INTEGRATED SQUARE OF A DENSITY: ASYMPTOTIC NORMALITY AND NON-REGULAR SEMI-PARAMETRIC INFERENCE

PAUL DELATTE

University of Southern California, *delatte@usc.edu*

We show that the optimal estimator of the integrated square of a density considered in Giné and Nickl (2008) remains asymptotically normal even when the density belongs to a Sobolev space of order $s \leq 1/4$. We demonstrate, moreover, that, while standard inferential methods break down, it is still possible to perform valid inference in that case. To show this, we connect a number of results for second-order U-statistics with n-dependent kernels, derive moment bounds for the estimator that are of independent interest, and construct a variance estimator whose simple structure may be transposed to similar problems. This paper is, as far as we know, the first to demonstrate the possibility of semi-parametric inference in the presence of an infinite-dimensional non-smooth nuisance parameter.

1 Introduction

1.1 Problem and results overview

Let X be a random variable with density f_0 with respect to some measure μ. Under integrability conditions on the density, the map $f_0 \mapsto \int f_0^2 \, d\mu$ is a well-defined functional commonly known as the integrated square of f_0. This quadratic functional has been the subject of much interest in statistics. This interest is first explained by the fact that $\int f_0^2 \, d\mu$ plays a prominent role in density estimation. Robins and van der Vaart (2006) used it, for instance, in the construction of adaptive confidence sets for density estimators. The interest in $\int f_0^2 \, d\mu$ is further explained by the fact this quadratic functional is a central object in information theory. It is, indeed, a natural measure of concentration for dominated probability measures which corresponds exactly to the Rényi entropy of order 2 after a log transform. Finally, and more fundamentally, the interest in the integrated square of a density is explained by the shaping role this functional has played in statistical theory. Since $\int f_0^2 \, d\mu$ directly rewrites as $\mathbb{E}\left[f_0(X)\right]$, this functional appears to be one of the simplest semi-parametric objects in statistics. This simplicity had made its study paradigmatic for semi-parametric estimation and its corresponding techniques – notably plug-in based ones with and without debiasing. Many fundamental lines of research in the field have been initiated through the study of this quadratic functional – be it the minimax theory for estimating smooth functionals with its beginning in Bickel and Ritov (1988) or the theory of adaptive semi-parametric estimation

This version: 6 October, 2025. A previous version of this manuscript was circulated under the title "Average density: weak limits and inference in non-regular semi-parametric problems".

with its start in Efromovich and Low (1996). These works as well as the paradigmatic simplicity of $\int f_0^2\, d\mu$ have made this functional a canonical reference in semi-parametric theory. As a result, the list of papers in which the integrated square of a density appears never stops growing: the functional is often used to investigate the properties of hard-to-analyze semi-parametric estimators – see, for instance, Leonenko, Pronzato, and Savani (2008) for a study of some plug-in nearest-neighbors estimators – or to probe general semi-parametric inferential procedures – see, for instance, Kennedy, Balakrishnan, and Wasserman (2020) for a bias testing problem in a causal inference context.

When $X_1, X_2, \ldots, X_n$ are independently and identically distributed (i.i.d.) random variables with common density f_0 with respect to the Lebesgue measure on the real line, Bickel and Ritov (1988) first unearthed strong optimality bounds for the estimation of $\int_{\mathbb{R}} f_0^2(x)\, dx$ and built an estimator reaching these bounds under mild conditions. Two main alternative estimators were subsequently proposed in the literature and were proved to reach the optimality bounds of Bickel and Ritov (1988) under even milder conditions: an elegant estimator based on orthogonal series in Laurent (1996) and a simple kernel estimator in Giné and Nickl (2008) of the form

$$\frac{2}{n(n-1)h_n} \sum_{1 \le i < j \le n} K\left(\frac{X_i - X_j}{h_n}\right).$$

These estimators were proved to be asymptotically normal at parametric rates with minimal variance when the density belonged to a Sobolev space of order $s > 1/4$ and to reach the minimax-optimal slower-than-parametric convergence rates when $s \le 1/4$. In a recent paper, Robins, Li, Tchetgen Tchetgen, and van der Vaart (2016) showed that the estimator of Laurent (1996) remained asymptotically normal when $s < 1/4$. The first contribution of our paper is to show that the estimator of Giné and Nickl (2008) is also asymptotically normal when $s \le 1/4$. For this purpose, we leverage the structure of the estimator as a second-order U-statistics with n-dependent kernels and resort to the central limit theorem for generalized quadratic forms of de Jong (1987). This is the traditional approach to handle problems of the form – see, for instance, Hall (1984), Hardle and Mammen (1993), or Cattaneo, Crump, and Jansson (2014b). The main technical challenge in our setting comes from appropriately bounding the moments of the estimator under the weak regularity conditions used in Giné and Nickl (2008). We show that it is possible to do so by only leveraging the integrability conditions of the density. The fact that these bounds do not depend on the smoothness of the density allows us to obtain the weak limits of the estimator by first varying the convergence rates of its bandwidth sequence for any given Sobolev order $s > 0$. By then applying the result to an optimal bandwidth sequence, we directly obtain the different asymptotic regimes for the estimator in terms of the value taken by s. Our proof strategy to obtain the weak limits of the estimator in Giné and Nickl (2008) thus differs from the one used in Robins, Li, Tchetgen Tchetgen, and van der Vaart (2016) for the estimator in Laurent (1996), but still delivers qualitatively similar results (while allowing us also to handle the corner case $s = 1/4$). By following a traditional approach but under low-regularity conditions, our proof thus helps connect weak convergence results for second-order U-statistics with n-dependent kernels.

