\
&= \mathbb{E}\left[\varphi(\tilde{\mathbf{Y}})\tilde{\mathcal{K}}_{\bar{\boldsymbol{\Lambda}}_{n,1}(\tilde{\mathbf{x}})}(\tilde{\mathbf{X}})\right]\\
&= \mathbb{E}\left[\tilde{\mathcal{K}}_{\bar{\boldsymbol{\Lambda}}_{n,1}(\tilde{\mathbf{x}})}\mathbb{E}\left[\varphi(\mathbf{Y}_1,\ldots,\mathbf{Y}_m)\mid(\mathbf{X}_1,\ldots,\mathbf{X}_m)=\tilde{\mathbf{x}}\right]\right]\\
&= \int_{\mathbb{S}^m_{d,1}} r^{(m)}(\varphi,\tilde{\mathbf{u}})\tilde{f}(\tilde{\mathbf{u}})\tilde{\mathcal{K}}_{\bar{\boldsymbol{\Lambda}}_{n,1}(\tilde{\mathbf{x}})}(\tilde{\mathbf{u}})\mathbf{d}\tilde{\mathbf{u}}\\
&= \int_{\mathbb{S}^m_{d,1}} \mathcal{R}(\varphi,\tilde{\mathbf{u}})\prod_{j=1}^m K_{(\alpha_j,\beta_j)}\left(\mathbf{u}_{i_j}\right)\mathbf{d}\tilde{\mathbf{u}}\\
&= \mathbb{E}[\mathcal{R}(\varphi,\tilde{\boldsymbol{\xi}}_{\tilde{\mathbf{x}}})],
\end{aligned}$$

where $\tilde{\boldsymbol{\xi}}_{\tilde{\mathbf{x}}}=(\xi_{\mathbf{x}_1},\ldots,\xi_{\mathbf{x}_m})$ such that $\xi_{\mathbf{x}_j}=\left(\xi_{x_{j_1}},\ldots,\xi_{x_{j_d}}\right)\sim$ Dirichlet$(\boldsymbol{\alpha}_j,\beta_j)$; $j=1,\ldots,m$. By a second-order Taylor expansion around $\tilde{\boldsymbol{\xi}}_{\tilde{\mathbf{x}}}=\tilde{\mathbf{x}}$, we get

$$\begin{aligned}
\mathbb{E}\left[\mathcal{R}(\varphi,\tilde{\boldsymbol{\xi}}_{\tilde{\mathbf{x}}})\right] &= \mathcal{R}(\varphi,\tilde{\mathbf{x}})+\sum_{i=1}^m\sum_{\ell=1}^d\frac{\partial\mathcal{R}(\varphi,\tilde{\mathbf{x}})}{\partial x_{i_\ell}}\mathbb{E}(\xi_{xi_\ell}-x_{i_\ell})+\frac{1}{2}\sum_{i=1}^m\sum_{\ell=1}^d\frac{\partial^2\mathcal{R}(\varphi,\underline{\tilde{\mathbf{x}}})}{\partial x_{i_\ell}^2}\mathbb{E}(\xi_{xi_\ell}-x_{i_\ell})^2\\
&\quad+\sum_{i,j=1,i\neq j}^m\ \sum_{\ell,r=1,\ell\neq r}^d\frac{\partial^2\mathcal{R}(\varphi,\underline{\tilde{\mathbf{x}}})}{\partial x_{i_\ell}\partial x_{j_r}}\mathbb{E}\left\{(\xi_{xi_\ell}-x_{i_\ell})(\xi_{xj_r}-x_{j_r})\right\},
\end{aligned}$$

for some $\underline{\tilde{\mathbf{x}}}$ joining $\tilde{\boldsymbol{\xi}}_{\tilde{\mathbf{x}}}$ and $\tilde{\mathbf{x}}$. Keep in mind that $\xi_{\mathbf{x}_j}$ are vectors whose components are beta-distributed, which implies that

$$\mathbb{E}\left[\xi_{x_{i_\ell}}\right]=x_{i_\ell}+\breve{b}\,(1-(d+1)x_{i_\ell})+O\left(\breve{b}^2\right),$$

$$\operatorname{Cov}\left(\xi_{x_{i_\ell}},\xi_{x_{j_r}}\right)=\breve{b}x_{i_\ell}\left(\mathbf{1}_{\{i_\ell=j_r\}}-x_{j_r}\right)+O\left(\breve{b}^2\right), \tag{A.47}$$

$$\mathbb{E}\left[\left(\xi_{x_{i_\ell}}-x_{i_\ell}\right)\left(\xi_{x_{j_r}}-x_{j_r}\right)\right]=\breve{b}x_{i_\ell}\left(\mathbf{1}_{\{i_\ell=j_r\}}-x_k\right)+O\left(\breve{b}^2\right). \tag{A.48}$$

Then, the Cauchy-Schwartz inequality, (A.47), and (A.48) yields uniformly on $\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}$

$$\begin{aligned}
\left|\mathbb{E}\left[u_{n,1}(\varphi,\tilde{\mathbf{x}})\right]-\mathcal{R}(\varphi,\tilde{\mathbf{x}})\right| &= \sum_{i=1}^m\sum_{\ell=1}^d O\left(\mathbb{E}\left|\xi_{xi_\ell}-x_{i_\ell}\right|\right)+\frac{1}{2}\sum_{i=1}^m\sum_{\ell=1}^d O\left(\mathbb{E}\left|\xi_{xi_\ell}-x_{i_\ell}\right|^2\right)\\
&\quad+\sum_{i,j=1,i\neq j}^m\ \sum_{\ell,r=1,\ell\neq r}^d O\left(\mathbb{E}\left|(\xi_{xi_\ell}-x_{i_\ell})(\xi_{xj_r}-x_{j_r})\right|\right)\\
&\leq \sum_{i=1}^m\sum_{\ell=1}^d O\left(\sqrt{\mathbb{E}\left|\xi_{xi_\ell}-x_{i_\ell}\right|^2}\right)+\sum_{i=1}^m\sum_{\ell=1} O(\breve{b})+\sum_{i=1}^m\sum_{\ell=1} O(\breve{b}^2)\\
&\leq O(\breve{b}^{1/2})+O(\breve{b})+O(\breve{b}^2)\\
&\leq O(\breve{b}^{1/2})(1+o(1)).
\end{aligned}$$

Consequently, we deduce that

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{d,1}}\left|\mathbb{E}\left(u_{n,1}(\varphi,\tilde{\mathbf{x}})\right)-\mathcal{R}(\varphi,\tilde{\mathbf{x}})\right|=O\left(\breve{b}^{1/2}\right), \tag{A.49}$$

which implies that for $\varphi \equiv 1$,

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^{m}} \left|\mathbb{E}\left(u_{n,1}(1,\tilde{\mathbf{x}})\right) - \tilde{f}(\tilde{\mathbf{x}})\right| = O\left(\breve{b}^{1/2}\right). \tag{A.50}$$

As an immediate consequence of the last two equations, we infer that

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^{m}} \left|\mathbb{E}\left(u_{n,1}(\varphi,\tilde{\mathbf{x}})\right) - r^{(m)}(\varphi,\tilde{\mathbf{x}})\mathbb{E}\left(u_{n,1}(1,\tilde{\mathbf{x}})\right)\right| = O\left(\breve{b}^{1/2}\right). \tag{A.51}$$

This completes the proof of the theorem. □

A.1.3. Proofs of the results of Section 3.3

Before we start the proofs, we will state some lemmas that are necessary to obtain the desired results. It is worth mentioning that we will follow the steps of [134] while making the appropriate changes to fit our general setting.

Lemma A.4. *Under assumptions **(A.1)–(A.4)**, and if* $\mathbb{E}\varphi^2 < \infty$, *the Hájek projection* $\widehat{U}_{n,1}$ *of* $\mathbb{U}$ *satisfies, as* $n \to \infty$:

(i)

$$\lim_{n\to\infty} \mathbb{E}\left[\sqrt{n\breve{b}^{d/2}}\left(\widehat{U}_{n,1} - \theta_n\right)\right]^2 = \sigma^2(\varphi),$$

where

$$\sigma^2(\varphi) := \sum_{i=1}^{m}\sum_{j=1}^{m} \mathbf{1}_{\{\mathbf{x}_i=\mathbf{x}_j\}} r_{ij}(\tilde{\mathbf{x}}) \int_{\mathbb{S}_{d,1}} K_{\alpha,\beta}^2(\mathbf{x},\mathbf{t})\mathrm{d}\mathbf{t}/f(\mathbf{x}_i) > 0, \tag{A.52}$$

(ii) *and if, in addition, assumption **(A.5)** is verified, we have*

$$\sqrt{n\breve{b}^{d/2}}\left(\widehat{U}_{n,1} - \theta_n\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(\varphi)\right). \tag{A.53}$$

In the following lemma, we show that $\mathbb{U}$ has the same asymptotic distribution as $\widehat{U}_{n,1}$.

Lemma A.5. *Under assumption **(A.1)–(A.6)**, we have, as* $n \to \infty$,

$$\sqrt{n\breve{b}^{d/2}}\left(\mathbb{U}_{n,1} - \theta_n\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(\varphi)\right). \tag{A.54}$$

Specification of $\sigma^2(\varphi)$ leads to the following lemma.

Lemma A.6. *[134] Under assumptions **(A.1)–(A.6)**, we have, as* $n \to \infty$,

$$\left(n\breve{b}^{d/2}\right)^{1/2}\left[\mathbb{U}_{n,1}(\varphi_1,\tilde{\mathbf{x}}) - \theta_n(\varphi_1), \mathbb{U}_{n,1}(\varphi_2,\tilde{\mathbf{x}}) - \theta_n(\varphi_2)\right] \xrightarrow{\mathcal{D}} \mathscr{N}(\mathbf{0},\Sigma),$$

with

$$\Sigma = \left[\begin{array}{cc} \sigma^2(\varphi_1,\varphi_1) & \sigma^2(\varphi_1,\varphi_2) \\ \sigma^2(\varphi_1,\varphi_2) & \sigma^2(\varphi_2,\varphi_2) \end{array}\right],$$

and where for two functions $g_1(\cdot)$ and $g_2(\cdot)$,

$$\begin{aligned}\sigma^2(g_1,g_2) &= \sum_{j=1}^{m}\sum_{l=1}^{m}\mathbf{1}_{\{x_j-x_l\}}r_{jl}^{g_1g_2}(\tilde{\mathbf{x}})\int K_{\alpha,\beta}^2(\mathbf{x},\mathbf{t})\mathrm{d}\mathbf{t}/f_{\mathbf{X}}\left(x_j\right),\\ r_{jl}^{g_1g_2}(\tilde{\mathbf{x}}) &= \mathbb{E}\left[g_1\left(\mathbf{Y}_1,\ldots,\mathbf{Y},\ldots,\mathbf{Y}_m\right)g_2\left(\mathbf{Y}_{m+1},\ldots,\mathbf{Y},\ldots,\mathbf{Y}_{2m}\right)\mid\cdots\right],\end{aligned}$$

with $\mathbf{Y}$ entering in the j^{th} and l^{th} positions.

Proof of Lemma A.4. We write

$$\begin{aligned}\theta_n &= \mathbb{E}[\mathbb{U}_{n,1}(\varphi,\tilde{\mathbf{x}})]\\ &= N^{-1}\int_{\mathbb{S}_{d,1}^m} r^{(m)}(\varphi,\tilde{\mathbf{t}})\tilde{f}(\tilde{\mathbf{t}})\tilde{\mathcal{K}}_{\bar{\Lambda}_{n,1}(\tilde{\mathbf{x}})}(\tilde{\mathbf{t}})\mathrm{d}\tilde{\mathbf{t}} \qquad \text{(A.55)}\\ &= N^{-1}\int_{\mathbb{S}_{d,1}^m} r^{(m)}(\varphi,\tilde{\mathbf{t}})\tilde{f}(\tilde{\mathbf{t}})\prod_{i=1}^{m}K_{(\alpha_i,\beta_i)}\left(\mathbf{t}_i\right)\mathrm{d}\tilde{\mathbf{t}}.\end{aligned}$$

The Hájek projection $\widehat{U}_{n,1}(\varphi,\tilde{\mathbf{x}})$ of $\mathbb{U}_{n,1}(\varphi,\tilde{\mathbf{x}})$ satisfies

$$\widehat{U}_{n,1}-\theta_n = n^{-1}\sum_{i=1}^{n}\bar{\varphi}_n(\mathbf{X}_i,\mathbf{Y}_i),$$

with

$$\bar{\varphi}_n(\mathbf{x},\mathbf{y}) = \sum_{j=1}^{m}\left[\varphi_{n,j}(\tilde{\mathbf{x}},\tilde{\mathbf{y}})-\theta_n\right],$$

and $\varphi_{n,j}(\mathbf{x},\mathbf{y})$ is defined by

$$\varphi_{n,j}(\mathbf{x},\mathbf{y}) = N^{-1}\int_{\mathbb{S}_{d,1}^m\times\mathbb{R}^{qm}}\varphi(\mathbf{Y}_1,\ldots,\mathbf{Y}_{j-1},\mathbf{y},\mathbf{Y}_{j+1},\ldots,\mathbf{Y}_m)\prod_{\substack{r=1\\ r\neq j}}^{m}K_{(\alpha_r,\beta_r)}\left(\mathbf{X}_r\right)K_{(\alpha_j,\beta_j)}\left(\mathbf{x}\right)\mathrm{d}\mathbb{P},$$

where $\mathbb{P}$ represents the underlying probability measure. By independence, we get

$$\begin{aligned}n\mathbb{E}\left(\widehat{U}_{n,1}-\theta_n\right)^2 &= \mathbb{E}\left[\bar{\varphi}_n{}^2(\mathbf{X},\mathbf{Y})\right]\\ &= \sum_{j=1}^{m}\sum_{l=1}^{m}\mathbb{E}\left[\varphi_{n,j}(\mathbf{X},\mathbf{Y})-\theta_n\right]\left[\varphi_{n,l}(\mathbf{X},\mathbf{Y})-\theta_n\right].\end{aligned}$$

Using the fact that $(\mathbf{X},\mathbf{Y})$, $(\mathbf{X}_i,\mathbf{Y}_i)_{1\leq i\leq 2m}$ are i.i.d, we get

$$\begin{aligned}&\mathbb{E}\left[\varphi_{n,j}(\mathbf{X},\mathbf{Y})\varphi_{n,l}(\mathbf{X},\mathbf{Y})\right]\\ =&N^{-2}\int_{\mathbb{S}_{d,1}^{2m}\times\mathbb{R}^{2qm}}\varphi(\mathbf{Y}_1,\ldots,\mathbf{Y}_{j-1},\mathbf{Y},\mathbf{Y}_{j+1},\ldots,\mathbf{Y}_m)\varphi(\mathbf{Y}_{m+1},\ldots,\mathbf{Y}_{m+l-1},\mathbf{Y},\mathbf{Y}_{m+l+1},\ldots,\mathbf{Y}_{2m})\\ &\times\prod_{\substack{r=1\\ r\neq j}}^{m}K_{(\alpha_r,\beta_r)}\left(\mathbf{X}_r\right)\prod_{\substack{s=1\\ s\neq l}}^{m}K_{(\alpha_s,\beta_s)}\left(\mathbf{X}_{m+s}\right)K_{(\alpha_j,\beta_j)}\left(\mathbf{X}\right)K_{(\alpha_l,\beta_l)}\left(\mathbf{X}\right)\mathrm{d}\mathbb{P}.\end{aligned}$$

By condition **(A.2)**, we have, for $\mathbf{x}_j \neq \mathbf{x}_l$

$$K_{(\alpha_j,\beta_j)}(\mathbf{X})\, K_{(\alpha_l,\beta_l)}(\mathbf{X}) \longrightarrow 0.$$

Consequently,

$$\mathbb{E}\left[\varphi_{n,j}(\mathbf{X},\mathbf{Y})\varphi_{n,l}(\mathbf{X},\mathbf{Y})\right] = 0 \ \text{ for all } \ \mathbf{x}_j \neq \mathbf{x}_l.$$

