

A nonparametric distribution-free test of independence among continuous random vectors based on L_1 -norm

NOUR-EDDINE BERRAHOU^{1,a}, SALIM BOUZEBDA^{2,c}  and LAHCEN DOUGE^{1,b}

¹ *Laboratoire de Modélisation des Systèmes Complexes, University Cadi Ayyad Marrakesh, Morocco, ^an.berrahou@uca.ac.ma, ^blahcen.douge@uca.ac.ma*

² *Université de technologie de Compiègne, LMAC (Laboratory of Applied Mathematics of Compiègne), ^csalim.bouzebda@utc.fr*

We propose a novel statistical test to assess the mutual independence of multidimensional random vectors. Our approach is based on the L_1 -distance between the joint density function and the product of the marginal densities associated with the presumed independent vectors. Under the null hypothesis, we employ Poissonization techniques to establish the asymptotic normal approximation of the corresponding test statistic, without imposing any regularity assumptions on the underlying Lebesgue density function, denoted as $f(\cdot)$. Remarkably, we observe that the limiting distribution of the L_1 -based statistics remains unaffected by the specific form of $f(\cdot)$. This unexpected outcome contributes to the robustness and versatility of our method. Moreover, our tests exhibit nontrivial local power against a subset of local alternatives, which converge to the null hypothesis at a rate of $n^{-1/2}h_n^{-d/4}$, $d \geq 2$, where n represents the sample size and h_n denotes the bandwidth. Finally, the theory is supported by a comprehensive simulation study to investigate the finite-sample performance of our proposed test. The results demonstrate that our testing procedure generally outperforms existing approaches across various examined scenarios.

Keywords: Asymptotic normality; distribution-free tests; independence test; kernel density function estimator; L_1 -distance; Poissonization.

MSC2020 subject classifications: 60F05; (60F15, 60F17, 62G07).

1. Introduction

One of the classical and important problems in statistics is testing the independence between two or more components of a random vector. Testing for mutual independence, which characterizes the structural relationships between random variables and is strictly stronger than pairwise independence, is a fundamental task in inference. Independent component analysis of a random vector, which consists of searching for a linear transformation that minimizes the statistical dependence between its components, is one of the fields in which mutual independence plays a central role; for instance, see [Bach and Jordan \(2003\)](#), [Chen and Bickel \(2006\)](#), [Samworth and Yuan \(2012\)](#) and [Matteson and Tsay \(2017\)](#). Testing independence also has many applications, including causal inference ([Pearl \(2009\)](#), [Peters et al. \(2014\)](#), [Pfister et al. \(2018\)](#), [Chakraborty and Zhang \(2019\)](#)), graphical modeling ([Lauritzen \(1996\)](#), [Gan, Narisetty and Liang \(2019\)](#)), linguistics ([Nguyen and Eisenstein \(2017\)](#)), clustering ([Székely and Rizzo, 2005](#)), dimension reduction ([Fukumizu, Bach and Jordan, 2004](#), [Sheng and Yin, 2016](#)). The traditional approach for testing independence is based on Pearson's correlation coefficient; for instance, refer to [Binet and Vaschide \(1897\)](#), [Pearson \(1920\)](#), [Spearman \(1904\)](#), [Kendall \(1938\)](#). However, its lack of robustness to outliers and departures from normality eventually led researchers to consider alternative nonparametric procedures. To overcome such a problem, a natural approach is to consider the functional difference between the empirical joint distribution and the product of the empirical marginal

distributions, see [Hoeffding \(1948\)](#), [Blum, Kiefer and Rosenblatt \(1961\)](#) and [Bouzebda \(2011\)](#). This approach can also use characteristic empirical functions; see [Csörgő \(1985\)](#). Inspired by the work of [Blum, Kiefer and Rosenblatt \(1961\)](#) and [Dugué \(1975\)](#), [Deheuvels \(1981\)](#) studied a test of multivariate independence based on the Möbius decomposition, generalized in [Bouzebda \(2014\)](#). [Rosenblatt \(1975\)](#), [Rosenblatt and Wahlen \(1992\)](#) and [Ahmad and Li \(1997\)](#) presented the application of nonparametric kernel estimation method to evaluate the independence of two random vectors. This method involves comparing the L_2 norm of the difference between the joint density and the product of their marginals. The work conducted by [Han, Chen and Liu \(2017\)](#) delves into two families of distribution-free test statistics, notably featuring Kendall’s tau (a rank-type U -statistic) and Spearman’s rho (a simple linear rank statistic), both serving as significant examples. [Leung and Drton \(2018\)](#) addressed the problem of testing the independence of m continuous variables when m is greater than the sample size n . The authors considered three types of test statistics constructed as the sums or sums of squares of pairwise rank correlations. [Drton, Han and Shi \(2020\)](#) investigated the testing of the mutual independence of a p -dimensional random vector, when $p = p_n$ increases to infinity. The authors derive maximum-type tests based on pairwise rank correlation measures. The rank correlations are presented in an elegant U -statistic framework and then aggregated to create maximum-type test statistics with a null distribution that converges to a nonstandard Gumbel distribution. Recent studies have focused on testing random vectors’ pairwise independence. The concept of distance covariance was introduced by [Székely, Rizzo and Bakirov \(2007\)](#) to quantify the dependence between two random vectors, and [Székely and Rizzo \(2009\)](#) investigated it further. This distance covariance is defined as a weighted L_2 -norm between the joint distribution’s characteristic function and the product of the marginal distribution’s characteristic functions. To test the independence of two random vectors, [Gretton et al. \(2005\)](#) proposed a kernel-based independence criterion called the Hilbert-Schmidt independence criterion (HSIC). Due to the “curse of dimensionality,” these tests may fail to detect the non-linear dependence when the random vectors are of high dimension. To address this issue, [Qiu, Xu and Zhu \(2023\)](#) proposed a general framework for testing the dependence of two random vectors by randomly selecting two subspaces composed of vector components. [Zhang, Gao and Ng \(2023\)](#) proposed a new class of independence measures based on the maximum mean discrepancy in Reproducing Kernel Hilbert Space. In the literature, additional methods for testing the independence of two multidimensional random variables have emerged, including those based on the L_1 -norm between the distribution of the vector and the product of the distributions of its components ([Gretton and Györfi \(2010\)](#)), on ranks of distances ([Heller, Heller and Gorfine, 2013](#)), on nearest neighbor methods ([Berrett and Samworth, 2019](#)) and on applying distance covariance to center-outward ranks and signs ([Shi, Drton and Han, 2022a](#)). In [Jin and Matteson \(2018\)](#), three distinct measures of mutual dependency are proposed. One of the approaches extends the concept of distance covariance from pairwise dependency to mutual dependence, and the remaining two measures are derived by summing the squared distance covariances. Finally, [Auddy, Deb and Nandy \(2024\)](#) considered the problem of testing for independence using the correlation from [Chatterjee \(2021\)](#).

A common idea to measure the difference between two functions is to consider the L_q -norm, $q \geq 1$. [Csörgő and Horváth \(1988\)](#) was the first to prove a central limit for the L_q -norm, $q \geq 1$, between a density function and its estimator, and [Horváth \(1991\)](#) introduced a Poissonization technique into the study of central limit theorems for the L_q -norm. It is noteworthy that the key Poissonization techniques date back to [Le Cam \(1970\)](#) and [Araujo and Giné \(1980\)](#). [Beirlant and Mason \(1995\)](#) developed a general method for deriving the asymptotic normality of the L_q -norm of empirical functionals without assuming restrictive regularity conditions. [Giné, Mason and Zaitsev \(2003\)](#) established the weak convergence of an L_1 -norm density estimator process indexed by kernel functions.

This work presents a different approach to the problem of testing the mutual independence of random vectors in arbitrary dimensions based on the L_1 -norm between the joint density and the product of p the marginal densities, for $p \geq 2$. The motivation for L_1 -norm is connected with the fact that $\int_{\mathbb{R}^d} |f(\tilde{\mathbf{x}}) -$

$g(\tilde{\mathbf{x}})|d\tilde{\mathbf{x}}$ is the total variation distance between the measures with densities $f(\cdot)$ and $g(\cdot)$. As a result, this criterion does not share any good qualities associated with the L_2 (see Remark 3.6 below) or L_∞ norms, both of which are mathematically simpler to work with; for instance, see Devroye (1987) and Scott (1992) for more details. Cox and Czanner (2016) pointed out that the choice of L_1 -norm for the comparison of curves poses several mathematical challenges, which are caused by the fact that the mapping $f \rightarrow \int_{\mathbb{R}^d} |f(\tilde{\mathbf{x}})|d\tilde{\mathbf{x}}$ is in general not (Hadamard-) differentiable. Since the Lebesgue density functions, by definition, sit in $L_1(\mathbb{R}^d, \mathcal{B}, \lambda)$, where λ denotes the Lebesgue measure and $\mathcal{B}$ is the class of all Borel sets of $\mathbb{R}^d$, Devroye and Györfi (1985) has repeatedly stated that the L_1 -norm is the most appropriate metric for characterizing the error in estimation between a density function $f(\cdot)$ and its estimator.

To our best knowledge, this is the first time that the general context L_1 -norm for testing the independence appeared in the literature and gives the main motivation of the present work by responding to the open problems mentioned in Gretton and Györfi (2010). The main contribution of this paper is to establish the asymptotic distribution of the proposed test statistic under the null hypothesis and under the local alternatives that converge to the null at the rate of $n^{-1/2}h_n^{-d/4}$. As an important feature, the L_1 -based tests are all model-free. Furthermore, no regularity requirements for the densities are necessary to demonstrate the asymptotic normality of our statistic, a desired attribute. We conduct simulations to determine the size and power of the test. We illustrate that the proposed test has superior power characteristics compared to existing tests for various analyzed situations. The proposed test encompasses all dependency types, including complicated dependence structures, particularly sinusoidal dependence. As it will be seen later, the problem requires much more than ‘simply’ combining ideas from existing papers. Delicate mathematical derivations will be required to cope with L_1 -based tests. We highlight that the use of the Poissonization techniques due Giné, Mason and Zaitsev (2003) is substantially more difficult in our context presenting the technical merits of our contribution, see Remark 3.4 below. More precisely, several problems arise when the L_1 -norm is used in nonparametric tests. Even when the sample consists of independent random vectors, the L_1 -norm introduces a dependence structure. This is the fundamental challenge. The solution to these problems is broken down into three distinct steps: Truncation, Poissonization, and dePoissonization devices are appropriate tools to reach the results. Particularly, Poissonization involves randomizing the sample size using a Poisson random variable, allowing the application of techniques requiring the independence of increments.

The layout of the article is as follows. Section 2 gives the notation and the definitions we need. Section 3 provides the asymptotic behavior of the L_1 -based test under the null hypothesis and gives the consistency against the alternative at the nominal level α . Section 4 provides the limiting law of the L_1 -based test under the contiguous hypothesis. We discuss a bandwidth choice for practical use in Section 5. Section 6 summarizes comprehensive simulation tests to evaluate the finite-sample properties of our approach across various dependency scenarios and compare it to several state-of-the-art approaches. A summary of the findings highlighting remaining open issues are given in Section 7. All proofs are deferred to Section A. Due to the lengthiness of the proofs, we restrict our analysis to the most important arguments. Finally, technical proofs can be found in the supplementary material Berrahou, Bouzebda and Douge (2024).

2. Notation and setting

We want to test the mutual independence of p continuous random vectors $\mathbb{X}^1, \dots, \mathbb{X}^p$ of dimensions $d_1, \dots, d_p$ respectively, based on the L_1 -distance of the discrepancy between the product of their densities and the joint density. Let $d = d_1 + \dots + d_p$ be the dimension of the random vector $(\mathbb{X}^{1\top}, \dots, \mathbb{X}^{p\top})^\top$,

$p \geq 1$, where $\mathbf{x}^\top$ denotes the transpose of $\mathbf{x}$. Accordingly, $\tilde{\mathbf{x}} = (\mathbf{x}_1^\top, \dots, \mathbf{x}_p^\top)^\top \in \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_p}$ is a vector of length d . Let $f(\cdot)$ denote the density function of $\tilde{\mathbf{X}} = (X_1, \dots, X_d)^\top = (\mathbb{X}^{1\top}, \dots, \mathbb{X}^{p\top})^\top$ with respect to the Lebesgue measure λ on $\mathbb{R}^d$. For each $l = 1, \dots, p$, let $f_l(\cdot)$ be the density function of $\mathbb{X}^l$ with respect to the Lebesgue measure λ_l on $\mathbb{R}^{d_l}$. Let us formulate the null hypothesis

$$\mathcal{H}_0 : \int_{\mathbb{R}^d} \left| f(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_l(\mathbf{x}_l) \right| d\tilde{\mathbf{x}} = 0, \text{ against the alternative } \mathcal{H}_1 : \int_{\mathbb{R}^d} \left| f(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_l(\mathbf{x}_l) \right| d\tilde{\mathbf{x}} > 0.$$

We next consider sequences $\{\tilde{\mathbf{X}}_i : i \geq 1\}$ and $\{\mathbb{X}_i^l : i \geq 1\}$, $l = 1, \dots, p$, of independent and identically distributed [iid] random copies of the random vector [rv] $\tilde{\mathbf{X}}$ and $\mathbb{X}^l$, $l = 1, \dots, p$, respectively. We now introduce the kernel functions $K_l(\cdot)$ defined on $\mathbb{R}^{d_l}$, $l = 1, \dots, p$; fulfilling the condition $(\mathbf{A}_1)$ below. For each $n \geq 1$, and for each choice of the bandwidth $h_n > 0$, we define the classical Akaike–Parzen–Rosenblatt [refer to [Akaike \(1954\)](#), [Parzen \(1962\)](#), [Rosenblatt \(1956\)](#)] kernel estimators of the density functions $f(\cdot)$ and $f_l(\cdot)$, $l = 1, \dots, p$, respectively, by

$$f_n(\tilde{\mathbf{x}}) = f_{n,h_n}(\tilde{\mathbf{x}}) = \frac{1}{nh_n^d} \sum_{i=1}^n \prod_{l=1}^p K_l\left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n}\right), \quad f_{n,l}(\mathbf{x}_l) = f_{n,h_n,l}(\mathbf{x}_l) = \frac{1}{nh_n^{d_l}} \sum_{i=1}^n K_l\left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n}\right).$$

[Parzen \(1962\)](#) has shown, under some assumptions on the kernel function, that $f_n(\cdot)$ is an asymptotically unbiased and consistent estimator for $f(\cdot)$ whenever $h_n \rightarrow 0$, $nh_n^d \rightarrow \infty$ and $\tilde{\mathbf{x}}$ is a continuity point of $f(\cdot)$. Under some additional assumptions on $f(\cdot)$ and h_n , he obtained an asymptotic normality result, too. The above Akaike–Parzen–Rosenblatt kernel estimators have been extensively studied in the literature, see, e.g., [Silverman \(1986\)](#), [Nadaraya \(1989\)](#), [Härdle \(1990\)](#), [Scott \(1992\)](#), [Wand and Jones \(1995\)](#), [Eggermont and LaRiccia \(2001\)](#), [Devroye and Lugosi \(2001\)](#), [Bouzebda and Nemouchi \(2023\)](#), [Bouzebda and Taachouche \(2023\)](#) and the references therein. To perform the test of the hypothesis $\mathcal{H}_0$, we consider the statistic

$$V_n := \int_{\mathbb{R}^d} \left| f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_{n,l}(\mathbf{x}_l) \right| d\tilde{\mathbf{x}}. \quad (2.1)$$

Remark 2.1. *Careful bookkeeping throughout the proofs in the paper could give an adapted set of conditions on possibly different bandwidths. We do not pursue this here to avoid an unnecessary increase in the technicality of the presentation. If one wants to use the vector bandwidths (see, in particular, Chapter 12 of [Devroye and Lugosi \(2001\)](#)), with obvious changes in notation, our results and their proofs remain true when h_n is replaced by a vector bandwidth $\mathbf{h}_n = (h_n^{(1)}, \dots, h_n^{(d)})$, where $\min_{1 \leq i \leq d} h_n^{(i)} > 0$. In this situation we set $h_n^d = \prod_{i=1}^d h_n^{(i)}$, and for any vector $\mathbf{v} = (v_1, \dots, v_d)$ we replace $\mathbf{v}/h_n$ by $(v_1/h_n^{(1)}, \dots, v_d/h_n^{(d)})$.*

3. Main results

Before presenting our results, we introduce some notation and the underlying assumptions. For $s \geq 1$, let $\mathcal{B}_s = \{\mathbf{u} \in \mathbb{R}^s : \|\mathbf{u}\| \leq 1\}$, where $\|\cdot\|$ is some norm on $\mathbb{R}^s$. Consider the following condition upon the kernels $K_l(\cdot)$, $l = 1, \dots, p$, that will be used below.

- $(\mathbf{A}_1)$ For each $l = 1, \dots, p$, $K_l(\cdot)$ is a bounded Lebesgue density function having support contained in the closed ball of radius $1/2$ centered at zero of $\mathbb{R}^{d_l}$.

Let Z_1 and Z_2 be independent standard normal random variables. In the sequel, the L_2 -norm of a kernel $K(\cdot)$ is denoted by $\|K\|_2$. For each $l = 1, \dots, p$, define the function

$$\rho_l(\mathbf{t}_l) = \frac{\int_{\mathbb{R}^{d_l}} K_l(\mathbf{u}_l) K_l(\mathbf{u}_l + \mathbf{t}_l) d\mathbf{u}_l}{\|K_l\|_2^2}, \quad \mathbf{t}_l \in \mathbb{R}^{d_l}.$$

Denote $\tilde{\mathbf{K}} = \prod_{l=1}^p \|K_l\|_2^2$ and, for each $\tilde{\mathbf{t}} = (\mathbf{t}_1^\top, \dots, \mathbf{t}_p^\top)^\top$, $\rho(\tilde{\mathbf{t}}) = \prod_{l=1}^p \rho_l(\mathbf{t}_l)$. Further, for $\tilde{\mathcal{B}} = \mathcal{B}_{d_1} \times \dots \times \mathcal{B}_{d_p}$, set

$$\sigma^2 := \tilde{\mathbf{K}} \int_{\tilde{\mathcal{B}}} \text{Cov} \left(\left| \sqrt{1 - \rho^2(\tilde{\mathbf{t}})} Z_1 + \rho(\tilde{\mathbf{t}}) Z_2 \right|, |Z_2| \right) d\tilde{\mathbf{t}}.$$

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 . Let $\stackrel{\mathcal{D}}{\rightarrow}$ denote the convergence in distribution. The main result to be proved here may now be stated precisely as follows.

Theorem 3.1. *Assume that $(\mathbf{A}_1)$ holds and let $h_n > 0$ be such that, as $n \rightarrow \infty$, $h_n \rightarrow 0$ and $nh_n^{3d} \rightarrow \infty$. Then, under $\mathcal{H}_0$, as $n \rightarrow \infty$, we have $\sqrt{n}(V_n - \mathbb{E}V_n) \stackrel{\mathcal{D}}{\rightarrow} \mathcal{N}(0, \sigma^2)$.*

Remark 3.1. *Note that the choice $h_n = \lambda n^{-\vartheta}$, $\lambda > 0$, $0 < \vartheta < 1/(3d)$, fulfills the theorem's conditions, see Section 5 for more discussions. The variance σ^2 has an interesting alternate representation. Making use of the formulas for the absolute moments of a bivariate normal random variable of Nabeya (1951), we have*

$$\sigma^2 = \tilde{\mathbf{K}} \int_{\tilde{\mathcal{B}}} \varphi(\rho(\tilde{\mathbf{t}})) d\tilde{\mathbf{t}}, \quad \text{where } \varphi(\rho) = \frac{2}{\pi} \left(\rho \arcsin \rho + \sqrt{1 - \rho^2} - 1 \right), \text{ for } \rho \in [-1, 1].$$

Remark 3.2. *It is noteworthy that Theorem 3.1 imposes any assumptions on the Lebesgue density. The bandwidth condition $nh_n^{3d} \rightarrow \infty$ is mainly due to the evaluation of the variance in Lemma B.6. It allows us also to provide some asymptotic approximations in Lemma B.7 and to approximate the centering term in Lemma B.12. This condition is considered in nonparametric regression testing proposed by Lee, Song and Whang (2013) to establish the asymptotic normality of the one-sided L_q -test statistic in testing functional inequalities, $1 \leq q < \infty$.*

Let us discuss how the process works so that our test can be carried out. Let, for each $l = 1, \dots, p$, C_l be a bounded Borel set such that $\mathbb{P}(\mathbb{X}^l \in C_l) > 1 - \varepsilon$ for some ε small enough and set, for a Borel set $A \in \mathbb{R}^d$,

$$V_n(A) := \int_A \left| f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_{n,l}(\mathbf{x}_l) \right| d\tilde{\mathbf{x}}, \quad \text{and} \quad V(A) := \int_A \left| f(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_l(\mathbf{x}_l) \right| d\tilde{\mathbf{x}}.$$

We denote $\tilde{C} = C_1 \times \dots \times C_p$. Making use of the arguments developed in Lemma B.2 in Berrahou, Bouzebda and Douge (2024), and Tchebychev's inequality, it follows that

$$\sqrt{n} [V_n - V_n(\tilde{C}) - \mathbb{E}(V_n - V_n(\tilde{C}))] \xrightarrow{\mathcal{P}} 0,$$

where $\xrightarrow{\mathcal{P}}$ stands for the convergence in probability. Therefore, an application of Theorem 3.1 gives

$$\sqrt{n}(V_n(\tilde{C}) - \mathbb{E}V_n(\tilde{C})) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2) \text{ as } n \rightarrow \infty.$$

The centering term $\sqrt{n}\mathbb{E}V_n(\tilde{C})$ depends on an unknown distribution. To circumvent this problem, we estimate $\sqrt{n}\mathbb{E}V_n(\tilde{C})$ by

$$a_n(\tilde{C}) := \mathbb{E}|Z_1| \int_{\tilde{C}} \sqrt{\hat{\mathcal{L}}_n(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}}, \quad (3.1)$$

where

$$\hat{\mathcal{L}}_n(\tilde{\mathbf{x}}) := \prod_{l=1}^p \hat{v}_{n,l}(\mathbf{x}_l) - \sum_{l=1}^p \hat{v}_{n,l}(\mathbf{x}_l) \prod_{j \neq l} g_{n,j}(\mathbf{x}_j) + (p-1) \prod_{l=1}^p g_{n,l}(\mathbf{x}_l),$$

where, for each $l = 1, \dots, p$,

$$\hat{v}_{n,l}(\mathbf{x}_l) := \frac{1}{nh_n^{2d_l}} \sum_{i=1}^n K_l^2\left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n}\right), \text{ and } g_{n,l}(\mathbf{x}_l) := \frac{1}{n(n-1)h_n^{2d_l}} \sum_{i=1}^n \sum_{j \neq i} K_l\left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n}\right) K_l\left(\frac{\mathbf{x}_l - \mathbb{X}_j^l}{h_n}\right).$$

Note that $\mathbb{E}|Z_1| = \sqrt{2/\pi} \approx 0.7978$. Our suggested test statistic is

$$T_n(\tilde{C}) := \frac{\sqrt{n}V_n(\tilde{C}) - a_n(\tilde{C})}{\sigma}.$$

Let $z_{1-\alpha}$ be the upper α quantile of the standard normal distribution. In this paper, we propose the following test: Reject $\mathcal{H}_0$ if and only if $T_n(\tilde{C}) > z_{1-\alpha}$. Now, for each $l = 1, \dots, p$, for each $\delta > 0$, let C_l^δ be the δ -neighborhood of C_l , i.e., $C_l^\delta = \{\mathbf{x}_l + \mathbf{u}_l : \mathbf{x}_l \in C_l, \|\mathbf{u}_l\| \leq \delta\}$. We need the following additional notation. Set, for each $a \in \{0, 1\}$,

$$\Gamma_{n,a}(\tilde{\mathbf{x}}) := f_n(\tilde{\mathbf{x}}) - a \sum_{l=1}^p f_{n,l}(\mathbf{x}_l) \prod_{j \neq l} \mathbb{E}f_{n,j}(\mathbf{x}_j) + (pa-1) \prod_{l=1}^p \mathbb{E}f_{n,l}(\mathbf{x}_l)$$

and

$$\Gamma_{\eta,a}(\tilde{\mathbf{x}}) := f_\eta(\tilde{\mathbf{x}}) - a \sum_{l=1}^p f_{\eta,l}(\mathbf{x}_l) \prod_{j \neq l} \mathbb{E}f_{\eta,j}(\mathbf{x}_j) + (pa-1) \prod_{l=1}^p \mathbb{E}f_{\eta,l}(\mathbf{x}_l),$$

where

$$f_\eta(\tilde{\mathbf{x}}) = \frac{1}{nh_n^d} \sum_{i=1}^\eta \prod_{l=1}^p K_l\left(\frac{\mathbf{x}_l - \mathbf{X}_i^l}{h_n}\right) \text{ and } f_{\eta,l}(\mathbf{x}_l) = \frac{1}{nh_n^{d_l}} \sum_{i=1}^\eta K_l\left(\frac{\mathbf{x}_l - \mathbf{X}_i^l}{h_n}\right),$$

and η is a Poisson random variable with mean n , independent of $\tilde{\mathbf{X}}, \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \dots$. Set, for a Borel set $A \in \mathbb{R}^d$ and for each $a \in \{0, 1\}$,

$$U_{n,a}(A) := \int_A |\Gamma_{n,a}(\tilde{\mathbf{x}})| d\tilde{\mathbf{x}} \text{ and } U_{\eta,a}(A) := \int_A |\Gamma_{\eta,a}(\tilde{\mathbf{x}})| d\tilde{\mathbf{x}}.$$

The following theorem shows that the test has an asymptotically valid size.