The second contribution of our paper is to tackle the problem of inference for the estimator in Giné and Nickl (2008) in view of its extended asymptotic normality. This problem was not considered by Robins, Li, Tchetgen Tchetgen, and van der Vaart (2016) for the estimator in Laurent (1996). In fact, it has never been handled before for any estimator of the integrated square of a density. We recover for the estimator in Giné and Nickl (2008) inferential results that are qualitatively similar to those obtained for other second-order U-statistics with n-dependent kernels – see, for instance, Hardle and Mammen (1993) and Cattaneo, Crump, and Jansson (2014b,a). In particular, we show that if the quadratic term of the estimator dominates, then the plug-in variance estimator is inconsistent and the non-parametric bootstrap fails, but it is still possible to construct consistent variance estimators by leveraging the quadratic nature of the problem. Compared to previous inferential results for U-statistics with n-dependent kernels, the main difference and challenge in our problem come again from its low-regularity structure. We obtain, as a result, what we believe is the first example in the literature of valid semi-parametric inference in the presence of an infinite-dimensional non-smooth nuisance parameter. Beyond the paradigmatic nature of this result, a number of practical implications directly follow from it. First, the rules we introduce can be used in practice to perform valid inference for the integrated square of a density when the density is known to be irregular ($s \leq 1/4$). This is, as far as we know, the first rules available for inference in this case. Moreover, as a consequence of the proof, we directly obtain new valid inferential rules for an extended range of bandwidth sequences when the density is known to be regular ($s > 1/4$). This can be used to motivate under-smoothing as a robust practice for inference on the integrated square of a density in the smooth case – a similar motivation for under-smoothing can be found in Cattaneo, Crump, and Jansson (2014b) for another smooth functional. In the course of this paper, we also build a new leave-one-out variance estimator by constructive methods whose simple structure may find applications in other inferential problem featuring second-order U-statistics with n-dependent kernels.

The rest of the paper is constructed as follows. In the remaining parts of the Introduction, we first introduce the estimator and the running hypotheses for the problem and then review Hoeffding's decomposition for U-statistics. In Section 2, we derive the moment bounds for the estimator and prove that the estimator remains asymptotically normal even when the density belongs to a Sobolev space of order $s \leq 1/4$. We obtain a first convergence result where the dominating term of the estimator depends on the rate of convergence of the bandwidth sequence of the estimator. In virtue of the assumption-lean moment bounds, no hypothesis on the smoothness of the density is needed in this case. We then use this result to derive the weak limit of the estimator when the Sobolev class for the density varies but the bandwidth sequence is fixed to the optimal one (trading off squared bias against variance). In Section 3, we tackle the problem of inference based on these newly derived weak limits. We show that while valid inference remains possible, many standard procedures break down in spite of asymptotic normality. By leveraging the structure of the problem, we first build a simple variance estimator and show its consistency by constructive methods in both the parametric and non-regular regimes. We then show that the plug-in variance estimator is inconsistent in the non-regular regime, but that a simple bias-correction restores consistency. These results are then used to show that the non-parametric bootstrap fails in the non-regular regime. Most of the proofs are collected in the Supplementary Material.

1.2 Estimator and hypotheses

Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables with common square-integrable density f_0 with respect to the Lebesgue measure on $\mathbb{R}$. The parameter of interest is the integrated square of the density f_0 given by

$$\theta_0 = \int_{\mathbb{R}} f_0^2(x)\, dx.$$

To estimate it, we consider the estimator in Giné and Nickl (2008) given by

$$U_n = \frac{2}{n(n-1)h_n} \sum_{1 \le i < j \le n} K\left(\frac{X_i - X_j}{h_n}\right),$$

where $K \colon \mathbb{R} \to \mathbb{R}$ is a smoothing kernel with associated bandwidth h_n. The estimator U_n is obtained by first plugging a kernel density estimator in the empirical counterpart to the moment condition defining θ_0 and then removing the diagonal elements in the double sum. The estimator U_n is directly seen to be a second-order U-statistics with n-dependent kernel k_n given by

$$k_n(X_i, X_j) = \frac{1}{h_n} K\left(\frac{X_i - X_j}{h_n}\right)$$

for any $1 \le i < j \le n$. The n-dependence of the kernel stems from the presence of h_n.

For our study of the estimator U_n, we reuse the assumptions considered in Giné and Nickl (2008) for the kernel function K and the density f_0. These assumptions are collected below in Assumption K and Assumption D(s), respectively. Fundamentally, the regularity of the problem is controlled by assuming that f_0 belongs to a Sobolev space and by varying its smoothness parameter $s > 0$: the lower the value of s, the less regular the density f_0.

To state these conditions, we introduce some standard notations. For $1 \le p < \infty$, we denote by $L^p = L^p(\mathbb{R}) = L^p(\mathbb{R}; \lambda)$ the space of p-integrable functions with respect to the Lebesgue measure λ and endow it with the p-norm $\|\phi\|_p^p = \int_{\mathbb{R}} \phi(x)^p\, dx$. For $\phi \in L^1$, we define the Fourier transform by $F\phi(u) = \int_{\mathbb{R}} e^{-iux} \phi(x)\, dx$ and we extend it by continuity to L^2.