In the case of $\mathbf{x}_j = \mathbf{x}_l$ and $\tilde{\mathbf{x}} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$ being a point of continuity for $r^{(m)}_{j,l}$, we have

$$\frac{\mathbb{E}^2\left[K_{(\alpha_j,\beta_j)}(\mathbf{X}_1)\right]}{\mathbb{E}\left[K^2_{(\alpha_j,\beta_j)}(\mathbf{X}_1)\right]}\mathbb{E}\left[\varphi_{n,j}(\mathbf{X},\mathbf{Y})\varphi_{n,l}(\mathbf{X},\mathbf{Y})\right] \longrightarrow r_{j,l}(\tilde{\mathbf{x}}), \tag{A.56}$$

by a differentiation argument. Now, provided that the density function $f(\cdot)$ is continuous and $f(\mathbf{x}_j) = f(\mathbf{x}_l) > 0$, by the fact that

$$\mathbb{E}\left[K_{(\alpha_j,\beta_j)}(\mathbf{X}_1)\right] = \mathbb{E}[f(\xi_{\mathbf{x}_j})],$$

where

$$\xi_{\mathbf{x}_j} = (\xi_1, \ldots, \xi_d) \sim \text{Dirichlet}\left(\frac{\mathbf{x}_j}{\breve{b}} + \mathbf{1}, \frac{\left(1 - \|\mathbf{x}_j\|_1\right)}{\breve{b}} + 1\right), \quad \mathbf{x}_j \in \mathbb{S}_{d,1},$$

and if $\boldsymbol{\gamma}_{\mathbf{x}} \sim \text{Dirichlet}\left(\frac{2\mathbf{x}_j}{\breve{b}} + \mathbf{1}, \frac{2\left(1 - \|\mathbf{x}_j\|_1\right)}{\breve{b}} + 1\right)$, then

$$\begin{aligned}\mathbb{E}\left[K^2_{(\alpha_j,\beta_j)}(\mathbf{X}_1)\right] &= A_b(\mathbf{x}_j)\mathbb{E}\left[f\left(\boldsymbol{\gamma}_{\mathbf{x}_j}\right)\right]\\ &= A_b(\mathbf{x}_j)\left(f(\mathbf{x}_j) + O\left(\breve{b}^{1/2}\right)\right),\end{aligned}$$

where

$$A_b(\mathbf{x}) := \frac{\Gamma\left(2\left(1 - \|\mathbf{x}\|_1\right)/\breve{b} + 1\right)\prod_{i=1}^{d}\Gamma\left(2x_i/\breve{b} + 1\right)}{\Gamma^2\left(\left(1 - \|\mathbf{x}\|_1\right)/\breve{b} + 1\right)\prod_{i=1}^{d}\Gamma^2\left(x_i/\breve{b} + 1\right)} \cdot \frac{\Gamma^2(1/\breve{b} + d + 1)}{\Gamma(2/\breve{b} + d + 1)}.$$

We readily infer that

$$\frac{\mathbb{E}^2\left[K_{(\alpha_j,\beta_j)}(\mathbf{X}_1)\right]}{\mathbb{E}\left[K^2_{(\alpha_j,\beta_j)}(\mathbf{X}_1)\right]} \sim \frac{\mathbb{E}[f(\xi_{\mathbf{x}_j})]}{A_b(\mathbf{x}_j)\left(f(\mathbf{x}_j) + o(1)\right)}.$$

Hence, by Lemma B.4, we have

$$\breve{b}^{1/2}\mathbb{E}\left[\varphi_{n,j}(\mathbf{X},\mathbf{Y})\varphi_{n,l}(\mathbf{X},\mathbf{Y})\right] = r_{j,l}(\tilde{\mathbf{x}})\frac{A_b(\mathbf{x}_j)\left(f(\mathbf{x}_j) + o(1)\right)}{f(\mathbf{x}_j)}.$$

Using the fact that $r^{(m)}(\varphi, \tilde{\mathbf{x}})$ is bounded in the neighborhood of $\tilde{\mathbf{x}}$, then $\theta_n, n \geq 1$ is bounded, therefore we have $\breve{b}^{1/2}\theta_n^2 \to 0$, from which we have

$$\lim_{n\to\infty}\mathbb{E}\left[\sqrt{n\breve{b}^{d/2}}\left(\widehat{U}_{n,1} - \theta_n\right)\right]^2 = \sigma^2(\varphi), \tag{A.57}$$

where $\sigma^2(\varphi)$ is defined in (A.52), which yields the assertion (i). Now, to prove (A.53), we only need to verify Lyapunov's condition for third moments, i.e.,

$$n^{-1/2}(\breve{b}^{d/2})^{3/2}\mathbb{E}\,|\bar{\varphi}_n(\mathbf{X},\mathbf{Y})|^3 \to 0. \tag{A.58}$$

Using the fact that

$$|a-b|^3 \leq 3(|a|^3 + |a|^2\,|b| + |a|\,|b|^2 + |b|^3),$$

an upper bound for the absolute third moment of $\bar{\varphi}_n(\mathbf{X},\mathbf{Y})$ is dominated by sums of the form

$$\mathbb{E}\left|\varphi_{n,i}(\mathbf{X},\mathbf{Y})\varphi_{n,j}(\mathbf{X},\mathbf{Y})\varphi_{n,l}(\mathbf{X},\mathbf{Y})\right|.$$

Following the same steps as in [134], we may restrict ourselves to triples (i,j,l) such that $\mathbf{x}_i = \mathbf{x}_j = \mathbf{x}_l$. Under **(A.5)**, we have

$$\mathbb{E}\left|\varphi_{n,i}(\mathbf{X},\mathbf{Y})\varphi_{n,j}(\mathbf{X},\mathbf{Y})\varphi_{n,l}(\mathbf{X},\mathbf{Y})\right| = O(n\breve{b}^{-d}),$$

taking into account **(A.1)**, which gives us the desired result. This concludes the proof. □

Proof of Lemma A.5. To study the asymptotic distribution of U_n, we need to bound the variance of $U_n - \widehat{U}_{n,1}$. To do that, it is sufficient to show that

$$(n\breve{b}^{d/2})^{1/2}\left[\mathbb{U}_{n,1} - \widehat{U}_{n,1}\right] \to 0 \ \text{ in } \ L^2.$$

As in [134], using the centered variance formula for a centered, or zero mean U-statistic of degree m, for $\mathbf{Z}_i, i \geq 1$ i.i.d., we have

$$\mathscr{V}_n = \frac{(n-m)!}{n!}\sum_{\mathbf{i}\in I(m,n)}\frac{\tilde{G}(\mathbf{Z}_{i_1},\ldots,\mathbf{Z}_{i_m})}{N},$$

with a non-necessary symmetric U-Kernel $\tilde{G}(\cdot)$, that is square-integrable, which gives us

$$\mathrm{Var}(\mathscr{V}_n)\left[\frac{(n-m)!}{n!}\right]^2\sum_{r=1}^{m}\frac{(n-r)!}{(n-2m+r)!}\sum^{(r)}\frac{I(\Delta_1,\Delta_2)}{N^2},$$

where Δ_1 and Δ_2 represent positions of some length $1 \leq r \leq m$, and

$$I\left(\tilde{\Delta}_1,\tilde{\Delta}_2\right) = \int \tilde{G}(z_1,\ldots,z_m)\,\tilde{G}(y_1,\ldots,y_m)\,F(\mathrm{d}z_1)\cdots F(\mathrm{d}z_{2m-r}),$$

with the y 's in position $\tilde{\Delta}_2$ coincide with the z 's in position $\tilde{\Delta}_1$ and are taken from $z_{m+1},\ldots,z_{2m-r}$ otherwise. Moreover, $\Sigma^{(r)}$ represents the summation over all positions $\tilde{\Delta}_1,\tilde{\Delta}_2$ with a cardinality of r, and $F(\cdot)$ denotes the common distribution function of the Z's. When considering $V_n = U_n - \hat{U}_n$ and recalling $\tilde{G}$ from [129] (in the symmetric case), we obtain

$$\Sigma^{(1)}I\left(\tilde{\Delta}_1,\tilde{\Delta}_2\right) = 0.$$

Furthermore, by **(A.6)**, we infer that

$$N^{-2}I\left(\tilde{\Delta}_1,\tilde{\Delta}_2\right) = O\left(\breve{b}^{-dr/2}\right) \quad \text{for each } 2 \leq r \leq m.$$

In conclusion, we have

$$\begin{aligned} n\breve{b}^{d/2}\operatorname{Var}\left(U_n-\hat{U}_n\right) &= O\left[n\breve{b}^{d/2}\sum_{r=2}^{m}\left[\begin{array}{c} n \\ m \end{array}\right]^{-1}\left[\begin{array}{c} m \\ r \end{array}\right]\left[\begin{array}{c} n-m \\ m-r \end{array}\right](\breve{b}^{d/2})^{-r}\right] \\ &= O\left[\sum_{r=2}^{m}\left(n\breve{b}^{d/2}\right)^{1-r}\right] = O\left[\left(n\breve{b}^{d/2}\right)^{-1}\right] = o(1). \end{aligned}$$

Hence, the proof is complete. □

Proof of Theorem 3.6. To obtain the desired result, we shall first apply the Cramér-Wold device to investigate the asymptotic behavior of the two-dimensional vector

$$\left(\mathbb{U}_{n,1}(\varphi_1,\tilde{\mathbf{x}})-\theta_n\left(\varphi_1\right), \mathbb{U}_{n,1}(\varphi_2,\tilde{\mathbf{x}})-\theta_n\left(\varphi_2\right)\right),$$

where $\mathbb{U}_{n,1}(\varphi_1,\tilde{\mathbf{x}})$ and $\mathbb{U}_{n,1}(\varphi_2,\tilde{\mathbf{x}})$ are U-statistics with U kernels $G_{\varphi_1,\tilde{\mathbf{x}},1}(\cdot)$ and $G_{\varphi_2,\tilde{\mathbf{x}},1}(\cdot)$ respectively, satisfying the smoothness assumptions of Lemma A.5. Let c_1 and c_2 be any two real numbers. We can see that

$$c_1\mathbb{U}_{n,1}(\varphi_1,\tilde{\mathbf{x}})+c_2\mathbb{U}_{n,1}(\varphi_2,\tilde{\mathbf{x}})=\mathbb{U}_{n,1}\left(c_1\varphi_1+c_2\varphi_2,\tilde{\mathbf{x}}\right)\equiv\mathbb{U}_{n,1}\left(\varphi,\tilde{\mathbf{x}}\right),$$

which means we can apply Lemma A.5. The limit distribution of $\widehat{r}_n^{(m)}(\varphi,\tilde{\mathbf{x}},\mathbf{b}_n)$ may now be easily deduced from Lemma A.6. We have

$$\widehat{r}_{n,1}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,1}(\tilde{\mathbf{x}}))=\frac{\mathbb{U}_{n,1}(\varphi,\tilde{\mathbf{x}})}{\mathbb{U}_{n,1}(1,\tilde{\mathbf{x}})}.$$

Let us define, as in [134],

$$g\left(x_1,x_2\right)=x_1/x_2 \text{ for } x_2\neq 0,$$

which means

$$D=\left[\frac{\partial g}{\partial x_1},\frac{\partial g}{\partial x_2}\right]=\left[x_2^{-1},-x_1x_2^{-2}\right].$$

Since, by continuity of $r^{(m)}(\varphi,\tilde{\mathbf{x}})$, we have

$$\mathbb{E}\left[\mathbb{U}_{n,1}(\varphi,\tilde{\mathbf{x}})\right]\to r^{(m)}(\varphi,\tilde{\mathbf{x}}),$$

and

$$\mathbb{E}\left[\mathbb{U}_{n,1}(1,\tilde{\mathbf{x}})\right]=1.$$

From Lemma A.6, we deduce that

$$\left(n\breve{b}^{d/2}\right)^{1/2}\left[\widehat{r}_{n,1}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,1}(\tilde{\mathbf{x}}))-\mathbb{E}\left[\mathbb{U}_{n,1}(\varphi,\tilde{\mathbf{x}})\right]\right]\to\mathcal{N}\left(0,\rho^2\right),$$

where

$$\rho^2=(1,-r^{(m)}(\varphi,\tilde{\mathbf{x}}))\Sigma\left[\begin{array}{c} 1 \\ -r^{(m)}(\varphi,\tilde{\mathbf{x}}) \end{array}\right],$$

and ρ^2 is defined in (3.10). This concludes the proof of Theorem 3.6. □

A.2. Proof of Section 4: Bernstein polynomials

Proof of Theorem 4.2. In order to prove Theorem 4.2, we need

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\mathbb{E}[\widehat{r}^{(1)}_{n,2}(\varphi,\mathbf{x})]-r^{(1)}(\varphi,\mathbf{x})\right| = O(\vartheta^{-1/2}), \tag{A.59}$$

and

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\widehat{r}^{(1)}_{n,2}(\varphi,\mathbf{x})-\mathbb{E}[\widehat{r}^{(1)}_{n,2}(\varphi,\mathbf{x})]\right| = O(\vartheta^{d-1/2}(n^{-1}\log n)^{1/2})\ \ a.s. \tag{A.60}$$

The result in (A.59) is established in the following proposition.

Proposition A.7. *Assume that condition **(C.2)** holds. We have, uniformly for* $\mathbf{x}\in\mathbb{S}_{d,1}$,

$$\mathbb{E}[\widehat{g}_{n,\vartheta}(\varphi,\mathbf{x})] = \mathcal{R}(\varphi,\mathbf{x})+\vartheta^{-1}\mathcal{L}(\mathbf{x})+o(\vartheta^{-1}),\quad \vartheta\longrightarrow\infty,$$

where

$$\mathcal{L}(\mathbf{x}) := \left(\frac{d(d-1)}{2\vartheta}\right)\mathcal{R}(\varphi,\mathbf{x})+\sum_{i=1}^{d}\left(\frac{1}{2}-x_i\right)\frac{\partial}{\partial x_i}\mathcal{R}(\varphi,\mathbf{x})+\frac{1}{2}\sum_{i,j=1}^{d}\left(x_i\mathbf{1}_{\{i=j\}}-x_ix_j\right)\frac{\partial^2}{\partial x_i\partial x_j}\mathcal{R}(\varphi,\mathbf{x}).$$

It is important to mention that this proposition closely follows Proposition 2 in [115]. However, it offers greater generality as it holds for any function $\varphi(\cdot)$ that satisfies condition (C.3), thereby extending its range of applications.

Proof of Proposition A.7. We first remark that

$$\begin{aligned}
\mathbb{E}[\widehat{g}_{n,\vartheta}(\varphi,\mathbf{x})] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\varphi(\mathbf{Y}_i)K_{\mathbf{x},\vartheta}(\mathbf{X}_i)\right]\\
&= \int_{\mathbb{S}_{d,1}} r^{(1)}(\varphi,\mathbf{u})K_{\mathbf{x},\vartheta}(\mathbf{u})f(\mathbf{u})\mathrm{d}\mathbf{u}\\
&= \sum_{\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}}\frac{(\vartheta-1+d)!}{(\vartheta-1)!}\int_{\mathbf{1}_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k}+\mathbf{1}}{\vartheta}\right]}} r^{(1)}(\varphi,\mathbf{u})f(\mathbf{u})\mathrm{d}\mathbf{u}P_{\mathbf{k},\vartheta-1}(\mathbf{x})\\
&= \sum_{\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}}\frac{(\vartheta-1+d)!}{(\vartheta-1)!}\int_{\mathbf{1}_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k}+\mathbf{1}}{\vartheta}\right]}} \mathcal{R}(\varphi,\mathbf{u})\mathrm{d}\mathbf{u}P_{\mathbf{k},\vartheta-1}(\mathbf{x}).
\end{aligned}$$

For all $\varepsilon>0$, there exists a decreasing sequence $(\delta_{n,\varepsilon})_n$, $0<\delta_{n,\varepsilon}\leq 1$ such that for all $\mathbf{x},\mathbf{y}\in\mathbb{S}_{d,1}$, $\|\mathbf{x}-\mathbf{y}\|\leq\delta_{n,\varepsilon}$. By a Taylor's expansion for any $\mathbf{k}$ such that $\|\mathbf{k}/\vartheta-\mathbf{x}\|_1=o(1)$, we obtain

$$\begin{aligned}
\vartheta^d\int_{\mathbf{1}_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k}+\mathbf{1}}{\vartheta}\right]}}\mathcal{R}(\varphi,\mathbf{u})\mathrm{d}\mathbf{u}-\mathcal{R}(\varphi,\mathbf{x}) =& \mathcal{R}(\varphi,\mathbf{k}/\vartheta)-\mathcal{R}(\varphi,\mathbf{x})+\frac{1}{2\vartheta}\sum_{i=1}^{d}\frac{\partial}{\partial x_i}\mathcal{R}(\varphi,\mathbf{k}/\vartheta)+O\left(\vartheta^{-2}\right)\\
=&\frac{1}{\vartheta}\sum_{i=1}^{d}(k_i-\vartheta x_i)\frac{\partial}{\partial x_i}\mathcal{R}(\varphi,\mathbf{x})+\frac{1}{2\vartheta}\sum_{i=1}^{d}\frac{\partial}{\partial x_i}\mathcal{R}(\varphi,\mathbf{x})\\
&+o\left(\vartheta^{-1}\right)+\frac{1}{2\vartheta^2}\sum_{i,j=1}^{d}(k_i-\vartheta x_i)\left(k_j-\vartheta x_j\right)\frac{\partial^2}{\partial x_i\partial x_j}\mathcal{R}(\varphi,\mathbf{x})(1+o(1))
\end{aligned}$$

$$=\frac{1}{\vartheta}\sum_{i=1}^{d}\left(k_i-(\vartheta-1)x_i\right)\frac{\partial}{\partial x_i}\mathcal{R}(\varphi,\mathbf{x})+\frac{1}{\vartheta}\sum_{i=1}^{d}\left(\frac{1}{2}-x_i\right)\frac{\partial}{\partial x_i}\mathcal{R}(\varphi,\mathbf{x})$$
$$+\frac{1}{2}\sum_{i,j=1}^{d}\left(\frac{k_i}{\vartheta}-x_i\right)\left(\frac{k_j}{\vartheta}-x_j\right)\frac{\partial^2}{\partial x_i\partial x_j}\mathcal{R}(\varphi,\mathbf{x})(1+o(1))+o\left(\vartheta^{-1}\right).$$