Theorem 3.2. Assume that $(\mathbf{A}_1)$ holds and, for each $l = 1, \dots, p$, there exists $\delta > 0$ such that $f_l(\cdot)$ is bounded on C_l^δ . Let $h_n > 0$ be such that, as $n \rightarrow \infty$, $h_n \rightarrow 0$ and $nh_n^{3d} \rightarrow \infty$. Then, under $\mathcal{H}_0$, we have $\lim_{n \rightarrow \infty} \mathbb{P}\{T_n(\tilde{C}) > z_{1-\alpha}\} = \alpha$.

The last result of this section shows the consistency of the test against the alternative at nominal level α .

Theorem 3.3. Assume that for each $l = 1, \dots, p$, $K_l(\cdot)$ is a density function. Let $h_n > 0$ be such that, as $n \rightarrow \infty$, $h_n \rightarrow 0$ and $nh_n^{3d} \rightarrow \infty$. If $\mathcal{H}_1$ is true, then, for any bounded subset $\tilde{D} = D_1 \times \dots \times D_p \subset \mathbb{R}^d$ such that $V(\tilde{D}) > 0$, we have $\lim_{n \rightarrow \infty} \mathbb{P}\{T_n(\tilde{D}) > z_{1-\alpha}\} = 1$.

Remark 3.3. Our results can be generalized by considering the following weighted family

$$V_n^{(1)} := \int_{\mathbb{R}^d} \left| f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_{n,l}(\mathbf{x}_l) \right| w(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}},$$

where $w : \mathbb{R}^d \rightarrow [0, \infty)$ is a nonnegative weight function, or, for $1 \leq q < \infty$,

$$V_n^{(q)} := \int_{\mathbb{R}^d} \left| f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_{n,l}(\mathbf{x}_l) \right|^q w(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}. \quad (3.2)$$

Let us recall a popular family of ϕ -divergences

$$D_\phi(\mathbb{P}, \mathbb{Q}) := \int_{\mathbb{R}^d} \phi\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{Q} \text{ if } \mathbb{P} \ll \mathbb{Q},$$

where $\phi : [0, \infty) \rightarrow (-\infty, \infty]$ is a convex function. $\mathbb{P} \ll \mathbb{Q}$ denotes that $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$. Well-known distance/divergence measures obtained by appropriately choosing ϕ include total variation distance ($\phi(t) = |t - 1|$). Let us define $\mathcal{P}_\lambda$ as the set of all probability measures, $\mathbb{P}$, that are absolutely continuous with respect to some σ -finite measure, λ on $\mathbb{R}^d$. For $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_\lambda$, let $\prod_{l=1}^p f_l = \frac{d\mathbb{P}}{d\lambda}$ and $f = \frac{d\mathbb{Q}}{d\lambda}$ be the Radon-Nikodym derivatives of $\mathbb{P}$ and $\mathbb{Q}$ with respect to λ , respectively. In the interesting paper [Sriperumbudur et al. \(2012\)](#), it is mentioned that ϕ -divergences and integral probability metrics are fundamentally different and intersect only at the total variation distance. In the last situation, we have

$$D^{TV}(\mathbb{P}, \mathbb{Q}) := \int_{\mathbb{R}^d} \left| \frac{d\mathbb{P}}{d\mathbb{Q}} - 1 \right| d\mathbb{Q} = \int_{\mathbb{R}^d} \left| \frac{\prod_{l=1}^p f_l}{f} - 1 \right| d\mathbb{Q}.$$

Let $\mathbb{Q}_n$ denote the empirical measure based on $\{\tilde{\mathbf{X}}_i : i \geq 1\}$. One can estimate $D_\phi(\mathbb{P}, \mathbb{Q})$ by

$$D_n^{TV}(\mathbb{P}, \mathbb{Q}) := \int_{\mathbb{R}^d} \left| \frac{\prod_{l=1}^p f_{n,l,i}}{f_{n,i}} - 1 \right| d\mathbb{Q}_n = \frac{1}{n} \sum_{i=1}^n \left| \frac{\prod_{l=1}^p f_{n,l,i}(\mathbf{x}_i^l)}{f_{n,i}(\mathbf{x}_i^1, \dots, \mathbf{x}_i^p)} - 1 \right|,$$

where

$$f_{n,i}(\tilde{\mathbf{x}}) = \frac{1}{(n-1)h_n^d} \sum_{j \neq i} \prod_{l=1}^p K_l\left(\frac{\mathbf{x}_l - \mathbf{x}_j^l}{h_n}\right), \quad f_{n,l,i}(\mathbf{x}_l) = \frac{1}{(n-1)h_n^{d_l}} \sum_{j \neq i} K_l\left(\frac{\mathbf{x}_l - \mathbf{x}_j^l}{h_n}\right).$$

[Moon \(2016\)](#), Chapter 4, investigated the statistic

$$D_{n,\phi}(\mathbb{P}, \mathbb{Q}) := \int_{\mathbb{R}^d} \phi \left(\frac{\prod_{l=1}^p f_{n,l,i}}{f_{n,i}} \right) d\mathbb{Q}_n.$$

If the densities are bounded and continuously differentiable up to order s , and $\phi(\cdot)$ is continuously differentiable with bounded derivatives. Under some additional assumptions on the kernels, ([Moon, 2016](#), Chapter 4) provided the bias of $D_{n,\phi}(\mathbb{P}, \mathbb{Q})$ is of the order $O(h_n^s + 1/(nh_n^d))$ and the variance of order $O(1/n)$. Making use of the Efron-Stein inequality, he also proved the asymptotic normality for $D_{n,\phi}(\mathbb{P}, \mathbb{Q})$. The investigation of the case $\phi(t) = |t - 1|$, can not treated by the same techniques and needs more intricate developments.

Remark 3.4. Regarding the difference between this work and of [Giné, Mason and Zaitsev \(2003\)](#), the first to be resolved is that our statistic is not an integral of a sum of centered independent variables. The initial task was to establish the asymptotic equivalence, in probability, of our truncated statistic $\sqrt{n}(V_n(\bar{C}) - \mathbb{E}V_n(\bar{C}))$ and $\sqrt{n}(U_{n,0}(\bar{C}) - \mathbb{E}U_{n,0}(\bar{C}))$, where $\bar{C} = C_1 \times \dots \times C_p$ and, for each $l = 1, \dots, p$, C_l is a bounded Borel set in $\mathbb{R}^{d_l}$ satisfying (B.3) and (B.4) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K} = \mathcal{K}_l := \{K_l, K_l^2, K_l^3\}$. This equivalence has been established through Lemmas B.3 to B.7, in [Berrahou, Bouzebda and Douge \(2024\)](#), so that certain techniques from [Giné, Mason and Zaitsev \(2003\)](#) are extended, in a nontrivial way, to the multivariate case to obtain the asymptotic normality of $\sqrt{n}(U_{n,0}(\bar{C}) - \mathbb{E}U_{n,0}(\bar{C}))$. Finally, we mention that the delicate approximation of $\mathbb{E}V_n(\bar{C})$ by $a_n(\bar{C})$, defined in (3.1), is not discussed in [Giné, Mason and Zaitsev \(2003\)](#).

Remark 3.5. Recall, in the classical kernel estimation, that the standardizing factor is $(nh_n^d)^{1/2}$ with $h_n \rightarrow 0$, indicating a lower rate of convergence. This is the cost incurred when drawing conclusions regarding the local quantities. The use of the L_1 -norm avoids such a lower rate of convergence.

Remark 3.6. In this remark, we will point out some differences between the L_1 and L_2 norms. We first recall the following example from Section 6.5 in [Devroye and Lugosi \(2001\)](#). Let $f(\cdot)$ be the uniform density on $[0, 1]$. Let ϵ be a small positive number, and assume that both density estimates ignore the data: $f_\epsilon(\cdot)$ denotes our catastrophic candidate, that is $1 - \epsilon$ on $[0, 1]$ and ϵ^3 on $[1, 1 + 1/\epsilon^2]$. Now, $g_\epsilon = 1$ on $[-\epsilon^2, 1 - \epsilon^2]$. Verify that $\int |g_\epsilon - f| dx = 2\epsilon^2 < 2\epsilon = \int |f_\epsilon - f| dx$, so that the choices are not even close. However, $\int (f_\epsilon - f)^2 dx = \epsilon^2 + \epsilon^4 < 2\epsilon^2 = \int (g_\epsilon - f)^2 dx$, so that minimizing the L_2 -norm, even with $f(\cdot)$ given, picks the wrong density from the set $\{f_\epsilon, g_\epsilon\}$. Hence, it can be concluded that criteria relying on the L_2 -norm are unsuitable for universal properties in density estimation. In Section 2.3.2.1 of [Scott \(1992\)](#), for pointwise estimation of $f(\cdot)$ by $\hat{f}(\cdot)$, it has been noted that one appeal of the L_1 -norm $\int |\hat{f} - f| dx$ is that it pays more attention to the tails of a density than the L_2 -norm, which de-emphasizes the relatively small density values thereby squaring. By the fact that the density function has inverse length as its unit, L_1 is a dimensionless quantity after integration. The L_2 -norm, on the other hand, retains the units of inverse length after integration of the squared density error. We highlight the fact that $0 \leq \int |f - g| dx \leq 2$ while $0 \leq \int (f - g)^2 dx \leq \infty$, where $f(\cdot)$ and $g(\cdot)$ are density functions; see Problem 7 in [Scott \(1992\)](#). Some asymptotic results for L_1 estimates by [Hall and Wand \(1988\)](#) and [Scott and Wand \(1991\)](#) support the notion that the practical differences between L_1 and L_2 criteria are reasonably small except in extreme situations. A classical approach for assessing independence involves the L_2 -norm between the joint density function with the product of the marginal density functions; for instance, see [Rosenblatt \(1975\)](#), [Rosenblatt and Wahlen \(1992\)](#) and [Ahmad and Li \(1997\)](#), that is defined in (3.2) for $p = 2$ and $q = 2$. It is important to mention that Rosenblatt's test

is not distribution-free. According to Rosenblatt, it has been observed that L_2 -tests relying on density estimates generally exhibit lower statistical power compared to tests based on sample distribution functions, [Hoeffding \(1948\)](#) and [Blum, Kiefer and Rosenblatt \(1961\)](#). This finding raises important concerns for many applications, refer to [Feuerverger \(1993\)](#). The [Blum, Kiefer and Rosenblatt \(1961\)](#) independence test is consistent when the appropriate assumptions are satisfied. However, two challenges arise when utilizing this statistic in a test. Estimating quantiles of the null distribution poses challenges. Furthermore, it is crucial to note that the accuracy of the empirical distribution function estimations deteriorates significantly as the dimensionality of the spaces $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$ grows. This limitation restricts the effectiveness of the statistic in a multivariate context; for instance, see [Gretton and Györfi \(2010\)](#) for more details and discussions. In [Remark 4.2](#), we provide some additional comments on important advantages of the L_1 -norm. It is seen that the L_1 test demonstrates superior performance compared to the L_2 test in [Section 6](#).

Remark 3.7. To the best of our knowledge, there exists a limited number of nonparametric methods for assessing independence that exhibit a nearly linear computational complexity. For further details, we refer to [Deb, Ghosal and Sen \(2020\)](#), [Auddy, Deb and Nandy \(2024\)](#), [Gorsky and Ma \(2022\)](#) and [Berrett \(2022\)](#). The [Gorsky and Ma](#)'s approach possesses several benefits, with the most prominent being its multivariate nature. Additionally, the computational cost of this approach is generally characterized by a time complexity of $O(n \log n)$ in relation to the sample size n . Moreover, it is worth noting that the test threshold remains exact at any sample size, rather than being limited to an asymptotic approximation. In most situations, the existing nonparametric independence tests typically require computation that scales at least quadratically with the sample size; for instance, we may refer to [Pfister et al. \(2018\)](#). In the last reference, it has been observed that the computational complexity of their approach can be reduced by employing linear time approximation techniques (large-scale approximations), as elucidated by the work of [Zhang et al. \(2018\)](#) in the context of pairwise HSIC. [Berrett, Kontoyiannis and Samworth \(2021\)](#) proposed statistical test based on a U -statistic in general setting where the computational complexity is $O(n^2)$. In [Deb and Sen \(2023\)](#), the time taken to compute optimal transport-based multivariate ranks is $O(n^3)$ in the worst case. The computational cost with that of permutation k -nearest-neighbour methods [Berrett and Samworth \(2019\)](#), which run in $O(Bkn \log n)$ time, where in permutation tests any test statistic must be calculated B times. In [Jin and Matteson \(2018\)](#), the proposed statistics have the time complexity $O(n^2)$, by considering the dimension is a constant. In our setting, the computational complexity is $O(m_n dn)$, where m_n is the number of points used in the integral approximation, that may be reduced using the idea in [Devroye, Györfi and Lugosi \(2002\)](#). Identifying certain transformations that can effectively mitigate the computational complexity associated with our test would be interesting. However, obtaining such transformations necessitates a distinct methodology from that employed in the existing literature, and we defer this issue for further investigation.

4. Local alternatives

We determine the power of the test in [\(2.1\)](#) against some sequences of local alternatives. Consider the following sequences of local alternatives converging to the null hypothesis at the rate $n^{-1/2}h_n^{-d/4}$,

$$H_{\delta,n} : f(\tilde{\mathbf{x}}) = \prod_{l=1}^p f_l(\mathbf{x}_l) + n^{-1/2}h_n^{-d/4}\delta(\tilde{\mathbf{x}}),$$

where $\delta(\cdot)$ is an integrable real function and its integral is zero. The next result gives the asymptotic distribution under the local alternatives $H_{\delta,n}$.

Theorem 4.1. Suppose that (A_1) holds and that $h_n \rightarrow 0$ and $nh_n^{3d} \rightarrow \infty$, as $n \rightarrow \infty$. Then, under $H_{\delta,n}$, we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \{T_n(\tilde{C}) > z_{1-\alpha}\} = 1 - \Phi(z_{1-\alpha} - \eta(\delta)),$$

where $\Phi(\cdot)$ denotes the cdf of the standard normal distribution and

$$\eta(\delta) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\tilde{C}} \delta_1(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} \quad \text{and} \quad \delta_1(\tilde{\mathbf{x}}) = (\delta(\tilde{\mathbf{x}}))^2 \left\{ \tilde{\mathbf{K}} \prod_{l=1}^p f_l(\mathbf{x}_l) \right\}^{-1/2}.$$

Remark 4.1. Following [Auddy, Deb and Nandy \(2024\)](#), let us consider the joint density of $\tilde{\mathbf{X}}$ defined by

$$f_r(\tilde{\mathbf{x}}) = (1-r) \prod_{l=1}^p f_l(\mathbf{x}_l) + r g(\tilde{\mathbf{x}}), \quad (4.1)$$

where $g(\cdot)$ is a density function with marginals $f_l(\cdot)$, $l = 1, \dots, p$, and $r \in [0, 1]$. For a sequence $\{r_n\}$ with $r_n \in [0, 1]$ for all $n \geq 1$, consider the family of joint density $f_{r_n}(\cdot)$ and the following testing problem : $H_0 : r_n = 0$ versus $H_{1,n} : r_n > 0$. Following the same steps in the proof of Theorem 4.1, as $r_n \rightarrow 0$, we obtain the limit of the power function

$$\lim_{n \rightarrow \infty} \mathbb{P} \{T_n(\tilde{C}) > z_{1-\alpha}\} = \begin{cases} \alpha, & \text{if } nh_n^{d/2} r_n^2 \rightarrow 0, \\ 1 - \Phi(z_{1-\alpha} - c \eta(g)), & \text{if } nh_n^{d/2} r_n^2 \rightarrow c, \\ 1, & \text{if } nh_n^{d/2} r_n^2 \rightarrow \infty, \end{cases}$$

where

$$\eta(g) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\tilde{C}} g_1(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} \quad \text{and} \quad g_1(\tilde{\mathbf{x}}) = \left(g(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_l(\mathbf{x}_l) \right)^2 \left\{ \tilde{\mathbf{K}} \prod_{l=1}^p f_l(\mathbf{x}_l) \right\}^{-1/2}.$$

[Auddy, Deb and Nandy \(2024\)](#) obtained a similar result in the bivariate case; for instance, see Corollary 3.1. In their result, the limiting behavior of the power depends on whether $n^{1/4} r_n^2$ converges to 0, ∞ or some number in $(0, \infty)$ while, for L_1 -test and L_2 -test, studied by [Rosenblatt \(1975\)](#), the power depends on the limit of $nh_n^{d/2} r_n^2$ ($d = 2$ for L_2 -test). Since h_n is chosen to satisfy $nh_n^d \rightarrow \infty$, then $n^{-1/2} h_n^{-d/4} = o(n^{-1/4})$, and we conclude that the Chatterjee's test, studied in [Auddy, Deb and Nandy \(2024\)](#), is powerless along the local alternative (4.1) compared with the L_1 -test. [Shi, Drton and Han \(2022b\)](#) showed that Hoeffding's test, Blum-Kiefer-Rosenblatt's test and Bergsma-Dassios-Yanagimoto's test are more powerful, with the optimal rate of order $O(n^{-1/2})$, than L_1 -test, L_2 -test and Chatterjee's test for the local alternative (4.1), but presenting some difficulties in estimating quantiles of the null distribution in high dimensional setting. The main question is now to find other local alternatives for which the L_1 -test is more powerful than the others.

Remark 4.2. This paper develops a general framework for testing the independence among continuous random vectors based on L_1 -norm. Our test achieves the following desirable properties:

- Full distribution-freeness. Numerous statistical tests utilize asymptotic distribution-freeness for computationally efficient distributional approximations that result in pointwise asymptotic control over their size. We note that some existing tests of multivariate independence are not distribution-free, which has both computational and theoretical repercussions.

- *Transformation invariance.* It is well known that L_1 -distance is invariant under any smooth monotone transformation; for instance, see [Devroye and Lugosi \(2001\)](#). In multivariate statistics, this kind of invariance is highly interesting.
- *Computational efficiency.* Modern applications require evaluating a dependence measure and the accompanying test to be as computationally efficient as possible, irrespective of their statistical characteristics. Therefore, the effectiveness of the numerical experiments must come first, we may refer to [Heller, Heller and Gorfine \(2013\)](#). Furthermore, distribution-free procedures can alleviate the computational cost of statistical problems.
- *Consistency under absolute continuity.* The unique condition for the consistency of our tests is that the underlying distributions are absolutely continuous, without the need for any moment requirements. This makes possible the nonparametric inference under heavy-tailed data-generating distributions such as stable laws [Yang \(2012\)](#) and Pareto distributions [Rizzo \(2009\)](#), and it distinguishes our tests from commonly used techniques like traditional distance covariance and energy statistic, for more discussion, refer to [Deb and Sen \(2023\)](#).

5. Bandwidth selection

There are basically no restrictions on the choice of the kernels $K_l(\cdot)$, $l = 1, \dots, p$, in our setup, apart from satisfying conditions (A_1) . In [Devroye and Lugosi \(2001\)](#), it is noticed that for large sample sizes, the shape of the optimal kernel is unique. For example, for $\mathbb{R}$, classical L_2 theory [Watson and Leadbetter \(1963\)](#) shows that for L_2 errors, among all positive kernels, the [Epanečnikov \(1969\)](#) kernel, $K(x) = \max(3/4(1 - x^2), 0)$ is best possible. For $\mathbb{R}^d$, [Deheuvels \(1977\)](#) showed the L_2 optimality of $\max((1 - \|\mathbf{x}\|^2)^d, 1)$, $\mathbf{x} \in \mathbb{R}^d$. For the L_1 -norm, there is evidence that the Epanečnikov kernel is also best among all positive kernels; refer to [Devroye and Lugosi \(2001\)](#), Chapters 16 and 17. The selection of the bandwidth, however, is more problematic. Although any choice of bandwidth h_n satisfying $h_n \rightarrow 0$ and $nh_n^{2d} \rightarrow \infty$ will deliver the asymptotic distribution in Theorem 3.1, in practice we need some guidance on how to select h_n . Notice that the quadratic criteria such as the ASE, the ISE, or the MISE are applicable when estimating the density as an element of the L_2 -space, and h_n can be selected using the cross-validation or plug-in approach; for instance, see [Sain, Baggerly and Scott \(1994\)](#), [Duong and Hazelton \(2005\)](#). [Hall and Wand \(1988\)](#) demonstrated that the under-smoothing approach yields confidence intervals with greater coverage accuracy than bias-reduced density estimators. [Hall and Wand \(1988\)](#) suggested to use $h_n = c 1.05 \hat{\sigma} n^{-1/5}$, where $0 < c < 1$ and $\hat{\sigma}$ is the sample standard deviation. The difficulty of selecting a bandwidth suited to point and interval estimates was examined in [Chan, Lee and Peng \(2010\)](#). They recommended selecting the local bandwidth with the highest optimal rate and then developing bias reduction estimators for both pointwise and interval estimation. [Matteson and Tsay \(2017\)](#) evaluated spacing bandwidth with low sensitivity to estimate independent components using a nonparametric probability integral transformation and distance covariance. In nonparametric testing, [Lee, Song and Whang \(2013\)](#) proposed a one-sided L_q approach in testing nonparametric functional inequalities. To evaluate the performance of their test, they used $h_n = c \hat{\sigma} n^{-1/5}$, where $0.75 \leq c \leq 2.5$. For testing a parametric model for conditional mean function against a nonparametric alternative, [Horowitz and Spokoiny \(2001\)](#) proposed an adaptive-rate-optimal rule. [Gao and Gijbels \(2008\)](#) proposed, utilizing the Edgeworth expansion of the asymptotic distribution of the test, to select the bandwidth such that the power function of the test is maximized while the size function is controlled. [Hall and Wand \(1988\)](#) looked at the asymptotic expansion of $\mathbb{E} \left\{ \int |f_{n,h} - f| dx \right\}$ and minimized the main asymptotic terms to obtain a recipe for h as a function of n , $f(\cdot)$ and the kernel function $K(\cdot)$. They then estimate the unknown quantity involving $f(\cdot)$ from the data and propose this as a plug-in bandwidth estimate. However, all other plug-in methods we are aware of are not universal bandwidths. While the plug-in

procedure displays an analytical solution, which depends on unknown quantities that need to be estimated, the double kernel is performed empirically. Notice also that one may use the maximum likelihood cross-validation method to determine the smoothing parameter; however, this procedure performs very poorly, as indicated in [Devroye \(1997\)](#). The double kernel method uses a pair of kernels, $K(\cdot)$ and $L(\cdot)$, and picks $H = \arg \min_h \int |f_{n,h} - g_{n,h}| dx$, where $f_{n,h}(\cdot)$ and $g_{n,h}(\cdot)$ are the kernel estimates with kernels $K(\cdot)$ and $L(\cdot)$, respectively. Assume that $d = 1$. If the characteristic functions of $K(\cdot)$ and $L(\cdot)$ do not coincide on an open interval about the origin, then the choice H is consistent, refer to [Devroye \(1989\)](#). The ideal bandwidth selection for nonparametric testing differs from that for nonparametric estimation because we must balance the test's size and power rather than the estimator's bias and variance. There are no methods for calculating the appropriate bandwidth for our test, and it is difficult to formulate a theory that provides the solution. The choice of bandwidth determines the sensitivity with which specific types of dependence can be identified and, thus, affects the practical performance of the test. Idealistically, we should select a bandwidth h that provides the best power for a given sample size, but deriving this process is intricate enough to need a separate study. Thus, for the present purposes, we provide the combinatorial procedure of [Devroye and Lugosi \(2001\)](#) that will be used. To this end, for fixed $\epsilon > 0$, define the [Yatracos \(1985\)](#) class $\mathcal{A}_\epsilon = \{\{\tilde{\mathbf{x}} : f(\tilde{\mathbf{x}}) > g(\tilde{\mathbf{x}})\} : f, g \in \mathcal{G}_\epsilon\}$, where for every $\epsilon > 0$, there exists a finite number N_ϵ of densities in $\mathcal{F}$, a prespecified class of densities, such that the L_1 balls of radius ϵ centered at these densities cover $\mathcal{F}$, that is, if these chosen densities are $\mathcal{G}_\epsilon = \{g_1, \dots, g_{N_\epsilon}\}$, then $\mathcal{F} \subseteq \cup_{i=1}^{N_\epsilon} B_{g_i, \epsilon}$. Let $m < n$, and define $\mathcal{A}_\Theta$ as the Yatracos class of subsets of $\mathbb{R}^d$ (corresponding to the family of density estimates $f_{n,\theta}, \theta \in \Theta \subset \mathbb{R}$) as the class of all sets of the form $A_{\theta_1, \theta_2} = \{\tilde{\mathbf{x}} : f_{n-m, \theta_1}(\tilde{\mathbf{x}}) > f_{n-m, \theta_2}(\tilde{\mathbf{x}})\}, \theta_1, \theta_2 \in \Theta$. We select a parameter θ_n from Θ by minimizing the distance $\Delta_\theta = \sup_{A \in \mathcal{A}_\Theta} \left| \int_A f_{n-m, \theta} dx - \mu_m(A) \right|$, over all $\theta \in \Theta$, where $\mu_m(\cdot)$ denotes the empirical measure defined by the subsample $\tilde{\mathbf{X}}_{n-m+1}, \dots, \tilde{\mathbf{X}}_n$. If the minimum does not exist, we select θ_n in such a way that $\Delta_{\theta_n} < \inf_{\theta^* \in \Theta} \Delta_{\theta^*} + 1/n$. Define $f_n(\cdot) = f_{n-m, \theta_n}(\cdot)$. If $\int f_{n-m, \theta} = 1$ for all $\theta \in \Theta$, then for the minimum distance estimate $f_n(\cdot)$ as defined above, from Theorem 10.1 of [Devroye and Lugosi \(2001\)](#), we have

$$\int |f_n - f| dx \leq 3 \inf_{\theta \in \Theta} \int |f_{n-m, \theta} - f| dx + 4\Delta + \frac{3}{n}, \text{ where } \Delta = \sup_{A \in \mathcal{A}_\Theta} \left| \int_A f dx - \mu_m(A) \right|. \quad (5.1)$$

Note that if the estimates do not satisfy the condition $\int f_{n-m, \theta} dx = 1$, (5.1) remains valid, but with the factors of “3” replaced by “5”. [Felber, Kohler and Krzyzak \(2015\)](#) proposed an adaptive method to estimate density based on data with small measurement errors inspired by the combinatorial method described before. A simulation study in [Felber, Kohler and Krzyzak \(2015\)](#) shows that the proposed estimator converges faster, in terms of the L_1 -norm than a standard kernel density estimator (Akaike–Parzen–Rosenblatt) to the unknown density.