Assumption K. *The kernel $K \colon \mathbb{R} \to \mathbb{R}$ satisfies:*

1. *K is symmetric and bounded;*
2. *$\int K(u)\, du = 1$;*
3. *$\int |K(u)||u|\, du < \infty$.*

Remark 1.1. By assuming K bounded and integrable, we have $K \in L^1 \cap L^\infty$. It follows then that $K \in L^p$ for $1 \le p \le \infty$. Indeed, given an arbitrary measure μ on a space X, define $A = \{x \in X : |f(x)| > 1\}$, then $\mu(A) \le \int_A |f(x)|\, d\mu \le \|f\|_1$ and $|f(x)|^p \le |f(x)|$ on $X \setminus A$, hence $\int_X |f(x)|^p\, d\mu \le \int_A |f(x)|^p\, d\mu + \int_{X \setminus A} |f(x)|^p\, d\mu \le \|f\|_\infty^p \|f\|_1 + \|f\|_1 < \infty$.

Assumption D(s). *The density f_0 satisfies:*

1. *f_0 is bounded;*
2. *$f_0 \in H_2^s$, where $H_2^s = W^{2,s}(\mathbb{R})$ is the Sobolev space of integrability $p = 2$ and of order s, that is,*

$$H_2^s = \left\{\phi \in L^2 : \|\phi\|_{2,s} = \|F\phi(\cdot)(1 + |\cdot|^2)^{s/2}\|_2 < \infty\right\}.$$

Remark 1.2. Since f_0 is a density for some random variable, we implicitly assume that $\int_{\mathbb{R}} f_0(x)\,dx = 1$ and so $f_0 \in L^1$. Boundedness of f_0 implies that $f_0 \in L^\infty$. Then $f_0 \in L^1 \cap L^\infty$, and so, by a similar argument as Remark 1.1, we have $f_0 \in L^p$ for any $1 \le p \le \infty$. It follows, in particular, that f_0 is square-integrable (which guarantees that θ_0 is well-defined).

Remark 1.3. It does not follow from these assumptions that f_0 is necessarily continuous. In particular, if $f_0 \in H_2^s$ with $s < 1/2$, then f_0 can be discontinuous. Indeed, the Sobolev embedding theorem only ensures continuity of $f_0 \in H_2^s$ if $s \ge 1/2$. This is not an issue since continuity is not needed. However, we will use repeatedly the integrability assumption that f_0 is in $L^1 \cap L^\infty$. For instance, we will make use of L^1-continuity, that is, the fact that if $f_0 \in L^1$, then

$$\lim_{|t|\to 0} \int |f_0(x+t) - f_0(x)|\,dx = 0.$$

Other results of the sort will be used – they are based on a density argument using the fact that compactly supported continuous functions are dense in L^1.

Remark 1.4. The smoothness and integrability conditions in Assumption D(s) are the one considered by Giné and Nickl (2008). Under Assumption D(s), the authors showed that the bias $B_n = \mathbb{E}[U_n] - \theta_0$ of U_n satisfied $B_n = O(h^{2s})$ where s is the Sobolev order for the density class. For completeness, the proof is reproduced in Section D of the Supplementary Material. It is important to note, however, that we are able in Section 2 to derive the weak limit of $\sigma(U_n)^{-1}(U_n - \mathbb{E}[U_n])$ without using the smoothness assumption in D(s) but only the integrability condition $f_0 \in L^1 \cap L^\infty$. It is only when looking at the centered quantity $\sigma(U_n)^{-1}(U_n - \theta_0)$ that the smoothness parameter s will play a role through the rate of decay of the bias. This has important consequences when compared to previous results in the literature as explained in the introductions of Section 2 and Section 3.

Remark 1.5 (On relaxing Assumption D(s)). The smoothness condition in Assumption D(s) is already (much) more general than those used in the literature on second-order U-statistics with n-dependent kernels – see, e.g., Hall (1984), Hall and Marron (1987), Hardle and Mammen (1993), or Cattaneo, Crump, and Jansson (2014b,a). It is the same smoothness assumption as considered in Laurent (1996). It can be extended at no cost to the slightly more general class considered in Laurent (2005) – see Section D of the Supplementary Material. The L^∞ integrability assumption in Assumption D(s) is more crucial, both for the bias and the weak limit – see Remark 2 in Giné and Nickl (2008) for a relaxation for the bias in the case of the Lipschitz class of Bickel and Ritov (1988).

Remark 1.6 (On multivariate extensions). We follow Giné and Nickl (2008) and focus on the one-dimensional case $d = 1$. This allows us to reuse their Fourier argument for handling the bias without modification. It also greatly simplifies notations when deriving the tedious moment bounds for the weak limits. This also allows us to work out direct arguments from which we can unearth a new simple variance estimator than can be applied to other similar problems. Extending our results to higher dimensions $d > 1$ is of interest, especially to investigate the effects of the order of the kernel on convergence and inference. These extensions are left for future research.

1.3 A preliminary Hoeffding decomposition

Most of the arguments we will make depend on the Hoeffding decomposition of the second-order U-statistics U_n. This is a well-known approach that dates back to Hoeffding (1948). Because the decomposition will be used repeatedly, we collect in the next lemma the different terms entering into the decomposition. It is useful to introduce the following notations

$$\begin{aligned} u_n^0 &= \mathbb{E}\left[U_n\right] = \mathbb{E}\left[k_n(X_i, X_j)\right], \\ u_n^1(X_i) &= \mathbb{E}\left[k_n(X_i, X_j)|X_i\right] \\ u_n^2(X_i, X_j) &= k_n(X_i, X_j), \end{aligned}$$

where $i \neq j$ are any two indexes.