If we multiply the last expression by $\vartheta^{-d}\cdot\frac{(\vartheta-1+d)!}{(\vartheta-1)!}P_{\mathbf{k},\vartheta-1}(\mathbf{x})$ and sum over all $\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}$, then considering the well-known identities

$$\sum_{\mathbf{k}\in\mathbb{N}_0^d\cap\vartheta\mathbb{S}_{d,1}}\left(\frac{k_i}{\vartheta}-x_i\right)P_{\mathbf{k},\vartheta}(\mathbf{x})=0, \tag{A.61}$$

and the fact that

$$\sum_{\mathbf{k}\in\mathbb{N}_0^d\cap\vartheta\mathbb{S}_{d,1}}\left(\frac{k_i}{\vartheta}-x_i\right)\left(\frac{k_j}{\vartheta}-x_j\right)P_{\mathbf{k},\vartheta}(\mathbf{x})=\frac{1}{\vartheta}\left(x_i\mathbf{1}_{\{i=j\}}-x_ix_j\right), \tag{A.62}$$

it yields

$$\mathbb{E}[\widehat{g}_{n,\vartheta}(\varphi,\mathbf{x})]-\left(1+\frac{d(d-1)}{2\vartheta}\right)\mathcal{R}(\varphi,\mathbf{x})$$
$$=0+\frac{1}{\vartheta}\sum_{i=1}^{d}\left(\frac{1}{2}-x_i\right)\frac{\partial}{\partial x_i}\mathcal{R}(\varphi,\mathbf{x})+\frac{1}{2\vartheta}\sum_{i,j=1}^{d}\left(x_i\mathbf{1}_{\{i=j\}}-x_ix_j\right)\frac{\partial^2}{\partial x_i\partial x_j}\mathcal{R}(\varphi,\mathbf{x})+o\left(\vartheta^{-1}\right),$$

assuming that $\|\mathbf{k}/\vartheta-\mathbf{x}\|_1=o(1)$ decays slowly enough to 0 that the contributions coming from outside the bulk are negligible. Hence, the proof of the proposition is complete. □

Next, based on the results of Proposition A.7, we can see that

$$\vartheta^d\int_{\mathbf{1}_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k}+1}{\vartheta}\right]}}\mathcal{R}(\varphi,\mathbf{u})d\mathbf{u} = \mathcal{R}(\varphi,\mathbf{k}/\vartheta)+O\left(\vartheta^{-1}\right),$$

which implies considering the Lipschitz continuity of $\mathcal{R}(\varphi,\cdot)$ uniformly for $\mathbf{x}\in\mathbb{S}_{d,1}$

$$\mathbb{E}[\widehat{g}_{n,\vartheta}(\varphi,\mathbf{x})]-\mathcal{R}(\varphi,\mathbf{x})=\sum_{i=1}^{d}O\left(\sum_{\mathbf{k}\in\mathbb{N}_0^d\cap\vartheta\mathbb{S}_{d,1}}\left|\frac{k_i}{\vartheta}-x_i\right|P_{\mathbf{k},\vartheta}(\mathbf{x})\right)+O\left(\vartheta^{-1}\right).$$

Finally, all it takes to get the bias term is to apply the Cauchy-Schwarz inequality in the last expression, followed by an application of (A.62), and we obtain

$$\begin{aligned}\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\mathbb{E}[\widehat{g}_{n,\vartheta}(\varphi,\mathbf{x})]-\mathcal{R}(\varphi,\mathbf{x})\right| &= O\left(\vartheta^{-d/2}\right)+O\left(\vartheta^{-1}\right)\\ &= O\left(\vartheta^{-1}\right).\end{aligned} \tag{A.63}$$

Use the fact that

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\frac{\widehat{g}_{n,\vartheta}(\varphi,\mathbf{x})}{f(\mathbf{x})}-r^{(1)}(\varphi,\mathbf{x})\right|\leq\frac{\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\widehat{g}_{n,\vartheta}(\varphi,\mathbf{x})-\mathcal{R}(\varphi,\mathbf{x})\right|}{\inf_{\mathbf{x}\in\mathbb{S}_{d,1}}f(\mathbf{x})},$$

and

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\frac{\widehat{f}_{n,\vartheta}(\mathbf{x})}{f(\mathbf{x})}-1\right| \leq \frac{\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}\left|\widehat{f}_{n,\vartheta}(\mathbf{x})-f(\mathbf{x})\right|}{\inf_{\mathbf{x}\in\mathbb{S}_{d,1}} f(\mathbf{x})}.$$

Combining (A.63) with the fact that $r^{(1)}(\varphi,\mathbf{x}) = \dfrac{\widehat{g}_{n,\vartheta}(\varphi,\mathbf{x})}{f(\mathbf{x})}\cdot\dfrac{f(\mathbf{x})}{\widehat{f}_{n,\vartheta}(\mathbf{x})}$, completes the proof of (A.59). Now, we only need to prove (A.60), noticing that

$$\widehat{g}_{n,\vartheta}(\varphi,\mathbf{x})-\mathbb{E}[\widehat{g}_{n,\vartheta}(\varphi,\mathbf{x})] = \frac{(\vartheta-1+d)!}{(\vartheta-1)!}\frac{1}{n}\sum_{i=1}^{n}\mathcal{Z}_{i,\vartheta},$$

where

$$\mathcal{Z}_{i,\vartheta} := \sum_{\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}}\left[\varphi(\mathbf{Y}_i)\mathbf{1}_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k+1}}{\vartheta}\right]}(\mathbf{X}_i)-\int_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k+1}}{\vartheta}\right]}\mathcal{R}(\varphi,\mathbf{u})d\mathbf{u}\right]P_{\mathbf{k},\vartheta-1}(\mathbf{x}),\quad i\in\{1,\ldots,n\}. \tag{A.64}$$

For every ϑ, the random variables $Y_{1,\vartheta},\ldots,Y_{n,\vartheta}$ are i.i.d. and centered. Set

$$\begin{aligned}\varphi^{(T)}(\mathbf{y}) &:= \varphi(\mathbf{y})\mathbf{1}_{\{|\varphi(\mathbf{y})|\leq\omega_{n,2}\}},\\ \varphi^{(R)}(\mathbf{y}) &:= \varphi(\mathbf{y})\mathbf{1}_{\{|\varphi(\mathbf{y})|>\omega_{n,2}\}},\end{aligned}$$

then, we have

$$\mathcal{Z}_{i,\vartheta} = \mathcal{Z}_{i,\vartheta}^{(T)}+\mathcal{Z}_{i,\vartheta}^{(R)},$$

where

$$\begin{aligned}\mathcal{Z}_{i,\vartheta}^{(T)} &:= \sum_{\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}}\left[\varphi(\mathbf{Y}_i)\mathbf{1}_{\{|\varphi(\mathbf{Y})|\leq\omega_{n,2}\}}\mathbf{1}_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k+1}}{\vartheta}\right]}(\mathbf{X}_i)-\int_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k+1}}{\vartheta}\right]}\mathcal{R}(\varphi^{(T)},\mathbf{u})d\mathbf{u}\right]P_{\mathbf{k},\vartheta-1}(\mathbf{x}),\\ \mathcal{Z}_{i,\vartheta}^{(R)} &:= \sum_{\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}}\left[\varphi(\mathbf{Y}_i)\mathbf{1}_{\{|\varphi(\mathbf{Y})|>\omega_{n,2}\}}\mathbf{1}_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k+1}}{\vartheta}\right]}(\mathbf{X}_i)-\int_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k+1}}{\vartheta}\right]}\mathcal{R}(\varphi^{(R)},\mathbf{u})d\mathbf{u}\right]P_{\mathbf{k},\vartheta-1}(\mathbf{x}).\end{aligned}$$

So we can write

$$\begin{aligned}\widehat{g}_{n,\vartheta}(\varphi,\mathbf{x})-\mathbb{E}[\widehat{g}_{n,\vartheta}(\varphi,\mathbf{x})] &= \frac{(\vartheta-1+d)!}{(\vartheta-1)!}\frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{Z}_{i,\vartheta}^{(T)}+\mathcal{Z}_{i,\vartheta}^{(R)}\right)\\ &= \left\{\widehat{g}_{n,\vartheta}(\varphi^{(T)},\mathbf{x})-\mathbb{E}[\widehat{g}_{n,\vartheta}(\varphi^{(T)},\mathbf{x})]\right\}+\left\{\widehat{g}_{n,\vartheta}(\varphi^{(R)},\mathbf{x})-\mathbb{E}[\widehat{g}_{n,\vartheta}(\varphi^{(R)},\mathbf{x})]\right\}\\ &:= \mathcal{T}_1+\mathcal{T}_2.\end{aligned}$$

To obtain the desired result, we need to prove that

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}|\mathcal{T}_1| = O(\vartheta^{d-1/2}(n^{-1}\log n)^{1/2})\ \ a.s, \tag{A.65}$$

and

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}}|\mathcal{T}_2| = o(1). \tag{A.66}$$

Note that the proof of $\mathcal{T}_2$ can be derived directly from the proof of the remainder part of the conditional U-statistics in the following section. Moving on to $\mathcal{T}_1$, we have

$$\operatorname{Var}\left[\widehat{g}_{n,\vartheta}(\varphi^{(T)},\mathbf{x})\right] = n^{-1}\left(\frac{(\vartheta-1+d)!}{(\vartheta-1)!}\right)^2 \mathbb{E}\left[(\mathcal{Z}_{1,\vartheta}^{(T)})^2\right],$$

where

$$\begin{aligned}\mathbb{E}\left[(\mathcal{Z}_{1,\vartheta}^{(T)})^2\right] = &\sum_{\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}} \int_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k+1}}{\vartheta}\right]} \mathbb{E}[(\varphi^{(T)}(\mathbf{Y}))^2 \mid \mathbf{X}=\mathbf{u}] f(\mathbf{u}) d\mathbf{u} P_{\mathbf{k},\vartheta-1}^2(\mathbf{x}) - \left(\frac{(\vartheta-1)!}{(\vartheta-1+d)!}\mathbb{E}[\widehat{g}_{n,\vartheta}(\varphi^{(T)},\mathbf{x})]\right)^2\\ \leq &\omega_{n,2}^2\left\{\sum_{\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}} \int_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k+1}}{\vartheta}\right]} f(\mathbf{u}) d\mathbf{u} P_{\mathbf{k},\vartheta-1}^2(\mathbf{x}) - \left(\frac{(\vartheta-1)!}{(\vartheta-1+d)!}\mathbb{E}[\widehat{f}_{n,\vartheta}(\mathbf{x})]\right)^2\right\}.\end{aligned}$$

Set $\varsigma_n := \sqrt{\dfrac{\log n}{n}}$, and

$$L_{n,\vartheta} := \max_{\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}} \frac{1}{n}\sum_{i=1}^n \left[\varphi(\mathbf{Y}_i)\mathbf{1}_{\{|\varphi(\mathbf{Y})|\leq\omega_{n,2}\}}\mathbf{1}_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k+1}}{\vartheta}\right]}(\mathbf{X}_i) - \int_{\left(\frac{\mathbf{k}}{\vartheta},\frac{\mathbf{k+1}}{\vartheta}\right]} \mathcal{R}(\varphi^{(T)},\mathbf{u}) d\mathbf{u}\right].$$

By a union bound on $\mathbf{k}\in\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1}$ (there are at most ϑ^d such points) and Bernstein's inequality, we have, for all $\rho>0$,

$$\mathbb{P}\left(\left|L_{n,\vartheta}\right| > \rho\vartheta^{-1/2}\varsigma_n\right) \leq \vartheta^d\cdot 2\exp\left(-\frac{n^2\rho^2\vartheta^{-1}\varsigma_n^2/2}{nc\omega_{n,2}^2\vartheta^{-1}+\frac{1}{3}\omega_{n,2}n\rho\vartheta^{-1/2}\varsigma_n}\right) \leq \vartheta^d\cdot 2n^{-\rho^2/(4c)}, \tag{A.67}$$

where the second inequality assumes that $\vartheta \leq n/\log n$ (equivalently, $\varsigma_n \leq \vartheta^{-1/2}$), and $c \geq \rho$ is a Lipschitz constant for $f(\cdot)$. If we choose $\rho = \rho(c,d) > 0$ large enough, then the righthand side of (A.67) is summable in n, and the Borel-Cantelli lemma implies

$$\sup_{\mathbf{x}\in\mathbb{S}_{d,1}} |\mathcal{T}_1| \leq O\left(\vartheta^{d-1/2}\varsigma_n\right) \quad a.s. \text{ as } n\to\infty.$$

Hence, the proof of Theorem 4.2 is complete. □

Proof of Theorem 4.4. Recall the notation (A.1). We will start by investigating the term $u_{n,2}^{(T)}(\varphi,\tilde{\mathbf{x}})$.

Truncated Part:

Notice that

$$\begin{aligned}\left|u_{n,2}^{(T)}(\varphi,\tilde{\mathbf{x}}) - \mathbb{E}\left(u_{n,2}^{(T)}(\varphi,\tilde{\mathbf{x}})\right)\right| &= \frac{(n-m)!}{n!}\left|\sum_{i\in I(m,n)}\left\{\varphi^{(T)}(\tilde{\mathbf{Y}}_{\mathbf{i}}\tilde{\mathcal{K}}_{\tilde{\mathbf{x}},\vartheta}(\tilde{\mathbf{X}}_i) - \mathbb{E}\left[\varphi^{(T)}(\tilde{\mathbf{Y}}_{\mathbf{i}}\tilde{\mathcal{K}}_{\tilde{\mathbf{x}},\vartheta}(\tilde{\mathbf{X}}_i)\right]\right\}\right|\\ &= \frac{(n-m)!}{n!}\left|\sum_{i\in I(m,n)}\left\{\mathcal{G}_{\varphi,\tilde{\mathbf{x}},2}^{(T)}(\tilde{\mathbf{X}}_{\mathbf{i}},\tilde{\mathbf{Y}}_{\mathbf{i}}) - \mathbb{E}\left[\mathcal{G}_{\varphi,\tilde{\mathbf{x}},2}^{(T)}(\tilde{\mathbf{X}}_{\mathbf{i}},\tilde{\mathbf{Y}}_{\mathbf{i}})\right]\right\}\right|\\ &= \frac{(n-m)!}{n!}\left|\sum_{i\in I(m,n)} H_{\vartheta}^{(T)}(\tilde{\mathbf{X}}_{\mathbf{i}},\tilde{\mathbf{Y}}_{\mathbf{i}})\right|,\end{aligned}$$

where

$$
\begin{aligned}
H_{\vartheta}^{(T)}(\tilde{\mathbf{X}}_{\mathbf{i}}, \tilde{\mathbf{Y}}_{\mathbf{i}}) &= \mathcal{G}_{\varphi,\tilde{\mathbf{x}},2}^{(T)}(\tilde{\mathbf{X}}_{\mathbf{i}}, \tilde{\mathbf{Y}}_{\mathbf{i}}) - \mathbb{E}\left[\mathcal{G}_{\varphi,\tilde{\mathbf{x}},2}^{(T)}(\tilde{\mathbf{X}}_{\mathbf{i}}, \tilde{\mathbf{Y}}_{\mathbf{i}})\right] \\
&= \left(\frac{(\vartheta-1+d)!}{(\vartheta-1)!}\right)^m \sum_{(\mathbf{k}_1,\ldots,\mathbf{k}_m)\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m} \left\{\varphi^{(T)}(\mathbf{Y}_{i_1},\ldots,\mathbf{Y}_{i_m}) \prod_{j=1}^m \mathbf{1}_{\left(\frac{\mathbf{k}_j}{\vartheta},\frac{\mathbf{k}_j+\mathbf{1}}{\vartheta}\right]}(\mathbf{X}_{i_j})\right. \\
&\quad \left. - \int_{\left(\frac{\mathbf{k}_1}{\vartheta},\frac{\mathbf{k}_1+\mathbf{1}}{\vartheta}\right]} \cdots \int_{\left(\frac{\mathbf{k}_m}{\vartheta},\frac{\mathbf{k}_m+\mathbf{1}}{\vartheta}\right]} \varphi^{(T)}(\tilde{\mathbf{y}}) \tilde{f}(\tilde{\mathbf{u}},\tilde{\mathbf{y}}) \mathbf{d}\tilde{\mathbf{u}} \mathbf{d}\tilde{\mathbf{y}}\right\} \prod_{j=1}^m P_{\mathbf{k}_j,\vartheta-1}(\mathbf{x}_j).
\end{aligned}
$$