6. Monte Carlo experiments

In this section, we examine the empirical performance of the L_1 -based independence test compared to alternative approaches. For this purpose, we assess L_1 -based test's size and power for various non-standard distributions. We assess the power and size of the L_1 test in comparison to other pre-existing tests enumerated below. The dHSIC test described by [Pfister et al. \(2018\)](#) and implemented in the R package dHSIC [[Pfister and Peters \(2017\)](#)]. The test based on ranks of distances introduced in [Heller, Heller and Gorfine \(2013\)](#) and implemented in the R package HHG [[Heller, Heller and Gorfine \(2013\)](#)]. The test based on the distance covariance discussed in [Székely and Rizzo \(2009\)](#) (will be called Dcov)

and implemented in the R package `energy` [Rizzo and Székely (2016)]. The mutual information-based test developed in Berrett and Samworth (2019) (will be called `Mintav`) and implemented in the R package `IndepTest` [Berrett, Grose and Samworth (2018)]. Three measures based on the generalized distance covariance, asymmetric measure R_n , symmetric measure S_n based on distance covariance and simplified complete measure Q_n^* based on incomplete V -statistics and implemented in the R package `EDMeasure` [Jin and Matteson (2018)]. Finally, the L_2 -test of independence proposed by Rosenblatt (1975). Rosenblatt and Wahlen (1992) showed that if the second order partial derivatives of the density function $f(\cdot)$ are bounded and uniformly continuous on $\mathbb{R}^2$,

$$h_n^{-1} \left(nh_n^2 \int_{\mathbb{R}^2} \left[f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^2 f_{n,l}(x_l) \right]^2 d\tilde{\mathbf{x}} - A(n) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2\vartheta^2) \text{ as } n \rightarrow \infty,$$

where, for $K(\cdot) = K_1(\cdot) = K_2(\cdot)$ be a bounded kernel,

$$A(n) = \left(\int_{\mathbb{R}} K^2(u_1) du_1 \right)^2 - h_n \int_{\mathbb{R}} K^2(u_1) du_1 \left(\int_{\mathbb{R}} [f_1^2(x) + f_2^2(x)]^2 dx \right),$$

and

$$\vartheta^2 = \int_{\mathbb{R}} f_1^2(x) dx \int_{\mathbb{R}} f_2^2(x) dx \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}^2} K(t_1)K(t_2)K(t_1+s_1)K(t_2+s_2) dt_1 dt_2 \right)^2 ds_1 ds_2.$$

To compare our test with the L_2 -test, we use the result above and approximate $A(n)$ and ϑ^2 by using Lemma B.12. We present the results of a Monte Carlo experiment for evaluating the size and power of each test in two-dimensional testing scenarios. We examine a collection of data-generating mechanisms designed to illustrate various types of possible dependence:

(i) We consider first the six simulated examples of unusual bivariate distributions presented in the Supplementary Material of Heller, Heller and Gorfine (2013) and labeled '4 Clouds', 'W', 'Diamond', 'Parabola', '2 Parabolas', and 'Circle'. Following Heller, Heller and Gorfine (2013) and Fan et al. (2017), for these examples, we simulated data under both dependence and independence structures. The models coded as `GEVmodel1` and `GEVmodel2` were considered in Fan et al. (2017).

- The dependence model for `GEVmodel1` is as follows: $X_1 = W + N_1$, $X_2 = X_1 + T + N_2$, where W is Weibull distributed, T is Fréchet distributed and N_1, N_2 are independent $\mathcal{N}(0, 0.2^2)$.
- The dependence model for `GEVmodel2` is as follows: $X_1 = W + N_1$ and $X_2 = 3W + N_2$.

These models have non-symmetric distributions with heavy tails.

- We include an example taken from Székely, Rizzo and Bakirov (2007) given by $X_2 = X_1 N_3$, with X_1 and N_3 being independent standard normal variables.

(ii) We consider the class of sinusoidally dependent data, proposed by Sejdinovic et al. (2013), for which the density function is expressed, for $\ell \in \mathbb{N}$ and $(x_1, x_2) \in [-\pi, \pi]^2$, by $f_\ell(x_1, x_2) = \frac{1}{4\pi^2} (1 + \sin(\ell x_1) \sin(\ell x_2))$. This class of densities is particularly interesting and is considered by Sejdinovic et al. (2013) as being challenging to detect dependence; intuitively, this is because as ℓ increases, the dependence becomes increasingly localized, while the marginal densities are uniform on $[-\pi, \pi]$ for each ℓ . This example was further discussed in Berrett and Samworth (2019) and Böttcher, Keller-Ressel and Schilling (2019).

(iii) For $\theta \in [0, 1]$ and $(x_1, x_2) \in [-1, 1]^2$, define the density function $f_\theta(x_1, x_2) = 4^{-1} (1 + \theta(1 - 2\mathbb{1}_{\{x_1^2 + x_2^2 \leq 2/\pi\}}))$.

- (iv) Let X_1 and Z be independent with $X_1 \sim U[-1, 1]$, the uniform law on $[-1, 1]$, and $Z \sim \mathcal{N}(0, 1)$, and for a parameter $\rho \in [0, \infty)$, let $X_2 = |X_1|^\rho Z$.
- (v) Let X_1 and Z be independent with $X_1 \sim \mathcal{N}(0, 1)$ and $Z \sim \mathcal{N}(0, 1)$, and for a parameter $\rho \in [0, 1)$. Let X_2 be defined by $X_2 = \rho X_1 + \sqrt{1 - \rho^2} Z$.

In implementing our test, the computing program codes are in R. The joint density estimate and its marginals are computed, without loss of generality, by using the univariate uniform kernel function, i.e., $K_l(u) = \mathbb{1}_{\{|u| \leq 1/2\}}$, $l = 1, \dots, p$, and the combinatorial procedure described above for choosing the bandwidth vector $\mathbf{h}_n$ discussed in Remark 2.1. This last method was applied only to find bandwidth vector $\mathbf{h}_n$ which allows us to estimate the joint density and then we use each component of $\mathbf{h}_n$ to give an estimate of their marginals and the function $\hat{\mathcal{L}}_n(\cdot)$. The support set $\tilde{C}$ is estimated from the sample, specifically the common support set of all functions used in our test statistic, in this case, is estimated by the set $\prod_{i=1}^d [\min_{1 \leq j \leq n} (X_{i,j}) - h_n^{(i)}/2, \max_{1 \leq j \leq n} (X_{i,j}) + h_n^{(i)}/2]$. The multidimensional integration is based on the SUBregion-Adaptive Vegas Algorithm developed in Hahn (2005) (called suave) and implemented in the R package `cubature`. We chose the bandwidth h for the L_2 test through the cross-validation process, employing the Gaussian kernel. We test the null hypothesis $\mathcal{H}_0$ with a significance level of $\alpha = 0.05$. We run 1,000 replications with the adaptive permutation size $B = 200$ for all empirical measures that require a permutation procedure, i.e., dHSIC, HHG, Mintav, R_n , S_n , Q_n^* and Dcov.

Table 1 represents size and power comparisons between L_1 and all testing procedures described above with sample size $n = 50$. In Setting (i), all the cited tests have levels close to the nominal level of 0.05 for independent simulated data. For examples with dependent data, the L_1 test shows the same or better power compared to L_2 , dHSIC, HHG and Mintav for ‘W’, ‘Parabola’, ‘2 Parabolas’ and ‘Circle’. We see also that for ‘4 Clouds’ where dHSIC, R_n , S_n , Q_n^* , Dcov perform poorly, L_1 , L_2 , HHG and Mintav have good power. R_n , S_n and Dcov present good performance only for ‘GEVmodel2’, ‘Parabola’, ‘W’ and Székely, Rizzo and Bakirov (2007)’s setting. While Q_n^* presents good performance only for ‘W’, ‘2 Parabola’, ‘GEVmodel2’ and Székely, Rizzo and Bakirov (2007)’s setting.

For the last three examples, we also compare our proposed test with all previously cited dependence measures, the Mintav test was proposed in Berrett and Samworth (2019), where it was shown that it outperforms many other tests for sinusoidal dependence (Setting (ii)) and underperforms in Setting (iv) when ρ increases. Figure 1 shows that, for $n = 200$, the L_1 test has solid performances for Settings (ii), (iii) and (iv). In Setting (ii), we note that the L_1 test keeps a better power than Mintav, even if ℓ increases, while the other tests have almost no power when $\ell \geq 2$. In Setting (iii), L_1 and Mintav provide good power when θ increases. The L_1 test outperforms the other tests. In Setting (iv), L_1 , HHG and dHSIC have similar power when ρ increases and performs better than Mintav. In the last Setting (v), the R_n and S_n tests outperform the other tests. These measures share similar properties and asymptotic distributions with distance covariance quickly detecting the dependence in the Gaussian setting. We highlight that our test necessitates the selection of bandwidth based on a priori information derived from the sample. Numerous tests of independence use a permutation or bootstrap strategy to determine the critical values, while other tests choose a posteriori the parameter with the greatest power. The aforementioned procedures render the decision uncertain, which is inconvenient for real-world data analysis applications. In conclusion, due to the wide variety of potential forms of dependence, there is no universally most powerful test. However, if the specific nature of the dependence is known beforehand, it may be feasible to create a tailored test that yields high power.

Table 1. Size and power of L_1 , L_2 , dHSIC, HHG, Mintav, R_n , S_n , Q_n^* and Dcov for some unusual relations with $n = 50$. Results based on 1,000 replications.

Distribution	Indep	L_1	L_2	dHSIC	HHG	Mintav	R_n	S_n	Q_n^*	Dcov
IndNorm	Yes	0.023	0.012	0.033	0.043	0.039	0.045	0.049	0.045	0.043
4 Clouds	Yes	0.055	0.059	0.056	0.057	0.042	0.052	0.055	0.041	0.047
4 Clouds	No	1	1	0.092	1	1	0.095	0.096	0.138	0.086
W	Yes	0.058	0.044	0.057	0.051	0.048	0.057	0.054	0.057	0.05
W	No	1	1	0.989	0.999	1	0.933	0.93	0.863	0.915
Parabola	Yes	0.041	0.028	0.053	0.05	0.04	0.052	0.065	0.058	0.051
Parabola	No	0.972	0.895	0.988	0.998	1	0.958	0.966	0.381	0.938
2 Parabolas	Yes	0.032	0.025	0.051	0.046	0.032	0.055	0.053	0.055	0.05
2 Parabolas	No	0.996	1	1	1	1	0.312	0.301	0.858	0.274
Circle	Yes	0.055	0.058	0.053	0.056	0.046	0.047	0.051	0.051	0.041
Circle	No	1	0.994	0.897	0.985	0.993	0.094	0.088	0.265	0.076
$X_2 = X_1 Z$	Yes	0.025	0.022	0.051	0.056	0.033	0.046	0.048	0.06	0.037
$X_2 = X_1 Z$	No	0.892	0.843	0.95	0.992	0.987	0.665	0.664	0.669	0.633
Diamond	Yes	0.033	0.017	0.044	0.054	0.03	0.035	0.035	0.046	0.034
Diamond	No	0.473	0.248	0.51	0.643	0.095	0.038	0.037	0.141	0.031
GEVmodel1	Yes	0.038	0.028	0.053	0.059	0.036	0.052	0.053	0.053	0.047
GEVmodel1	No	0.851	0.786	0.863	0.937	0.916	0.355	0.356	0.731	0.337
GEVmodel2	Yes	0.023	0.014	0.053	0.052	0.035	0.052	0.04	0.051	0.042
GEVmodel2	No	0.216	0.099	0.268	0.401	0.356	0.645	0.648	0.293	0.624

7. Concluding remarks

The notion of independence is of the highest significance in the domain of probability and statistics. It sets probability apart from being only a part of measure theory and forms the foundation for statistical theory and the approach practitioners take to modelling. Statisticians often need to determine the realism of assumptions of independence. This is vital for assessing whether certain theoretical aspects of procedures would hold and for evaluating the fit quality of a statistical model, one can refer to [Berrett, Kontoyiannis and Samworth \(2021\)](#). Although many existing tests for independence focus on two random vectors, they frequently do not have clear methods to evaluate mutual independence among more than two random vectors with different dimensions. This paper presents a new method to assess the independence of several random vectors, regardless of their dimensionality. Our method uses the L_1 -distance between the joint density and the product of the marginal densities. Under the null hypothesis, we use the Poissonization techniques to find, for the first time, the asymptotic normal approximation of the corresponding statistic. We do this without making assumptions about the regularity of the underlying Lebesgue density $f(\cdot)$. Also, and this was a surprise, the limiting distribution

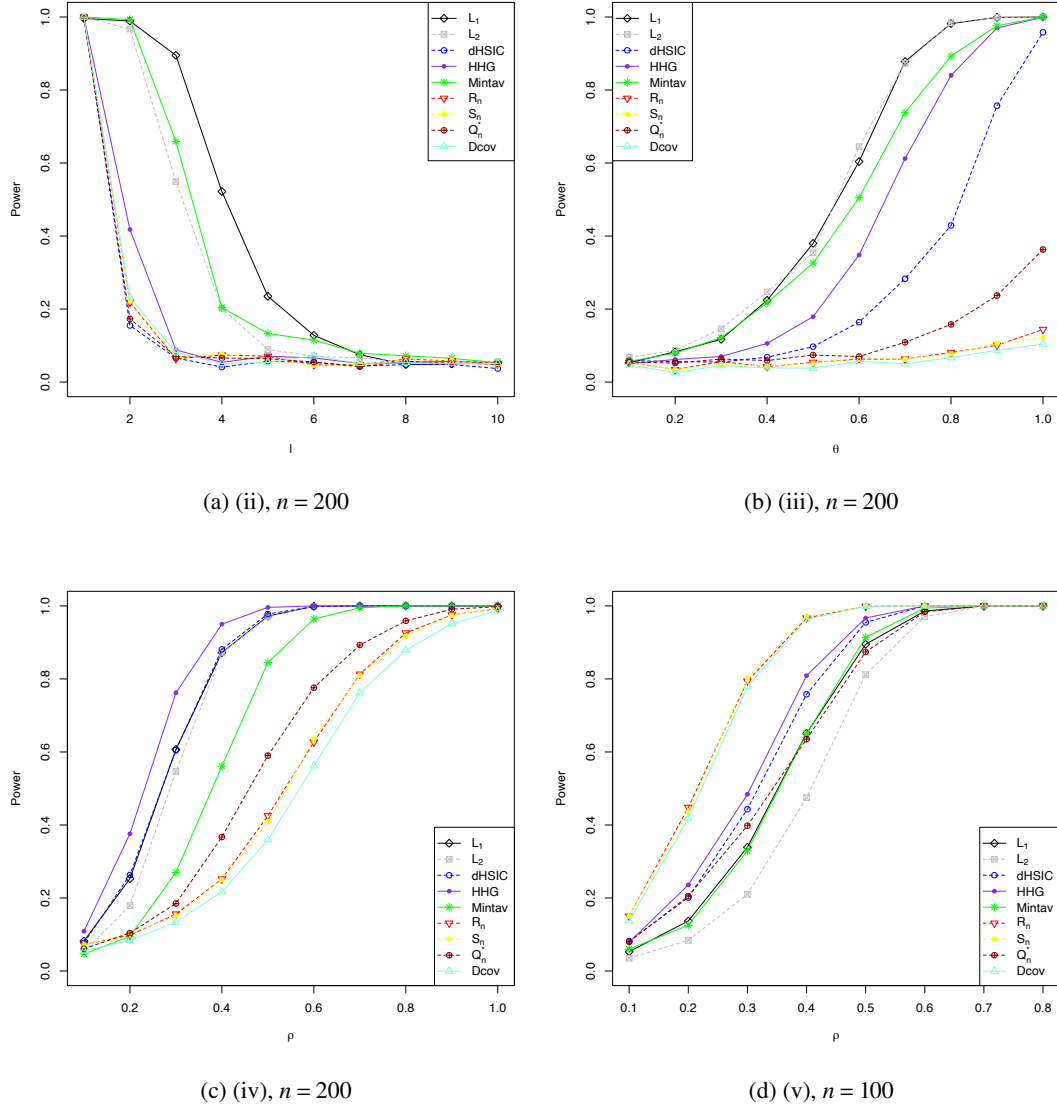


Figure 1: Power curves for the different tests

of the statistics based on L_1 -distance does not depend on $f(\cdot)$. We show that the tests have nontrivial local power against a subset of local alternatives that converge to the null at the rate of $n^{-1/2}h_n^{-d/4}$. Lastly, simulations are used to study how the tests behave for a moderate sample size. One aspect that remains unexplored in this article is the best selection of the smoothing parameters to maximize the power of the proposed tests. The subject at hand holds significant importance and warrants dedicated research effort. We defer this matter to a forthcoming investigation. Multiple avenues exist for developing our method further. When we look at modern machine learning algorithms, we may need to deal

with kernel density estimation with complicated kernels that are not chosen by the user and may even be irregular and asymmetric on Riemannian manifolds with Riemann integrable kernels. A challenging task is to consider this setting. It would be interesting to extend the present work to the problem of testing conditional independence in an incomplete data setting that requires nontrivial mathematics that goes well beyond the scope of the present paper. In the context of the serially dependent data, a future research direction would be to investigate the problem of testing independence as such that was examined in this work.

Appendix A: Proof of Theorems

In this section, we give the proof of all theoretical results in this work. The proof of Theorem 3.1 is quite involved. For the sake of readability, the proof of Theorem 3.1 is presented with a sequence of lemmas (given in Berrahou, Bouzebda and Douge (2024)) providing the necessary techniques at every demonstration step. Lemma B.1 shows that Borel sets exist, allowing the use of the truncation procedure and providing some technical results on convolutions which are crucial for the proof. To prove the statements of this lemma, we suppose that the kernel functions have compact supports. Lemma B.2 is useful to investigate the behavior of the difference between the normalized and truncated statistic $\sqrt{n}V_n(A)$ and the normalized statistic $\sqrt{n}V_n$, for any Borel set A of $\mathbb{R}^d$, and shows that this difference is asymptotically negligible for large n and large set A . Lemma B.3 and Lemma B.7 show that $\sqrt{n}V_n(\bar{C})$ is asymptotically equivalent to $\sqrt{n} \int_{\bar{C}} |f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p \mathbb{E}f_{n,l}(\mathbf{x}_l)| d\tilde{\mathbf{x}}$. The problem in dealing with V_n is its Poissonization¹ does not allow obtaining the independence of increments for disjoint sets, which are required in establishing Lemma B.6 by using a type of Berry-Esseen approximations for sums of independent random variables, due to Sweeting (1977), for easy reference, see Theorem C.1. Intermediate results about the Poissonization step are given in five lemmas. In this respect, Lemma B.4 establishes the asymptotic equivalence between the expectation of Poissonized and the non-Poissonized terms. Lemma B.6 allows us to derive the result of Lemma B.7 and provides the limit variance of the Poissonized version of $\sqrt{n}V_n(\bar{C})$. Lemma B.10 gives the asymptotic distribution of the Poissonized and normalized version of $\int_{\bar{C}} |f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p \mathbb{E}f_{n,l}(\mathbf{x}_l)| d\tilde{\mathbf{x}}$. This result follows from Lemma B.8 and Lemma B.9 by applying Theorem 1 of Shergin (1990) to the sum of m -dependent random fields deduced by partitioning the integral into integrals on small disjoint domains. Lemma B.11 is devoted to the dePoissonization step.

Proof of Theorem 3.1. Notice that, by Lemma B.1, for each $l = 1, \dots, p$, there exists a sequence of Borel sets $\{C_{l,k}\}_{k \geq 1}$ in $\mathbb{R}^{d_l}$ such that, for each $k \geq 1$, $C_{l,k}$ has finite Lebesgue measure and satisfies (B.3) and (B.4) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K} = \mathcal{K}_l$, and

$$\lim_{k \rightarrow \infty} \int_{C_{l,k}^c} f_l(\mathbf{u}_l) d\mathbf{u}_l = 0, \quad \text{where } C_{l,k}^c \text{ is the complement set of } C_{l,k}. \quad (\text{A.1})$$

Let $\bar{C}_k = C_{1,k} \times \dots \times C_{p,k}$ for $k \geq 1$. By Lemma B.11, for each $k \geq 1$, as $n \rightarrow \infty$, we have

$$\frac{\sqrt{n}(V_n(\bar{C}_k) - \mathbb{E}V_n(\bar{C}_k))}{\sqrt{n\text{Var}(U_{\eta,0}(\bar{C}_k))}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

¹The interest of Poissonization relies on the nice properties of Poisson processes, namely the independence of their increments and the behavior of their moments. These properties considerably simplify calculations, to be more precise, if η is a Poisson random variable independent of the i.i.d. sequence $\{X_i : i \in \mathbb{N}\}$, $X_0 = 0$, and if $A_k, k \in \mathbb{N}$, are disjoint measurable sets, then the processes $\sum_{i=0}^{\eta} \mathbb{1}(X_i \in A_k) \delta_{X_i}$, $k = 1, 2, \dots$, are independent.

and, by (B.36), we infer $\lim_{n \rightarrow \infty} n \text{Var}(U_{\eta,0}(\tilde{C}_k)) = \sigma^2 \prod_{l=1}^p \int_{C_{l,k}} f_l(\mathbf{u}_l) d\mathbf{u}_l$. By Lemma B.2, for each $k \geq 1$, it readily follows that

$$\limsup_{n \rightarrow \infty} n \text{Var}(V_n(\tilde{C}_k^c)) \leq 64 \int_{\tilde{C}_k^c} f(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} \leq 64 \sum_{l=1}^p \int_{\tilde{C}_{l,k}^c} f_l(\mathbf{u}_l) d\mathbf{u}_l.$$

Now, by (A.1) and Theorem 4.2 of Billingsley (1999), we deduce that, as $n \rightarrow \infty$, $\sqrt{n}(V_n - \mathbb{E}V_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$. We obtain the result of Theorem 3.1, as sought. ■

Proof of Theorem 3.2. Introduce $L_n(\cdot)$ by setting

$$L_n(\tilde{\mathbf{x}}) = \prod_{l=1}^p K_l\left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n}\right) - \sum_{l=1}^p K_l\left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n}\right) \prod_{j \neq l} \mathbb{E}K_j\left(\frac{\mathbf{x}_j - \mathbb{X}^j}{h_n}\right) + (p-1) \prod_{l=1}^p \mathbb{E}K_l\left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n}\right).$$

By the statement (C.1), we get, for some constant $\Upsilon_1 > 0$,

$$\left| \sqrt{n} \mathbb{E}U_{n,1}(\tilde{C}) - \mathbb{E}|Z| \int_{\tilde{C}} \sqrt{\mathcal{L}_{n,1}(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}} \right| \leq \frac{\Upsilon_1}{\sqrt{n} h_n^{3d}} \int_{\tilde{C}} \frac{\mathbb{E}|L_n(\tilde{\mathbf{x}})|^3}{\mathcal{L}_{n,1}(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}}, \quad (\text{A.2})$$

where $\mathcal{L}_{n,1}(\tilde{\mathbf{x}}) = n \text{Var}(\Gamma_{n,1}(\tilde{\mathbf{x}}))$. This last bound is, since, for each $l = 1, \dots, p$, $K_l(\cdot)$ is bounded, $\lambda(\tilde{C}) < \infty$ and $h_n^{-2d} \text{Var}(L_n(\tilde{\mathbf{x}})) = \mathcal{L}_{n,1}(\tilde{\mathbf{x}})$, less than or equals to $\frac{v_2}{\sqrt{n} h_n^d}$ for some constant $v_2 > 0$. Now, since, for each $l = 1, \dots, p$, $f_l(\cdot)$ is bounded on C_l^δ for some $\delta > 0$, thus, by arguing as in the proof of Lemma B.3, we get

$$\sqrt{n} \mathbb{E}|V_n(\tilde{C}) - U_{n,1}(\tilde{C})| = O\left(\frac{1}{\sqrt{n} h_n^d}\right). \quad (\text{A.3})$$

Combining (A.2)-(A.3) and Lemma B.12, we obtain $\sqrt{n}(V_n(\tilde{C}) - a_n(\tilde{C})) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$ as $n \rightarrow \infty$. Therefore, for n large enough, we obtain readily

$$\mathbb{P}\left\{\frac{\sqrt{n}}{\sigma}(V_n(\tilde{C}) - a_n(\tilde{C})) \leq z_{1-\alpha}\right\} \rightarrow 1 - \alpha.$$

Thus, the α -level test rejects the null hypothesis if $V_n(\tilde{C}) > a_n(\tilde{C}) + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$. This proves that the test has an asymptotic error probability equal to α . ■

Proof of Theorem 3.3. Making use of the triangular inequality, we readily obtain

$$\begin{aligned} |V_n(\tilde{D}) - V(\tilde{D})| &\leq \int_{\tilde{D}} \left| f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_{n,l}(\mathbf{x}_l) - \left(\mathbb{E}f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p \mathbb{E}f_{n,l}(\mathbf{x}_l) \right) \right| d\tilde{\mathbf{x}} \\ &\quad + \int_{\tilde{D}} \left| f(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_l(\mathbf{x}_l) - \left(\mathbb{E}f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p \mathbb{E}f_{n,l}(\mathbf{x}_l) \right) \right| d\tilde{\mathbf{x}}. \end{aligned}$$