Lemma 1.1 (Hoeffding Decomposition of U_n)**.** *The statistics U_n admits the following Hoeffding decomposition*

$$U_n = \mathbb{E}\left[U_n\right] + 2L_n + W_n \tag{1.1}$$

where

$$\mathbb{E}\left[U_n\right] = u_n^0,$$

$$L_n = \frac{1}{n}\sum_{i=1}^{n}\left[u_n^1(X_i) - u_n^0\right],$$

and

$$W_n = \binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[u_n^2(X_i, X_j) - u_n^1(X_i) - u_n^1(X_j) + u_n^0\right].$$

Proof. This follows from Theorem 1 in Section 1.6. in Lee (1990). In this case, the proof is simpler. The equality follows directly by expanding the terms. The fact that $2L_n$ is a L^2-projection follows by verifying that $\mathbb{E}\left[(U_n - 2L_n)\sum_{i=1}^{n} g_i(X_i)\right] = 0$. □

Lemma 1.2 (Variance of U_n)**.** *The variance of U_n is given by*

$$\operatorname{Var} U_n = \frac{4}{n}\operatorname{Var}(u_n^1(X_1)) + \frac{2}{n(n-1)}\operatorname{Var}(u_n^2(X_1, X_2) - u_n^1(X_1) - u_n^1(X_2)).$$

Proof. This follows directly from Theorem 4 in Section 1.6. in Lee (1990). For completeness, we rapidly sketch the proof. By construction, $2L_n$ and W_n are uncorrelated. In particular, we have

$$\operatorname{Var} U_n = 4\operatorname{Var} L_n + \operatorname{Var} W_n. \tag{1.2}$$

Since the components of L_n are i.i.d. (as measurable functions of X_i), we have

$$\operatorname{Var} L_n = \frac{1}{n}\operatorname{Var}(u_n^1(X_1) - u_n^0) = \frac{1}{n}\operatorname{Var}(u_n^1(X_1)). \tag{1.3}$$

Since the components of W_n are uncorrelated for any four indexes $i < j, k < l$ such that at least three are different, we have

$$\begin{aligned}
\operatorname{Var} W_n &= \frac{2}{n(n-1)}\operatorname{Var}(u_n^2(X_1, X_2) - u_n^1(X_1) - u_n^1(X_2) + u_n^0) \\
&= \frac{2}{n(n-1)}\operatorname{Var}(u_n^2(X_1, X_2) - u_n^1(X_1) - u_n^1(X_2)) \\
&= \frac{2}{n(n-1)}\Big[\operatorname{Var}(u_n^2(X_1, X_2)) - 2\operatorname{Var}(u_n^1(X_1))\Big],
\end{aligned} \tag{1.4}$$

where the last equality follows since $\operatorname{cov}(u_n^2(X_1, X_2), u_n^1(X_1)) = \operatorname{Var}(u_n^1(X_1))$ and $\operatorname{cov}(u_n^1(X_1), u_n^1(X_2)) = 0$ (by independence of X_1 and X_2). □

From the proof, we get the alternative expression

$$\operatorname{Var} U_n = \frac{4}{n}\operatorname{Var}(u_n^1(X_1)) + \frac{2}{n(n-1)}\Big[\operatorname{Var}(u_n^2(X_1, X_2)) - 2\operatorname{Var}(u_n^1(X_1))\Big].$$

We also get the two useful inequalities

$$\operatorname{Var} L_n \le \frac{1}{n}\mathbb{E}\,[(u_n^1(X_1))^2],$$

and

$$\operatorname{Var} W_n \le \frac{2}{n(n-1)}\operatorname{Var}(u_n^2(X_1, X_2)) \le \frac{2}{n(n-1)}\mathbb{E}\,[(u_n^2(X_1, X_2))^2].$$

2 Asymptotic normality in the non-regular regime

2.1 Weak limits with varying bandwidth rates and arbitrary smoothness

In this section, we show that the estimator U_n remains asymptotically normal whether its linear term dominates or its quadratic term dominates. To show this, we derive the weak limits of the estimator U_n for different convergence rates of the bandwidth sequence h_n to 0 as $n \to \infty$ while keeping the smoothness of the density fixed to some arbitrary level $s > 0$. This translates into the following assumptions:

1. Assumption K;

2. Assumption D(s) for some arbitrary $s > 0$.

Under these assumptions, we show in Corollary 2.3 that whenever:

- $nh_n \to \infty$: the linear term of U_n dominates and a standard central limit theorem delivers asymptotic normality with variance the semi-parametric lower bound;
- $nh_n \to 0$: the quadratic term of U_n dominates and a central limit theorem for quadratic forms delivers asymptotic normality of U_n; in this case, the variance depends on the kernel and convergence is slower-than-parametric and depends on the bandwidth sequence.