It suffices to consider

$$
\begin{aligned}
\mathcal{L}_{\vartheta,n} &:= \max_{\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m} \frac{(n-m)!}{n!} \sum_{i\in I(m,n)} \omega_{n,2} \left(\prod_{j=1}^m \mathbf{1}_{\left(\frac{\mathbf{k}_j}{\vartheta},\frac{\mathbf{k}_j+\mathbf{1}}{\vartheta}\right]}(\mathbf{X}_{i_j}) - \int_{\left(\frac{\mathbf{k}_1}{\vartheta},\frac{\mathbf{k}_1+\mathbf{1}}{\vartheta}\right]} \cdots \int_{\left(\frac{\mathbf{k}_m}{\vartheta},\frac{\mathbf{k}_m+\mathbf{1}}{\vartheta}\right]} \tilde{f}(\tilde{\mathbf{u}}) \mathbf{d}\tilde{\mathbf{u}}\right) \\
&= \max_{\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m} \frac{(n-m)!}{n!} \sum_{i\in I(m,n)} Z_i.
\end{aligned}
$$

We have $\mathbb{E}(Z_i) = 0$, and by condition **(C.2)**,

$$
\begin{aligned}
\mathrm{Var}(Z_i) &= \mathbb{E}(Z_i^2) \\
&= \omega_{n,2}^2 \int_{\left(\frac{\mathbf{k}_1}{\vartheta},\frac{\mathbf{k}_1+\mathbf{1}}{\vartheta}\right]} \cdots \int_{\left(\frac{\mathbf{k}_m}{\vartheta},\frac{\mathbf{k}_m+\mathbf{1}}{\vartheta}\right]} \tilde{f}(\tilde{\mathbf{u}}) \mathbf{d}\tilde{\mathbf{u}} - \omega_{n,2}^2 \left(\int_{\left(\frac{\mathbf{k}_1}{\vartheta},\frac{\mathbf{k}_1+\mathbf{1}}{\vartheta}\right]} \cdots \int_{\left(\frac{\mathbf{k}_m}{\vartheta},\frac{\mathbf{k}_m+\mathbf{1}}{\vartheta}\right]} \tilde{f}(\tilde{\mathbf{u}}) \mathbf{d}\tilde{\mathbf{u}}\right)^2 \\
&\le \omega_{n,2}^2 C_0 \int_{\left(\frac{\mathbf{k}_1}{\vartheta},\frac{\mathbf{k}_1+\mathbf{1}}{\vartheta}\right]} \cdots \int_{\left(\frac{\mathbf{k}_m}{\vartheta},\frac{\mathbf{k}_m+\mathbf{1}}{\vartheta}\right]} 1 \cdot \mathbf{d}\tilde{\mathbf{u}} - \omega_{n,2}^2 C_0^2 \left(\int_{\left(\frac{\mathbf{k}_1}{\vartheta},\frac{\mathbf{k}_1+\mathbf{1}}{\vartheta}\right]} \cdots \int_{\left(\frac{\mathbf{k}_m}{\vartheta},\frac{\mathbf{k}_m+\mathbf{1}}{\vartheta}\right]} 1 \cdot \mathbf{d}\tilde{\mathbf{u}}\right)^2 \\
&\le \omega_{n,2}^2 C_0 \left(\vartheta^{-dm} - C_0 \vartheta^{-2dm}\right) \\
&\le \omega_{n,2}^2 C_0 \vartheta^{-dm}.
\end{aligned}
$$

Applying Bernstein's inequality on

$$
H(\tilde{\mathbf{X}}) = \omega_{n,2} \prod_{i=1}^m \mathbf{1}_{\left(\frac{\mathbf{k}_j}{\vartheta},\frac{\mathbf{k}_j+\mathbf{1}}{\vartheta}\right]}(\mathbf{X}_j),
$$

we have

$$
\|H(\tilde{\mathbf{X}})\|_\infty \le \omega_{n,2},
$$

$$
\sigma^2 = \mathrm{Var}[H(\tilde{\mathbf{X}})] \le \omega_{n,2}^2 C_0 \vartheta^{-dm},
$$

where $C_{\delta,dm}$ is a Lipschitz constant of the cumulative distribution function of $\tilde{\mathbf{X}}$. We obtain, for $\kappa > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\max_{\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m} \left|u_{n,2}(\varphi,\tilde{\mathbf{x}}) - \mathbb{E}\left[u_{n,2}(\varphi,\tilde{\mathbf{x}})\right]\right| > \varepsilon_0 \vartheta^{-m/2} \varsigma_n\right) \\
\le& 2\vartheta^{md} \exp\left(-\frac{[n/m]\varepsilon_0^2 \vartheta^{-m} \varsigma_n^2}{2\omega_{n,2}^2 C_0 \vartheta^{-dm} - \frac{2}{3} 2\omega_{n,2} \varepsilon_0 \vartheta^{-m/2} \varsigma_n}\right)
\end{aligned}
$$

$$\leq 2\vartheta^{md} n^{-1-\kappa}.$$

Remainder Part:

Recall that

$$u_{n,2}^{(R)}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}\left[u_{n,2}^{(R)}(\varphi, \tilde{\mathbf{x}})\right] = \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I(m,n)} \mathcal{G}_{\varphi,\tilde{\mathbf{x}},2}^{(R)}(\tilde{\mathbf{X}}_{\mathbf{i}}, \tilde{\mathbf{Y}}_{\mathbf{i}}) - \mathbb{E}\left[\mathcal{G}_{\varphi,\tilde{\mathbf{x}},2}^{(R)}(\tilde{\mathbf{X}}_{\mathbf{i}}, \tilde{\mathbf{Y}}_{\mathbf{i}})\right].$$

Now, using the fact that for $\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right| > \omega_{n,2}$, we have $(\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right| / \omega_{n,2})^{1+\gamma} > 1$, which implies that

$$\begin{aligned}
\left|\mathbb{E}[u_{n,2}^{(R)}(\varphi, \tilde{\mathbf{x}})]\right| &\leq \mathbb{E}[\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right| \mathbf{1}_{\{|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})|>\omega_{n,2}\}} \tilde{\mathcal{K}}_{\tilde{\mathbf{x}},\vartheta}(\tilde{\mathbf{X}}_{\mathbf{i}})] \\
&\leq \mathbb{E}[\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right| \left(\frac{\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right|}{\omega_{n,2}}\right)^{1+\gamma} \mathbf{1}_{\{|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})|>\omega_{n,2}\}} \tilde{\mathcal{K}}_{\tilde{\mathbf{x}},\vartheta}(\tilde{\mathbf{X}}_{\mathbf{i}})] \\
&\leq \omega_{n,2}^{-(1+\gamma)} \mathbb{E}[\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right|^{2+\gamma} \tilde{\mathcal{K}}_{\tilde{\mathbf{x}},\vartheta}(\tilde{\mathbf{X}}_{\mathbf{i}})], \qquad \text{(A.68)}
\end{aligned}$$

where, by Assumption **(C.3)**, we have

$$\begin{aligned}
\mathbb{E}\left[\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right|^{2+\gamma} \tilde{\mathcal{K}}_{\tilde{\mathbf{x}},\vartheta}(\tilde{\mathbf{X}}_{\mathbf{i}})\right] &= \mathbb{E}\left\{\mathbb{E}\left(\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right|^{2+\gamma} \mid \tilde{\mathbf{X}}_{\mathbf{i}}\right) \tilde{\mathcal{K}}_{\tilde{\mathbf{x}},\vartheta}(\tilde{\mathbf{X}}_{\mathbf{i}})\right\} \\
&= \int_{\mathbb{S}_{d,1}^m} \mathbb{E}\left(\left|\varphi(\tilde{\mathbf{Y}})\right|^{2+\gamma} \mid \tilde{\mathbf{X}} = \tilde{\mathbf{u}}\right) \tilde{f}(\tilde{\mathbf{u}}) \tilde{\mathcal{K}}_{\tilde{\mathbf{x}},\vartheta}(\tilde{\mathbf{u}}) \mathbf{d}\tilde{\mathbf{u}} \leq C_1. \qquad \text{(A.69)}
\end{aligned}$$

Therefore, if we set of $\omega_{n,2} = (\vartheta^{-m/2}\varsigma_n)^{-1/(1+\gamma)}$, it follows that $\left|\mathbb{E}\left[u_{n,2}^{(R)}(\varphi, \tilde{\mathbf{x}})\right]\right| \leq O(\vartheta^{-m/2}\varsigma_n)$ uniformly on $\tilde{\mathbf{x}} \in \mathbb{S}_{d,1}^m$. Consequently, Markov's inequality gives us

$$\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{d,1}^m} \left|u_{n,2}^{(R)}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}\left[u_{n,2}^{(R)}(\varphi, \tilde{\mathbf{x}})\right]\right| = O_{\mathbb{P}}(\vartheta^{-m/2}\varsigma_n). \qquad \text{(A.70)}$$

Hence, the proof is complete. □

Proof of Theorem 4.5. The proof of Theorem 4.5 is similar to the proof of Theorem 3.3. Using the decomposition (A.2) and the fact that for some positive constants $c_1, c_2 > 0$, we readily infer

$$\begin{aligned}
\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{d,1}^m} \left|u_{n,2}(1, \tilde{\mathbf{x}})\right| &= c_1 \qquad \text{a.s.,} \\
\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{d,1}^m} \left|\mathbb{E}\left(u_{n,2}(1, \tilde{\mathbf{x}})\right)\right| &= c_2, \\
\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{d,1}^m} \left|\mathbb{E}\left(u_{n,2}(\varphi, \tilde{\mathbf{x}})\right)\right| &= O(1).
\end{aligned}$$

Set $\varsigma_{n,2} := \vartheta^{m(d-1/2)}(n^{-1}\log n)^{1/2}$. By Theorem 4.4, for some $c'' > 0$, we get with probability 1:

$$\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{d,1}^m} \frac{\left|\widehat{r}_{n,2}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,2}(\tilde{\mathbf{x}})) - \widehat{\mathbb{E}}\left(\widehat{r}_{n,2}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,2}(\tilde{\mathbf{x}})\right)\right|}{\varsigma_{n,2}} \leq \sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{d,1}^m} \frac{(\mathscr{I}_{2,1})}{\varsigma_{n,2}} + \sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{d,1}^m} \frac{(\mathscr{I}_{2,2})}{\varsigma_{n,2}} \leq c''.$$

Hence, the proof is complete. □

Proof of Theorem 4.6. We have

$$\begin{aligned}\left|\widehat{\mathbb{E}}\left[\widehat{r}_{n,2}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,2}(\tilde{\mathbf{x}}))\right]-r^{(m)}(\varphi,\tilde{\mathbf{x}})\right| &= \left|\frac{\mathbb{E}\left(u_{n,2}(\varphi,\tilde{\mathbf{x}})\right)}{\mathbb{E}\left(u_{n,2}(1,\tilde{\mathbf{x}})\right)}-r^{(m)}(\varphi,\tilde{\mathbf{x}})\right| \\ &= \frac{1}{\left|\mathbb{E}\left(u_{n,2}(1,\tilde{\mathbf{x}})\right)\right|}\left|\mathbb{E}\left(u_{n,2}(\varphi,\tilde{\mathbf{x}})\right)-r^{(m)}(\varphi,\tilde{\mathbf{x}})\mathbb{E}\left(u_{n,2}(1,\tilde{\mathbf{x}})\right)\right|.\end{aligned}$$

It suffices to control

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\left|\mathbb{E}\left(u_{n,2}(\varphi,\tilde{\mathbf{x}})\right)-\mathcal{R}(\varphi,\tilde{\mathbf{x}})\right|=O\left(\vartheta^{-m/2}\right), \tag{A.71}$$

and that

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\left|\mathbb{E}\left(u_{n,2}(1,\tilde{\mathbf{x}})\right)-\tilde{f}(\tilde{\mathbf{x}})\right|=O\left(\vartheta^{-m/2}\right). \tag{A.72}$$

Let us start with (A.71), and we have the following

Proposition A.8. *Assume that condition **(C.2)** holds. We have, uniformly for $\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m$,*

$$\mathbb{E}\left[u_{n,2}(\varphi,\tilde{\mathbf{x}})\right]=\mathcal{R}(\varphi,\tilde{\mathbf{x}})+\vartheta^{-m}\mathcal{L}_m(\tilde{\mathbf{x}})+o(\vartheta^{-m}),\quad \vartheta\longrightarrow\infty,$$

where

$$\begin{aligned}\mathcal{L}_m(\tilde{\mathbf{x}}) &:= \left(\frac{d(d-1)}{2\vartheta}\right)^m\mathcal{R}(\varphi,\tilde{\mathbf{x}})+\sum_{i=1}^m\sum_{\ell=1}^d\left(\frac{1}{2}-x_{i_\ell}\right)\frac{\partial\mathcal{R}(\varphi,\tilde{\mathbf{x}})}{\partial x_{i_\ell}} \\ &\quad+\frac{1}{2}\sum_{i,j=1}^m\sum_{\ell,r=1}^d\left(x_{i_\ell}\mathbf{1}_{\{i_\ell=j_r\}}-x_{i_\ell}x_{j_r}\right)\frac{\partial^2\mathcal{R}(\varphi,\tilde{\mathbf{x}})}{\partial x_{i_\ell}\partial x_{j_r}}.\end{aligned}$$

Proof of Proposition A.8. Let us start with

$$\begin{aligned}\mathbb{E}\left[u_{n,2}(\varphi,\tilde{\mathbf{x}})\right] &= \int_{\mathbb{S}_{d,1}^m}r^{(m)}(\varphi,\tilde{\mathbf{u}})\tilde{f}(\tilde{\mathbf{u}})\tilde{\mathcal{K}}_{\tilde{\mathbf{x}},\vartheta}(\tilde{\mathbf{u}})\mathbf{d}\tilde{\mathbf{u}} \\ &= \int_{\mathbb{S}_{d,1}^m}\mathcal{R}(\varphi,\tilde{\mathbf{u}})\tilde{\mathcal{K}}_{\tilde{\mathbf{x}},\vartheta}(\tilde{\mathbf{u}})\mathbf{d}\tilde{\mathbf{u}} \\ &= \left(\frac{(\vartheta-1+d)!}{(\vartheta-1)!}\right)^m\sum_{(\mathbf{k}_1,\ldots,\mathbf{k}_m)\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m}\int_{\left(\frac{\mathbf{k}_1}{\vartheta},\frac{\mathbf{k}_1+\mathbf{1}}{\vartheta}\right]}\cdots\int_{\left(\frac{\mathbf{k}_m}{\vartheta},\frac{\mathbf{k}_m+\mathbf{1}}{\vartheta}\right]}\mathcal{R}(\varphi,\tilde{\mathbf{u}})\mathbf{d}\tilde{\mathbf{u}}\prod_{j=1}^mP_{\mathbf{k}_j,\vartheta-1}(\mathbf{x}_j).\end{aligned}$$