We now observe that

$$\begin{aligned} & \left| f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_{n,l}(\mathbf{x}_l) - \left(\mathbb{E} f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p \mathbb{E} f_{n,l}(\mathbf{x}_l) \right) \right| \\ & \leq |f_n(\tilde{\mathbf{x}}) - \mathbb{E} f_n(\tilde{\mathbf{x}})| + \sum_{j=1}^p |f_{n,j}(\mathbf{x}_j) - \mathbb{E} f_{n,j}(\mathbf{x}_j)| \prod_{l=1}^{j-1} \mathbb{E} f_{n,l}(\mathbf{x}_l) \prod_{l=j+1}^p f_{n,l}(\mathbf{x}_l), \end{aligned}$$

and

$$\begin{aligned} & \left| f(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_l(\mathbf{x}_l) - \left(\mathbb{E} f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p \mathbb{E} f_{n,l}(\mathbf{x}_l) \right) \right| \\ & \leq |f(\tilde{\mathbf{x}}) - \mathbb{E} f_n(\tilde{\mathbf{x}})| + \sum_{j=1}^p |f_j(\mathbf{x}_j) - \mathbb{E} f_{n,j}(\mathbf{x}_j)| \prod_{l=1}^{j-1} \mathbb{E} f_{n,l}(\mathbf{x}_l) \prod_{l=j+1}^p f_l(\mathbf{x}_l). \end{aligned}$$

Now, for each $l = 1, \dots, p$, we have

$$\int_{\mathbb{R}^{d_l}} f_{n,l}(\mathbf{x}_l) d\mathbf{x}_l = 1 \quad \text{and} \quad \int_{\mathbb{R}^{d_l}} \mathbb{E} f_{n,l}(\mathbf{x}_l) d\mathbf{x}_l \rightarrow \int_{\mathbb{R}^{d_l}} f_l(\mathbf{x}_l) d\mathbf{x}_l = 1.$$

Then, for n large enough, there exists a constant $C > 0$ such that

$$\begin{aligned} \mathbb{E}|V_n(\tilde{D}) - V(\tilde{D})| & \leq \int_{\tilde{D}} \mathbb{E}|f_n(\tilde{\mathbf{x}}) - \mathbb{E} f_n(\tilde{\mathbf{x}})| d\tilde{\mathbf{x}} + C \sum_{j=1}^p \int_{D_j} \mathbb{E}|f_{n,j}(\mathbf{x}_j) - \mathbb{E} f_{n,j}(\mathbf{x}_j)| d\mathbf{x}_j \\ & \quad + \int_{\mathbb{R}^d} |f(\tilde{\mathbf{x}}) - \mathbb{E} f_n(\tilde{\mathbf{x}})| d\tilde{\mathbf{x}} + C \sum_{j=1}^p \int_{\mathbb{R}^{d_j}} |f_j(\mathbf{x}_j) - \mathbb{E} f_{n,j}(\mathbf{x}_j)| d\mathbf{x}_j. \end{aligned}$$

Making use of the Cauchy-Schwartz and Young's inequalities, we infer that

$$\begin{aligned} \int_{\tilde{D}} \mathbb{E}|f_n(\tilde{\mathbf{x}}) - \mathbb{E} f_n(\tilde{\mathbf{x}})| d\tilde{\mathbf{x}} & \leq \int_{\tilde{D}} \left(\text{Var}(f_n(\tilde{\mathbf{x}})) \right)^{1/2} d\tilde{\mathbf{x}} \leq \frac{(\mu(\tilde{D}))^{1/2}}{\sqrt{nh_n^d}} \left(\int_{\tilde{D}} \left\{ \prod_{l=1}^p K_l \right\}^2 * f(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} \right)^{1/2} \\ & = O\left(\frac{1}{\sqrt{nh_n^d}} \right). \end{aligned}$$

Likewise, for each $j = 1, \dots, p$, we have $\int_{D_j} \mathbb{E}|f_{n,j}(\mathbf{x}_j) - \mathbb{E} f_{n,j}(\mathbf{x}_j)| d\mathbf{x}_j = O\left(\frac{1}{\sqrt{nh_n^{d_j}}} \right)$. On the other hand, by Theorem 1 in Chapter 2 of [Devroye and Györfi \(1985\)](#), we readily infer

$$\int_{\mathbb{R}^d} |f(\tilde{\mathbf{x}}) - \mathbb{E} f_n(\tilde{\mathbf{x}})| d\tilde{\mathbf{x}} \rightarrow 0 \quad \text{and} \quad \int_{\mathbb{R}^{d_j}} |f_j(\mathbf{x}_j) - \mathbb{E} f_{n,j}(\mathbf{x}_j)| d\mathbf{x}_j \rightarrow 0, \quad j = 1, \dots, p.$$

Using the fact that $n^{-1/2} h^{-d/2} \rightarrow 0$, we conclude that $V_n(\tilde{D}) \xrightarrow{\mathcal{P}} V(\tilde{D})$. To complete the proof of the theorem, it suffices to show that $\mathbb{P}\left(V_n(\tilde{D}) < n^{-1/2}(\sigma z_{1-\alpha} + a_n(\tilde{D}))\right) \rightarrow 0$. Now, we can check easily

that $a_n(\tilde{D}) = O(h^{-d/2})$, then for n large enough, there exists some constant $\nu > 0$ in such a way that $\frac{1}{\sqrt{n}}(\sigma z_{1-\alpha} + a_n(\tilde{D})) - V(\tilde{D}) < -\nu$. Therefore, we infer that

$$\mathbb{P}\left(V_n(\tilde{D}) < \frac{1}{\sqrt{n}}(\sigma z_{1-\alpha} + a_n(\tilde{D}))\right) \leq \mathbb{P}(V_n(\tilde{D}) - V(\tilde{D}) < -\nu) \leq \mathbb{P}(|V_n(\tilde{D}) - V(\tilde{D})| > \nu) \rightarrow 0.$$

Hence, we obtain the desired conclusion of Theorem 3.3. $\blacksquare$

Proof of Theorem 4.1. Notice that

$$\begin{aligned} T_n(\tilde{C}) &= \sqrt{n}\sigma^{-1}V_n(\tilde{C}) - \sigma^{-1}a_n(\tilde{C}) \\ &= \sqrt{n}\sigma^{-1}(V_n(\tilde{C}) - \mathbb{E}V_n(\tilde{C})) + \sqrt{n}\sigma^{-1}(\mathbb{E}V_n(\tilde{C}) - n^{-1/2}a_n(\tilde{C})). \end{aligned} \quad (\text{A.4})$$

To establish that the first term on the right side of (A.4) converges to $\mathcal{N}(0, 1)$ under $H_{\delta, n}$, we follow the proof of Theorem 3.1 with some slight differences. Note that in this proof, we have not assumed independence. So for this reason we will use a type of Berry-Esseen bound for sums of independent random variables due to Sweeting (1977), the theorem in Pinelis (1994) and Theorem 2 on page 63 of Stein (1970). Consider the second term in (A.4). Under $H_{\delta, n}$, since $nh_n^{3d} \rightarrow \infty$, by using the arguments of the proof of Lemma B.4, we get

$$\lim_{n \rightarrow \infty} \sqrt{n}\mathbb{E}U_{n,1}(\tilde{C}) - \int_{\tilde{C}} \mathbb{E}|\gamma_{n,1}(\tilde{\mathbf{x}})Z_1 + \delta_n(\tilde{\mathbf{x}})| d\tilde{\mathbf{x}} = 0,$$

where $\gamma_{n,1}(\tilde{\mathbf{x}}) = \sqrt{\mathcal{L}_{n,1}(\tilde{\mathbf{x}})}$ and $\delta_n(\tilde{\mathbf{x}}) = h_n^{-d/4} \int_{\mathbb{R}^d} \delta(\tilde{\mathbf{x}} - \mathbf{u}h_n) \left\{ \prod_{l=1}^p K_l(\mathbf{u}_l) \right\} d\mathbf{u}$. Now we have

$$\begin{aligned} &\mathbb{E}|\gamma_{n,1}(\tilde{\mathbf{x}})Z_1 + \delta_n(\tilde{\mathbf{x}})| - \mathbb{E}|\gamma_{n,1}(\tilde{\mathbf{x}})Z_1| \\ &= 2\delta_n(\tilde{\mathbf{x}})\Phi\left(\frac{\delta_n(\tilde{\mathbf{x}})}{\gamma_{n,1}(\tilde{\mathbf{x}})}\right) + 2\gamma_{n,1}(\tilde{\mathbf{x}})\phi\left(\frac{\delta_n(\tilde{\mathbf{x}})}{\gamma_{n,1}(\tilde{\mathbf{x}})}\right) - \delta_n(\tilde{\mathbf{x}}) - 2\gamma_{n,1}(\tilde{\mathbf{x}})\phi(0), \end{aligned}$$

where $\phi(\cdot)$ denotes the pdf of the standard normal distribution. By a Taylor expansion, we infer

$$\begin{aligned} &\mathbb{E}|\gamma_{n,1}(\tilde{\mathbf{x}})Z_1 + \delta_n(\tilde{\mathbf{x}})| - \mathbb{E}|\gamma_{n,1}(\tilde{\mathbf{x}})Z_1| = 2\delta_n(\tilde{\mathbf{x}}) \left(\Phi(0) + \frac{\delta_n(\tilde{\mathbf{x}})}{\gamma_{n,1}(\tilde{\mathbf{x}})}\phi(0) + O\left(\frac{\delta_n(\tilde{\mathbf{x}})}{\gamma_{n,1}(\tilde{\mathbf{x}})}\right)^2 \right) - \delta_n(\tilde{\mathbf{x}}) \\ &\quad + 2\gamma_{n,1}(\tilde{\mathbf{x}}) \left(\sum_{k=0}^3 \frac{1}{k!} \left(\frac{\delta_n(\tilde{\mathbf{x}})}{\gamma_{n,1}(\tilde{\mathbf{x}})} \right)^k \phi^{(k)}(0) + o\left\{ \left(\frac{\delta_n(\tilde{\mathbf{x}})}{\gamma_{n,1}(\tilde{\mathbf{x}})} \right)^3 \right\} \right) - 2\gamma_{n,1}(\tilde{\mathbf{x}})\phi(0). \end{aligned}$$

By Lemma B.1, Lemma B.4 and (B.15) we obtain, as $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \sigma^{-1} \left(\sqrt{n}\mathbb{E}U_{n,1}(\tilde{C}) - \int_{\tilde{C}} \mathbb{E}|\gamma_{n,1}(\tilde{\mathbf{x}})Z_1| d\tilde{\mathbf{x}} \right) = \eta(\delta),$$

this when combined with Lemma B.3 and Lemma B.12 implies that $\sqrt{n}\sigma^{-1}(\mathbb{E}V_n(\tilde{C}) - n^{-1/2}a_n(\tilde{C}))$ converges in probability to $\eta(\delta)$, as $n \rightarrow \infty$, which completes the proof of Theorem 4.1. $\blacksquare$

Conflicts 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.

Acknowledgments

The authors would like to thank the Editor-in-Chief, an Associate-Editor, and two anonymous referees for their constructive remarks, which resulted in a substantial improvement of the work's original form and a more sharply focused presentation.

Supplementary material

The supplementary material in [Berrahou, Bouzebda and Douge \(2024\)](#) contains the proofs of all technical lemmas.

Appendix B: Technical lemmas

The following lemma will be crucial for the proof of Theorem 3.1.

Lemma B.1. *Let $g(\cdot)$ be a Lebesgue density function on $\mathbb{R}^s$, $s \geq 1$, and $\mathcal{K}$ be a finite class of bounded real-valued functions $K(\cdot)$ with compact support. Then, for each $K \in \mathcal{K}$, we have*

$$|K_{h_n} * g(\mathbf{z}) - J(K)g(\mathbf{z})| \rightarrow 0 \quad \text{as } h_n \rightarrow 0 \text{ for almost all } \mathbf{z} \in \mathbb{R}^s, \quad (\text{B.1})$$

where

$$J(K) = \int_{\mathbb{R}^s} K(\mathbf{u}) d\mathbf{u} \quad \text{and} \quad K_{h_n} * g(\mathbf{z}) := h_n^{-s} \int_{\mathbb{R}^s} K\left(\frac{\mathbf{z} - \mathbf{v}}{h_n}\right) g(\mathbf{v}) d\mathbf{v}.$$

Moreover, for all $0 < \varepsilon < 1$, there exist $M, \nu > 0$ and a Borel set C of finite Lebesgue measure such that

$$C \subset [-M + \nu, M - \nu]^s, \quad \int_{\mathbb{R}^s \setminus [-M, M]^s} g(\mathbf{v}) d\mathbf{v} = \mu > 0, \quad \int_C g(\mathbf{v}) d\mathbf{v} > 1 - \varepsilon, \quad (\text{B.2})$$

$$g(\cdot) \text{ is bounded, continuous and bounded away from 0 on } C \quad (\text{B.3})$$

and, for each $K \in \mathcal{K}$,

$$\sup_{\mathbf{z} \in C} |K_{h_n} * g(\mathbf{z}) - J(K)g(\mathbf{z})| \rightarrow 0 \quad \text{as } h_n \rightarrow 0. \quad (\text{B.4})$$

Proof of Lemma B.1. The first statement is just a simple case of Theorem 3 in Chapter 2 of [Devroye \(1987\)](#). The other statements can be proved in the same way as in Lemma 6.1 of [Giné, Mason and Zaitsev \(2003\)](#). Since $g(\cdot)$ is Lebesgue integrable, the integral $\int_{\mathbb{R}^s \setminus [-M, M]^s} g(\mathbf{v}) d\mathbf{v}$ is continuous in M and converges to zero as $M \rightarrow \infty$. We may find $M > 0$ and $\nu > 0$ so that

$$\int_{\mathbb{R}^s \setminus [-M, M]^s} g(\mathbf{v}) d\mathbf{v} = \frac{\varepsilon}{8} \quad \text{and} \quad \int_{\mathbb{R}^s \setminus [-M + \nu, M - \nu]^s} g(\mathbf{v}) d\mathbf{v} = \frac{\varepsilon}{4}.$$

The existence of the desired set $C \subset [-M + \nu, M - \nu]^s$ can be inferred from Lusin's theorem followed by Egorov's theorem, for instance, see [Dudley \(1989\)](#), Theorems 7.5.1 and 7.5.2. $\blacksquare$

The next lemma provides a truncation step and will play an instrumental role in the sequel.

Lemma B.2. Assume that assumption (A_1) holds. If $h_n \rightarrow 0$ and $nh_n^d \rightarrow \infty$, then, for all Borel set $A \in \mathbb{R}^d$,

$$\limsup_{n \rightarrow \infty} n \text{Var}(V_n(A)) \leq 64 \int_A f(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}.$$

Proof of Lemma B.2. Let $\mathbb{E}_0$ be the expectation $\mathbb{E}$ and let, for each $i = 1, \dots, n$, $\mathbb{E}_i$ be the conditional expectation given $(\mathbb{X}_1^1, \dots, \mathbb{X}_i^1)$. Set, for each $i = 1, \dots, n$,

$$\Upsilon_i := \mathbb{E}_i V_n(A) - \mathbb{E}_{i-1} V_n(A) \text{ and } \Phi_i := \mathbb{E}_i(V_n(A) - V_{n,-i}(A)),$$

where

$$\begin{aligned} V_{n,-i}(A) := \int_A & \left| \frac{1}{nh_n^d} \sum_{k \neq i} \prod_{l=1}^p K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_k^l}{h_n} \right) \right. \\ & \left. - \frac{1}{n^p h_n^d} \sum_{k \neq i} K_1 \left(\frac{\mathbf{x}_1 - \mathbb{X}_k^1}{h_n} \right) \prod_{l=2}^p \sum_{k=1}^n K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_k^l}{h_n} \right) \right| d\tilde{\mathbf{x}}. \end{aligned}$$

Clearly, we have

$$\sum_{i=1}^n \Upsilon_i = V_n(A) - \mathbb{E} V_n(A) \quad \text{and} \quad \mathbb{E}_{i-1} \Upsilon_i = 0, \quad i = 1, \dots, n. \quad (\text{B.5})$$

Also observe that, for each $i = 1, \dots, n$, $\Upsilon_i = \Phi_i - \mathbb{E}_{i-1} \Phi_i$ and

$$\begin{aligned} |\Phi_i| & \leq \mathbb{E}_i |V_n(A) - V_{n,-i}(A)| \\ & \leq \frac{2}{nh_n^d} \int_A K_1 \left(\frac{\mathbf{x}_1 - \mathbb{X}_i^1}{h_n} \right) \prod_{l=2}^p \mathbb{E}_i K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n} \right) d\tilde{\mathbf{x}} \\ & =: \Psi_i. \end{aligned} \quad (\text{B.6})$$

Hence, for each $i = 1, \dots, n$,

$$|\Upsilon_i| \leq |\Phi_i| + \mathbb{E}_{i-1} |\Phi_i| \leq \Psi_i + \mathbb{E} \Psi_i := \xi_i. \quad (\text{B.7})$$

For each $i = 1, \dots, n$, let $\tilde{\xi}_i$ be independent copy of ξ_i and let (ε_i) be an i.i.d. sequence of Rademacher variables ($\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = 1/2$), independent of the sequence $(\xi_i, \tilde{\xi}_i)$. Using (B.5)-(B.7), the convexity of $y = x^2$ and the fact that, for each $i = 1, \dots, n$, ξ_i is independent of $(\mathbb{X}_1^1, \dots, \mathbb{X}_{i-1}^1)$, it follows from Theorem 3.1 in [Berger \(1991\)](#) that

$$\mathbb{E} \left(\sum_{i=1}^n \Upsilon_i \right)^2 \leq \mathbb{E} \left(\sum_{i=1}^n \varepsilon_i (\xi_i + \tilde{\xi}_i) \right)^2 \leq 4n \mathbb{E} (\Psi_1 + \mathbb{E} \Psi_1)^2 \leq 16n \mathbb{E} \Psi_1^2. \quad (\text{B.8})$$

Combining (B.5)-(B.6) with (B.8), we obtain

$$n \mathbb{E} [V_n(A) - \mathbb{E} V_n(A)]^2 \leq 64 \int_A \mathbb{E} f_n(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}.$$

By Theorem 1 in Chapter 2 of Devroye (1987), we infer that, as $n \rightarrow \infty$,

$$\int_A |\mathbb{E} f_n(\tilde{\mathbf{x}}) - f(\tilde{\mathbf{x}})| d\tilde{\mathbf{x}} \rightarrow 0,$$

which completes the proof of the lemma. $\blacksquare$

Choose, for each $l = 1, \dots, p$, any bounded Borel set $C_l \subset \mathbb{R}^{d_l}$ satisfying (B.3) and (B.4) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K} = \mathcal{K}_l := \{K_l, K_l^2, K_l^3\}$. Clearly, for each $l = 1, \dots, p$, for each $K \in \mathcal{K}_l$, for all large enough n uniformly in $\mathbf{z}_l \in C_l$ and for some constant $D_0 > 0$,

$$\sup_{\mathbf{z}_l \in C_l} \int_{\mathbb{R}^{d_l}} K\left(\frac{\mathbf{z}_l - \mathbf{u}_l}{h_n}\right) f_l(u_l) d\mathbf{u}_l \leq D_0 h_n^{d_l}. \quad (\text{B.9})$$

Let us recall, for each $a \in \{0, 1\}$,

$$\Gamma_{n,a}(\tilde{\mathbf{x}}) := f_n(\tilde{\mathbf{x}}) - a \sum_{l=1}^p f_{n,l}(\mathbf{x}_l) \prod_{j \neq l} \mathbb{E} f_{n,j}(\mathbf{x}_j) + (pa - 1) \prod_{l=1}^p \mathbb{E} f_{n,l}(\mathbf{x}_l)$$

and set, for $\bar{C} = C_1 \times \dots \times C_p$,

$$U_{n,a}(\bar{C}) := \int_{\bar{C}} |\Gamma_{n,a}(\tilde{\mathbf{x}})| d\tilde{\mathbf{x}}.$$

The next lemma shows that $\sqrt{n}(V_n(\bar{C}) - \mathbb{E} V_n(\bar{C}))$ is asymptotically equivalent in probability to $\sqrt{n}(U_{n,1}(\bar{C}) - \mathbb{E} U_{n,1}(\bar{C}))$.

Lemma B.3. *Suppose that, for each $l = 1, \dots, p$, C_l satisfies (B.3) and (B.4) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K}(\cdot) = \mathcal{K}_l(\cdot)$. If $h_n \rightarrow 0$ and $nh_n^d \rightarrow \infty$, then*

$$\lim_{n \rightarrow \infty} \sqrt{n} \mathbb{E} |V_n(\bar{C}) - U_{n,1}(\bar{C})| = 0.$$

Proof of Lemma B.3. Notice that

$$\begin{aligned} & \left| f_n(\tilde{\mathbf{x}}) - \prod_{l=1}^p f_{n,l}(\mathbf{x}_l) - \Gamma_{n,1}(\tilde{\mathbf{x}}) \right| \\ & \leq \sum_{\{I \subsetneq I_p, |I_p \setminus I| \geq 2\}} \prod_{l \in I} \mathbb{E} f_{n,l}(\mathbf{x}_l) \prod_{j \in I_p \setminus I} |f_{n,j}(\mathbf{x}_j) - \mathbb{E} f_{n,j}(\mathbf{x}_j)|, \end{aligned}$$

where $I_p = \{1, \dots, p\}$, $|I|$ denotes the cardinality of the set I and the empty product is defined to be 1. Clearly, we have

$$\begin{aligned} \mathbb{E} |V_n(\bar{C}) - U_{n,1}(\bar{C})| & \leq \sum_{\{I \subsetneq I_p, |I_p \setminus I| \geq 2\}} \prod_{l \in I_p \setminus I} \int_{C_l} \mathbb{E} |f_{n,l}(\mathbf{x}_l) - \mathbb{E} f_{n,l}(\mathbf{x}_l)| d\mathbf{x}_l \\ & \leq \sum_{\{I \subsetneq I_p, |I_p \setminus I| \geq 2\}} \prod_{l \in I_p \setminus I} \int_{C_l} (\mathbb{E} [f_{n,l}(\mathbf{x}_l) - \mathbb{E} f_{n,l}(\mathbf{x}_l)]^2)^{1/2} d\mathbf{x}_l. \quad (\text{B.10}) \end{aligned}$$

Now, observe that, for each $l = 1, \dots, p$ and each $\mathbf{x}_l \in C_l$,

$$\mathbb{E} [f_{n,l}(\mathbf{x}_l) - \mathbb{E} f_{n,l}(\mathbf{x}_l)]^2 \leq \frac{1}{nh_n^{2d_l}} \int_{\mathbb{R}^{d_l}} K_l^2 \left(\frac{\mathbf{x}_l - \mathbf{u}_l}{h_n} \right) f_l(\mathbf{u}_l) d\mathbf{u}_l.$$

Therefore, by (B.9), for each $l = 1, \dots, p$, we have

$$\sup_{\mathbf{x}_l \in C_l} \mathbb{E} [f_{n,l}(\mathbf{x}_l) - \mathbb{E} f_{n,l}(\mathbf{x}_l)]^2 = O \left(\frac{1}{nh_n^{d_l}} \right). \quad (\text{B.11})$$

Inequality (B.10) together with (B.11) give, for some constant $M_0 > 0$,

$$\begin{aligned} \sqrt{n} \mathbb{E} |V_n(\bar{C}) - U_{n,1}(\bar{C})| &\leq M_0 \sum_{\{I \subsetneq I_p, |I_p \setminus I| \geq 2\}} \frac{1}{n^{(|I_p \setminus I| - 1)/2} \prod_{l \in I_p \setminus I} h_n^{d_l/2}} \\ &\leq \frac{M_0}{\sqrt{nh_n^d}} \sum_{\{I \subsetneq I_p, |I_p \setminus I| \geq 2\}} \prod_{l \in I} h_n^{d_l/2}. \end{aligned}$$

This completes the proof of Lemma B.3. ■

The next step in the proof of the asymptotic normality theorem is what is known as Poissonization. Let η be a Poisson random variable with mean n , independent of $\tilde{\mathbf{X}}, \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \dots$, and set

$$f_\eta(\tilde{\mathbf{x}}) = \frac{1}{nh_n^d} \sum_{i=1}^{\eta} \prod_{l=1}^p K_l \left(\frac{\mathbf{x}_l - \mathbf{X}_i^l}{h_n} \right) \text{ and } f_{\eta,l}(\mathbf{x}_l) = \frac{1}{nh_n^{d_l}} \sum_{i=1}^{\eta} K_l \left(\frac{\mathbf{x}_l - \mathbf{X}_i^l}{h_n} \right),$$

where the empty sum is defined to be zero. For each $l = 1, \dots, p$, notice that

$$\mathbb{E} f_{\eta,l}(\mathbf{x}_l) = \mathbb{E} f_{n,l}(\mathbf{x}_l) = h_n^{-d_l} \int_{\mathbb{R}^{d_l}} K_l \left(\frac{\mathbf{x}_l - \mathbf{u}_l}{h_n} \right) f_l(\mathbf{u}_l) d\mathbf{u}_l \quad (\text{B.12})$$

and

$$v_{n,l}(\mathbf{x}_l) := n \text{Var}(f_{\eta,l}(\mathbf{x}_l)) = h_n^{-2d_l} \int_{\mathbb{R}^{d_l}} K_l^2 \left(\frac{\mathbf{x}_l - \mathbf{u}_l}{h_n} \right) f_l(\mathbf{u}_l) d\mathbf{u}_l. \quad (\text{B.13})$$

Set, for each $a \in \{0, 1\}$,

$$\Gamma_{\eta,a}(\tilde{\mathbf{x}}) := f_\eta(\tilde{\mathbf{x}}) - a \sum_{l=1}^p f_{\eta,l}(\mathbf{x}_l) \prod_{j \neq l} \mathbb{E} f_{n,j}(\mathbf{x}_j) + (pa - 1) \prod_{l=1}^p \mathbb{E} f_{n,l}(\mathbf{x}_l).$$