We can further show that if:

- $nh_n \to (0, \infty)$: the linear term and the quadratic term of U_n have the same order, the weak limit is still normal at the parametric rate, but the asymptotic variance depends on the kernel;

This is a standard behavior for second-order U-statistics with n-dependent kernels that is well-understood – see, for instance, Robins, Li, Tchetgen Tchetgen, and van der Vaart (2016). However, our proof follows a more traditional approach where the weak convergence when the quadratic term dominates is handled by resorting to the central limit theorem of Hall (1984) as generalized in de Jong (1987) – the same approach was notably used in Hardle and Mammen (1993) and Cattaneo, Crump, and Jansson (2014b) for some other smooth functionals also estimated by some second-order U-statistics with n-dependent kernels. The main challenge in our setting comes from bounding the moments of U_n under the weak regularity conditions of Giné and Nickl (2008). We show that it is possible to bound these moments sufficiently by using a density argument that only leverages the integrability conditions for the density. The qualitative nature of the resulting bounds may be of independent interest for other integral functionals. This low-regularity setting differs markedly from the other results in the literature that use de Jong (1987) – it will bear a few interesting fruits later. We collect the bounds in Lemma 2.3, Lemma 2.4, and Lemma 2.5. Before stating them, we characterize the asymptotic variance of U_n in Lemma 2.1 and Lemma 2.2 – this will be needed to obtain a closed form for the weak limit.

Lemma 2.1.

$$\lim_{n\to\infty} n\mathrm{Var}\, L_n = \int_{\mathbb{R}} f_0(x)^3\, dx - \left(\int_{\mathbb{R}} f_0(x)^2\, dx\right)^2.$$

Proof. For simplicity, we write $h_n = h$. From Equation (1.3), we have

$$n\mathrm{Var}\, L_n = \mathbb{E}\,[(u_n^1(X_1))^2] - \mathbb{E}\,[u_n^1(X_1)]^2.$$

We first have

$$\begin{aligned}\mathbb{E}\,[(u_n^1(X_1))^2] &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} K_h(x-y) f_0(y)\, dy\right)^2 f_0(x)\, dx \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} K_h(u) f_0(x-u)\, du\right)^2 f_0(x)\, dx.\end{aligned}$$

Since $K \in L^1$ and $f_0 \in L^2$, we have by the mollification theorem (see Theorem 8.14. in Folland (1999)) that $\int_{\mathbb{R}} K_h(u) f_0(x-u)\, du$ converges in L^2 to f_0 as $h \to 0$. That is,

$$\lim_{h \to 0} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} K_h(u) f_0(x-u)\, du - f_0(x) \right)^2 f_0(x)\, dx = 0.$$

Then, by continuity of the norm, we directly get that

$$\lim_{h \to 0} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} K_h(u) f_0(x-u)\, du \right)^2 f_0(x)\, dx = \int_{\mathbb{R}} f_0(x)^3\, dx.$$

For the limit of $\mathbb{E}\left[(u_n^1(X_1))\right]^2$, we can directly invoke a density argument that extends L^1-continuity. It is proved in Section B of the Supplementary Material. We first have

$$\begin{aligned} \mathbb{E}\left[(u_n^1(X_1))\right] &= \int_{\mathbb{R}} \int_{\mathbb{R}} K_h(x-y) f_0(y)\, dy\, f_0(x)\, dx \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} K(u) f_0(x-uh)\, du\, f_0(x)\, dx \\ &= \int_{\mathbb{R}} K(u) \int_{\mathbb{R}} f_0(x-uh) f_0(x)\, dx\, du. \end{aligned} \tag{2.1}$$

Then, as proved in Section B of the Supplementary Material, we have

$$\lim_{h \to 0} \int_{\mathbb{R}} |f_0(x-uh) f_0(x) - f_0(x)^2|\, dx = 0.$$

Then, by dominated convergence, we have

$$\begin{aligned} \lim_{h \to 0} \int_{\mathbb{R}} K(u) \int_{\mathbb{R}} f_0(x-uh) f_0(x)\, dx\, du &= \int_{\mathbb{R}} K(u) \int_{\mathbb{R}} f_0(x)^2\, dx\, du \\ &= \int_{\mathbb{R}} f_0(x)^2\, dx. \end{aligned}$$

This concludes the proof by composition of limits. □

Remark 2.1. This result already appears in Theorem 1 in Giné and Nickl (2008) but our proofs differ. Their proof is based on showing mean squared convergence of L_n towards an i.i.d. sum with $Y_i = f_0(X_i) - \int_{\mathbb{R}} f_0(x)^2\, dx$ and then obtaining convergence in variance from the triangular inequality and continuity of the norm. Our proof is more direct and is provided because it is based on a density argument that will be used repeatedly in this paper. The result would be direct if f_0 were continuous and compactly supported, but we did not assume continuity nor compact support. However, functions in L^1 are "approximately" such in the sense that continuous compactly supported functions are dense in L^1. This is this approximation that is used in proving the result in Section B of the Supplementary Material as well as the mollification theorem and L^1-continuity.