By a Taylor's expansion for any $\tilde{\mathbf{k}}$ such that $\|\frac{\tilde{\mathbf{k}}}{\vartheta}-\tilde{\mathbf{x}}\|_1=o(1)$, we obtain

$$\vartheta^{dm}\int_{\left(\frac{\tilde{\mathbf{k}}}{\vartheta},\frac{\tilde{\mathbf{k}}+\tilde{\mathbf{1}}}{\vartheta}\right]}\mathcal{R}(\varphi,\tilde{\mathbf{u}})\mathbf{d}\tilde{\mathbf{u}}-\mathcal{R}(\varphi,\tilde{\mathbf{x}}) \tag{A.73}$$

$$\begin{aligned}&=\mathcal{R}(\varphi,\tilde{\mathbf{k}}/\vartheta)-\mathcal{R}(\varphi,\tilde{\mathbf{x}})+\frac{1}{2\vartheta^d}\sum_{i=1}^m\sum_{\ell=1}^d\frac{\partial\mathcal{R}(\varphi,\tilde{\mathbf{k}}/\vartheta)}{\partial x_{i_\ell}}+O(\vartheta^{-2d}) \\ &=\frac{1}{\vartheta^d}\sum_{i=1}^m\sum_{\ell=1}^d(k_{i_\ell}-\vartheta x_{i_\ell})\frac{\partial\mathcal{R}(\varphi,\tilde{\mathbf{x}})}{\partial x_{i_\ell}}+\frac{1}{2\vartheta^d}\sum_{i=1}^m\sum_{\ell=1}^d\frac{\partial\mathcal{R}(\varphi,\tilde{\mathbf{x}})}{\partial x_{i_\ell}}+o(\vartheta^{-d})\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2\vartheta^{2d}}\sum_{i,j=1}^{m}\sum_{\ell,r=1}^{d}(k_{i_\ell}-\vartheta x_{i_\ell})(k_{j_r}-\vartheta x_{j_r})\frac{\partial^2\mathcal{R}(\varphi,\tilde{\mathbf{x}})}{\partial x_{i_\ell}\partial x_{j_r}}(1+o(1))\\
=&\frac{1}{\vartheta^d}\sum_{i=1}^{m}\sum_{\ell=1}^{d}(k_{i_\ell}-(\vartheta-1)x_{i_\ell})\frac{\partial\mathcal{R}(\varphi,\tilde{\mathbf{x}})}{\partial x_{i_\ell}}+\frac{1}{\vartheta^d}\sum_{i=1}^{m}\sum_{\ell=1}^{d}(\frac{1}{2}-x_{i_\ell})\frac{\partial\mathcal{R}(\varphi,\tilde{\mathbf{x}})}{\partial x_{i_\ell}}\\
&+\frac{1}{2}\sum_{i,j=1}^{m}\sum_{\ell,r=1}^{d}\left(\frac{k_{i_\ell}}{\vartheta}-x_{i_\ell}\right)\left(\frac{k_{j_r}}{\vartheta}-x_{j_r}\right)\frac{\partial^2\mathcal{R}(\varphi,\tilde{\mathbf{x}})}{\partial x_{i_\ell}\partial x_{j_r}}(1+o(1))+o(\vartheta^{-d}).
\end{aligned}
$$

Multiplying the last expression by $\vartheta^{-dm}\cdot\left(\frac{(\vartheta-1+d)!}{(\vartheta-1)!}\right)^m\prod_{j=1}^{m}P_{\mathbf{k}_j,\vartheta-1}(\mathbf{x}_j)$, summing over all $\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m$, and using the same principle as the identities (A.61) and (A.62) leads to

$$
\begin{aligned}
&\mathbb{E}\left[u_{n,2}(\varphi,\tilde{\mathbf{x}})\right]-\left(1+\frac{d(d-1)}{2\vartheta}\right)^m\mathcal{R}(\varphi,\tilde{\mathbf{x}})\\
=&0+\frac{1}{\vartheta^m}\sum_{i=1}^{m}\sum_{\ell=1}^{d}(\frac{1}{2}-x_{i_\ell})\frac{\partial\mathcal{R}(\varphi,\tilde{\mathbf{x}})}{\partial x_{i_\ell}}+\frac{1}{2\vartheta}\sum_{i,j=1}^{m}\sum_{\ell,r=1}^{d}\left(x_{i_\ell}\mathbf{1}_{\{i_\ell=j_r\}}-x_{i_\ell}x_{j_r}\right)\frac{\partial^2\mathcal{R}(\varphi,\tilde{\mathbf{x}})}{\partial x_{i_\ell}\partial x_{j_r}}+o\left(\vartheta^{-m}\right),
\end{aligned}
$$

assuming that $\|\frac{\tilde{\mathbf{k}}}{\vartheta}-\tilde{\mathbf{x}}\|_1=o(1)$ decays slowly enough to 0 so that the contributions coming from outside the bulk are negligible. Hence, the proof of the proposition is complete. □

Now, recall that based on (A.73), we get that

$$
\begin{aligned}
\vartheta^{dm}\int_{\left(\frac{\tilde{\mathbf{k}}}{\vartheta},\frac{\tilde{\mathbf{k}}+\tilde{\mathbf{1}}}{\vartheta}\right]}\mathcal{R}(\varphi,\tilde{\mathbf{u}})\mathbf{d}\tilde{\mathbf{u}} &= \mathcal{R}(\varphi,\tilde{\mathbf{k}}/\vartheta)+\frac{1}{2\vartheta^d}\sum_{i=1}^{m}\sum_{\ell=1}^{d}\frac{\partial\mathcal{R}(\varphi,\tilde{\mathbf{k}}/\vartheta)}{\partial x_{i_\ell}}+O(\vartheta^{-2d})\\
&= \mathcal{R}(\varphi,\tilde{\mathbf{k}}/\vartheta)+O(\vartheta^{-d}),
\end{aligned}
$$

and multiplying the last expression by

$$\vartheta^{-dm}\cdot\left(\frac{(\vartheta-1+d)!}{(\vartheta-1)!}\right)^m\prod_{j=1}^{m}P_{\mathbf{k}_j,\vartheta-1}(\mathbf{x}_j)$$

and summing over all

$$\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m,$$

we obtain

$$\mathbb{E}\left[u_{n,2}(\varphi,\tilde{\mathbf{x}})\right]=\vartheta^{-dm}\left(\frac{(\vartheta-1+d)!}{(\vartheta-1)!}\right)^m\sum_{\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m}\mathcal{R}(\varphi,\tilde{\mathbf{k}}/\vartheta)\prod_{j=1}^{m}P_{\mathbf{k}_j,\vartheta-1}(\mathbf{x}_j)+O(\vartheta^{-d(m+1)}),$$

which implies that together with the fact that the function $\mathcal{R}(\varphi,\cdot)$ is Lipscitz continuous

$$
\begin{aligned}
&\mathbb{E}\left[u_{n,2}(\varphi,\tilde{\mathbf{x}})\right]-\mathcal{R}(\varphi,\tilde{\mathbf{x}})\\
=&\vartheta^{-dm}\left(\frac{(\vartheta-1+d)!}{(\vartheta-1)!}\right)^m\sum_{\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m}\mathcal{R}(\varphi,\tilde{\mathbf{k}}/\vartheta)\prod_{j=1}^{m}P_{\mathbf{k}_j,\vartheta-1}(\mathbf{x}_j)-\mathcal{R}(\varphi,\tilde{\mathbf{x}})+O(\vartheta^{-d(m+1)})
\end{aligned}
$$

$$\leq \vartheta^{-dm}\left(\frac{(\vartheta-1+d)!}{(\vartheta-1)!}\right)^m \sum_{\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m} \left\{\mathcal{R}(\varphi,\tilde{\mathbf{k}}/\vartheta)-\mathcal{R}(\varphi,\tilde{\mathbf{x}})\right\}\prod_{j=1}^m P_{\mathbf{k}_j,\vartheta-1}(\mathbf{x}_j)+O(\vartheta^{-d(m+1)})$$

$$\leq \vartheta^{-dm}\left(\frac{(\vartheta-1+d)!}{(\vartheta-1)!}\right)^m \sum_{\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m} \|\frac{\tilde{\mathbf{k}}}{\vartheta}-\tilde{\mathbf{x}}\|_1\prod_{j=1}^m P_{\mathbf{k}_j,\vartheta-1}(\mathbf{x}_j)+O(\vartheta^{-d(m+1)})$$

$$\leq \vartheta^{-dm}\left(\frac{(\vartheta-1+d)!}{(\vartheta-1)!}\right)^m \sum_{\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m} \sum_{j=1}^m\sum_{\ell=1}^d \left|k_{j\ell}-\vartheta x_{j\ell}\right|\prod_{j=1}^m P_{\mathbf{k}_j,\vartheta-1}(\mathbf{x}_j)+O(\vartheta^{-d(m+1)})$$

$$= O\left(\sum_{j=1}^m\sum_{\ell=1}^d \sum_{\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m}\left|k_{j\ell}-\vartheta x_{j\ell}\right|\prod_{j=1}^m P_{\mathbf{k}_j,\vartheta-1}(\mathbf{x}_j)\right)+O(\vartheta^{-d(m+1)}).$$

Finally, an application of the Cauchy-Schwarz inequality combined with the identity (A.62) gives us uniformly for $\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m$,

$$\begin{aligned}&\mathbb{E}\left[u_{n,2}(\varphi,\tilde{\mathbf{x}})\right]-\mathcal{R}(\varphi,\tilde{\mathbf{x}})\\ =&O\left(\sum_{j=1}^m\sum_{\ell=1}^d\sqrt{\sum_{\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m}\left|k_{j\ell}-\vartheta x_{j\ell}\right|^2\prod_{j=1}^m P_{\mathbf{k}_j,\vartheta-1}(\mathbf{x}_j)}\sqrt{\sum_{\tilde{\mathbf{k}}\in(\mathbb{N}_0^d\cap(\vartheta-1)\mathbb{S}_{d,1})^m}\prod_{j=1}^m P_{\mathbf{k}_j,\vartheta-1}(\mathbf{x}_j)}\right)+O(\vartheta^{-d(m+1)})\\ =&O(\vartheta^{-dm/2})+O(\vartheta^{-d(m+1)})=O(\vartheta^{-dm/2}).\end{aligned}$$

Therefore, we obtain

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\left|\mathbb{E}\left(u_{n,2}(\varphi,\tilde{\mathbf{x}})\right)-\mathcal{R}(\varphi,\tilde{\mathbf{x}})\right|=O\left(\vartheta^{-md/2}\right).$$

It remains to prove (A.72), which can easily be verified by simply taking $\varphi\equiv 1$ in the above equation. Combining the previously obtained results,

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\left|\mathbb{E}\left(u_{n,2}(\varphi,\tilde{\mathbf{x}})\right)-r^{(m)}(\varphi,\tilde{\mathbf{x}})\mathbb{E}\left(u_{n,2}(1,\tilde{\mathbf{x}})\right)\right|=O\left(\vartheta^{-md/2}\right),\tag{A.74}$$

and if we suppose that $\inf_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\tilde{f}(\tilde{\mathbf{x}})>0$, we can infer that

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{d,1}^m}\left|\widehat{\mathbb{E}}\left[\widehat{r}_{n,2}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\mathbf{\Lambda}}_{n,2}(\tilde{\mathbf{x}}))\right]-r^{(m)}(\varphi,\tilde{\mathbf{x}})\right|=O\left(\vartheta^{-md/2}\right).$$

Hence, the proof is complete. □

A.3. Proof of Section 5: Beta kernels

Let $\mathbf{A}_h, h=1,\ldots,N_n^d$ be the h-th sub-hyper-rectangle. Also let $\mathbf{x}_h$ be the most distant point in $\mathbf{A}_h$ from the origin, that is, $\mathbf{x}_h := \arg\max_{\mathbf{x}\in\mathbf{A}_h}\|\mathbf{x}\|$. Suppose that the design point $\mathbf{x}$ falls into $\mathbf{A}_h$. Then, for all $\tilde{\mathbf{x}}=(\mathbf{x}_1,\ldots,\mathbf{x}_m)$, we denote $\tilde{\mathbf{x}}_h=(\mathbf{x}_{1,h},\ldots,\mathbf{x}_{m,h})$ such that $\tilde{\mathbf{x}}_h := \arg\max_{\tilde{\mathbf{x}}\in\mathbf{A}_h^m}\|\tilde{\mathbf{x}}\|$. For $\tilde{\mathbf{x}}=(\mathbf{x}_1,\ldots,\mathbf{x}_m)\in\mathbb{S}_{\tilde{\mathbf{x}}}:=\prod_{i=1}^m\mathbb{S}_{\mathbf{X}_i}$, where

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

the boundary parameters $\boldsymbol{\eta}_i := (\eta_{i_1}, \ldots, \eta_{i_d})$ either are fixed or shrink to zero at a suitable rate. For each $1 \leq i \leq m$, we divide every edge of the d-hyper-rectangles $\mathbb{S}_{\mathbf{X}_i}$ into N_n evenly spaced grids, resulting in N_n^d identical sub-hyper-rectangles. For any $\tilde{\mathbf{x}} = (\mathbf{x}_1, \ldots, \mathbf{x}_m) \in \mathbb{S}_{\mathbf{X}}^m$, there exists $\boldsymbol{\ell}(\tilde{\mathbf{x}}) = (\ell(\mathbf{x}_1), \ldots, \ell(\mathbf{x}_m))$ such that for all $1 \leq i \leq m$, $1 \leq \ell(\mathbf{x}_i) \leq N_n^d$, and

$$\tilde{\mathbf{x}} \in \prod_{i=1}^{m} \mathbf{A}(\mathbf{x}_{\ell(\mathbf{x}_i)}) \text{ such that } \mathbf{x}_{\ell(\mathbf{x}_i)} := \arg \max_{\mathbf{x} \in \mathbf{A}(\mathbf{x}_{\ell(\mathbf{x}_i)})} \|\mathbf{x}\|.$$

For each $\tilde{\mathbf{x}} \in \mathbb{S}_{\mathbf{X}}^m$, we consider the U-statistic as

$$\begin{aligned}\left|u_{n,3}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}[u_{n,3}(\varphi, \tilde{\mathbf{x}})]\right| &\leq \left|u_{n,3}(\varphi, \tilde{\mathbf{x}}) - u_{n,3}(\varphi, \tilde{\mathbf{x}}_{\boldsymbol{\ell}(\mathbf{x})})\right| + \left|\mathbb{E}[u_{n,3}(\varphi, \tilde{\mathbf{x}}_{\boldsymbol{\ell}(\mathbf{x})})] - \mathbb{E}[u_{n,3}(\varphi, \tilde{\mathbf{x}})]\right| \\ &\quad + \left|u_{n,3}(\varphi, \tilde{\mathbf{x}}_{\boldsymbol{\ell}(\mathbf{x})}) - \mathbb{E}[u_{n,3}(\varphi, \tilde{\mathbf{x}}_{\boldsymbol{\ell}(\mathbf{x})})]\right|.\end{aligned}$$

Before proceeding, we borrow a few lemmas from [79], all of which are key building blocks for the technical proofs below. Throughout, θ_{x_j} denotes a beta random variable so that

$$\theta_{x_j} \overset{\mathcal{D}}{=} \text{Beta}\left\{x_j/b_j + 1, \left(1 - x_j\right)/b_j + 1\right\}.$$

Lemma A.9. *Let θ_{x_j} and θ_{x_k} be independent for $j \neq k$. Then, as $n \to \infty$, we have*

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

Lemma A.10. *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 $n \to \infty$, we have*

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

Lemma A.11. *Under the same condition as in Lemma A.10, as $n \to \infty$, we have*

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

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

A.3.1. Proof of Theorem 5.1

Proof of Theorem 5.1. To establish this theorem, we'll have to truncate the conditional U-statistic. First, let's introduce the following notation:

$$\phi_n = \sqrt{\frac{(\log n/n)}{\prod_{j=1}^{m}\left(\prod_{i=1}^{d} b_{j_i} \eta_{j_i}\right)}},$$

$$\begin{aligned}\omega_{n,3} &= \phi_n^{-1/(1+\gamma)},\\ N_n &= \phi_n^{-(1+\frac{1}{1+\gamma})}\left(\prod_{j=1}^{m}\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)\right)^{-\frac{1}{2}}\left(\sum_{j=1}^{m}\sum_{i=1}^{d}\frac{1}{b_{j_i}^2}\right).\end{aligned}$$

From (A.1), we know that we can write

$$\begin{aligned}u_{n,3}(\varphi,\tilde{\mathbf{x}}) &= u_{n,\ell}^{(m)}\left(\mathcal{G}_{\varphi,\tilde{\mathbf{x}},3}^{(T)}\right) + u_{n,\ell}^{(m)}\left(\mathcal{G}_{\varphi,\tilde{\mathbf{x}},3}^{(R)}\right)\\ &= u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}) + u_{n,3}^{(R)}(\varphi,\tilde{\mathbf{x}}).\end{aligned}$$

Using the same truncation technique we used in the previous sections' proof, we need to establish the results for the truncated and remainder parts.