Clearly, for each $a \in \{0, 1\}$, we have

$$\begin{aligned} k_{n,a}(\tilde{\mathbf{x}}) &:= n \text{Var}(\Gamma_{\eta,a}(\tilde{\mathbf{x}})) \\ &= \prod_{l=1}^p v_{n,l}(\mathbf{x}_l) - a \sum_{l=1}^p v_{n,l}(\mathbf{x}_l) \prod_{j \neq l} (\mathbb{E} f_{n,j}(\mathbf{x}_j))^2 + p(p-1)a \prod_{l=1}^p (\mathbb{E} f_{n,l}(\mathbf{x}_l))^2 \end{aligned} \quad (\text{B.14})$$

and

$$\begin{aligned}\mathcal{L}_{n,a}(\tilde{\mathbf{x}}) &:= n\text{Var}(\Gamma_{n,a}(\tilde{\mathbf{x}})) \\ &= \prod_{l=1}^p v_{n,l}(\mathbf{x}_l) - a \sum_{l=1}^p v_{n,l}(\mathbf{x}_l) \prod_{j \neq l}^p (\mathbb{E} f_{n,j}(\mathbf{x}_j))^2 + (pa - 1) \prod_{l=1}^p (\mathbb{E} f_{n,l}(\mathbf{x}_l))^2.\end{aligned}$$

We now observe that

$$\begin{aligned}& \left| h_n^d k_{n,a}(\tilde{\mathbf{x}}) - \tilde{\mathbf{K}} \prod_{l=1}^p f_l(\mathbf{x}_l) \right| \\ & \leq \sum_{I \subsetneq I_p} \prod_{l \in I} \|K_l\|_2^2 f_l(\mathbf{x}_l) \prod_{j \in I_p \setminus I} |h_n^{d_j} v_{n,j}(\mathbf{x}_j) - \|K_j\|_2^2 f_j(\mathbf{x}_j)| \\ & \quad + a h_n^d \left\{ \sum_{l=1}^p v_{n,l}(\mathbf{x}_l) \prod_{j \neq l}^p (\mathbb{E} f_{n,j}(\mathbf{x}_j))^2 + p(p-1) \prod_{l=1}^p (\mathbb{E} f_{n,l}(\mathbf{x}_l))^2 \right\}.\end{aligned}$$

Combining this last observation with (B.12)-(B.13) and the fact that, for $l = 1, \dots, p$, C_l satisfies (B.3) and (B.4) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K}(\cdot) = \mathcal{K}_l(\cdot)$, we obtain

$$\lim_{n \rightarrow \infty} \sup_{\tilde{\mathbf{x}} \in \bar{C}} \left| h_n^d k_{n,a}(\tilde{\mathbf{x}}) - \tilde{\mathbf{K}} \prod_{l=1}^p f_l(\mathbf{x}_l) \right| = 0. \quad (\text{B.15})$$

Thus, for all large enough n uniformly in $\tilde{\mathbf{x}} \in \bar{C}$ and for some constants $D_1 > 0$ and $D_2 > 0$, we have

$$D_1 \leq h_n^d k_{n,a}(\tilde{\mathbf{x}}) \leq D_2. \quad (\text{B.16})$$

Therefore, by (B.9) and (B.16), we get

$$\sup_{\tilde{\mathbf{x}} \in \bar{C}} \left| \sqrt{\mathcal{L}_{n,a}(\tilde{\mathbf{x}})} - \sqrt{k_{n,a}(\tilde{\mathbf{x}})} \right| \leq |2pa - p^2a - 1| \sup_{\tilde{\mathbf{x}} \in \bar{C}} \frac{\prod_{l=1}^p (\mathbb{E} f_{n,l}(\mathbf{x}_l))^2}{\sqrt{k_{n,a}(\tilde{\mathbf{x}})}} = O\left(\sqrt{h_n^d}\right).$$

Set, for each $a \in \{0, 1\}$,

$$U_{\eta,a}(\bar{C}) := \int_{\bar{C}} |\Gamma_{\eta,a}(\tilde{\mathbf{x}})| d\tilde{\mathbf{x}}.$$

The next lemma shows that, for each $a \in \{0, 1\}$, $\sqrt{n}\mathbb{E}U_{n,a}(\bar{C})$ is asymptotically equivalent to $\sqrt{n}\mathbb{E}U_{\eta,a}(\bar{C})$.

Lemma B.4. *Suppose that, for each $l = 1, \dots, p$, C_l satisfies (B.3) and (B.4) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K}(\cdot) = \mathcal{K}_l(\cdot)$. If $h_n \rightarrow 0$ and $\sqrt{n}h_n^d \rightarrow \infty$, as $n \rightarrow \infty$, then, for each $a \in \{0, 1\}$,*

$$\lim_{n \rightarrow \infty} \sqrt{n}\mathbb{E}U_{\eta,a}(\bar{C}) - \mathbb{E}|Z_1| \int_{\bar{C}} \sqrt{k_{n,a}(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}} = 0, \quad (\text{B.17})$$

and

$$\lim_{n \rightarrow \infty} \sqrt{n} \mathbb{E} U_{n,a}(\bar{C}) - \mathbb{E}|Z_1| \int_{\bar{C}} \sqrt{k_{n,a}(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}} = 0. \quad (\text{B.18})$$

Proof of Lemma B.4. Let η_1 denote a Poisson random variable with mean 1, independent of $\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \dots$, and set

$$\begin{aligned} \mathcal{Y}_{a,n}(\tilde{\mathbf{x}}) = \frac{1}{h_n^d \sqrt{k_{n,a}(\tilde{\mathbf{x}})}} & \left[\sum_{i \leq \eta_1} \left\{ \prod_{l=1}^p K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n} \right) \right. \right. \\ & - a \sum_{l=1}^p K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n} \right) \prod_{j \neq l} \mathbb{E} K_j \left(\frac{\mathbf{x}_j - \mathbb{X}_i^j}{h_n} \right) \Big\} \\ & \left. + (pa - 1) \prod_{l=1}^p \mathbb{E} K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n} \right) \right]. \end{aligned}$$

Observe that $\text{Var}[\mathcal{Y}_{a,n}(\tilde{\mathbf{x}})] = 1$. Using the statement (2.8) of Lemma 2.3 of [Giné, Mason and Zaitsev \(2003\)](#) and c_r -inequality, we get

$$\mathbb{E} |\mathcal{Y}_{a,n}(\tilde{\mathbf{x}})|^3 \leq D_3 \frac{h_n^{-\frac{3}{2}d} \prod_{l=1}^p \mathbb{E} K_l^3 \left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n} \right)}{(h_n^d k_{n,a}(\tilde{\mathbf{x}}))^{3/2}},$$

where $D_3 > 0$ is a constant. Using (B.9) with (B.16), we obtain, for all large enough n and for some constant $D_4 > 0$,

$$\sup_{\tilde{\mathbf{x}} \in \bar{C}} \mathbb{E} |\mathcal{Y}_{a,n}(\tilde{\mathbf{x}})|^3 \leq D_4 h_n^{-\frac{d}{2}}. \quad (\text{B.19})$$

Let $\mathcal{Y}_{a,n}^{(1)}(\tilde{\mathbf{x}}), \dots, \mathcal{Y}_{a,n}^{(n)}(\tilde{\mathbf{x}})$ be independent copies of $\mathcal{Y}_{a,n}(\tilde{\mathbf{x}})$. Clearly, we have

$$T_{\eta,a}(\tilde{\mathbf{x}}) := \frac{\sqrt{n} \Gamma_{\eta,a}(\tilde{\mathbf{x}})}{\sqrt{k_{n,a}(\tilde{\mathbf{x}})}} \stackrel{\mathcal{D}}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathcal{Y}_{a,n}^{(i)}(\tilde{\mathbf{x}}),$$

where $\stackrel{\mathcal{D}}{=}$ stands for equality in distribution. Therefore, by (C.1) we have

$$\sup_{\tilde{\mathbf{x}} \in \bar{C}} \left| \sqrt{n} \frac{\mathbb{E} |\Gamma_{\eta,a}(\tilde{\mathbf{x}})|}{\sqrt{k_{n,a}(\tilde{\mathbf{x}})}} - \mathbb{E}|Z_1| \right| \leq \frac{D_5}{\sqrt{n}} \sup_{\tilde{\mathbf{x}} \in \bar{C}} \mathbb{E} |\mathcal{Y}_{a,n}(\tilde{\mathbf{x}})|^3, \quad (\text{B.20})$$

where D_5 is a universal positive constant. Now, by (B.16), in combination with (B.19) and (B.20), we get

$$\left| \int_{\bar{C}} \left\{ \sqrt{n} \mathbb{E} |\Gamma_{\eta,a}(\tilde{\mathbf{x}})| - \mathbb{E}|Z_1| \sqrt{k_{n,a}(\tilde{\mathbf{x}})} \right\} d\tilde{\mathbf{x}} \right| = O \left(\frac{1}{\sqrt{n} h_n^d} \right).$$

Similarly, we obtain, by the statement (C.1),

$$\left| \int_{\bar{C}} \left\{ \sqrt{n} \mathbb{E} |\Gamma_{n,a}(\tilde{\mathbf{x}})| - \mathbb{E} |Z_1| \sqrt{\mathcal{L}_{n,a}(\tilde{\mathbf{x}})} \right\} d\tilde{\mathbf{x}} \right| = O\left(\frac{1}{\sqrt{n}h_n^d}\right),$$

which by (B.17) implies

$$\left| \int_{\bar{C}} \left\{ \sqrt{n} \mathbb{E} |\Gamma_{n,a}(\tilde{\mathbf{x}})| - \mathbb{E} |Z_1| \sqrt{k_{n,a}(\tilde{\mathbf{x}})} \right\} d\tilde{\mathbf{x}} \right| = O\left(\frac{1}{\sqrt{n}h_n^d} + \sqrt{h_n^d}\right).$$

This completes the proof. $\blacksquare$

Lemma B.5. *Whenever $h_n \rightarrow 0$ and, for $l = 1, \dots, p$, C_l satisfies (B.3) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$, we have*

$$\mathbb{1}_{\bar{C}}(\tilde{\mathbf{x}} + h_n \tilde{\mathbf{t}}) \text{ converges in measure to } \mathbb{1}_{\bar{C}}(\tilde{\mathbf{x}}) = 1 \text{ on } \bar{C} \times \tilde{\mathcal{B}}. \quad (\text{B.21})$$

Proof of Lemma B.5. Notice that

$$\int_{\bar{C}} \int_{\tilde{\mathcal{B}}} \mathbb{1}_{\bar{C}}(\tilde{\mathbf{x}} + h_n \tilde{\mathbf{t}}) d\tilde{\mathbf{t}} d\tilde{\mathbf{x}} = \frac{1}{h_n^d} \int_{\bar{C}} \int_{\tilde{\mathcal{B}}} \mathbb{1}_{\tilde{\mathcal{B}}}\left(\frac{\tilde{\mathbf{x}} - \tilde{\mathbf{y}}}{h_n}\right) d\tilde{\mathbf{y}} d\tilde{\mathbf{x}}.$$

Now, by (B.1), applied to $K(\tilde{\mathbf{t}}) = \frac{1}{\lambda(\tilde{\mathcal{B}})} \mathbb{1}_{\tilde{\mathcal{B}}}(\tilde{\mathbf{t}})$ and $f(\tilde{\mathbf{x}}) = \frac{1}{\lambda(\bar{C})} \mathbb{1}_{\bar{C}}(\tilde{\mathbf{x}})$, for almost every $\tilde{\mathbf{x}}$, we have

$$\frac{1}{h_n^d} \int_{\mathbb{R}^d} \mathbb{1}_{\bar{C}}(\tilde{\mathbf{y}}) \mathbb{1}_{\tilde{\mathcal{B}}}\left(\frac{\tilde{\mathbf{x}} - \tilde{\mathbf{y}}}{h_n}\right) d\tilde{\mathbf{y}} \rightarrow \lambda(\tilde{\mathcal{B}}) \mathbb{1}_{\bar{C}}(\tilde{\mathbf{x}}).$$

Thus, by the dominated convergence theorem, we infer that

$$\frac{1}{h_n^d} \int_{\bar{C}} \int_{\tilde{\mathcal{B}}} \mathbb{1}_{\tilde{\mathcal{B}}}\left(\frac{\tilde{\mathbf{x}} - \tilde{\mathbf{y}}}{h_n}\right) d\tilde{\mathbf{y}} d\tilde{\mathbf{x}} \rightarrow \lambda(\tilde{\mathcal{B}}) \lambda(\bar{C}),$$

which, in other words, says

$$\begin{aligned} & (\lambda \times \lambda) \{(\tilde{\mathbf{x}}, \tilde{\mathbf{t}}) \in \bar{C} \times \tilde{\mathcal{B}} : 1 - \mathbb{1}_{\bar{C}}(\tilde{\mathbf{x}} + h_n \tilde{\mathbf{t}}) \neq 0\} \\ &= \lambda(\tilde{\mathcal{B}}) \lambda(\bar{C}) - \int_{\bar{C}} \int_{\tilde{\mathcal{B}}} \mathbb{1}_{\bar{C}}(\tilde{\mathbf{x}} + h_n \tilde{\mathbf{t}}) d\tilde{\mathbf{t}} d\tilde{\mathbf{x}} \rightarrow 0, \end{aligned}$$

yielding to (B.21). Hence the proof of Lemma B.5 is complete. $\blacksquare$

Lemma B.6. *Suppose that, for each $l = 1, \dots, p$, C_l satisfies (B.3) and (B.4) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K}(\cdot) = \mathcal{K}_l(\cdot)$. If $h_n \rightarrow 0$ and $nh_n^{3d} \rightarrow \infty$, as $n \rightarrow \infty$, then, for all $a, b \in \{0, 1\}$, we have*

$$\lim_{n \rightarrow \infty} n \text{Cov}(U_{\eta,a}(\bar{C}), U_{\eta,b}(\bar{C})) = \sigma^2 \prod_{l=1}^p \int_{C_l} f_l(\mathbf{x}_l) d\mathbf{x}_l.$$

Proof of Lemma B.6. For any Borel subset A of $\mathbb{R}^{2d}$, set

$$\sigma_{n,a,b}(\bar{C}, A) := \int_{\bar{C}} \int_{\bar{C}} \text{Cov}(|T_{\eta,a}(\tilde{\mathbf{x}})|, |T_{\eta,b}(\tilde{\mathbf{y}})|) \mathbb{1}_A(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \sqrt{k_{n,a}(\tilde{\mathbf{x}})k_{n,b}(\tilde{\mathbf{y}})} d\tilde{\mathbf{x}} d\tilde{\mathbf{y}}.$$

Notice that

$$\begin{aligned} n\text{Cov}(U_{\eta,a}(\bar{C}), U_{\eta,b}(\bar{C})) &= n \int_{\bar{C}} \int_{\bar{C}} \text{Cov}(|\Gamma_{\eta,a}(\tilde{\mathbf{x}})|, |\Gamma_{\eta,b}(\tilde{\mathbf{y}})|) d\tilde{\mathbf{x}} d\tilde{\mathbf{y}} \\ &= \int_{\bar{C}} \int_{\bar{C}} \text{Cov}(|T_{\eta,a}(\tilde{\mathbf{x}})|, |T_{\eta,b}(\tilde{\mathbf{y}})|) \sqrt{k_{n,a}(\tilde{\mathbf{x}})k_{n,b}(\tilde{\mathbf{y}})} d\tilde{\mathbf{x}} d\tilde{\mathbf{y}} \\ &= \sum_{I \subset I_p} \sigma_{n,a,b}(\bar{C}, A_n(I)), \end{aligned}$$

where, for each $I \subset I_p = \{1, \dots, p\}$,

$$A_n(I) := \bigcap_{l \in I} \left\{ (\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) : \|\mathbf{x}_l - \mathbf{y}_l\| \leq h_n \right\} \bigcap \bigcap_{j \in I_p \setminus I} \left\{ (\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) : \|\mathbf{x}_j - \mathbf{y}_j\| > h_n \right\}.$$

Here, the empty intersection is defined to be $\mathbb{R}^{2d}$. We will show that, as $n \rightarrow \infty$,

$$\sigma_{n,a,b}(\bar{C}, A_n(I_p)) \rightarrow \sigma^2 \prod_{l=1}^p \int_{C_l} f_l(\mathbf{x}_l) d\mathbf{x}_l \quad (\text{B.22})$$

and

$$\sum_{I \subsetneq I_p} \sigma_{n,a,b}(\bar{C}, A_n(I)) \rightarrow 0, \quad (\text{B.23})$$

which will complete the proof of the lemma. First, consider (B.22). By (B.15), in combination with the fact that $\lambda(\bar{C}) < \infty$ and

$$\int_{\bar{C}} \int_{\bar{C}} \mathbb{1}_{A_n(I_p)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) d\tilde{\mathbf{x}} d\tilde{\mathbf{y}} \leq \beta_0 h_n^d \lambda(\bar{C}),$$

for some constant $\beta_0 > 0$, we see that

$$\sigma_{n,a,b}(\bar{C}, A_n(I_p)) = \bar{\sigma}_{n,a,b}(\bar{C}, A_n(I_p)) + o(1),$$

where

$$\begin{aligned} \bar{\sigma}_{n,a,b}(\bar{C}, A_n(I_p)) &= \tilde{\mathbf{K}} \int_{\bar{C}} \int_{\bar{C}} \mathbb{1}_{A_n(I_p)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \text{Cov}(|T_{\eta,a}(\tilde{\mathbf{x}})|, |T_{\eta,b}(\tilde{\mathbf{y}})|) \\ &\quad \times h_n^{-d} \prod_{l=1}^p \sqrt{f_l(\mathbf{x}_l) f_l(\mathbf{y}_l)} d\tilde{\mathbf{x}} d\tilde{\mathbf{y}}. \end{aligned}$$

Now, let $(Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}}), Z_{n,2}^{(a,b)}(\tilde{\mathbf{y}}))$, $\tilde{\mathbf{x}}, \tilde{\mathbf{y}} \in \mathbb{R}^d$, be a mean zero bivariate Gaussian process such that, for each $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \in \mathbb{R}^{2d}$, $(Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}}), Z_{n,2}^{(a,b)}(\tilde{\mathbf{y}}))$ and $(T_{\eta,a}(\tilde{\mathbf{x}}), T_{\eta,b}(\tilde{\mathbf{y}}))$ have the same covariance structure.

In particular, we have

$$(Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}}), Z_{n,2}^{(a,b)}(\tilde{\mathbf{y}})) \stackrel{\mathcal{D}}{=} \left(\sqrt{1 - (\rho_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}))^2} Z_1 + \rho_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) Z_2, Z_2 \right),$$

where

$$\begin{aligned} \rho_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) &:= \mathbb{E}[T_{\eta,a}(\tilde{\mathbf{x}})T_{\eta,b}(\tilde{\mathbf{y}})] \\ &= \left[\prod_{l=1}^p w_{n,l}(\mathbf{x}_l, \mathbf{y}_l) + p(p-1)ab \prod_{l=1}^p \mathbb{E}f_{n,l}(\mathbf{x}_l)\mathbb{E}f_{n,l}(\mathbf{y}_l) \right. \\ &\quad \left. + (ab - a - b) \sum_{l=1}^p w_{n,l}(\mathbf{x}_l, \mathbf{y}_l) \prod_{j \neq l} \mathbb{E}f_{n,j}(\mathbf{x}_j)\mathbb{E}f_{n,j}(\mathbf{y}_j) \right] \\ &\quad \times \frac{1}{\sqrt{k_{n,a}(\tilde{\mathbf{x}})k_{n,b}(\tilde{\mathbf{y}})}}, \end{aligned} \tag{B.24}$$

where, for each $l = 1, \dots, p$,

$$w_{n,l}(\mathbf{x}_l, \mathbf{y}_l) := h_n^{-2d_l} \mathbb{E} \left[K_l \left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n} \right) K_l \left(\frac{\mathbf{y}_l - \mathbb{Y}^l}{h_n} \right) \right].$$

Set

$$\begin{aligned} \bar{\tau}_{n,a,b}(\bar{C}, A_n(I_p)) &= \tilde{\mathbf{K}} \int_{\bar{C}} \int_{\bar{C}} \mathbb{1}_{A_n(I_p)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \text{Cov}(|Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}})|, |Z_{n,2}^{(a,b)}(\tilde{\mathbf{y}})|) \\ &\quad \times h_n^{-d} \prod_{l=1}^p \sqrt{f_l(\mathbf{x}_l)f_l(\mathbf{y}_l)} d\tilde{\mathbf{x}} d\tilde{\mathbf{y}}, \end{aligned}$$

which by the change of variables $\tilde{\mathbf{y}} = \tilde{\mathbf{x}} + h_n \tilde{\mathbf{t}}$ equals

$$\int_{\bar{C}} \int_{\bar{\mathcal{B}}} g_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{t}}) d\tilde{\mathbf{x}} d\tilde{\mathbf{t}},$$

where

$$\begin{aligned} g_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{t}}) &:= \tilde{\mathbf{K}} \mathbb{1}_{\bar{C}}(\tilde{\mathbf{x}}) \mathbb{1}_{\bar{C}}(\tilde{\mathbf{x}} + h_n \tilde{\mathbf{t}}) \text{Cov}(|Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}})|, |Z_{n,2}^{(a,b)}(\tilde{\mathbf{x}} + h_n \tilde{\mathbf{t}})|) \\ &\quad \times \prod_{l=1}^p \sqrt{f_l(\mathbf{x}_l)f_l(\mathbf{x}_l + h_n \mathbf{t}_l)}. \end{aligned}$$

We will show that, as $n \rightarrow \infty$,

$$\bar{\tau}_{n,a,b}(\bar{C}, A_n(I_p)) \rightarrow \sigma^2 \prod_{l=1}^p \int_{C_l} f_l(\mathbf{x}_l) d\mathbf{x}_l \tag{B.25}$$

and then, as $n \rightarrow \infty$, we have

$$\bar{\tau}_{n,a,b}(\bar{C}, A_n(I_p)) - \bar{\sigma}_{n,a,b}(\bar{C}, A_n(I_p)) \rightarrow 0, \tag{B.26}$$

which will complete the proof of (B.22). Now, consider (B.25). Applying (B.1) of Lemma B.1, with $g(\cdot) = f_l(\cdot)$ and $K(\cdot) = K_l(\cdot + \mathbf{t}_l)K_l(\cdot + \mathbf{s}_l)$, $t_l, s_l \in \mathbb{R}^{d_l}$, we get, for each $(\mathbf{t}_l, \mathbf{s}_l)$, as $n \rightarrow \infty$, for almost every $\mathbf{x}_l \in \mathbb{R}^{d_l}$, hence for almost every $\mathbf{x}_l \in C_l$,

$$\begin{aligned} h_n^{-d_l} \mathbb{E} \left[K_l \left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n} + \mathbf{t}_l \right) K_l \left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n} + \mathbf{s}_l \right) \right] \\ \rightarrow f_l(\mathbf{x}_l) \int K_l(\mathbf{u}_l + \mathbf{t}_l) K_l(\mathbf{u}_l + \mathbf{s}_l) d\mathbf{u}_l. \end{aligned} \quad (\text{B.27})$$

Moreover, we get with $g(\cdot) = f_l(\cdot)$ and $K(\cdot) = K_l(\cdot + \mathbf{t}_l)$, as $n \rightarrow \infty$, for almost every $\mathbf{x}_l \in C_l$,

$$h_n^{-d_l} \mathbb{E} \left[K_l \left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n} + \mathbf{t}_l \right) \right] \rightarrow f_l(\mathbf{x}_l). \quad (\text{B.28})$$

Now, by (B.12)-(B.14) in combination with (B.27) and (B.28), we get, for each $\tilde{\mathbf{t}}$, as $n \rightarrow \infty$, for almost every $\tilde{\mathbf{x}} \in \tilde{C}$,

$$h_n^d k_{n,a}(\tilde{\mathbf{x}} + h_n \tilde{\mathbf{t}}) \rightarrow \tilde{\mathbf{K}} \prod_{l=1}^p f_l(\mathbf{x}_l). \quad (\text{B.29})$$

Thus, by (B.24), (B.27) and (B.29), we have, for each $\tilde{\mathbf{t}}$ and almost every $\tilde{\mathbf{x}} \in \tilde{C}$, as $n \rightarrow \infty$,

$$\rho_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{x}} + h_n \tilde{\mathbf{t}}) \rightarrow \prod_{l=1}^p \rho_l(\mathbf{t}_l) = \rho(\tilde{\mathbf{t}}),$$

thus we obtain

$$\text{Cov} \left(|Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}})|, |Z_{n,2}^{(a,b)}(\tilde{\mathbf{x}} + h_n \tilde{\mathbf{t}})| \right) \rightarrow \text{Cov} \left(\left| \sqrt{1 - \rho^2(\tilde{\mathbf{t}})} Z_1 + \rho(\tilde{\mathbf{t}}) Z_2 \right|, |Z_2| \right).$$

Combining this with Lemma B.5 and the continuity of $f_l(\cdot)$ on C_l , $l = 1, \dots, p$, we readily conclude that $g_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{t}})$ converges in measure on $\tilde{C} \times \tilde{\mathcal{B}}$ to

$$\tilde{\mathbf{K}} \mathbb{1}_{\tilde{C}}(\tilde{\mathbf{x}}) \text{Cov} \left(\left| \sqrt{1 - \rho^2(\tilde{\mathbf{t}})} Z_1 + \rho(\tilde{\mathbf{t}}) Z_2 \right|, |Z_2| \right) \prod_{l=1}^p f_l(\mathbf{x}_l).$$

Since, for each $l = 1, \dots, p$, $f_l(\cdot)$ is bounded on C_l , the function $g_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{t}})$ is for all $n \geq 1$ uniformly bounded on $\tilde{C} \times \tilde{\mathcal{B}}$. Thus, we get by the Lebesgue bounded convergence theorem, as $n \rightarrow \infty$, $\bar{\tau}_{a,b,n}(\tilde{C}, A_n(I_p))$ converges to

$$\sigma^2 \prod_{l=1}^p \int_{C_l} f_l(\mathbf{x}_l) d\mathbf{x}_l.$$

This completes the proof of (B.25). Now, we consider the relation (B.26). Set

$$G_n(\tilde{\mathbf{x}}, \tilde{\mathbf{t}}) := \tilde{\mathbf{K}} \mathbb{1}_{\tilde{C}}(\tilde{\mathbf{x}}) \mathbb{1}_{\tilde{C}}(\tilde{\mathbf{x}} + h_n \tilde{\mathbf{t}}) \prod_{l=1}^p \sqrt{f_l(\mathbf{x}_l) f_l(\mathbf{x}_l + h_n \mathbf{t}_l)}.$$