Lemma 2.2.

$$\lim_{n \to \infty} \binom{n}{2} h_n \operatorname{Var} W_n = \int_{\mathbb{R}} f_0(x)^2\, dx \int_{\mathbb{R}} K(u)^2\, du.$$

Proof. For simplicity, we write $h_n = h$. From Equation (1.4), we have

$$\binom{n}{2}\operatorname{Var} W_n = \operatorname{Var}(u_n^2(X_1, X_2)) - 2\operatorname{Var}(u_n^1(X_1)).$$

From the proof of Lemma 2.1, we know that $\operatorname{Var}(u_n^1(X_1)) = O(1) = o(h^{-1})$ and $\mathbb{E}\,[u_n^2(X_1, X_2)]^2 = \mathbb{E}\,[u_n^1(X_1)]^2 = O(1) = o(h^{-1})$. It remains to handle $\mathbb{E}\,[(u_n^2(X_1, X_2))^2]$. We have

$$\begin{aligned} h\mathbb{E}\,[(u_n^2(X_1, X_2))^2] &= h\int_{\mathbb{R}}\int_{\mathbb{R}}(K_h(x-y))^2 f_0(y)\,dy\, f_0(x)\,dx \\ &= h\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{1}{h}K(u)^2 f_0(x-uh)\,du\, f_0(x)\,dx \\ &= \int_{\mathbb{R}} K(u)^2\int_{\mathbb{R}} f_0(x-uh) f_0(x)\,dx\,du \end{aligned}$$

As in Equation (2.1), we can conclude by using the result proved in Section B of the Supplementary Material and dominated convergence that

$$\lim_{h\to 0}\int_{\mathbb{R}} K(u)^2\int_{\mathbb{R}} f_0(x-uh) f_0(x)\,dx\,du = \int_{\mathbb{R}} f_0(x)^2\,dx\int_{\mathbb{R}} K(u)^2\,du.$$

□

From this results, it follows directly that $\operatorname{Var}(\sqrt{n}L_n) = O(1)$ and $\operatorname{Var}(\sqrt{n}W_n) = O((nh_n)^{-1})$, so that the dominating terms in the Hoeffding decomposition depends on $\lim_{n\to\infty} nh_n$. To obtain the weak limit, we have to be more precise and bound higher moments of L_n and W_n. This is the objective of Lemma 2.3, Lemma 2.4, and Lemma 2.5 which are proved in Section A.1. of the Supplementary Material. The proofs make use of a recurring density argument that is stated and proved in Section B of the Supplementary Material. This argument is similar to the one used to prove Lemma 2.1 and Lemma 2.2.

Lemma 2.3. *Let $i, j, k \in \{1, 2 \dots, n\}$ with $i \neq j \neq k$. Let $q, r \geq 1$ be integers. Then*

1.
$$\mathbb{E}\,[|u_n^1(X_i)|^q] = O(1);$$

2.
$$\mathbb{E}\,[|u_n^2(X_i, X_j)|^q] = O(h^{-q+1});$$

3.
$$\mathbb{E}\,[|u_n^2(X_i, X_j)|^r |u_n^1(X_i)|^q] = O(h^{-r+1/2});$$

4.
$$\mathbb{E}\,[|u_n^2(X_i, X_j)|^r |u_n^2(X_i, X_k)|^q] = O(h^{-r-q+2});$$

Lemma 2.4. *Let C_2^n denote the set of all pairs (i, j) with $i < j$, $1 \leq i, j \leq n$. For $(i, j) \in C_2^n$, let $r_{i,j} \geq 0$ be non-negative integers. For $k \in \{1, 2, \dots, n\}$, let $s_k \geq 0$ be non-negative integers. Suppose $r_{i,j} \geq 1$ for at least one pair $(i, j) \in C_2^n$. Suppose that $s_k \geq 1$ for at least one index*

$k \in \{1, 2, \ldots, n\}$. Consider the product $\Pi_{(i,j)\in C_2^n}(u_n^2(X_i, X_j))^{r_{i,j}}$. Denote $l \in \{1, 2, \ldots, n\}$ the number of indexes i such that $r_{i,j} \neq 0$ for at least one index $j \in \{1, 2, \ldots, i-1, i+1, \ldots, n\}$. Then

5.

$$\mathbb{E}\,[\Pi_{(i,j)\in C_2^n}|u_n^2(X_i, X_j)|^{r_{i,j}}] = O(h^{-\sum_{(i,j)} r_{i,j}+l-1});$$

6.

$$\mathbb{E}\,[\Pi_{(i,j)\in C_2^n}|u_n^2(X_i, X_j)|^{r_{i,j}}\Pi_{k=1}^n|u_n^1(X_k)|^{s_k}] = O(h^{-\sum_{(i,j)} r_{i,j}+\frac{l-1}{2}});$$

By using two nested density arguments, result (3.) and (6.) of last lemmas can probably be improved to $O(|h|^{-r+1})$ and $O(h^{-\sum_{(i,j)} r_{i,j}+l-1})$, respectively, but the result is not needed to prove the main result of this section and so we only resort to a cruder bound based on Holder's inequality. As shown in next lemma, the price to pay is a cruder and more cumbersome bound for the higher moments of W_n, which remains nevertheless sufficient for our purpose.

Lemma 2.5. *Define*

$$\begin{aligned} l(X_i) &= u_n^1(X_i) - u_n^0, \\ w(X_i, X_j) &= u_n^2(X_i, X_j) - u_n^1(X_i) - u_n^1(X_j) + u_n^0 \end{aligned}$$

for $1 \le i, j \le n$, $i \neq j$. Let $i, j, k \in \{1, 2 \ldots, n\}$ with $i \neq j \neq k$. Let $q, r \geq 1$ be integers. Then

1.

$$\mathbb{E}\,[|l(X_i)|^q] = O(1);$$

2.

$$\mathbb{E}\,[|w(X_i, X_j)|^q] = O(h^{-q+1});$$

3.

$$\mathbb{E}\,[|w(X_i, X_j)|^r|w(X_i, X_k)|^q] = O(h^{-r-q+2});$$

4.

$$\mathbb{E}\,[\Pi_{(i,j)\in C_2^n}|w(X_i, X_j)|^{r_{i,j}}] = O(h^{-\sum_{(i,j)} r_{i,j}+\frac{l+1}{2}} \vee h^{-\sum_{(i,j)} r_{i,j}+l-1});$$

where the C_2^n and l are defined as in Lemma 2.4 for w instead of u_n^2.