Truncated Part:

Let us remark that

$$\begin{aligned}\left|u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}) - \mathbb{E}\left(u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}})\right)\right| &= \frac{(n-m)!}{n!}\left|\sum_{i\in I(m,n)}\left\{\mathcal{G}_{\varphi,\tilde{\mathbf{x}},3}^{(T)}(\tilde{\mathbf{X}}_{\mathbf{i}},\tilde{\mathbf{Y}}_{\mathbf{i}}) - \mathbb{E}\left[\mathcal{G}_{\varphi,\tilde{\mathbf{x}},3}^{(T)}(\tilde{\mathbf{X}}_{\mathbf{i}},\tilde{\mathbf{Y}}_{\mathbf{i}})\right]\right\}\right|\\ &= \frac{(n-m)!}{n!}\left|\sum_{i\in I(m,n)} H^{(T)}(\tilde{\mathbf{X}}_{\mathbf{i}},\tilde{\mathbf{Y}}_{\mathbf{i}})\right|,\end{aligned}$$

where

$$H^{(T)}(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}) = \mathcal{G}_{\varphi,\tilde{\mathbf{x}},3}^{(T)}(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}) - \mathbb{E}\left[\mathcal{G}_{\varphi,\tilde{\mathbf{x}},3}^{(T)}(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})\right].$$

We apply Lemma B.2 on the function $H^{(T)}(\cdot,\cdot)$. Throughout the rest of the proof, we suppose that the function $\mathcal{G}_{\varphi,\tilde{\mathbf{x}},3}^{(T)}$ is symmetric. Moreover, by Lemma A.10, for a sufficiently large n, we readily infer

$$\begin{aligned}\left|H^{(T)}(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})\right| &\le 2\omega_{n,3}\left(\frac{9}{4\sqrt{\pi}}\right)^{dm}\prod_{j=1}^{m}\left\{\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)^{-\frac{1}{2}}\right\}\\ &\le 2\left(\frac{9}{4\sqrt{\pi}}\right)^{dm}\frac{\phi_n^{2-1/(1+\gamma)}}{\log(n)} := C_H.\end{aligned}$$

We also remark that

$$\theta = \mathbb{E}[H^{(T)}(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})] = 0.$$

One can easily derive

$$\begin{aligned}\sigma^2 &= \operatorname{Var}(H^{(T)}(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})) \le \mathbb{E}[H^{(T)}(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})^2]\\ &\le \int_{[0,1]^{dm}}\mathbb{E}\left[\left|\varphi^{(T)}(\tilde{\mathbf{Y}})\right|^2 \mid \tilde{\mathbf{X}}=\tilde{\mathbf{u}}\right]\tilde{f}(\tilde{\mathbf{u}})\tilde{\mathcal{K}}^2_{\tilde{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{u}})\mathbf{d}\tilde{\mathbf{u}}.\end{aligned}$$

Using Lyapunov's inequality (2.7), and condition **(C.3)**, for $C_0, C_1 \ge 1$, we have

$$\begin{aligned}\mathbb{E}\left[\left|\varphi^{(T)}(\tilde{\mathbf{Y}})\right|^2 \mid \tilde{\mathbf{X}}=\tilde{\mathbf{u}}\right]\tilde{f}(\tilde{\mathbf{u}}) &\le \left\{\mathbb{E}\left[\left|\varphi^{(T)}(\tilde{\mathbf{Y}})\right|^{2+\gamma} \mid \tilde{\mathbf{X}}=\tilde{\mathbf{u}}\right]\tilde{f}(\tilde{\mathbf{u}})\right\}^{2/(2+\gamma)}\left\{\tilde{f}(\tilde{\mathbf{u}})\right\}^{\gamma/(2+\gamma)}\\ &\le C_1^{2/(2+\gamma)}C_0^{\gamma/(2+\gamma)} \le C_0C_1.\end{aligned}$$

In addition, recall that

$$K^2_{\alpha,\beta}(u) = \frac{B\{2x/b+1, 2(1-x)/b+1\}}{b^2\{x/b+1, (1-x)/b+1\}} \frac{u^{2x/b}(1-u)^{2(1-x)/b}}{B\{2x/b+1, 2(1-x)/b+1\}} \mathbf{1}_{\{u\in[0,1]\}}.$$

By Lemma of [40], the first term is bounded by $b^{-1/2}(1+b)^{3/2}/\{2\sqrt{\pi}\sqrt{x(1-x)}\}$ for a sufficiently large n. The second term is the pdf of Beta $\{2x/b+1, 2(1-x)/b+1\}$. Therefore, we derive

$$\sigma^2 \leq C_0C_1 \prod_{j=1}^{m} \left\{ \prod_{i=1}^{d} \frac{b_{j_i}^{-1/2}\left(1+b_{j_i}\right)^{3/2}}{2\sqrt{\pi}\sqrt{x_{j_i}\left(1-x_{j_i}\right)}} \right\} \leq C_0C_1 \prod_{j=1}^{m} \left\{ \prod_{i=1}^{d} \frac{b_{j_i}^{-1/2}\left(1+b_{j_i}\right)^{3/2}}{2\sqrt{\pi}\sqrt{\eta_{j_i}\left(1-\eta_{j_i}\right)}} \right\}.$$

For a sufficiently large $n, b_{j_1}, \ldots, b_{j_d}$ and $\eta_{j_1}, \ldots, \eta_{j_d}$,$1 \leq j \leq m$ are no greater than $1/2$, and thus

$$\begin{aligned}
\sigma^2 &\leq \prod_{j=1}^{m} \sqrt{\prod_{j=1}^{d} b_{j_i}\eta_{j_i} C_0C_1} \left(\frac{3}{4}\sqrt{\frac{3}{\pi}}\right)^{dm} \\
&\leq n\prod_{j=1}^{m} \sqrt{\prod_{j=1}^{d} b_{j_i}\eta_{j_i} C_0C_1} \left(\frac{3}{4}\sqrt{\frac{3}{\pi}}\right)^{dm} \\
&\leq \frac{\phi_n^2}{\log(n)} C_0C_1 \left(\frac{3}{4}\sqrt{\frac{3}{\pi}}\right)^{dm} \\
&\leq \frac{\phi_n^2}{\log(n)} \rho^2,
\end{aligned}$$

where $\rho^2 := C_0C_1\left(\frac{3}{4}\sqrt{\frac{3}{\pi}}\right)^{dm}$. For any $\varepsilon > 0$ and n large enough, we get that

$$\begin{aligned}
\mathbb{P}\left(\left|u^{(T)}_{n,3}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}\left(u^{(T)}_{n,3}(\varphi, \tilde{\mathbf{x}})\right)\right| > \varepsilon\rho\phi_n\right) &\leq 2\exp\left[-\frac{[n/m]\rho^2\phi_n^2\varepsilon^2}{2\sigma^2 + \frac{2}{3}C_H\rho\varepsilon\phi_n}\right] \\
&\leq 2\exp\left[-\frac{\varepsilon^2\log(n)}{2\left\{1+\frac{2}{3}\left(\frac{9}{4\sqrt{\pi}}\right)^{dm}\frac{\varepsilon\phi_n^{1-1/(1+\gamma)}}{\rho}\right\}}\right].
\end{aligned}$$

Taking into account $\phi_n = o(1)$ and $\frac{2}{3}\left(\frac{9}{4\sqrt{\pi}}\right)^{dm}\frac{\varepsilon\phi_n^{1-1/(1+\gamma)}}{\rho} \leq 1$ for sufficiently large n, it follows that

$$\mathbb{P}\left(\left|u^{(T)}_{n,3}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}\left(u^{(T)}_{n,3}(\varphi, \tilde{\mathbf{x}})\right)\right| > \varepsilon\rho\phi_n\right) \leq 2\exp\left[-\frac{\varepsilon^2\log(n)}{2(1+1)}\right] = 2n^{-\frac{\varepsilon^2}{4}}. \tag{A.75}$$

On the other hand, we have

$$\mathbb{P}\left(\sup_{\tilde{\mathbf{x}}\in\mathbb{S}^m_{\mathbf{X}}}\left|u^{(T)}_{n,3}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}\left(u^{(T)}_{n,3}(\varphi, \tilde{\mathbf{x}})\right)\right| > 2\varepsilon\rho\phi_n\right)$$

$$\leq \mathbb{P}\Bigg(\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\Big|u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}) - u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}_{\boldsymbol{\ell}(\mathbf{x})})$$
$$+\mathbb{E}[u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}_{\boldsymbol{\ell}(\tilde{\mathbf{x}})})] - \mathbb{E}[u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}})]\Big| > \varepsilon\rho\phi_n\Bigg)$$
$$+\mathbb{P}\Bigg(\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\Big|u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}_{\boldsymbol{\ell}(\tilde{\mathbf{x}})}) - \mathbb{E}[u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}_{\boldsymbol{\ell}(\tilde{\mathbf{x}})})]\Big| > \varepsilon\rho\phi_n\Bigg). \tag{A.76}$$

We highlight that

$$\begin{aligned}\left|u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}) - u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}_{\boldsymbol{\ell}(\tilde{\mathbf{x}})})\right| &\leq \frac{(n-m)!}{n!}\sum_{\mathbf{i}\in I(m,n)}\left|\mathcal{G}_{\varphi,\tilde{\mathbf{x}},3}^{(T)}(\tilde{\mathbf{X}}_{\mathbf{i}},\tilde{\mathbf{Y}}_{\mathbf{i}}) - \mathcal{G}_{\varphi,\tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})},3}(\tilde{\mathbf{X}}_{\mathbf{i}},\tilde{\mathbf{Y}}_{\mathbf{i}})\right|\\ &\leq \frac{(n-m)!}{n!}\sum_{\mathbf{i}\in I(m,n)}\left|\varphi^{(T)}(\tilde{\mathbf{Y}}_{\mathbf{i}})\tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{X}}_{\mathbf{i}}) - \varphi^{(T)}(\tilde{\mathbf{Y}}_{\mathbf{i}})\tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})})}(\tilde{\mathbf{X}}_{\mathbf{i}})\right|\\ &\leq \frac{(n-m)!}{n!}\sum_{\mathbf{i}\in I(m,n)}\left|\varphi^{(T)}(\tilde{\mathbf{Y}}_{\mathbf{i}})\right|\left|\tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{X}}_{\mathbf{i}}) - \tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})})}(\tilde{\mathbf{X}}_{\mathbf{i}})\right|.\end{aligned}$$

Hence, the rate of

$$\sup_{\tilde{\mathbf{x}}\in\mathbf{A}_h^m}\left|u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}) - \mathbb{E}\left[u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}_{\boldsymbol{\ell}(\tilde{\mathbf{x}})})\right]\right|$$

is determined by

$$\left|\varphi^{(T)}(\tilde{\mathbf{Y}}_{\mathbf{i}})\right|\left|\tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{X}}_{\mathbf{i}}) - \tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})})}(\tilde{\mathbf{X}}_{\mathbf{i}})\right|.$$

By the mean-value theorem, we have

$$\left|\tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{X}}_{\mathbf{i}}) - \tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})})}(\tilde{\mathbf{X}}_{\mathbf{i}})\right| \leq \sup_{(\tilde{\mathbf{x}},\tilde{\mathbf{u}})\in\mathbf{A}_h^m\times[0,1]^{dm}}\left\|\nabla\left\{\tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\mathbf{u})\right\}\right\|\sup_{\tilde{\mathbf{x}}\in\mathbf{A}_h^m}\left\|\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})}\right\|,$$

for some $\underline{\tilde{\mathbf{x}}}$ joining $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}_{\boldsymbol{\ell}(\tilde{\mathbf{x}})}$. For $k = 1,\ldots,m$, observe that

$$\begin{aligned}\left|\frac{\partial\tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\mathbf{u})}{\partial\mathbf{x}_k}\right| &\leq \left\{\prod_{j=1,j\neq k}^{m}K_{\mathbf{\Lambda}_{n,3}(\mathbf{x}_j)}(\mathbf{u}_j)\right\}\left|\frac{\partial K_{\mathbf{\Lambda}_{n,3}(\mathbf{x}_k)}(\mathbf{u}_k)}{\partial\mathbf{x}_k}\right|\\ &\leq \left\{\prod_{j=1,j\neq k}^{m}\left(\prod_{i=1}^{d}K_{\breve{\alpha}_{j_i},\breve{\beta}_{j_i}}\left(u_{j_i}\right)\right)\right\}\left|\frac{\partial K_{\mathbf{\Lambda}_{n,3}(\mathbf{x}_k)}(\mathbf{u}_k)}{\partial\mathbf{x}_k}\right|\\ &\leq \prod_{j=1,j\neq k}^{m}O\left\{\left(\prod_{i=1}^{d}b_{j_i}\eta_{j_i}\right)^{-\frac{1}{2}}\right\}\left|\frac{\partial K_{\mathbf{\Lambda}_{n,3}(\mathbf{x}_k)}(\mathbf{u}_k)}{\partial\mathbf{x}_k}\right|,\end{aligned}$$

where, by Lemmas A.10 and A.11, for $\ell = 1,\ldots,d$, we have

$$\left|\frac{\partial K_{\mathbf{\Lambda}_{n,3}(\mathbf{x}_k)}(\mathbf{u}_k)}{\partial\mathbf{x}_{k_\ell}}\right| \leq \left\{\prod_{i=1,i\neq\ell}^{d}K_{\breve{\alpha}_{k_i},\breve{\beta}_{k_i}}\left(u_{k_i}\right)\right\}\left|\frac{\partial K_{\breve{\alpha}_{k_\ell},\breve{\beta}_{k_\ell}}\left(u_{k_\ell}\right)}{\partial x_{k_\ell}}\right| = O\left\{\left(\prod_{i=1}^{d}b_{k_i}\eta_{k_i}\right)^{-\frac{1}{2}}\frac{1}{b_{k_\ell}^2}\right\},$$

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

$$\begin{aligned}\prod_{j=1, j\neq k}^{m} K_{\mathbf{\Lambda}_{n,3}(\mathbf{x}_j)}(\mathbf{u}_j) &= \prod_{j=1, j\neq k}^{m} \left\{ \prod_{i=1}^{d} K_{\breve{\alpha}_{j_i}, \breve{\beta}_{j_i}}\left(u_{j_i}\right) \right\} \\ &\leq \prod_{j=1, j\neq k}^{m} O\left\{ \left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i} \right)^{-\frac{1}{2}} \right\},\end{aligned}$$

which implies

$$\sup_{(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) \in \mathbf{A}_h^m \times [0,1]^{dm}} \left\| \nabla \left\{ \tilde{\mathcal{K}}_{\tilde{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{u}}) \right\} \right\| = O\left\{ \prod_{j=1}^{m} \left\{ \left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i} \right)^{-\frac{1}{2}} \right\} \left(\sum_{j=1}^{m} \sum_{i=1}^{d} \frac{1}{b_{j_i}^2} \right) \right\}. \tag{A.77}$$

Using the fact that $\sup_{\tilde{\mathbf{x}} \in \mathbf{A}_h^m} \left\| \tilde{\mathbf{x}} - \tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})} \right\| = O(N_n^{-m})$, it follows that

$$\begin{aligned}\left|\varphi^{(T)}(\tilde{\mathbf{Y}}_{\mathbf{i}})\right| \left| \tilde{\mathcal{K}}_{\tilde{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{X}}_{\mathbf{i}}) - \tilde{\mathcal{K}}_{\tilde{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})})}(\tilde{\mathbf{X}}_{\mathbf{i}}) \right| &\leq O\left\{ \omega_{n,3} N_n^{-m} \prod_{j=1}^{m} \left\{ \left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i} \right)^{-\frac{1}{2}} \right\} \left(\sum_{j=1}^{m} \sum_{i=1}^{d} \frac{1}{b_{j_i}^2} \right) \right\} \\ &= O(\phi_n),\end{aligned} \tag{A.78}$$

uniformly on $(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) \in \mathbf{A}_h^m \times [0,1]^{dm}$. Next, making use of (A.78), we have

$$\left| \mathbb{E}[u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})})] - \mathbb{E}[u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}})] \right| = \left| \mathbb{E}\left[u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})}) - u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}}) \right] \right| \tag{A.79}$$

$$\leq \mathbb{E}\left| \left[u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})}) - u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}}) \right] \right|. \tag{A.80}$$

Just like in the bounded scenario, the progression from (A.79) to (A.80) arises from Jensen's inequality and certain properties of the absolute value function. We can deduce that

$$\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{\mathbf{X}}^m} \left| \mathbb{E}[u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})})] - \mathbb{E}[u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}})] \right| = O(\phi_n).$$

For sufficiently large n and each $m \geq 2$, for some $\varepsilon > 0$, we infer that

$$\mathbb{P}\left(\sup_{\tilde{\mathbf{x}} \in \mathbb{S}_{\mathbf{X}}^m} \left| u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}}) - u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}}_{\ell(\mathbf{x})}) \;\; + \mathbb{E}[u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})})] - \mathbb{E}[u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}})] \right| > \varepsilon\rho\phi_n \right) = 0.$$

Continue, now, with (A.76), by imposing that the kernel function $\mathcal{G}_{\varphi, \tilde{\mathbf{x}}_\ell, 3}^{(T)}(\cdot)$ is symmetric and the U-statistic is decomposed according to [80] decomposition, that is,

$$\begin{aligned}u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})}) - \mathbb{E}[u_{n,3}^{(T)}(\varphi, \tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})})] &= \sum_{q=1}^{m} \frac{m!}{(m-q)!} u_{n,3}^{(q)}\left(\pi_{q,m}(\mathcal{G}_{\varphi, \tilde{\mathbf{x}}_\ell, 3}^{(T)}) \right) \\ &= m u_{n,3}^{(1)}\left(\pi_{1,m}(\mathcal{G}_{\varphi, \tilde{\mathbf{x}}_\ell, 3}^{(T)}) \right) + \sum_{q=2}^{m} \frac{m!}{(m-q)!} u_{n,3}^{(q)}\left(\pi_{q,m}(\mathcal{G}_{\varphi, \tilde{\mathbf{x}}_\ell, 3}^{(T)}) \right).\end{aligned} \tag{A.81}$$