Notice that, since, for each $l = 1, \dots, p$, $f_l(\cdot)$ is bounded on C_l ,

$$\int_{\bar{C}} \int_{\bar{\mathcal{B}}} G_n(\tilde{\mathbf{x}}, \tilde{\mathbf{t}}) d\tilde{\mathbf{x}} d\tilde{\mathbf{t}} \leq \beta_1, \quad (\text{B.30})$$

where β_1 is a positive constant. We see that

$$\begin{aligned} & |\bar{\tau}_{n,a,b}(\bar{C}, A_n(I_p)) - \bar{\sigma}_{n,a,b}(\bar{C}, A_n(I_p))| \\ & \leq \int_{\bar{C}} \int_{\bar{\mathcal{B}}} |\mathbb{E}|Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}})|\mathbb{E}|Z_{n,2}^{(a,b)}(\tilde{\mathbf{x}} + h_n\tilde{\mathbf{t}})| - \mathbb{E}|T_{\eta,a}(\tilde{\mathbf{x}})|\mathbb{E}|T_{\eta,b}(\tilde{\mathbf{x}} + h_n\tilde{\mathbf{t}})||G_n(\tilde{\mathbf{x}}, \tilde{\mathbf{t}}) d\tilde{\mathbf{x}} d\tilde{\mathbf{t}} \\ & \quad + \int_{\bar{C}} \int_{\bar{\mathcal{B}}} |\mathbb{E}|Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}})Z_{n,2}^{(a,b)}(\tilde{\mathbf{x}} + h_n\tilde{\mathbf{t}})| - \mathbb{E}|T_{\eta,a}(\tilde{\mathbf{x}})T_{\eta,b}(\tilde{\mathbf{x}} + h_n\tilde{\mathbf{t}})||G_n(\tilde{\mathbf{x}}, \tilde{\mathbf{t}}) d\tilde{\mathbf{x}} d\tilde{\mathbf{t}} \\ & := \zeta_{n,a,b}(1) + \zeta_{n,a,b}(2). \end{aligned}$$

First, using (B.30) and again the statement (C.1) with (B.19), we obtain

$$\zeta_{n,a,b}(1) = O\left(\frac{1}{\sqrt{nh_n^d}}\right).$$

Choose any $0 < \varepsilon < 1$ and set

$$E_n(\varepsilon) = \{(\tilde{\mathbf{x}}, \tilde{\mathbf{t}}) : 1 - (\rho_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{x}} + h_n\tilde{\mathbf{t}}))^2 \geq \varepsilon\}.$$

Now, we infer that

$$\begin{aligned} \zeta_{n,a,b}(2) & \leq \int_{\bar{C}} \int_{\bar{\mathcal{B}}} |1 - \mathbb{E}|Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}})Z_{n,2}^{(a,b)}(\tilde{\mathbf{x}} + h_n\tilde{\mathbf{t}})||\mathbb{1}_{E_n^c(\varepsilon)}(\tilde{\mathbf{x}}, \tilde{\mathbf{t}})G_n(\tilde{\mathbf{x}}, \tilde{\mathbf{t}}) d\tilde{\mathbf{x}} d\tilde{\mathbf{t}} \\ & \quad + \int_{\bar{C}} \int_{\bar{\mathcal{B}}} |1 - \mathbb{E}|T_{\eta,a}(\tilde{\mathbf{x}})T_{\eta,b}(\tilde{\mathbf{x}} + h_n\tilde{\mathbf{t}})||\mathbb{1}_{E_n^c(\varepsilon)}(\tilde{\mathbf{x}}, \tilde{\mathbf{t}})G_n(\tilde{\mathbf{x}}, \tilde{\mathbf{t}}) d\tilde{\mathbf{x}} d\tilde{\mathbf{t}} \\ & \quad + \int_{\bar{C}} \int_{\bar{\mathcal{B}}} |\mathbb{E}|Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}})Z_{n,2}^{(a,b)}(\tilde{\mathbf{x}} + h_n\tilde{\mathbf{t}})| - \mathbb{E}|T_{\eta,a}(\tilde{\mathbf{x}})T_{\eta,b}(\tilde{\mathbf{x}} + h_n\tilde{\mathbf{t}})|| \\ & \quad \times \mathbb{1}_{E_n(\varepsilon)}(\tilde{\mathbf{x}}, \tilde{\mathbf{t}})G_n(\tilde{\mathbf{x}}, \tilde{\mathbf{t}}) d\tilde{\mathbf{x}} d\tilde{\mathbf{t}} \\ & := \bar{\zeta}_{n,1}^{(a,b)}(2, \varepsilon) + \bar{\zeta}_{n,2}^{(a,b)}(2, \varepsilon) + \zeta_n^{(a,b)}(2, \varepsilon) \\ & := \bar{\zeta}_n^{(a,b)}(2, \varepsilon) + \zeta_n^{(a,b)}(2, \varepsilon). \end{aligned}$$

To bound $\bar{\zeta}_n^{(a,b)}(2, \varepsilon)$, we use the elementary fact that if X and Y are mean-zero and variance-one random variables with $\rho = \mathbb{E}(XY)$, then $1 - \mathbb{E}|XY| \leq 1 - |\rho| \leq 1 - \rho^2$, in combination with (B.30), to get

$$\bar{\zeta}_n^{(a,b)}(2, \varepsilon) \leq 2\varepsilon\beta_1.$$

Next, using the inequality (B.30) and the statement (C.2) along with the statement (B.19), we obtain

$$\zeta_n^{(a,b)}(2, \varepsilon) = O\left(\frac{1}{\sqrt{nh_n^d}}\right).$$

Thus, for any $0 < \varepsilon < 1$, we have

$$\limsup_{n \rightarrow \infty} |\bar{\tau}_{n,a,b}(\bar{C}, A_n(I_p)) - \bar{\sigma}_{n,a,b}(\bar{C}, A_n(I_p))| \leq 2\varepsilon\beta_1,$$

which yields the statement (B.26). This finishes the proof of (B.22). Now we turn to (B.23). Let us show that, for each $I \subsetneq I_p$, as $n \rightarrow \infty$,

$$\sigma_{n,a,b}(\bar{C}, A_n(I)) \rightarrow 0. \quad (\text{B.31})$$

Set, for each $I \subsetneq I_p$,

$$\begin{aligned} \tau_{n,a,b}(\bar{C}, A_n(I)) &:= \int_{\bar{C}} \int_{\bar{C}} \text{Cov}(|Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}})|, |Z_{n,2}^{(a,b)}(\tilde{\mathbf{y}})|) \mathbb{1}_{A_n(I)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \\ &\quad \times \sqrt{k_{n,a}(\tilde{\mathbf{x}})k_{n,b}(\tilde{\mathbf{y}})} d\tilde{\mathbf{x}} d\tilde{\mathbf{y}}. \end{aligned}$$

We will show that, for each $I \subsetneq I_p$, as $n \rightarrow \infty$,

$$\tau_{n,a,b}(\bar{C}, A_n(I)) \rightarrow 0 \quad (\text{B.32})$$

and then, as $n \rightarrow \infty$,

$$\tau_{n,a,b}(\bar{C}, A_n(I)) - \sigma_{n,a,b}(\bar{C}, A_n(I)) \rightarrow 0, \quad (\text{B.33})$$

which completes the proof of (B.31). Notice that, for each $I \subsetneq I_p$ and each $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \in A_n(I)$,

$$\begin{aligned} \rho_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) &= \left[p(p-1)ab \prod_{l=1}^p \mathbb{E}f_{n,l}(\mathbf{x}_l) \mathbb{E}f_{n,l}(\mathbf{y}_l) \right. \\ &\quad \left. + (ab - a - b) \sum_{l \in I} w_{n,l}(\mathbf{x}_l, \mathbf{y}_l) \times \prod_{j \neq l} \mathbb{E}f_{n,j}(\mathbf{x}_j) \mathbb{E}f_{n,j}(\mathbf{y}_j) \right] / \sqrt{k_{n,a}(\tilde{\mathbf{x}})k_{n,b}(\tilde{\mathbf{y}})}. \end{aligned}$$

Therefore, by (B.9), (B.16) and the fact that, for each $l = 1, \dots, p$, K_l is bounded, we get, for each $I \subsetneq I_p$, for all large enough n uniformly in $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \in A_n(I) \cap \bar{C} \times \bar{C}$ and for some constant $\beta_2 > 0$,

$$|\rho_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})| \leq \beta_2 \prod_{l \in I_p \setminus I} h_n^{d_l}. \quad (\text{B.34})$$

Now, by using Nabeya (1951) formulas, we have

$$\begin{aligned} \text{Cov}\left(|Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}})|, |Z_{n,2}^{(a,b)}(\tilde{\mathbf{y}})|\right) &= \text{Cov}\left(|\sqrt{1 - (\rho_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}))^2} Z_1 + \rho_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) Z_2|, |Z_2|\right) \\ &:= \varphi(\rho_{n,a,b}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})), \end{aligned}$$

where

$$\varphi(\rho) = \frac{2}{\pi} \left(\rho \arcsin \rho + \sqrt{1 - \rho^2} - 1 \right), \quad \rho \in [-1, 1],$$

which in turn, for each $\rho \in [-1, 1]$, is less than or equal to $\frac{(\pi-2)\rho^2}{\pi}$. Thus, for each $I \subsetneq I_p$, for all large enough n uniformly in $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \in A_n(I) \cap \bar{C} \times \bar{C}$ and for some constant $\beta_3 > 0$, by (B.34) we get

$$\text{Cov}(|Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})|, |Z_{n,2}^{(a,b)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})|) \leq \beta_3 \prod_{l \in I_p \setminus I} h_n^{2d_l}.$$

By change of variables $\mathbf{y}_l = \mathbf{x}_l + h_n \mathbf{t}_l$ for each $l \in I$, for some constant $\beta_4 > 0$, we obtain

$$\int_{\bar{C}} \int_{\bar{C}} \mathbb{1}_{A_n(I)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) d\tilde{\mathbf{x}} d\tilde{\mathbf{y}} \leq \beta_4 \prod_{l \in I} h_n^{d_l}.$$

Hence, by (B.16) we have

$$\int_{\bar{C}} \int_{\bar{C}} \mathbb{1}_{A_n(I)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \sqrt{k_{n,a}(\tilde{\mathbf{x}})k_{n,b}(\tilde{\mathbf{y}})} d\tilde{\mathbf{x}} d\tilde{\mathbf{y}} \leq \beta_5 \prod_{l \in I_p \setminus I} h_n^{-d_l} \quad (\text{B.35})$$

for some $\beta_5 > 0$. This completes the proof of (B.32). We next evaluate the difference in (B.33). We can see that

$$\begin{aligned} & |\tau_{n,a,b}(\bar{C}, A_n(I)) - \sigma_{n,a,b}(\bar{C}, A_n(I))| \\ & \leq \int_{\bar{C}} \int_{\bar{C}} |\mathbb{E}|Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}})|\mathbb{E}|Z_{n,2}^{(a,b)}(\tilde{\mathbf{y}})| - \mathbb{E}|T_{\eta,a}(\tilde{\mathbf{x}})|\mathbb{E}|T_{\eta,b}(\tilde{\mathbf{y}})|| \\ & \quad \times \mathbb{1}_{A_n(I)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \sqrt{k_{n,a}(\tilde{\mathbf{x}})k_{n,b}(\tilde{\mathbf{y}})} d\tilde{\mathbf{x}} d\tilde{\mathbf{y}} \\ & \quad + \int_{\bar{C}} \int_{\bar{C}} |\mathbb{E}|Z_{n,1}^{(a,b)}(\tilde{\mathbf{x}})Z_{n,2}^{(a,b)}(\tilde{\mathbf{y}})| - \mathbb{E}|T_{\eta,a}(\tilde{\mathbf{x}})T_{\eta,b}(\tilde{\mathbf{y}})|| \\ & \quad \times \mathbb{1}_{A_n(I)}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \sqrt{k_{n,a}(\tilde{\mathbf{x}})k_{n,b}(\tilde{\mathbf{y}})} d\tilde{\mathbf{x}} d\tilde{\mathbf{y}} \\ & := \xi_{n,a,b}(1) + \xi_{n,a,b}(2). \end{aligned}$$

To bound $\xi_{n,a,b}(1)$, we use (B.35) and (C.1) in combination with (B.19), to get

$$\xi_{n,a,b}(1) = O\left(\frac{1}{\sqrt{nh_n^{3d}}}\right).$$

Next, we use (B.35) and (C.2) with (B.19) and (B.34) to get

$$\xi_{n,a,b}(2) = O\left(\frac{1}{\sqrt{nh_n^{3d}}}\right),$$

which gives (B.33). This completes the proof of Lemma B.6. ■

In the proof of the next lemmas, we shall apply Lemma B.6 with $a = b = 0$. Note that in this situation, whenever $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \in A_n(I)$ with $I \subsetneq I_p$, the random variables $|\Gamma_{\eta,0}(\tilde{\mathbf{x}})|$ and $|\Gamma_{\eta,0}(\tilde{\mathbf{y}})|$ are independent.

This follows from the fact that being functions of independent increments of a Poisson process. Therefore, $\sigma_{n,0,0}(\bar{C}, A_n(I)) = 0$ for all n and each $I \subsetneq I_p$, which implies that

$$\lim_{n \rightarrow \infty} n \text{Var}(U_{n,0}(\bar{C})) = \sigma^2 \prod_{l=1}^p \int_{C_l} f_l(\mathbf{x}_l) d\mathbf{x}_l \quad (\text{B.36})$$

can be obtained under the bandwidth condition $nh_n^d \rightarrow \infty$ instead of $nh_n^{3d} \rightarrow \infty$ as $n \rightarrow \infty$.

Lemma B.7. *Suppose that, for each $l = 1, \dots, p$, C_l satisfies (B.2)-(B.3)-(B.4) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K}(\cdot) = \mathcal{K}_l(\cdot)$. If $h_n \rightarrow 0$ and $nh_n^{3d} \rightarrow \infty$, then we have*

$$\sqrt{n} [U_{n,1}(\bar{C}) - U_{n,0}(\bar{C}) - \mathbb{E}(U_{n,1}(\bar{C}) - U_{n,0}(\bar{C}))] \xrightarrow{\mathcal{P}} 0.$$

Proof of Lemma B.7. For each $l = 1, \dots, p$, we can find a measurable partition $C_{l,1}, \dots, C_{l,k_l}$ of C_l so that

$$0 < \int_{C_{l,j_l}^{h/2}} f_l(\mathbf{u}_l) d\mathbf{u}_l < 1/(2p),$$

for each $j_l = 1, \dots, k_l$ and all $h_n > 0$ small enough, where $C_{l,j_l}^{h/2}$ is the $h/2$ -neighborhood of C_{l,j_l} . We will apply Lemma 2.1 of [Giné, Mason and Zaitsev \(2003\)](#) to the semigroup D generated by the point masses,

$$D = \left\{ 0, \sum_{i=1}^n \delta_{\tilde{\mathbf{x}}_i} : n \in \mathbb{N}, \tilde{\mathbf{x}}_i \in \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_p} \right\},$$

with the σ -algebra $\mathcal{D}$ generated by the functions

$$f_{n,B_{j_1}, \dots, j_p}(\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n) = \sum_{i=1}^n \mathbb{1}_{\{\delta_{\tilde{\mathbf{x}}_i} \in B_{j_1}, \dots, j_p\}} \delta_{\tilde{\mathbf{x}}_i}, n \in \mathbb{N},$$

where, for each $(j_1, \dots, j_p) \in \{1, \dots, k_1\} \times \dots \times \{1, \dots, k_p\} =: J_{k_1, \dots, k_p}$,

$$B_{j_1, \dots, j_p} = \left\{ \delta_{\tilde{\mathbf{x}}} : \tilde{\mathbf{x}} \in \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_{l-1}} \times C_{l,j_l}^{h/2} \times \mathbb{R}^{d_{l+1}} \times \dots \times \mathbb{R}^{d_p}, \text{ for some } l = 1, \dots, p \right\}.$$

It is easy to see that, for any measurable function $G : \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_p} \mapsto \mathbb{R}$, the map $\mu \mapsto \int G d\mu$ is $\mathcal{D}$ -measurable [just note that

$$f_{n,B_{j_1}, \dots, j_p}^{-1} \left\{ \mu \in D, \int G d\mu \leq t \right\} = \left\{ (\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n) : \sum_{i=1}^n \mathbb{1}_{\{\delta_{\tilde{\mathbf{x}}_i} \in B_{j_1}, \dots, j_p\}} G(\tilde{\mathbf{x}}_i) \leq t \right\}$$

is measurable set of $(\mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_p})^n$]. The function $H = H_{n,j_1, \dots, j_p}$ has the form

$$\begin{aligned} & H_{n,j_1, \dots, j_p} \left(\sum_{i=1}^n \mathbb{1}_{\{\delta_{\tilde{\mathbf{x}}_i} \in B_{j_1}, \dots, j_p\}} \delta_{\tilde{\mathbf{x}}_i} \right) \\ &= \left\{ \bar{H}_{n,j_1, \dots, j_p}, \mathbf{1} \left(\sum_{i=1}^n \mathbb{1}_{\{\delta_{\tilde{\mathbf{x}}_i} \in B_{j_1}, \dots, j_p\}} \delta_{\tilde{\mathbf{x}}_i} \right) - \bar{H}_{n,j_1, \dots, j_p}, \mathbf{0} \left(\sum_{i=1}^n \mathbb{1}_{\{\delta_{\tilde{\mathbf{x}}_i} \in B_{j_1}, \dots, j_p\}} \delta_{\tilde{\mathbf{x}}_i} \right) \right\} \end{aligned}$$

$$-nh_n^d \mathbb{E} \left(U_{\eta,1}(\bar{C}_{j_1,\dots,j_p}) - U_{\eta,0}(\bar{C}_{j_1,\dots,j_p}) \right) \Bigg\}^2,$$

where $\bar{C}_{j_1,\dots,j_p} = C_{1,j_1} \times \dots \times C_{p,j_p}$ and, for each $a \in \{0, 1\}$, we can therefore write

$$\begin{aligned} & \bar{H}_{n,j_1,\dots,j_p,a} \left(\sum_{i=1}^n \mathbb{1}_{\{\delta_{\mathbf{x}_i} \in B_{j_1,\dots,j_p}\}} \delta_{\mathbf{x}_i} \right) \\ &= \int_{\bar{C}_{j_1,\dots,j_p}} \left| \sum_{i=1}^n \left\{ \prod_{l=1}^p K_l \left(\frac{\mathbf{x}_l - \mathbf{x}_{i,l}}{h_n} \right) - a \sum_{l=1}^p K_l \left(\frac{\mathbf{x}_l - \mathbf{x}_{i,l}}{h_n} \right) \prod_{j \neq i} \mathbb{E} K_j \left(\frac{\mathbf{x}_j - \mathbb{X}^j}{h_n} \right) \right\} \right. \\ & \quad \left. \times \mathbb{1}_{\{\delta_{\mathbf{x}_i} \in B_{j_1,\dots,j_p}\}} + n(pa-1) \prod_{l=1}^p \mathbb{E} K_l \left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n} \right) \right| d\mathbf{x}, \end{aligned}$$

where we recall $\mathbf{x}_i = (x_{i,1}^\top, \dots, x_{i,p}^\top)^\top \in \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_p}$. Now, $H_{n,j_1,\dots,j_p}$ when considered as a function on $(\mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_p})^n$ can be shown to be Borel measurable, which in this setup is equivalent to be $\mathcal{D}$ -measurable. We now get from Lemma 2.1 of [Giné, Mason and Zaitsev \(2003\)](#), for each $(j_1, \dots, j_p) \in J_{k_1,\dots,k_p}$,

$$\mathbb{E} H_{n,j_1,\dots,j_p} \left(\sum_{i=1}^n \mathbb{1}_{\{\delta_{\mathbf{x}_i} \in B_{j_1,\dots,j_p}\}} \delta_{\mathbf{x}_i} \right) \leq 2 \mathbb{E} H_{n,j_1,\dots,j_p} \left(\sum_{i=1}^{\eta} \mathbb{1}_{\{\delta_{\mathbf{x}_i} \in B_{j_1,\dots,j_p}\}} \delta_{\mathbf{x}_i} \right),$$

with η is a Poisson random variable with mean n , independent of $\tilde{\mathbf{X}}, \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2 \dots$. Therefore, by assumption (A_1) , for each $(j_1, \dots, j_p) \in J_{k_1,\dots,k_p}$, we have

$$\begin{aligned} & n \mathbb{E} \left[\left\{ U_{n,1}(\bar{C}_{j_1,\dots,j_p}) - U_{n,0}(\bar{C}_{j_1,\dots,j_p}) - \mathbb{E}(U_{\eta,1}(\bar{C}_{j_1,\dots,j_p}) - U_{\eta,0}(\bar{C}_{j_1,\dots,j_p})) \right\}^2 \right] \\ & \leq 2n \text{Var} \left(U_{\eta,1}(\bar{C}_{j_1,\dots,j_p}) - U_{\eta,0}(\bar{C}_{j_1,\dots,j_p}) \right). \end{aligned}$$

Using Lemma B.6 with $\bar{C} = \bar{C}_{j_1,\dots,j_p}$, as $n \rightarrow \infty$, we get

$$n \text{Var} \left(U_{\eta,1}(\bar{C}_{j_1,\dots,j_p}) - U_{\eta,0}(\bar{C}_{j_1,\dots,j_p}) \right) \rightarrow 0$$

and thus, for each $(j_1, \dots, j_p) \in J_{k_1,\dots,k_p}$,

$$\sqrt{n} \left[U_{n,1}(\bar{C}_{j_1,\dots,j_p}) - U_{n,0}(\bar{C}_{j_1,\dots,j_p}) - \mathbb{E}(U_{\eta,1}(\bar{C}_{j_1,\dots,j_p}) - U_{\eta,0}(\bar{C}_{j_1,\dots,j_p})) \right] \xrightarrow{\mathcal{P}} 0$$

which, when coupled with (B.17), (B.18) and, for each $l = 1, \dots, p$, $C_{l,1}, \dots, C_{l,k_l}$ being a partition of C_l , completes the proof of Lemma B.7. $\blacksquare$

Suppose that, for each $l = 1, \dots, p$, C_l satisfies Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K}(\cdot) = \mathcal{K}_l(\cdot)$. For all $0 < \varepsilon < 1$ and each $l = 1, \dots, p$, let M_l, ν_l and μ_l be the numbers from (B.2), namely

$$C_l \subset [-M_l + \nu_l, M_l - \nu_l]^{d_l}, \quad 0 < \mu_l = \int_{\mathbb{R}^{d_l} \setminus [-M_l, M_l]^{d_l}} f_l(\mathbf{u}_l) d\mathbf{u}_l < 1$$

and

$$\int_{C_l} f_l(\mathbf{u}_l) d\mathbf{u}_l > 1 - \varepsilon.$$

Assume that n is large enough such that $h_n \leq \min(v_1, \dots, v_p, M_1/2, \dots, M_p/2)$. Define, for $l = 1, \dots, p$, $m_{n,l} = \lfloor M_l/h_n \rfloor$, where $\lfloor u \rfloor$ denotes the integer part, i.e., $\lfloor u \rfloor \leq u < \lfloor u \rfloor + 1$, and $h_{n,l} = M_l/m_{n,l}$. For $l = 1, \dots, p$, we have $M_l/(2h_n) \leq m_{n,l} \leq M_l/h_n$ and thus $h_n \leq h_{n,l} \leq 2h_n$. Next, consider the regular grid given by

$$\begin{aligned} A_{\mathbf{i}} &= [x_{1,i_1}, x_{1,i_1+1}] \times \cdots \times [x_{1,i_d}, x_{1,i_d+1}] \\ &\quad \times \cdots \times [x_{p,i_d-d_{p+1}}, x_{p,i_d-d_{p+1}+1}] \times \cdots \times [x_{p,i_d}, x_{p,i_d+1}], \end{aligned}$$

where $\mathbf{i} = (i_1, \dots, i_d) \in \mathbb{Z}^d$ and, for all $i \in \mathbb{Z}$, $x_{l,i} = ih_{n,l}$. Define

$$R_{\mathbf{i}} = A_{\mathbf{i}} \cap \mathcal{M},$$

where $\mathcal{M} = [-M_1, M_1]^{d_1} \times \cdots \times [-M_p, M_p]^{d_p}$. With $l_n = \{\mathbf{i} : R_{\mathbf{i}} \neq \emptyset\}$, we see that $\{R_{\mathbf{i}} : \mathbf{i} \in l_n \subset \mathbb{Z}^d\}$ constitutes a partition of $\mathcal{M}$ such that, for each $\mathbf{i} \in l_n$,

$$\lambda(R_{\mathbf{i}}) \leq (2h_n)^d \quad \text{and} \quad |l_n| \leq 2^d h_n^{-d} \prod_{l=1}^p M_l^{d_l}. \quad (\text{B.37})$$

Set, for any $\mathbf{i} \in l_n$,

$$\mathcal{X}_{\mathbf{i},n} := \frac{\int_{R_{\mathbf{i}} \cap \bar{C}} \Delta_n(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}}{\sqrt{n \text{Var}(U_{\eta,0}(\bar{C}))}},$$

where

$$\Delta_n(\tilde{\mathbf{x}}) := \sqrt{n} \left[|\Gamma_{\eta,0}(\tilde{\mathbf{x}})| - \mathbb{E}|\Gamma_{\eta,0}(\tilde{\mathbf{x}})| \right].$$