We are now ready to state the main result of this section, a central limit theorem for the vector with elements properly standardized terms L_n and W_n, from which the weak limit of $(\operatorname{Var} U_n)^{-1/2}(U_n - \mathbb{E}\,[U_n])$ can be directly derived. The proof is relegated to Section A.2. of the Supplementary Material. The idea is to use the previously derived moment bounds to apply the central limit theorem for quadratic forms of de Jong (1987).

Proposition 2.1. *If $n^2 h_n \to \infty$, then the terms L_n and W_n in the Hoeffding decomposition (1.1) converges jointly in distribution to a bivariate normal distribution*

$$\begin{pmatrix} \sqrt{n} L_n \\ \sqrt{\binom{n}{2} h_n} W_n \end{pmatrix} \rightsquigarrow \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_L^2 & 0 \\ 0 & \sigma_W^2 \end{pmatrix}\right]$$

where

$$\sigma_L^2 = \int_{\mathbb{R}} f_0(x)^3\,dx - \left(\int_{\mathbb{R}} f_0(x)^2\,dx\right)^2,$$
$$\sigma_W^2 = \int_{\mathbb{R}} f_0(x)^2\,dx \int_{\mathbb{R}} K(u)^2\,du.$$

From the Hoeffding decomposition for U_n, Lemma 2.1, and Lemma 2.2, we can directly derive the weak limit of $(\operatorname{Var} U_n)^{-1/2}(U_n - \mathbb{E}[U_n])$ from the previous result. The proof can be found in Section A.2. of the Supplementary Material.

Corollary 2.2. *If $n^2 h_n \to \infty$, then*

$$(\operatorname{Var} U_n)^{-1/2}(U_n - \mathbb{E}[U_n]) \rightsquigarrow N(0,1).$$

We can now make use of the bias result in Giné and Nickl (2008) – see Section D of the Supplementary Material – to obtain the weak limits of $(\operatorname{Var} U_n)^{-1/2}(U_n - \theta_0)$. The proof is relegated to Section A.2. of the Supplementary Material.

Corollary 2.3. *1. If $nh_n \to \infty$ and $nh_n^{4s} \to 0$, then*

$$\sqrt{n}(U_n - \theta_0) \rightsquigarrow N(0, 4\sigma_L^2).$$

2. If $nh_n \to C \in (0,\infty)$ and $nh_n^{4s} \to 0$, then

$$\sqrt{n}(U_n - \theta_0) \rightsquigarrow N\big(0, 4\sigma_L^2 + \frac{2}{C}\sigma_W^2\big).$$

3. If $n^2 h_n \to \infty$, $nh_n \to 0$, and $nh_n^{2s+\frac{1}{2}} \to 0$, then

$$\sqrt{\binom{n}{2} h_n}\,(U_n - \theta_0) \rightsquigarrow N(0, \sigma_W^2).$$

Remark 2.2. Conditions in (1.) and (2.) can only hold if $s > 1/4$. The last two conditions in (3.) have the following relation: if $s > 1/4$, then $nh_n \to 0$ implies $nh_n^{2s+\frac{1}{2}} \to 0$; if $s < 1/4$, then $nh_n^{2s+\frac{1}{2}} \to 0$ implies $nh_n \to 0$. For (2.), it would be of interest to relax the conditions to $nh_n^{4s} \to C' \in (0,\infty)$, but it is not clear if such a result holds: the bias result of Giné and Nickl (2008) only gives $B_n = O(h^{2s})$ – see Section D of the Supplementary Material.

2.2 Weak limits with varying smoothness and optimal bandwidth sequence

The asymptotic normality of U_n obtained in Corollary 2.2 can be applied directly when an optimal bandwidth sequence h_n is used in U_n. Indeed, if h_n satisfies Assumption OB below, then we necessarily have $n^2 h_n \to \infty$. In this case, the dominating term for the weak limit depends on the order $s > 0$ of the Sobolev class to which the density f_0 belongs: if $s > 1/4$, the linear term dominates; if $s = 1/4$, the linear term and the quadratic term have the same order; if $s < 1/4$, the quadratic term dominates. We then recover an asymptotic regime whose form is more traditional and directly comparable to the results in Bickel and Ritov (1988), Laurent (1996), Giné and Nickl (2008), or Robins, Li, Tchetgen Tchetgen, and van der Vaart (2016). It is seen, in particular, to directly extend the asymptotic normality of U_n first obtained in the parametric regime ($s > 1/4$) in Giné and Nickl (2008) to the non-regular regime ($s \leq 1/4$). We collect this result in Corollary 2.4 and prove it formally in Section A.2. of the Supplementary Material. In addition to Assumption OB, we suppose again that Assumption K and Assumption D(s) hold for some $s > 0$.

Assumption OB. *Given Assumption D(s) for some $s > 0$, the bandwidth sequence h_n satisfies*

$$0 < h_n = Cn^{-\frac{2}{4s+1}}.$$

for some constant $C > 0$.