Let us first start with the linear term. We have

$$mu_{n,3}^{(1)}\left(\pi_{1,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)})\right)=\frac{m}{n}\sum_{j=1}^{n}\pi_{1,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)})(\tilde{\mathbf{X}}_i,\tilde{\mathbf{Y}}_i).$$

From Hoeffding's projection (A.81), we have

$$\begin{aligned}\pi_{1,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)})(\mathbf{x},\mathbf{y}) &= \left\{\mathbb{E}\left[\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)}\left((\mathbf{x},\mathbf{X}_2,\ldots,\mathbf{X}_m),(\mathbf{y},\mathbf{Y}_2,\ldots,\mathbf{Y}_m)\right)\right]-\mathbb{E}[\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)}\left(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}\right)]\right\}\\ &= \left\{\mathbb{E}[\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)}\left(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}\right)|(\mathbf{X}_1,\mathbf{Y}_1)=(\mathbf{x},\mathbf{y})]-\mathbb{E}[\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)}\left(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}\right)]\right\}.\end{aligned}$$

Set

$$Z_i^{(T)}=\pi_{1,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)})(\tilde{\mathbf{X}}_i,\tilde{\mathbf{Y}}_i).$$

It's evident that $Z_i^{(T)}$ are independent and identically distributed random variables with a mean of zero, and

$$\sigma^2\leq\frac{\phi_n^2}{\log(n)}\rho^2.$$

Making use of (A.75) and an application of Bernstein's inequality, for some $\varepsilon>0$, yields

$$\begin{aligned}\mathbb{P}\left(\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\left|u_{n,3}^{(1)}\left(\pi_{1,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)})\right)\right|>\varepsilon\rho\phi_n\right) &\leq\sum_{i=1}^{N_n^d}\mathbb{P}\left(\max_{1\leq\ell_i\leq N_n^d}\left|u_{n,3}^{(1)}\left(\pi_{1,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)})\right)\right|>\varepsilon\rho\phi_n\right)\\ &\leq N_n^d\max_{1\leq\ell_i\leq N_n^d}\mathbb{P}\left(\left|u_{n,3}^{(1)}\left(\pi_{1,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)})\right)\right|>\varepsilon\rho\phi_n\right)\\ &=O\left(N_n^d n^{-\frac{\varepsilon^2}{4}}\right).\end{aligned}\tag{A.82}$$

Moving to the nonlinear term, we will prove that for $2\leq q\leq m$:

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\frac{\binom{m}{q}\left|u_{n,3}^{(q)}\left(\pi_{q,m}\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)}\right)\right|}{\phi_n}=o_{\mathbb{P}}(1),$$

which implies that, for $1\leq i\leq m$ and $\boldsymbol{\ell}=(\ell_1,\ldots,\ell_m)$:

$$\max_{1\leq\ell_i\leq N_n^d}\frac{\binom{m}{q}\left|u_n^{(q)}\left(\pi_{q,m}\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)}\right)\right|}{\phi_n}=o_{\mathbb{P}}(1).$$

To prove the abovementioned equation, we need to apply Proposition 1 of [5] (see Lemma B.3). We can see that $\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)}$ is bounded by $\left(\frac{9}{4\sqrt{\pi}}\right)^{dm}\frac{\phi_n^{2-1/(1+\gamma)}}{\log(n)}$, hence for $\varepsilon>0$, we have

$$\mathbb{P}\left(n^{1/2}\left|\sum_{q=2}^{m}\frac{m!}{(m-q)!}u_{n,3}^{(q)}\left(\pi_{q,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_\ell,3}^{(T)})\right)\right|>\varepsilon\rho\phi_n\right)$$

$$=\mathbb{P}\left(\left|\sum_{q=2}^{m}\frac{m!}{(m-q)!}u_{n,3}^{(q)}\left(\pi_{q,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_{\ell},3}^{(T)})\right)\right|>n^{-1/2}\varepsilon\rho\phi_n\right)$$
$$=\mathbb{P}\left(\left|\sum_{q=2}^{m}\frac{m!}{(m-q)!}u_{n,3}^{(q)}\left(\pi_{q,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_{\ell},3}^{(T)})\right)\right|>\varepsilon_0\rho\phi_n\right),$$

where $\varepsilon_0=\dfrac{\varepsilon}{\sqrt{n}}$. Now for $t=\varepsilon\rho\phi_n$, Lemma B.3 gives

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{q=2}^{m}\frac{m!}{(m-q)!}u_{n,3}^{(q)}\left(\pi_{q,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_{\ell},3}^{(T)})\right)\right|>\varepsilon_0\rho\phi_n\right)&\leq 2\exp\left(-\frac{t(n-1)^{1/2}}{2^{m+2}m^{m+1}\frac{1}{2}C_H}\right)\\&\leq 2\exp\left(-\frac{\varepsilon\rho\phi_n(n-1)^{1/2}}{2^{m+2}m^{m+1}\frac{1}{2}C_H}\right)\\&\leq 2\exp\left(-\frac{\varepsilon(n-1)^{1/2}\log(n)}{2^{m+2}m^{m+1}(\frac{1}{\sqrt{3}})^{dm}\phi_n^{1-1/(1+\gamma)}}\right).\end{aligned}$$

By the last result, it follows that there exists $\varepsilon>0$ in such a way that

$$\mathbb{P}\left(\left|\sum_{q=2}^{m}\binom{m}{q}u_{n,3}^{(q)}\left(\pi_{q,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_{\ell},3}^{(T)})\right)\right|>\varepsilon_0\rho\phi_n\right)\leq n^{-\varepsilon_0/2C_6},$$

where

$$C_6=2^{m+2}m^{m+1}\left(\frac{1}{\sqrt{3}}\right)^{dm}\phi_n^{1-1/(1+\gamma)}.$$

Therefore, for each $\varepsilon_0>0$, $1\leq i\leq m$ and $\ell=(\ell_1,\ldots,\ell_m)$, we infer that

$$\begin{aligned}&\mathbb{P}\left(\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\left|\sum_{q=2}^{m}\binom{m}{q}u_{n,3}^{(q)}\left(\pi_{q,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_{\ell},3}^{(T)})\right)\right|>\varepsilon_0\rho\phi_n\right)\\&\leq N_n^d\max_{1\leq\ell_i\leq N_n^d}\mathbb{P}\left(\left|\sum_{q=2}^{m}\binom{m}{q}u_{n,3}^{(q)}\left(\pi_{q,m}(\mathcal{G}_{\varphi,\tilde{\mathbf{x}}_{\ell},3}^{(T)})\right)\right|>\varepsilon_0\rho\phi_n\right)\\&\leq N_n^d n^{-m(\varepsilon_0/2C_6)}.\end{aligned}\tag{A.83}$$

By combining (A.75) and (A.83), for some $\varepsilon>0$, it follows that

$$\mathbb{P}\left(\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\left|u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}_{\boldsymbol{\ell}(\tilde{\mathbf{x}})})-\mathbb{E}[u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}_{\boldsymbol{\ell}(\tilde{\mathbf{x}})})]\right|>\varepsilon\rho\phi_n\right)=O(N_n^d n^{-m\varepsilon^2/4}),\tag{A.84}$$

which implies for $\varepsilon=2\sqrt{5d}$, as $n\to\infty$,

$$N_n^d n^{-m\varepsilon^2/4}=\phi_n^{-p(1+\frac{1}{1+\gamma})}\left(\prod_{j=1}^{m}\left(\prod_{i=1}^{d}b_{j_i}\eta_{j_i}\right)\right)^{-\frac{d}{2}}\left(\sum_{j=1}^{m}\sum_{i=1}^{d}\frac{1}{b_{j_i}^2}\right)^{d}n^{-5md}$$

$$= \left[(\log n)^{-5m} \phi_n^{10m-(1+\frac{1}{1+\gamma})} \left(\prod_{j=1}^{m} \left(\prod_{i=1}^{d} \eta_{j_i} \right) \right)^{-\frac{1}{2}} \left(\sum_{j=1}^{m} \sum_{i=1}^{d} \left(\prod_{k=1,k\neq j}^{m} \left(\prod_{\ell=1,\ell\neq i}^{d} b_{j_i} \right) \right)^{\frac{5m-1}{2}} b_{j_i}^{5(m-1)/2} \right) \right]^{d} \to 0.$$

Hence, the proof is complete. □

Reminder Part:

Notice that

$$u_{n,3}^{(R)}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}\left[u_{n,3}^{(R)}(\varphi, \tilde{\mathbf{x}})\right] = \frac{(n-m)!}{n!} \sum_{\mathbf{i}\in I(m,n)} \mathcal{G}_{\varphi,\tilde{\mathbf{x}},3}^{(R)}(\tilde{\mathbf{X}}_{\mathbf{i}}, \tilde{\mathbf{Y}}_{\mathbf{i}}) - \mathbb{E}\left[\mathcal{G}_{\varphi,\tilde{\mathbf{x}},3}^{(R)}(\tilde{\mathbf{X}}_{\mathbf{i}}, \tilde{\mathbf{Y}}_{\mathbf{i}})\right].$$

Now, using the fact that for $\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right| > \omega_{n,3}$, we have $(\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right| / \omega_{n,3})^{1+\gamma} > 1$, which implies that

$$\begin{aligned} \left|\mathbb{E}[u_{n,3}^{(R)}(\varphi, \tilde{\mathbf{x}})]\right| &\leq \mathbb{E}[\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right| \mathbf{1}_{\{|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})|>\omega_{n,3}\}} \tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{X}}_{\mathbf{i}})] \\ &\leq \mathbb{E}[\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right| \left(\frac{\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right|}{\omega_{n,3}}\right)^{1+\gamma} \mathbf{1}_{\{|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})|>\omega_{n,3}\}} \tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{X}}_{\mathbf{i}})] \\ &\leq \omega_{n,3}^{-(1+\gamma)} \mathbb{E}[\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right|^{2+\gamma} \tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{X}}_{\mathbf{i}})], \end{aligned} \quad (A.85)$$

where, by Assumption **(C.3)** and the fact that $\tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\cdot)$ is the density function of the product of dm independent beta random variables $\boldsymbol{\theta}_{\mathbf{x}_i} := (\theta_{\mathbf{x}_1}, \ldots, \theta_{\mathbf{x}_m}) \in [0,1]^{dm}$, $i = 1, \ldots, dm$, we have

$$\begin{aligned} \mathbb{E}\left[\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right|^{2+\gamma} \tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{X}}_{\mathbf{i}})\right] &= \mathbb{E}\left\{\mathbb{E}\left(\left|\varphi(\tilde{\mathbf{Y}}_{\mathbf{i}})\right|^{2+\gamma} \mid \tilde{\mathbf{X}}_{\mathbf{i}}\right) \tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{X}}_{\mathbf{i}})\right\} \\ &= \int_{[0,1]^{dm}} \mathbb{E}\left(\left|\varphi(\tilde{\mathbf{Y}})\right|^{2+\gamma} \mid \tilde{\mathbf{X}} = \tilde{\mathbf{u}}\right) \tilde{f}(\tilde{\mathbf{u}}) \tilde{\mathcal{K}}_{\bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{u}}) \mathbf{d}\tilde{\mathbf{u}} \leq C_1. \end{aligned} \quad (A.86)$$

Hence, by the definition of $\omega_{n,3}$, $\left|\mathbb{E}\left[u_{n,3}^{(R)}(\varphi, \tilde{\mathbf{x}})\right]\right| \leq O(\phi_n)$ uniformly on $\tilde{\mathbf{x}} \in \mathbb{S}_{\mathbf{X}}^m$. Consequently, Markov's inequality gives us

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \left|u_{n,3}^{(R)}(\varphi, \tilde{\mathbf{x}}) - \mathbb{E}\left[u_{n,3}^{(R)}(\varphi, \tilde{\mathbf{x}})\right]\right| = O_{\mathbb{P}}(\phi_n). \quad (A.87)$$

Hence, the proof is complete.

Proof of Theorem 5.2. Similar to the proof of the bias terms in the previous sections, we know that based on (A.2) :

$$\left|\widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}})) - \widehat{\mathbb{E}}\left(\widehat{r}_{n,3}^{(m)}(\varphi, \tilde{\mathbf{x}}; \bar{\mathbf{\Lambda}}_{n,3}(\tilde{\mathbf{x}}))\right)\right| \leq \mathscr{I}_{3,1} + \mathscr{I}_{3,2}. \quad (A.88)$$

In addition, we have for some positive constants $c_1, c_2 > 0$,

$$\begin{aligned} \sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \left|u_{n,3}(1, \tilde{\mathbf{x}})\right| &= c_1 \qquad \text{a.s.}, \\ \sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \left|\mathbb{E}\left(u_{n,3}(1, \tilde{\mathbf{x}})\right)\right| &= c_2, \\ \sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \left|\mathbb{E}\left(u_{n,3}(\varphi, \tilde{\mathbf{x}})\right)\right| &= O(1). \end{aligned}$$

Hence by Theorem 5.1, for some $c'' > 0$, we get with probability 1:

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \frac{\left|\widehat{r}_{n,3}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\boldsymbol{\Lambda}}_{n,3}(\tilde{\mathbf{x}})) - \widehat{\mathbb{E}}\left(\widehat{r}_{n,3}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\boldsymbol{\Lambda}}_{n,3}(\tilde{\mathbf{x}}))\right)\right|}{\phi_n} \leq \sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \frac{(\mathscr{I}_{3,1})}{\phi_n} + \sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \frac{(\mathscr{I}_{3,2})}{\phi_n} \leq c''.$$

Hence, the proof is complete. □

Proof of Theorem 5.3. To obtain the desired results, we need to prove that:

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \left|\mathbb{E}\left\{u_{n,3}(\varphi,\tilde{\mathbf{x}})\right\} - \mathcal{R}(\varphi,\tilde{\mathbf{x}})\right| = O\left(\sum_{j=1}^{m}\sum_{i=1}^{d} b_{j_i}\right).$$

We first remark that

$$\begin{aligned}\mathbb{E}\left[u_{n,3}(\varphi,\tilde{\mathbf{x}})\right] &= \int_{[0,1]^{dm}} r^{(m)}(\varphi,\tilde{\mathbf{t}})\tilde{f}(\tilde{\mathbf{u}})\tilde{\mathcal{K}}_{\bar{\boldsymbol{\Lambda}}_{n,3}(\tilde{\mathbf{x}})}(\tilde{\mathbf{u}})\mathbf{d}\tilde{\mathbf{u}} \\ &= \mathbb{E}\left[\mathcal{R}(\varphi,\boldsymbol{\theta}_{\tilde{\mathbf{x}}})\right],\end{aligned}$$

where $\boldsymbol{\theta}_{\tilde{\mathbf{x}}} = (\theta_{\mathbf{x}_1},\ldots,\theta_{\mathbf{x}_m}) \in [0,1]^{dm}$. Following the same reasoning as the proof of Theorem 4.6, a second-order Taylor expansion around $\boldsymbol{\theta}_{\tilde{\mathbf{x}}} = \tilde{\mathbf{x}}$ gives us:

$$\begin{aligned}\mathbb{E}\left[\mathcal{R}(\varphi,\boldsymbol{\theta}_{\tilde{\mathbf{x}}})\right] &= \mathcal{R}(\varphi,\tilde{\mathbf{x}}) + \sum_{i=1}^{m}\sum_{\ell=1}^{d}\frac{\partial\mathcal{R}(\varphi,\tilde{\mathbf{x}})}{\partial x_{i_\ell}}\mathbb{E}(\theta_{xi_\ell} - x_{i_\ell}) + \frac{1}{2}\sum_{i=1}^{m}\sum_{\ell=1}^{d}\frac{\partial^2\mathcal{R}(\varphi,\underline{\tilde{\mathbf{x}}})}{\partial x_{i_\ell}^2}\mathbb{E}(\theta_{xi_\ell} - x_{i_\ell})^2 \\ &\quad + \sum_{i,j=1,i\neq j}^{m}\sum_{\ell,r=1,\ell\neq r}^{d}\frac{\partial^2\mathcal{R}(\varphi,\underline{\tilde{\mathbf{x}}})}{\partial x_{i_\ell}\partial x_{j_r}}\mathbb{E}\left\{(\theta_{xi_\ell} - x_{i_\ell})(\theta_{xj_r} - x_{j_r})\right\},\end{aligned}$$

for some $\underline{\tilde{\mathbf{x}}}$ joining $\boldsymbol{\theta}_{\tilde{\mathbf{x}}}$ and $\tilde{\mathbf{x}}$. Taking into account condition **(C.2)** and Lemma A.9 implies that