Lemma B.8. *Whenever $h_n \rightarrow 0$, $nh_n^d \rightarrow \infty$ and, for each $l = 1, \dots, p$, C_l satisfies (B.2)-(B.3)-(B.4) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K}(\cdot) = \mathcal{K}_l(\cdot)$, there exists a constant γ_0 such that, uniformly in $\mathbf{i}$ and for all n sufficiently large,*

$$\mathbb{E}|\mathcal{X}_{\mathbf{i},n}|^3 \leq \gamma_0 h_n^{3d/2}. \quad (\text{B.38})$$

Proof of Lemma B.8. Observe that

$$\left(n \text{Var}(U_{\eta,0}(\bar{C})) \right)^{3/2} \mathbb{E}|\mathcal{X}_{\mathbf{i},n}|^3 \leq \int_{I_{\mathbf{i},n}} \mathbb{E}|\Delta_n(\tilde{\mathbf{x}}) \Delta_n(\tilde{\mathbf{x}}_1) \Delta_n(\tilde{\mathbf{x}}_2)| d\tilde{\mathbf{x}} d\tilde{\mathbf{x}}_1 d\tilde{\mathbf{x}}_2,$$

where $I_{\mathbf{i},n} = (R_{\mathbf{i}} \cap \bar{C})^3$. It is obvious to see that

$$\mathbb{E}|\Delta_n(\tilde{\mathbf{x}}) \Delta_n(\tilde{\mathbf{x}}_1) \Delta_n(\tilde{\mathbf{x}}_2)| \leq \mathbb{E} \left[|\Delta_n(\tilde{\mathbf{x}})| + |\Delta_n(\tilde{\mathbf{x}}_1)| + |\Delta_n(\tilde{\mathbf{x}}_2)| \right]^3,$$

which by c_r -inequality is, for some constant $\gamma_1 > 0$, less than or equal to

$$\gamma_1 n^{3/2} \left[\mathbb{E} |f_\eta(\tilde{\mathbf{x}}) - \mathbb{E} f_n(\tilde{\mathbf{x}})|^3 + \mathbb{E} |f_\eta(\tilde{\mathbf{x}}_1) - \mathbb{E} f_n(\tilde{\mathbf{x}}_1)|^3 + \mathbb{E} |f_\eta(\tilde{\mathbf{x}}_2) - \mathbb{E} f_n(\tilde{\mathbf{x}}_2)|^3 \right].$$

By Lemma 2.3 of [Giné, Mason and Zaitsev \(2003\)](#), we have, for any $\tilde{\mathbf{x}} \in \bar{C}$,

$$\begin{aligned} & n^{3/2} \mathbb{E} |f_\eta(\tilde{\mathbf{x}}) - \mathbb{E} f_n(\tilde{\mathbf{x}})|^3 \\ & \leq \left(\frac{45}{\log 3} \right)^3 \max \left\{ \left[\frac{1}{h_n^{2d}} \prod_{l=1}^p \mathbb{E} K_l^2 \left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n} \right) \right]^{3/2}, \frac{1}{\sqrt{n} h_n^{3d}} \prod_{l=1}^p \mathbb{E} K_l^3 \left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n} \right) \right\}, \end{aligned}$$

which by (B.9) is, for n large enough and uniformly in $\tilde{\mathbf{x}} \in \bar{C}$, is bounded by

$$\gamma_2 \left[\frac{1}{h_n^{3d/2}} + \frac{1}{\sqrt{n} h_n^{2d}} \right],$$

where γ_2 is some positive constant. Thus, uniformly in $\tilde{\mathbf{x}} \in \bar{C}$ and for n large enough, we have

$$\mathbb{E} |\Delta_n(\tilde{\mathbf{x}}) \Delta_n(\tilde{\mathbf{x}}_1) \Delta_n(\tilde{\mathbf{x}}_2)| \leq 3\gamma_1 \gamma_2 \left[\frac{1}{h_n^{3d/2}} + \frac{1}{\sqrt{n} h_n^{2d}} \right].$$

This implies that, for some constant $\gamma_3 > 0$, uniformly in $\mathbf{i}$ and for all large enough n ,

$$\begin{aligned} \int_{I_{i,n}} \mathbb{E} |\Delta_n(\tilde{\mathbf{x}}) \Delta_n(\tilde{\mathbf{x}}_1) \Delta_n(\tilde{\mathbf{x}}_2)| d\tilde{\mathbf{x}} d\tilde{\mathbf{x}}_1 d\tilde{\mathbf{x}}_2 & \leq 3(2^{3d}) \gamma_1 \gamma_2 \left[h_n^{3d/2} + \frac{h_n^d}{\sqrt{n}} \right] \\ & \leq \gamma_3 h_n^{3d/2}. \end{aligned}$$

Now, (B.38) follows from (B.36). ■

Recall that η is a Poisson random variable with mean n , independent of $\tilde{\mathbf{X}}, \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \dots$. Define

$$S_n = \sum_{\mathbf{i} \in I_n} X_{\mathbf{i},n},$$

$$\mathcal{U}_n = \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{\eta} \mathbb{1}_{\{\tilde{\mathbf{x}}_i \in \mathcal{M}\}} - n\mathbb{P}(\tilde{\mathbf{X}} \in \mathcal{M}) \right]$$

and

$$\mathcal{V}_n = \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{\eta} \mathbb{1}_{\{\tilde{\mathbf{x}}_i \notin \mathcal{M}\}} - n\mathbb{P}(\tilde{\mathbf{X}} \notin \mathcal{M}) \right].$$

It is obvious to see that the random vector $(S_n, \mathcal{U}_n)$ and the random variable $\mathcal{V}_n$ are independent. Observe also that

$$\text{Var}(S_n) = 1 \quad \text{and} \quad \text{Var}(\mathcal{U}_n) = \prod_{l=1}^p (1 - \mathbf{u}_l). \quad (\text{B.39})$$

Lemma B.9. *Whenever $h_n \rightarrow 0$, $nh_n^d \rightarrow \infty$ and, for each $l = 1, \dots, p$, C_l satisfies (B.3) and (B.4) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K}(\cdot) = \mathcal{K}_l(\cdot)$, there exists a constant τ such that, for all n sufficiently large,*

$$|\text{Cov}(S_n, \mathcal{U}_n)| \leq \frac{\tau}{\sqrt{nh_n^d}}. \quad (\text{B.40})$$

Proof of Lemma B.9. Observe that

$$\left(n \text{Var}(U_{\eta,0}(\bar{C}))\right)^{1/2} |\text{Cov}(S_n, \mathcal{U}_n)| = \left| \text{Cov}\left(\sqrt{n} \int_{\bar{C}} |\Gamma_{0,\eta}(\tilde{\mathbf{x}})| d\tilde{\mathbf{x}}, \mathcal{U}_n\right) \right|.$$

In order to get (B.40), it is sufficient, by using (B.36), to prove that there exists a constant τ_1 such that, for all n sufficiently large,

$$\left| \text{Cov}\left(\sqrt{n} \int_{\bar{C}} |\Gamma_{\eta,0}(\tilde{\mathbf{x}})| d\tilde{\mathbf{x}}, \mathcal{U}_n\right) \right| \leq \frac{\tau_1}{\sqrt{nh_n^d}}. \quad (\text{B.41})$$

Now, for any $\tilde{\mathbf{x}} \in \bar{C}$,

$$\left(\frac{\sqrt{n} \Gamma_{\eta,0}(\tilde{\mathbf{x}})}{\sqrt{k_{n,0}(\tilde{\mathbf{x}})}}, \frac{\mathcal{U}_n}{\sqrt{\mathbb{P}(\tilde{\mathbf{x}} \in \mathcal{M})}} \right) \stackrel{\mathcal{D}}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathcal{Y}_{0,n}^{(i)}(\tilde{\mathbf{x}}), \mathcal{U}^{(i)}),$$

where $(\mathcal{Y}_{0,n}^{(i)}(\tilde{\mathbf{x}}), \mathcal{U}^{(i)})_{1 \leq i \leq n}$ are independent and $(\mathcal{Y}_{0,n}(\tilde{\mathbf{x}}), \mathcal{U})$ - identically distributed random vectors with

$$\mathcal{Y}_{0,n}(\tilde{\mathbf{x}}) = \frac{1}{h_n^d \sqrt{k_{n,0}(\tilde{\mathbf{x}})}} \left[\sum_{j \leq \eta_1} \prod_{l=1}^p K_l\left(\frac{\mathbf{x}_l - \mathbb{X}_j^l}{h_n}\right) - \prod_{l=1}^p \mathbb{E} K_l\left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n}\right) \right]$$

and

$$\mathcal{U} = \left[\sum_{j \leq \eta_1} \mathbb{1}_{\{\tilde{\mathbf{x}}_i \in \mathcal{M}\}} - \mathbb{P}(\tilde{\mathbf{x}} \in \mathcal{M}) \right] / \sqrt{\mathbb{P}(\tilde{\mathbf{x}} \in \mathcal{M})}.$$

Here, as above, η_1 denotes a Poisson random variable with mean 1, independent of $\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2 \dots$. Notice that $\mathbb{E} \mathcal{Y}_{0,n}(\tilde{\mathbf{x}}) = \mathbb{E} \mathcal{U} = 0$, $\text{Var}(\mathcal{Y}_{0,n}(\tilde{\mathbf{x}})) = \text{Var}(\mathcal{U}) = 1$ and

$$\begin{aligned} |\text{Cov}(\mathcal{Y}_{0,n}(\tilde{\mathbf{x}}), \mathcal{U})| &= \left| \frac{\mathbb{E} \left[\prod_{l=1}^p K_l\left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n}\right) \mathbb{1}_{\{\tilde{\mathbf{x}} \in \mathcal{M}\}} \right]}{h_n^d \sqrt{k_{n,0}(\tilde{\mathbf{x}})} \sqrt{\mathbb{P}(\tilde{\mathbf{x}} \in \mathcal{M})}} \right| \\ &\leq \frac{\prod_{l=1}^p \mathbb{E} K_l\left(\frac{\mathbf{x}_l - \mathbb{X}^l}{h_n}\right)}{h_n^d \sqrt{k_{n,0}(\tilde{\mathbf{x}})} \sqrt{\mathbb{P}(\tilde{\mathbf{x}} \in \mathcal{M})}}. \end{aligned}$$

Therefore, by (B.16) and (B.9), this last term is, for all large enough n uniformly in $\tilde{\mathbf{x}} \in \bar{C}$ for some real number $\tau_2 > 0$, less than or equal to $\tau_2 \sqrt{h_n^d}$, which, in turn, is less than or equal to ε for all large enough n and any $0 < \varepsilon < 1$. This fact, in combination with (B.19), gives, by using (C.3) and Lemma 2.3 of Giné, Mason and Zaitsev (2003), that, for some constant $\tau_3 > 0$, uniformly in $\tilde{\mathbf{x}} \in \bar{C}$,

$$\begin{aligned} & \left| \text{Cov}(\sqrt{n}|f_\eta(\tilde{\mathbf{x}}) - \mathbb{E}f_n(\tilde{\mathbf{x}})|, \mathcal{U}_n) \right| \\ &= \left| \text{Cov} \left(\frac{\left| \sum_{i=1}^n \mathcal{Y}_{0,n}^{(i)}(\tilde{\mathbf{x}}) \right|}{\sqrt{n}}, \frac{\sum_{i=1}^n \mathcal{U}^{(i)}}{\sqrt{n}} \right) \right| \sqrt{k_{n,0}(\tilde{\mathbf{x}}) \mathbb{P}(\tilde{\mathbf{X}} \in \mathcal{M})} \\ &\leq \frac{\tau_3}{\sqrt{nh_n^d}} \sqrt{k_{n,0}(\tilde{\mathbf{x}})}, \end{aligned}$$

which, when combined with (B.16) and $\lambda(\bar{C}) < \infty$, gives (B.41). This finishes the proof of Lemma B.9. $\blacksquare$

Lemma B.10. *Suppose that, for each $l = 1, \dots, p$, C_l satisfies (B.3) and (B.4) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K}(\cdot) = \mathcal{K}_l(\cdot)$. If $h_n \rightarrow 0$ and $\sqrt{nh_n^d} \rightarrow \infty$, then*

$$(S_n, \mathcal{U}_n) \xrightarrow{\mathcal{D}} \left(Z_1, \prod_{l=1}^p \sqrt{(1 - \mathbf{u}_l)} Z_2 \right) \quad (\text{B.42})$$

as $n \rightarrow \infty$, where Z_1 and Z_2 are independent standard normal random variables.

Proof of Lemma B.10. We will show that, for any ξ_1 and ξ_2 , as $n \rightarrow \infty$,

$$\xi_1 S_n + \xi_2 \mathcal{U}_n \xrightarrow{\mathcal{D}} \xi_1 Z_1 + \xi_2 \prod_{l=1}^p \sqrt{(1 - \mathbf{u}_l)} Z_2.$$

For each $\mathbf{i} \in l_n$, set

$$\mathcal{U}_{\mathbf{i},n} := \frac{1}{\sqrt{n}} \left[\sum_{j=1}^n \mathbb{1}_{\{\tilde{\mathbf{X}}_j \in R_{\mathbf{i}}\}} - n \mathbb{P}(\tilde{\mathbf{X}} \in R_{\mathbf{i}}) \right]$$

and

$$\mathcal{Z}_{\mathbf{i},n} := \xi_1 \mathcal{X}_{\mathbf{i},n} + \xi_2 \mathcal{U}_{\mathbf{i},n}.$$

By c_r -inequality, for all $r \geq 1$, we have

$$\mathbb{E}|\mathcal{Z}_{\mathbf{i},n}|^r \leq 2^{r-1} \left(|\xi_1|^r \mathbb{E}|\mathcal{X}_{\mathbf{i},n}|^r + |\xi_2|^r \mathbb{E}|\mathcal{U}_{\mathbf{i},n}|^r \right). \quad (\text{B.43})$$

Now, by Hölder inequality, for all $2 < r < 3$ and all $\mathbf{i} \in l_n$,

$$\mathbb{E}|\mathcal{X}_{\mathbf{i},n}|^r \leq (\mathbb{E}|\mathcal{X}_{\mathbf{i},n}|^3)^{r/3},$$

which by (B.37) and (B.38) implies that, for all $2 < r < 3$

$$\sum_{\mathbf{i} \in l_n} \mathbb{E}|\mathcal{X}_{\mathbf{i},n}|^r \leq 2^d \gamma_0^{r/3} h_n^{d(r-2)/2} \prod_{l=1}^p M_l^{d_l} \rightarrow 0. \quad (\text{B.44})$$

Set now

$$p_{\mathbf{i},n} = \mathbb{P}(\tilde{\mathbf{X}} \in R_{\mathbf{i}}).$$

By Lemma 2.3 in [Giné, Mason and Zaitsev \(2003\)](#), we can find a constant $\zeta > 0$ such that

$$\begin{aligned} \sum_{\mathbf{i} \in l_n} \mathbb{E}|\mathcal{U}_{\mathbf{i},n}|^r &\leq \zeta n^{-r/2} \sum_{\mathbf{i} \in l_n} \left((np_{\mathbf{i},n})^{r/2} + np_{\mathbf{i},n} \right) \\ &\leq \zeta \max_{\mathbf{i} \in l_n} \left((p_{\mathbf{i},n})^{(r-2)/2} + n^{(2-r)/2} \right) \rightarrow 0. \end{aligned} \quad (\text{B.45})$$

Combining (B.43)-(B.45), we obtain, for all $2 < r < 3$,

$$\lim_{n \rightarrow \infty} \sum_{\mathbf{i} \in l_n} \mathbb{E}|\mathcal{Z}_{\mathbf{i},n}|^r = 0.$$

Moreover, note that $\{\mathcal{Z}_{\mathbf{i},n} : \mathbf{i} \in l_n\}$ is a triangular array of mean zero one-dependent random fields. Consequently, by the statements (B.39) and (B.40), it follows that

$$\text{Var} \left(\sum_{\mathbf{i} \in l_n} \mathcal{Z}_{\mathbf{i},n} \right) = \text{Var}(\xi_1 S_n + \xi_2 \mathcal{U}_n) \rightarrow \xi_1^2 + \xi_2^2 \prod_{l=1}^p (1 - \mathbf{u}_l) \quad \text{as } n \rightarrow \infty.$$

Thus, by Theorem 1 of [Shergin \(1990\)](#), we can infer that

$$\sum_{\mathbf{i} \in l_n} \mathcal{Z}_{\mathbf{i},n} \xrightarrow{\mathcal{D}} \xi_1 Z_1 + \xi_2 \prod_{l=1}^p \sqrt{(1 - \mathbf{u}_l)} Z_2 \quad \text{as } n \rightarrow \infty.$$

Since

$$\sum_{\mathbf{i} \in l_n} \mathcal{Z}_{\mathbf{i},n} = \xi_1 S_n + \xi_2 \mathcal{U}_n,$$

the statement (B.42) is proved by the Cramér-Wold device, for instance, see, [Billingsley \(1999\)](#). ■

Lemma B.11. *Suppose that, for each $l = 1, \dots, p$, C_l satisfies (B.3) and (B.4) of Lemma B.1 with $g(\cdot) = f_l(\cdot)$ and $\mathcal{K}(\cdot) = \mathcal{K}_l(\cdot)$. If $h_n \rightarrow 0$ and $nh_n^{3d} \rightarrow \infty$, then*

$$\frac{\sqrt{n}(V_n(\bar{C}) - \mathbb{E}V_n(\bar{C}))}{\sqrt{n \text{Var}(U_{\eta,0}(\bar{C}))}} \xrightarrow{\mathcal{D}} Z \quad \text{as } n \rightarrow \infty, \quad (\text{B.46})$$

where Z is a standard normal random variable.

Proof of Lemma B.11. Note that

$$S_n = \frac{\sqrt{n}(U_{\eta,0}(\bar{C}) - \mathbb{E}U_{\eta,0}(\bar{C}))}{\sqrt{n\text{Var}(U_{\eta,0}(\bar{C}))}}.$$

Indeed, conditioned on $\eta = n$, we have

$$S_n = \frac{\sqrt{n}(U_{n,0}(\bar{C}) - \mathbb{E}U_{n,0}(\bar{C}))}{\sqrt{n\text{Var}(U_{n,0}(\bar{C}))}}.$$

By Lemma 2.4 of [Giné, Mason and Zaitsev \(2003\)](#) refer also to [Beirlant and Mason \(1995\)](#) and Lemma B.10, we conclude that

$$\frac{\sqrt{n}(U_{n,0}(\bar{C}) - \mathbb{E}U_{n,0}(\bar{C}))}{\sqrt{n\text{Var}(U_{n,0}(\bar{C}))}} \xrightarrow{\mathcal{D}} Z.$$

Therefore, by (B.36) and Lemma B.4, we have

$$\frac{\sqrt{n}(U_{n,0}(\bar{C}) - \mathbb{E}U_{n,0}(\bar{C}))}{\sqrt{n\text{Var}(U_{n,0}(\bar{C}))}} \xrightarrow{\mathcal{D}} Z.$$

Assertion (B.46) now follows from Lemmas B.3 and B.7 in combination with (B.36). Hence the proof is complete. $\blacksquare$

Lemma B.12. Assume that (A_1) holds and, for each $l = 1, \dots, p$, there exists $\delta > 0$ such that $f_l(\cdot)$ is bounded on C_l^δ . If $h_n \rightarrow 0$ and $nh_n^{3d} \rightarrow \infty$, then we have

$$\lim_{n \rightarrow \infty} \mathbb{E} \int_{\bar{C}} \left| \sqrt{\hat{\mathcal{L}}_n(\tilde{\mathbf{x}})} - \sqrt{\mathcal{L}_{n,1}(\tilde{\mathbf{x}})} \right| d\tilde{\mathbf{x}} = 0. \quad (\text{B.47})$$

Proof of Lemma B.12. Notice that

$$\mathcal{L}_{n,1}(\tilde{\mathbf{x}}) = \sum_{\{I \subsetneq I_p, |I_p \setminus I| \geq 2\}} \prod_{l \in I} (\mathbb{E}f_{n,l}(\mathbf{x}_l))^2 \prod_{j \in I_p \setminus I} [v_{n,j}(\mathbf{x}_j) - (\mathbb{E}f_{n,j}(\mathbf{x}_j))^2]$$

and

$$\hat{\mathcal{L}}_n(\tilde{\mathbf{x}}) = \sum_{\{I \subsetneq I_p, |I_p \setminus I| \geq 2\}} \prod_{l \in I} g_{n,l}(\mathbf{x}_l) \prod_{j \in I_p \setminus I} [\hat{v}_{n,j}(\mathbf{x}_j) - g_{n,j}(\mathbf{x}_j)].$$

Set, for each $I \subsetneq I_p$,

$$\Theta_{n,1,I}(\tilde{\mathbf{x}}) := \left| \prod_{j \in I_p \setminus I} [v_{n,j}(\mathbf{x}_j) - (\mathbb{E}f_{n,j}(\mathbf{x}_j))^2] - \prod_{j \in I_p \setminus I} [\hat{v}_{n,j}(\mathbf{x}_j) - g_{n,j}(\mathbf{x}_j)] \right|$$

and

$$\Theta_{n,2,I}(\tilde{\mathbf{x}}) := \left| \prod_{l \in I} g_{n,l}(\mathbf{x}_l) - \prod_{l \in I} (\mathbb{E} f_{n,l}(\mathbf{x}_l))^2 \right|.$$

It is obvious to see that

$$\begin{aligned} |\mathcal{L}_{n,1}(\tilde{\mathbf{x}}) - \hat{\mathcal{L}}_n(\tilde{\mathbf{x}})| &\leq \sum_{\{I \subsetneq I_P, |I_P \setminus I| \geq 2\}} \prod_{l \in I} g_{n,l}(\mathbf{x}_l) \Theta_{n,1,I}(\tilde{\mathbf{x}}) \\ &\quad + \sum_{\{I \subsetneq I_P, |I_P \setminus I| \geq 2\}} \prod_{j \in I_P \setminus I} v_{n,j}(\mathbf{x}_j) \Theta_{n,2,I}(\tilde{\mathbf{x}}) \\ &=: \Psi_{n,1}(\tilde{\mathbf{x}}) + \Psi_{n,2}(\tilde{\mathbf{x}}). \end{aligned} \quad (\text{B.48})$$

For each $I \subsetneq I_P$, observe that

$$\mathbb{E} \Theta_{n,1,I}(\tilde{\mathbf{x}}) \leq \sum_{A \subsetneq I_P \setminus I} \prod_{l \in A} v_{n,l}(\mathbf{x}_l) \prod_{j \in I_P \setminus (I \cup A)} \left\{ \mathbb{E} |\hat{v}_{n,j}(\mathbf{x}_j) - v_{n,j}(\mathbf{x}_j)| + \mathbb{E} |g_{n,j}(\mathbf{x}_j) - (\mathbb{E} f_{n,j}(\mathbf{x}_j))^2| \right\} \quad (\text{B.49})$$

and

$$\mathbb{E} \Theta_{n,2,I}(\tilde{\mathbf{x}}) \leq \sum_{A \subsetneq I} \prod_{l \in A} (\mathbb{E} f_{n,l}(\mathbf{x}_l))^2 \prod_{j \in I \setminus A} \mathbb{E} \left| g_{n,j}(\mathbf{x}_j) - (\mathbb{E} f_{n,l}(\mathbf{x}_j))^2 \right|. \quad (\text{B.50})$$

Now, for each $l \in I_P$, we have

$$\begin{aligned} \mathbb{E} |\hat{v}_{n,l}(\mathbf{x}_l) - v_{n,l}(\mathbf{x}_l)| &\leq \left(\mathbb{E} |\hat{v}_{n,l}(\mathbf{x}_l) - v_{n,l}(\mathbf{x}_l)|^2 \right)^{1/2} \\ &\leq \frac{1}{\sqrt{n} h_n^{2d_l}} \left(\int_{\mathbb{R}^{d_l}} K_l^4 \left(\frac{\mathbf{x}_l - \mathbf{u}_l}{h_n} \right) f_l(\mathbf{u}_l) d\mathbf{u}_l \right)^{1/2}. \end{aligned}$$

Since, for each $l \in I_P$, $K_l(\cdot)$ is bounded on $\mathbb{R}^{d_l}$ and $f_l(\cdot)$ is bounded on $C_l^{h_n/2}$ for n large enough, we obtain

$$\sup_{\mathbf{x}_l \in C_l} \mathbb{E} |\hat{v}_{n,l}(\mathbf{x}_l) - v_{n,l}(\mathbf{x}_l)| = O \left(\frac{1}{\sqrt{n} h_n^{3d_l}} \right). \quad (\text{B.51})$$

Observe that

$$\begin{aligned} &\mathbb{E} \left| g_{n,l}(\mathbf{x}_l) - (\mathbb{E} f_{n,l}(\mathbf{x}_l))^2 \right| \\ &\leq \frac{1}{n(n-1)h_n^{2d_l}} \mathbb{E} \left| \sum_{i=1}^n \sum_{j \neq i} K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n} \right) \left[K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_j^l}{h_n} \right) - \mathbb{E} K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_j^l}{h_n} \right) \right] \right| \\ &\quad + \frac{1}{n(n-1)h_n^{2d_l}} \mathbb{E} \left| \sum_{i=1}^n \sum_{j \neq i} \mathbb{E} K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_j^l}{h_n} \right) \left[K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n} \right) - \mathbb{E} K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n} \right) \right] \right|. \end{aligned}$$

For each $l \in I_p$, since $K_l(\cdot)$ is bounded on $\mathbb{R}^{d_l}$, it follows that, for some constant $\mathfrak{H}_0 > 0$, we have

$$\begin{aligned} \mathbb{E} \left| g_{n,l}(\mathbf{x}_l) - (\mathbb{E} f_{n,l}(\mathbf{x}_l))^2 \right| &\leq \frac{\mathfrak{H}_0}{nh_n^{2d_l}} \mathbb{E} \left| \sum_{i=1}^n \left[K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n} \right) - \mathbb{E} K_l \left(\frac{\mathbf{x}_l - \mathbb{X}_i^l}{h_n} \right) \right] \right| \\ &\leq \frac{\mathfrak{H}_0}{\sqrt{n} h_n^{2d_l}} \left(\int_{\mathbb{R}^{d_l}} K_l^2 \left(\frac{\mathbf{x}_l - \mathbf{u}_l}{h_n} \right) f_l(\mathbf{u}_l) d\mathbf{u}_l \right)^{1/2}. \end{aligned}$$

Since, for each $l \in I_p$, $f_l(\cdot)$ is bounded on $C_l^{h_n/2}$ for n large enough, we have

$$\sup_{\mathbf{x}_l \in C_l} \mathbb{E} |g_{n,l}(\mathbf{x}_l) - (\mathbb{E} f_{n,l}(\mathbf{x}_l))^2| = O \left(\frac{1}{\sqrt{n} h_n^{3d_l}} \right). \quad (\text{B.52})$$

By the same arguments, we get

$$\sup_{\mathbf{x}_l \in C_l} v_{n,l}(\mathbf{x}_l) \leq \frac{\mathfrak{H}_1}{h_n^{d_l}}, \quad \sup_{\mathbf{x}_l \in C_l} \mathbb{E} g_{n,l}(\mathbf{x}_l) \leq \mathfrak{H}_1 \quad \text{and} \quad \sup_{\mathbf{x}_l \in C_l} \mathbb{E} f_{n,l}(\mathbf{x}_l) \leq \mathfrak{H}_2, \quad (\text{B.53})$$

for some positive constants $\mathfrak{H}_1$, $\mathfrak{H}_2$ and $\mathfrak{H}_3$. Therefore, by (B.49)-(B.53), we infer that

$$\sup_{\tilde{\mathbf{x}} \in \tilde{C}} \mathbb{E} \Psi_{n,1}(\tilde{\mathbf{x}}) = O \left(\frac{1}{\sqrt{n} h_n^{3d}} \right). \quad (\text{B.54})$$

Similarly, by (B.50) in combination with (B.52) and (B.53), we obtain

$$\sup_{\tilde{\mathbf{x}} \in \tilde{C}} \mathbb{E} \Psi_{n,2}(\tilde{\mathbf{x}}) = O \left(\frac{1}{\sqrt{n} h_n^{3d}} \right). \quad (\text{B.55})$$

By using (B.48), (B.54) and (B.55) in combination with the fact that $\tilde{C}$ is bounded, we get

$$\begin{aligned} \mathbb{E} \int_{\tilde{C}} \left| \sqrt{\mathcal{L}_{n,1}(\tilde{\mathbf{x}})} - \sqrt{\hat{\mathcal{L}}_n(\tilde{\mathbf{x}})} \right| d\tilde{\mathbf{x}} &\leq \int_{\tilde{C}} \sqrt{\mathbb{E} |\mathcal{L}_{n,1}(\tilde{\mathbf{x}}) - \hat{\mathcal{L}}_n(\tilde{\mathbf{x}})|} d\tilde{\mathbf{x}} \\ &= O \left(\frac{1}{\sqrt{n} h_n^{3d}} \right). \end{aligned}$$

We so obtain (B.47), as sought. ■

Appendix C

For completeness, we recall the following fact (Fact 6.1 of [Giné, Mason and Zaitsev \(2003\)](#)), which follows from Theorem 1 of [Sweeting \(1977\)](#) and is related to the classical Berry-Esseen theorem.

Theorem C.1. Let $(\omega, \zeta), (\omega_1, \zeta_1), (\omega_2, \zeta_2), \dots$ be a sequence of i.i.d. random vectors such that each component has variance 1, mean 0, and finite absolute moments of the third order. Further, let $(\bar{Z}_1, \bar{Z}_2)$ be bivariate normal with mean vector 0, $\text{Var}(\bar{Z}_1) = \text{Var}(\bar{Z}_2) = 1$, and $\text{Cov}(\bar{Z}_1, \bar{Z}_2) = \text{Cov}(\omega, \zeta) = \rho$. Then there exist universal positive constants A_1, A_2 and A_3 such that

$$\left| \mathbb{E} \left| \frac{\sum_{i=1}^n \zeta_i}{\sqrt{n}} \right| - \mathbb{E}|\bar{Z}_1| \right| \leq \frac{A_1}{\sqrt{n}} \mathbb{E}|\zeta|^3, \quad (\text{C.1})$$

and, whenever $\rho^2 < 1$,

$$\left| \mathbb{E} \left| \frac{\sum_{i=1}^n \omega_i}{\sqrt{n}} \cdot \frac{\sum_{i=1}^n \zeta_i}{\sqrt{n}} \right| - \mathbb{E}|\bar{Z}_1 \bar{Z}_2| \right| \leq \frac{A_2}{(1 - \rho^2)^{3/2}} \frac{1}{\sqrt{n}} \left(\mathbb{E}|\omega|^3 + \mathbb{E}|\zeta|^3 \right), \quad (\text{C.2})$$

and

$$\left| \mathbb{E} \left[\left| \frac{\sum_{i=1}^n \omega_i}{\sqrt{n}} \right| \left| \frac{\sum_{i=1}^n \zeta_i}{\sqrt{n}} \right| \right] \right| \leq \frac{A_3}{(1 - \rho^2)^{3/2}} \frac{1}{\sqrt{n}} \left(\mathbb{E}|\omega|^3 + \mathbb{E}|\zeta|^3 \right). \quad (\text{C.3})$$

References

- AHMAD, I. A. and LI, Q. (1997). Testing independence by nonparametric kernel method. *Statist. Probab. Lett.* **34** 201–210. [https://doi.org/10.1016/S0167-7152\(96\)00183-6](https://doi.org/10.1016/S0167-7152(96)00183-6)
- AKAIKE, H. (1954). An approximation to the density function. *Ann. Inst. Statist. Math., Tokyo* **6** 127–132. <https://doi.org/10.1007/bf02900741>
- ARAUJO, A. and GINÉ, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-Chichester-Brisbane.
- AUDDY, A., DEB, N. and NANDY, S. (2024). Exact detection thresholds and minimax optimality of Chatterjee’s correlation coefficient. *Bernoulli* **30** 1640–1668. <https://doi.org/10.3150/23-bej1648>
- BACH, F. R. and JORDAN, M. I. (2003). Kernel independent component analysis. *J. Mach. Learn. Res.* **3** 1–48. <https://doi.org/10.1162/153244303768966085>
- BEIRLANT, J. and MASON, D. M. (1995). On the asymptotic normality of L_p -norms of empirical functionals. *Math. Methods Statist.* **4** 1–19.
- BERGER, E. (1991). Majorization, exponential inequalities and almost sure behavior of vector-valued random variables. *Ann. Probab.* **19** 1206–1226. <https://doi.org/10.1214/aop/1176990341>
- BERRAHOU, N., BOUZEBDA, S. and DOUGE, L. (2024). Supplementary materials to “A nonparametric distribution-free test of independence among continuous random vectors based on L_1 -norm”.
- BERRETT, T. B. (2022). Discussion of ‘Multi-scale Fisher’s independence test for multivariate dependence’ [4472834]. *Biometrika* **109** 589–592. <https://doi.org/10.1093/biomet/asac023>
- BERRETT, T. B., GROSE, D. J. and SAMWORTH, R. J. (2018). *IndepTest: Nonparametric Independence Tests Based on Entropy Estimation*. R Package version 0.2.0. Available at <https://cran.r-project.org/web/packages/IndepTest/index.html>.
- BERRETT, T. B., KONTOYIANNIS, I. and SAMWORTH, R. J. (2021). Optimal rates for independence testing via U -statistic permutation tests. *Ann. Statist.* **49** 2457–2490. <https://doi.org/10.1214/20-aos2041>
- BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. <https://doi.org/10.1093/biomet/asz024>
- BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, second ed. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York A Wiley-Interscience Publication. <https://doi.org/10.1002/9780470316962>
- BINET, A. and VASCHIDE, N. (1897). Corrélation des épreuves physiques. *L’Année psychologique* **4** 142–172. <https://doi.org/10.3406/psy.1897.2892>

- BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **32** 485–498. <https://doi.org/10.1214/aoms/1177705055>
- BÖTTCHER, B., KELLER-RESSEL, M. and SCHILLING, R. L. (2019). Distance multivariate: new dependence measures for random vectors. *Ann. Statist.* **47** 2757–2789. <https://doi.org/10.1214/18-AOS1764>
- BOUZEBDA, S. (2011). Some new multivariate tests of independence. *Math. Methods Statist.* **20** 192–205. <https://doi.org/10.3103/S1066530711030021>
- BOUZEBDA, S. (2014). General tests of independence based on empirical processes indexed by functions. *Stat. Methodol.* **21** 59–87. <https://doi.org/10.1016/j.stamet.2014.03.001>
- BOUZEBDA, S. and NEMOUCHI, B. (2023). Weak-convergence of empirical conditional processes and conditional U -processes involving functional mixing data. *Stat. Inference Stoch. Process.* **26** 33–88. <https://doi.org/10.1007/s11203-022-09276-6>
- BOUZEBDA, S. and TAACHOUCHE, N. (2023). On the variable bandwidth kernel estimation of conditional U -statistics at optimal rates in sup-norm. *Phys. A* **625** Paper No. 129000, 72. <https://doi.org/10.1016/j.physa.2023.129000>
- CHAKRABORTY, S. and ZHANG, X. (2019). Distance metrics for measuring joint dependence with application to causal inference. *J. Amer. Statist. Assoc.* **114** 1638–1650. <https://doi.org/10.1080/01621459.2018.1513364>
- CHAN, N.-H., LEE, T. C. M. and PENG, L. (2010). On nonparametric local inference for density estimation. *Comput. Statist. Data Anal.* **54** 509–515. <https://doi.org/10.1016/j.csda.2009.09.021>
- CHATTERJEE, S. (2021). A new coefficient of correlation. *J. Amer. Statist. Assoc.* **116** 2009–2022. <https://doi.org/10.1080/01621459.2020.1758115>
- CHEN, A. and BICKEL, P. J. (2006). Efficient independent component analysis. *Ann. Statist.* **34** 2825–2855. <https://doi.org/10.1214/0090536060000000939>
- COX, T. F. and CZANNER, G. (2016). A practical divergence measure for survival distributions that can be estimated from Kaplan-Meier curves. *Stat. Med.* **35** 2406–2421. <https://doi.org/10.1002/sim.6868>
- CSÖRGŐ, S. (1985). Testing for independence by the empirical characteristic function. *J. Multivariate Anal.* **16** 290–299. [https://doi.org/10.1016/0047-259X\(85\)90022-3](https://doi.org/10.1016/0047-259X(85)90022-3)
- CSÖRGŐ, M. and HORVÁTH, L. (1988). Central limit theorems for L_p -norms of density estimators. *Probab. Theory Related Fields* **80** 269–291. <https://doi.org/10.1007/BF00356106>
- DEB, N., GHOSAL, P. and SEN, B. (2020). Measuring Association on Topological Spaces Using Kernels and Geometric Graphs. arXiv:2010.01768.
- DEB, N. and SEN, B. (2023). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.* **118** 192–207. <https://doi.org/10.1080/01621459.2021.1923508>
- DEHEUVELS, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés. II. *Publ. Inst. Statist. Univ. Paris* **22** 1–23.
- DEHEUVELS, P. (1981). A Kolmogorov-Smirnov type test for independence and multivariate samples. *Rev. Roumaine Math. Pures Appl.* **26** 213–226.
- DEVROYE, L. (1987). *A Course in Density Estimation. Progress in Probability and Statistics* **14**. Birkhäuser Boston, Inc., Boston, MA. <https://doi.org/10.1002/bimj.4710300618>
- DEVROYE, L. (1989). The double kernel method in density estimation. *Ann. Inst. H. Poincaré Probab. Statist.* **25** 533–580.
- DEVROYE, L. (1997). Universal smoothing factor selection in density estimation: theory and practice. *Test* **6** 223–320. With discussion and a rejoinder by the author. <https://doi.org/10.1007/BF02564701>
- DEVROYE, L. and GYÖRFI, L. (1985). *Nonparametric density estimation. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics*. John Wiley & Sons, Inc., New York The L_1 view.
- DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (2002). A note on robust hypothesis testing. *IEEE Trans. Inform. Theory* **48** 2111–2114. <https://doi.org/10.1109/TIT.2002.1013154>
- DEVROYE, L. and LUGOSI, G. (2001). *Combinatorial Methods in Density Estimation. Springer Series in Statistics*. Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4613-0125-7>
- DRTON, M., HAN, F. and SHI, H. (2020). High-dimensional consistent independence testing with maxima of rank correlations. *Ann. Statist.* **48** 3206–3227. <https://doi.org/10.1214/19-AOS1926>
- DUDLEY, R. M. (1989). *Real Analysis and Probability. The Wadsworth & Brooks/Cole Mathematics Series*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.

- DUGUÉ, D. (1975). Sur des tests d'indépendance "indépendants de la loi". *C. R. Acad. Sci. Paris Sér. A-B* **281** A1103–A1104.
- DUONG, T. and HAZELTON, M. L. (2005). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scand. J. Statist.* **32** 485–506. <https://doi.org/10.1111/j.1467-9469.2005.00445.x>
- EGGERMONT, P. P. B. and LARICCIA, V. N. (2001). *Maximum Penalized Likelihood Estimation. Vol. I. Springer Series in Statistics*. Springer-Verlag, New York. Density estimation. <https://doi.org/10.1007/978-1-0716-1244-6>
- EPSNEČNIKOV, V. A. (1969). Nonparametric estimation of a multidimensional probability density. *Teor. Veroyatnost. i Primenen.* **14** 156–162. <https://doi.org/10.1137/1114019>
- FAN, Y., LAFAYE DE MICHEAUX, P., PENEV, S. and SALOPEK, D. (2017). Multivariate nonparametric test of independence. *J. Multivariate Anal.* **153** 189–210. <https://doi.org/10.1016/j.jmva.2016.09.014>
- FELBER, T., KOHLER, M. and KRZYŻAK, A. (2015). Adaptive density estimation from data with small measurement errors. *IEEE Trans. Inform. Theory* **61** 3446–3456. <https://doi.org/10.1109/TIT.2015.2421297>
- FEUERVERGER, A. (1993). A consistent test for bivariate dependence. *Int. Stat. Rev.* **61** 419–433. <https://doi.org/10.2307/1403753>
- FUKUMIZU, K., BACH, F. R. and JORDAN, M. I. (2004). Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *J. Mach. Learn. Res.* **5** 73–99. <https://doi.org/10.1162/153244303768966111>
- GAN, L., NARISSETTY, N. N. and LIANG, F. (2019). Bayesian regularization for graphical models with unequal shrinkage. *J. Amer. Statist. Assoc.* **114** 1218–1231. <https://doi.org/10.1080/01621459.2018.1482755>
- GAO, J. and GIJBELS, I. (2008). Bandwidth selection in nonparametric kernel testing. *J. Amer. Statist. Assoc.* **103** 1584–1594. <https://doi.org/10.1198/016214508000000968>
- GINÉ, E., MASON, D. M. and ZAITSEV, A. Y. (2003). The L_1 -norm density estimator process. *Ann. Probab.* **31** 719–768. <https://doi.org/10.1214/aop/1048516534>
- GORSKY, S. and MA, L. (2022). Multi-scale Fisher's independence test for multivariate dependence. *Biometrika* **109** 569–587. <https://doi.org/10.1093/biomet/asac013>
- GRETTON, A. and GYÖRFI, L. (2010). Consistent nonparametric tests of independence. *J. Mach. Learn. Res.* **11** 1391–1423.
- GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. (2005). Measuring statistical dependence with Hilbert-Schmidt norms. In *Algorithmic learning theory. Lecture Notes in Comput. Sci.* **3734** 63–77. Springer, Berlin. https://doi.org/10.1007/11564089_7
- HAHN, T. (2005). Cuba—a library for multidimensional numerical integration. *Comput. Phys. Comm.* **168** 78–95. <https://doi.org/10.1016/j.cpc.2005.01.010>
- HALL, P. and WAND, M. P. (1988). Minimizing L_1 distance in nonparametric density estimation. *J. Multivariate Anal.* **26** 59–88. [https://doi.org/10.1016/0047-259X\(88\)90073-5](https://doi.org/10.1016/0047-259X(88)90073-5)
- HAN, F., CHEN, S. and LIU, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika* **104** 813–828. <https://doi.org/10.1093/biomet/asx050>
- HÄRDLE, W. (1990). *Applied Nonparametric Regression. Econometric Society Monographs* **19**. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CCOL0521382483>
- HELLER, R., HELLER, Y. and GORFINE, M. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika* **100** 503–510. <https://doi.org/10.1093/biomet/ass070>
- HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Statist.* **19** 546–557. <https://doi.org/10.1214/aoms/1177730150>
- HOROWITZ, J. L. and SPOKOINY, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69** 599–631. <https://doi.org/10.1111/1468-0262.00207>
- HORVÁTH, L. (1991). On L_p -norms of multivariate density estimators. *Ann. Statist.* **19** 1933–1949. <https://doi.org/10.1214/aos/1176348379>
- JIN, Z. and MATTESON, D. S. (2018). Generalizing distance covariance to measure and test multivariate mutual dependence via complete and incomplete V-statistics. *J. Multivariate Anal.* **168** 304–322. <https://doi.org/10.1016/j.jmva.2018.08.006>
- KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika* **30** 81–93. <https://doi.org/10.1093/biomet/30.1-2.81>
- LAURITZEN, S. L. (1996). *Graphical Models. Oxford Statistical Science Series* **17**. The Clarendon Press, Oxford University Press, New York Oxford Science Publications. <https://doi.org/10.1093/oso/9780198522195.001.0001>

- LE CAM, L. (1970). Remarques sur le théorème limite central dans les espaces localement convexes. In *Les probabilités sur les structures algébriques (Actes Colloq. Internat. CNRS, No. 186, Clermont-Ferrand, 1969)* 233–249. Éditions Centre Nat. Recherche Sci., Paris Avec commentaire en anglais par R. M. Dudley.
- LEE, S., SONG, K. and WHANG, Y.-J. (2013). Testing functional inequalities. *J. Econometrics* **172** 14–32. <https://doi.org/10.1016/j.jeconom.2012.08.006>
- LEUNG, D. and DRTON, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.* **46** 280–307. <https://doi.org/10.1214/17-AOS1550>
- MATTESON, D. S. and TSAY, R. S. (2017). Independent component analysis via distance covariance. *J. Amer. Statist. Assoc.* **112** 623–637. <https://doi.org/10.1080/01621459.2016.1150851>
- MOON, K. R. (2016). *Nonparametric Estimation of Distributional Functionals and Applications*. Thesis (Ph.D.)—University of Michigan.
- NABEYA, S. (1951). Absolute moments in 2-dimensional normal distribution. *Ann. Inst. Statist. Math., Tokyo* **3** 2–6. <https://doi.org/10.1007/bf02949770>
- NADARAYA, E. A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves. Mathematics and its Applications (Soviet Series)* **20**. Kluwer Academic Publishers Group, Dordrecht. <https://doi.org/10.1007/978-94-009-2583-0>
- NGUYEN, D. and EISENSTEIN, J. (2017). A kernel independence test for geographical language variation. *Computational Linguistics* **43** 567–592. https://doi.org/10.1162/COLI_a_00293
- PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076. <https://doi.org/10.1214/aoms/1177704472>
- PEARL, J. (2009). *Causality*, Second ed. Cambridge University Press, Cambridge Models, reasoning, and inference. <https://doi.org/10.1017/CBO9780511803161>
- PEARSON, K. (1920). Notes on the history of correlation. *Biometrika* **13** 25–45. <https://doi.org/10.2307/2331722>
- PETERS, J., MOOIJ, J. M., JANZING, D. and SCHÖLKOPF, B. (2014). Causal discovery with continuous additive noise models. *J. Mach. Learn. Res.* **15** 2009–2053. <https://doi.org/10.15496/publikation-1672>
- PFISTER, N. and PETERS, J. (2017). dHSIC: *Independence Testing via Hilbert-Schmidt Independence Criterion*. R Package version 2.1. Available at <https://cran.r-project.org/web/packages/dHSIC/index.html>.
- PFISTER, N., BÜHLMANN, P., SCHÖLKOPF, B. and PETERS, J. (2018). Kernel-based tests for joint independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 5–31. <https://doi.org/10.1111/rssb.12235>
- PINELIS, I. (1994). On a majorization inequality for sums of independent random vectors. *Statist. Probab. Lett.* **19** 97–99. [https://doi.org/10.1016/0167-7152\(94\)90139-2](https://doi.org/10.1016/0167-7152(94)90139-2)
- QIU, T., XU, W. and ZHU, L. (2023). Independence tests with random subspace of two random vectors in high dimension. *J. Multivariate Anal.* **195** Paper No. 105160. <https://doi.org/10.1016/j.jmva.2023.105160>
- RIZZO, M. L. (2009). New goodness-of-fit tests for Pareto distributions. *Astin Bull.* **39** 691–715. <https://doi.org/10.2143/AST.39.2.2044654>
- RIZZO, M. and SZÉKELY, G. (2016). E-Statistics: Multivariate inference via the energy of data. R package version 1.7-0. Available at <https://cran.r-project.org/web/packages/energy/index.html>.
- ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837. <https://doi.org/10.1214/aoms/1177728190>
- ROSENBLATT, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3** 1–14. <https://doi.org/10.1214/aos/1176342996>
- ROSENBLATT, M. and WAHLEN, B. E. (1992). A nonparametric measure of independence under a hypothesis of independent components. *Statist. Probab. Lett.* **15** 245–252. [https://doi.org/10.1016/0167-7152\(92\)90197-D](https://doi.org/10.1016/0167-7152(92)90197-D)
- SAIN, S. R., BAGGERLY, K. A. and SCOTT, D. W. (1994). Cross-validation of multivariate densities. *J. Amer. Statist. Assoc.* **89** 807–817. <https://doi.org/10.1080/01621459.1994.10476814>
- SAMWORTH, R. J. and YUAN, M. (2012). Independent component analysis via nonparametric maximum likelihood estimation. *Ann. Statist.* **40** 2973–3002. <https://doi.org/10.1214/12-AOS1060>
- SCOTT, D. W. (1992). *Multivariate Density Estimation. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. John Wiley & Sons Inc., New York. Theory, practice, and visualization, A Wiley-Interscience Publication. <https://doi.org/10.1002/9780470316849>
- SCOTT, D. W. and WAND, M. P. (1991). Feasibility of multivariate density estimates. *Biometrika* **78** 197–205. <https://doi.org/10.1093/biomet/78.1.197>

- SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. <https://doi.org/10.1214/13-AOS1140>
- SHENG, W. and YIN, X. (2016). Sufficient dimension reduction via distance covariance. *J. Comput. Graph. Statist.* **25** 91–104. <https://doi.org/10.1080/10618600.2015.1026601>
- SHERGIN, V. V. (1990). The central limit theorem for finitely dependent random variables. In *Probability theory and mathematical statistics, Vol. II (Vilnius, 1989)* 424–431. “Mokslas”, Vilnius. <https://doi.org/10.1007/BF01261286>
- SHI, H., DRTON, M. and HAN, F. (2022a). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.* **117** 395–410. <https://doi.org/10.1080/01621459.2020.1782223>
- SHI, H., DRTON, M. and HAN, F. (2022b). On the power of Chatterjee’s rank correlation. *Biometrika* **109** 317–333. <https://doi.org/10.1093/biomet/asab028>
- SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Applied Probability*. Chapman & Hall, London. <https://doi.org/10.1007/978-1-4899-3324-9>
- SPEARMAN, C. (1904). The proof and measurement of association between two things. *American journal of Psychology* **15** 72–101. <https://doi.org/10.2307/1422689>
- SRIPERUMBUDUR, B. K., FUKUMIZU, K., GRETTON, A., SCHÖLKOPF, B. and LANCKRIET, G. R. G. (2012). On the empirical estimation of integral probability metrics. *Electron. J. Stat.* **6** 1550–1599. <https://doi.org/10.1214/12-EJS722>
- STEIN, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions. Princeton Mathematical Series, No. 30*. Princeton University Press, Princeton, N.J. <https://doi.org/10.1515/9781400883882>
- SWEETING, T. J. (1977). Speeds of convergence for the multidimensional central limit theorem. *Ann. Probability* **5** 28–41. <https://doi.org/10.1214/aop/1176995888>
- SZÉKELY, G. J. and RIZZO, M. L. (2005). Hierarchical clustering via joint between-within distances: extending Ward’s minimum variance method. *J. Classification* **22** 151–183. <https://doi.org/10.1007/s00357-005-0012-9>
- SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. <https://doi.org/10.1214/009053607000000505>
- SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. <https://doi.org/10.1214/09-AOAS312>
- WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing. Monographs on Statistics and Applied Probability* **60**. Chapman and Hall Ltd., London. <https://doi.org/10.1007/978-1-4899-4493-1>
- WATSON, G. S. and LEADBETTER, M. R. (1963). On the estimation of the probability density. I. *Ann. Math. Statist.* **34** 480–491. <https://doi.org/10.1214/aoms/1177704159>
- YANG, G. (2012). *The energy goodness-of-fit test for univariate stable distributions*. ProQuest LLC, Ann Arbor, MI Thesis (Ph.D.)—Bowling Green State University.
- YATRACOS, Y. G. (1985). Rates of convergence of minimum distance estimators and Kolmogorov’s entropy. *Ann. Statist.* **13** 768–774. <https://doi.org/10.1214/aos/1176349553>
- ZHANG, W., GAO, W. and NG, H. K. T. (2023). Multivariate tests of independence based on a new class of measures of independence in Reproducing Kernel Hilbert Space. *J. Multivariate Anal.* **195** Paper No. 105144. <https://doi.org/10.1016/j.jmva.2022.105144>
- ZHANG, Q., FILIPPI, S., GRETTON, A. and SEJDINOVIC, D. (2018). Large-scale kernel methods for independence testing. *Stat. Comput.* **28** 113–130. <https://doi.org/10.1007/s11222-016-9721-7>