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \left|\mathbb{E}\left\{u_{n,3}(\varphi,\tilde{\mathbf{x}})\right\} - \mathcal{R}(\varphi,\tilde{\mathbf{x}})\right| = O\left(\sum_{j=1}^{m}\sum_{i=1}^{d} b_{j_i}\right). \tag{A.89}$$

Taking $\varphi \equiv 1$ in the above equation gives us

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \left|\mathbb{E}\left\{u_{n,3}(1,\tilde{\mathbf{x}})\right\} - \tilde{f}(\tilde{\mathbf{x}})\right| = O\left(\sum_{j=1}^{m}\sum_{i=1}^{d} b_{j_i}\right). \tag{A.90}$$

Combining these two results leads to

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \left|\mathbb{E}\left(u_{n,3}(\varphi,\tilde{\mathbf{x}})\right) - r^{(m)}(\varphi,\tilde{\mathbf{x}})\mathbb{E}\left(u_{n,3}(1,\tilde{\mathbf{x}})\right)\right| = O\left(\sum_{j=1}^{m}\sum_{i=1}^{d} b_{j_i}\right), \tag{A.91}$$

and if we suppose that $\inf_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \tilde{f}(\tilde{\mathbf{x}}) > 0$, we can infer that

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m} \left|\widehat{\mathbb{E}}\left[\widehat{r}_{n,3}^{(m)}(\varphi,\tilde{\mathbf{x}};\bar{\boldsymbol{\Lambda}}_{n,3}(\tilde{\mathbf{x}}))\right] - r^{(m)}(\varphi,\tilde{\mathbf{x}})\right| = O\left(\sum_{j=1}^{m}\sum_{i=1}^{d} b_{j_i}\right).$$

Hence, the proof is complete. □

Proof of Theorem 5.5. Using the notation established in the proof of Theorem 5.1, and employing a reasoning akin to that of [79], we proceed to redefine $\omega_{n,3}$ and N_n,

$$\omega_{n,3} := n^{\frac{1+\varepsilon}{2+\gamma}} \text{ and } N_n := n^{1+\varepsilon}\left(\prod_{j=1}^{m}\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)\right)^{-\frac{1}{2}}\left(\sum_{j=1}^{m}\sum_{i=1}^{d}\frac{1}{b_{j_i}^2}\right),$$

for an arbitrarily small ε. In order to prove Theorem 5.5, we need to demonstrate that

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\left|u_{n,3}(\varphi,\tilde{\mathbf{x}}) - \mathbb{E}\left\{u_{n,3}(\varphi,\tilde{\mathbf{x}})\right\}\right| = O\left(\sqrt{\frac{(\log n/n)}{\prod_{j=1}^{m}\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)}}\right) \; a.s. \tag{A.92}$$

Similar to the proof of Theorem 5.1, for the reminder part, (A.85) and (A.86) give us

$$\left|\mathbb{E}[u_{n,3}^{(R)}(\varphi,\tilde{\mathbf{x}})]\right| \leq \omega_{n,3}^{-(1+\gamma)}C_1 = n^{-(1+\varepsilon)\left(\frac{1+\gamma}{2+\gamma}\right)}C_1 \leq O(\phi_n). \tag{A.93}$$

Moreover, by **(C.3)** and Markov's inequality, we infer readily that

$$\sum_{n=1}^{\infty}\mathbb{P}\left(|\varphi(\tilde{\mathbf{Y}}_n)| > \omega_{n,3}\right) < \sum_{n=1}^{\infty}\frac{\mathbb{E}(|\varphi(\tilde{\mathbf{Y}}_n)|^{2+\gamma})}{\omega_{n,3}^{2+\gamma}} = \mathbb{E}(|\varphi(\tilde{\mathbf{Y}}_n)|^{2+\gamma})\sum_{n=1}^{\infty}\frac{1}{n^{1+\varepsilon}} < \infty.$$

Using the Borel-Cantelli lemma for a sufficiently large n, this gives us $|\varphi(\tilde{\mathbf{Y}}_n)| \leq \omega_{n,3}$ with probability 1. This implies that $\left|\varphi(\tilde{\mathbf{Y}}_i)\right| \leq \omega_{n,3}$ for any $i \leq n$ with probability 1 for a sufficiently large n. It follows that $u_{n,3}^{(R)}(\varphi,\tilde{\mathbf{x}}) = 0$ with probability 1, that is,

$$\left|u_{n,3}^{(R)}(\varphi,\tilde{\mathbf{x}}) - \mathbb{E}\left[u_{n,3}^{(R)}(\varphi,\tilde{\mathbf{x}})\right]\right| = O(\phi_n) \; a.s.,$$

uniformly on $\tilde{\mathbf{x}} \in \mathbb{S}_{\mathbf{X}}^m$. Next, observe that

$$\omega_{n,3}N_n^{-m}\prod_{j=1}^{m}\left\{\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)^{-\frac{1}{2}}\right\}\left(\sum_{j=1}^{m}\sum_{i=1}^{d}\frac{1}{b_{j_i}^2}\right) = n^{-m(1+\varepsilon)\left(\frac{1+\gamma}{2+\gamma}\right)} \leq O\left(\phi_n\right).$$

Hence, we infer that

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\left|\mathbb{E}[u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}_{\ell(\tilde{\mathbf{x}})})] - \mathbb{E}[u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}})]\right| = O\left\{\omega_{n,3}N_n^{-m}\prod_{j=1}^{m}\left\{\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)^{-\frac{1}{2}}\right\}\left(\sum_{j=1}^{m}\sum_{i=1}^{d}\frac{1}{b_{j_i}^2}\right)\right\} = O(\phi_n).$$

Then, (A.75) and (A.83) hold for a sufficiently large n. In addition, **(C.4')** implies that

$$\prod_{j=1}^{m}\left\{\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)^{-\frac{1}{2}}\right\}\left(\sum_{j=1}^{m}\sum_{i=1}^{d}\frac{1}{b_{j_i}^2}\right) = O\left\{n^{\frac{1}{1-\kappa}}\left(\frac{\left(\prod_{j=1}^{m}\left\{\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)^{-\frac{1}{2}}\right\}\right)^{\frac{\kappa}{2}}}{\log(n)}\right)^{\frac{1}{1-\kappa}}\right\} \leq O\left(n^{\frac{1}{1-\kappa}}\right),$$

where the last inequality holds because $\left(\prod_{j=1}^{m}\left\{\left(\prod_{i=1}^{d} b_{j_i}\eta_{j_i}\right)^{-\frac{1}{2}}\right\}\right)^{\frac{\kappa}{2}}$ / $\log(n)$ is bounded. Then, picking

$$K = 2\sqrt{(d+1)(1+\varepsilon)+d/(1-\kappa)},$$

yields $N_n^d n^{-K^2/4} = O\left\{n^{-(1+\varepsilon)}\right\}$, so that

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\left|u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}) - \mathbb{E}\left(u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}})\right)\right| > \varepsilon\rho\phi_n\right) \leq \sum_{n=1}^{\infty} O\left(\frac{1}{n^{1+\varepsilon}}\right) < \infty.$$

Therefore, by the Borel-Cantelli lemma, we obtain

$$\sup_{\tilde{\mathbf{x}}\in\mathbb{S}_{\mathbf{X}}^m}\left|u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}}) - \mathbb{E}\left(u_{n,3}^{(T)}(\varphi,\tilde{\mathbf{x}})\right)\right| = O(\phi_n) \quad a.s.$$

Hence, the proof is complete. □

Proof of Theorem 5.6. The proof of Theorem 5.6 is done in the same fashion as the proof of Theorem 5.2, combining (A.2) with the results of Theorem 5.5. □

Proof of Theorem 5.7. The proof of Theorem 5.7 is the same as the proof of Theorem 5.3. □

B. Appendix-2

This appendix contains supplementary information that is an essential part of providing a more comprehensive understanding of the paper.

Lemma B.1 (Lemma 2.2.9, [143]). *Let $X_1,\ldots,X_n$ be independent random variables with bounded ranges $[-M, M]$ and zero means. Then,*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| > t\right) \leq 2\exp\left\{-\frac{t^2}{2(v+Mt/3)}\right\},$$

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

Lemma B.2 (Theorem A. page 201, [129]). *Let f be a symmetric function taking its variables from $\mathbb{S}_{d,1}$ satisfying $\|f\|_\infty \leq c$,*

$$\mathbb{E}f(X_1,\ldots,X_m) = \theta,$$

and

$$\sigma^2 = Var\left(f(X_1,\ldots,X_m)\right),$$

then for $t > 0$ and $n \geq m$, we have:

$$\mathbb{P}\left\{|u_{n,\ell}^{(m)}(f) - \theta| \geq t\right\} \leq \exp\left\{-\frac{[n/m]t^2}{2\sigma^2 - \frac{2}{3}ct}\right\}.$$

Lemma B.3 (Proposition.1, [5])**.** *If $G : S^m \to \mathbb{R}$ is a measurable symmetric function with $\|G\|_\infty = b$, then*

$$\mathbb{P}\left\{n^{1/2}\left|\sum_{j=2}^{m}\binom{m}{j}u_n^{(j)}\left(\pi_{j,m}G\right)\right| \geqslant t\right\} \leqslant 2\exp\left(-\frac{t(n-1)^{1/2}}{2^{m+2}m^{m+1}b}\right).$$

Lemma B.4 (Lemma 1, [117])**.** *We have, as $b \to 0$ and uniformly for $\mathbf{x} \in \mathbb{S}_{d,1}$,*

$$0 < A_b(\mathbf{x}) \le \frac{b^{(d+1)/2}(1/b+d)^{d+1/2}}{(4\pi)^{d/2}\sqrt{(1-\|\mathbf{x}\|_1)\prod_{i=1}^{d}x_i}}(1+O(b)).$$

Furthermore, for any subset $\emptyset \neq \mathcal{J} \subseteq [d]$, and any $\kappa \in (0,\infty)^d$,

$$A_b(\mathbf{x}) = \begin{cases} \breve{b}^{-d/2}\psi(\mathbf{x})\,(1+O_s(b))\,, \\ \text{if } x_i/b \to \infty,\ \ \forall i \in [d] \text{ and } (1-\|\mathbf{x}\|_1)/b \to \infty, \\ \\ b^{-(d+|\mathcal{J}|)/2}\psi_{\mathcal{J}}(\mathbf{x})\prod_{i\in\mathcal{J}}\dfrac{\Gamma(2\kappa_i+1)}{2^{2\kappa_i+1}\Gamma^2(\kappa_i+1)}\cdot(1+O_{\kappa,\mathbf{x}}(b))\,, \\ \text{if } x_i/b \to \kappa_i,\ \ \forall i \in \mathcal{J} \text{ and } x_i/b \to \infty,\ \ \forall i \in [d]\backslash\mathcal{J},\ \text{and } (1-\|\mathbf{x}\|_1)/b \to \infty, \end{cases}$$

where $\psi(\cdot)$ and $\psi_{\mathcal{J}}(\cdot)$ are defined for every subset of indices $\mathcal{J} \subseteq [d]$, by

$$\psi(\mathbf{x}) := \psi_{\emptyset}(\mathbf{x}) \quad \text{and} \quad \psi_{\mathcal{J}}(\mathbf{x}) := \left[(4\pi)^{d-|\mathcal{J}|}\cdot(1-\|\mathbf{x}\|_1)\prod_{i\in[d]\backslash\mathcal{J}}x_i\right]^{-1/2}. \tag{B.1}$$

Lemma B.5 (Lemma 2, [117])**.** *If $\alpha_1,\ldots,\alpha_d,\beta \ge 2$, then*

$$\sup_{\mathbf{x}\in\mathcal{S}_d} K_{\alpha,\beta}(\mathbf{x}) \le \sqrt{\frac{\|\boldsymbol{\alpha}\|_1+\beta-1}{(\beta-1)\prod_{i\in[d]}(\alpha_i-1)}}\,(\|\boldsymbol{\alpha}\|_1+\beta-d-1)^d\,.$$

Lemma B.6 (Lemma 3, [117])**.** *If $\alpha_1,\ldots,\alpha_d,\beta \ge 2$, then for all $\mathbf{x} \in \mathrm{Int}\,(\mathbb{S}_{d,1})$,*

$$\left|\frac{\partial}{\partial\alpha_j}K_{\alpha,\beta}(\mathbf{x})\right| \le \left\{\left|\log\left(\|\boldsymbol{\alpha}\|_1+\beta\right)\right| + \left|\log\left(\alpha_j\right)\right| + \left|\log x_j\right|\right\}\cdot\sqrt{\frac{\|\boldsymbol{\alpha}\|_1+\beta-1}{(\beta-1)\prod_{i=1}^{d}(\alpha_i-1)}}\,(\|\boldsymbol{\alpha}\|_1+\beta-d-1)^d\,,$$

$$\left|\frac{\partial}{\partial\beta}K_{\alpha,\beta}(\mathbf{x})\right| \le \left\{\left|\log\left(\|\boldsymbol{\alpha}\|_1+\beta\right)\right| + |\log(\beta)| + \left|\log\left(1-\|\mathbf{x}\|_1\right)\right|\right\}\cdot\sqrt{\frac{\|\boldsymbol{\alpha}\|_1+\beta-1}{(\beta-1)\prod_{i=1}^{d}(\alpha_i-1)}}\,(\|\boldsymbol{\alpha}\|_1+\beta-d-1)^d\,.$$

Lemma B.7 (Lemma 4, [117])**.** *If $\alpha_1,\ldots,\alpha_d,\beta,\alpha'_1,\ldots,\alpha'_d,\beta' \ge 2$, and $\mathbf{X}$ is F distributed with a bounded density f supported on $\mathbb{S}_{d,1}$, then*

$$\begin{aligned}&\mathbb{E}\left[\left|K_{\alpha',\beta'}(\boldsymbol{X}) - K_{\alpha,\beta}(\boldsymbol{X})\right|\right] \\ \le& 3(d+1)\|f\|_\infty\sqrt{\frac{\|\boldsymbol{\alpha}\vee\boldsymbol{\alpha}'\|_1+(\beta\vee\beta')-1}{((\beta\wedge\beta')-1)\prod_{i\in[d]}\left(\left(\alpha_i\wedge\alpha'_i\right)-1\right)}}\cdot\left(\|\boldsymbol{\alpha}\vee\boldsymbol{\alpha}'\|_1+(\beta\vee\beta')-d-1\right)^d\end{aligned}$$

$$\cdot \log\left(\|\boldsymbol{\alpha} \vee \boldsymbol{\alpha}'\|_1 + (\beta \vee \beta')\right) \cdot \left\|(\boldsymbol{\alpha}', \beta') - (\boldsymbol{\alpha}, \beta)\right\|_\infty,$$

where $\boldsymbol{\alpha} \vee \boldsymbol{\alpha}' := \left(\max\left\{\alpha_i, \alpha'_i\right\}\right)_{i\in[d]}, \beta \vee \beta' := \max\{\beta, \beta'\}$*, and* $\beta \wedge \beta' := \min\{\beta, \beta'\}$*. Furthermore, let*

$$\mathbb{S}_{d,1}(\delta) := \{\mathbf{x} \in \mathbb{S}_{d,1} : 1 - \|\mathbf{x}\|_1 \geq \delta \text{ and } x_i \geq \delta \forall i \in [d]\}, \quad \delta > 0.$$

Then, for $0 < \delta \leq e^{-1}$*, we have*

$$\max_{\mathbf{x}\in\mathbb{S}_{d,1}(\delta)} \left|K_{\boldsymbol{\alpha}',\beta'}(\mathbf{x}) - K_{\boldsymbol{\alpha},\beta}(\mathbf{x})\right|$$
$$\leq 3(d+1)\|f\|_\infty |\log\delta| \cdot \sqrt{\frac{\|\boldsymbol{\alpha} \vee \boldsymbol{\alpha}'\|_1 + (\beta \vee \beta') - 1}{((\beta \wedge \beta') - 1)\prod_{i\in[d]}\left(\alpha_i \wedge \alpha'_i\right) - 1}} \cdot (\|\boldsymbol{\alpha} \vee \boldsymbol{\alpha}'\|_1 + (\beta \vee \beta') - d - 1)^d$$
$$\cdot \log\left(\|\boldsymbol{\alpha} \vee \boldsymbol{\alpha}'\|_1 + (\beta \vee \beta')\right) \cdot \left\|(\boldsymbol{\alpha}', \beta') - (\boldsymbol{\alpha}, \beta)\right\|_\infty.$$