Corollary 2.4. *Suppose that the bandwidth sequence h_n satisfies Assumption OB.*

1. If $s > 1/4$, then

$$\sqrt{n}(U_n - \theta_0) \rightsquigarrow N(0, 4\sigma_L^2).$$

2. If $s = 1/4$, then

$$\sqrt{n}(U_n - \mathbb{E}[U_n]) \rightsquigarrow N\left(0, 4\sigma_L^2 + \frac{2}{C}\sigma_W^2\right).$$

3. If $s < 1/4$, then

$$\sqrt{\binom{n}{2}h_n}(U_n - \mathbb{E}[U_n]) \rightsquigarrow N(0, \sigma_W^2).$$

Remark 2.3. Under Assumption OB, the bias of U_n is negligible when $s > 1/4$, and so the weak limit can be centered at the true value θ_0 in this case. When $s \leq 1/4$, the estimator is not necessarily unbiased, since then we only have $\sqrt{n}B_n = O(1)$ and $\sqrt{\binom{n}{2}h_n}B_n = O(1)$ by the bias result in Giné and Nickl (2008) – see Section D of the Supplementary Material. The asymptotic negligibility of the bias when $s \leq 1/4$ can still be obtained from Corollary 2.2 by changing the bandwidth sequence in Assumption OB to a sub-optimal under-smoothed one. This is made explicit and leveraged in next section when we consider the problem of inference in the case $s \leq 1/4$.

Remark 2.4. The proof is of interest for the theory of second-order U-statistics with n-dependent kernels. We show, indeed, that it is possible by following the more traditional approach of Hall (1984) generalized in the central limit theorem of de Jong (1987) to recover weak convergence results that are qualitatively similar to those obtained in Robins, Li, Tchetgen Tchetgen, and van der Vaart (2016) in the sense that they can be expressed in terms of the smoothness of the nuisance parameter. This connection is possible in virtue of the moment bounds we obtained for U_n from the integrability of the nuisance parameter and not its smoothness. This is an essential difference with the previous results on U-statistics with n-dependent kernels that used de Jong (1987) – see, for instance, Hardle and Mammen (1993) and Cattaneo, Crump, and Jansson (2014b).

3 Semi-parametric inference in the non-regular regime

In this section, we tackle the problem of inference for the integrated square of a density using the estimator U_n of Giné and Nickl (2008) and its extended asymptotic normality proved in Section 2. The results we obtain not only extend the ranges of cases for which it is possible to quantify uncertainty around an optimal estimator of $\int_{\mathbb{R}} f_0^2(x)\, dx$ but also provides the first example in the literature of a semi-parametric problem where inference is possible in spite of an infinite-dimensional non-smooth nuisance parameter. We recover, moreover, many standard inferential results for second-order U-statistics with n-dependent kernels but in a low-regularity setting.

The idea is to start from the primitive weak limit derived in Corollary 2.2 with the hope of extending the range of valid inference from cases when the linear term of U_n dominates ($nh_n \to \infty$) to cases when the quadratic term of U_n dominates ($nh_n \to 0$ provided $n^2 h_n \to \infty$). For this approach to be valid, we first need to ensure asymptotic negligibility of the bias. From Remark 2.3, this is seen to hold under two different sets of conditions compatible with $n^2 h_n \to \infty$, either:

Assumption NB1. *Assumption D(s) holds with $s > 1/4$ and the bandwidth sequence satisfies $nh_n^{4s} \to 0$;*

or:

Assumption NB2. *Assumption D(s) holds with $s \leq 1/4$ and the bandwidth sequence satisfies $nh_n^{2s+\frac{1}{2}} \to 0$.*

We can then consider using result 3 in Corollary 2.3 for inference under two new sets of non-regular cases corresponding to Assumption NB1 and Assumption NB2, respectively:

(A) the density is regular enough ($s > 1/4$) and the estimator belongs to a wide class of under-smoothed estimators (with bandwidths from $nh_n \to 0$ to $n^2 h_n \to \infty$);

(B) the density is irregular ($s \leq 1/4$) and the estimator belongs to a slightly narrower class of under-smoothed estimators (with bandwidths from $nh_n^{2s+\frac{1}{2}} \to 0$ to $n^2 h_n \to \infty$).

For either of these two cases, the non-regular asymptotic normality corresponding to result 3 in Corollary 2.3 can be used for inference since in each case we both have that $nh_n \to 0$ and $nh_n^{2s+\frac{1}{2}} \to 0$. Under this non-regular regime, we start by showing that it is possible to construct a simple leave-one-out estimator of the variance that is consistent – see Proposition 3.1. We then prove that the plug-in variance estimator is inconsistent in this case but that consistency can be restored by appropriate bias-correction – see Proposition 3.2 and Proposition 3.3. We finally derive the failure of the non-parametric bootstrap in this non-regular regime – see Proposition 3.4.

Remark 3.1. Because Assumption NB1 is also compatible with results 1 and 2 of Corollary 2.3 (when $nh_n \to C \in (0, +\infty]$) while Assumption NB2 is not as explained in Remark 2.3, we should work under Assumption NB1 in the rest of this section for additional generality: it allows us, indeed, to consider the cases when $nh_n \to \infty$ and when $nh_n \to C \in (0, +\infty)$ while Assumption

NB2 does not. It is important to note, however, that all the results we derive in this section for the non-regular case (A) under Assumption NB1 (that is, when $nh_n \to 0$) hold equivalently for the non-regular case (B) under NB2 since in both cases we have $nh_n \to 0$ and $nh_n^{2s+\frac{1}{2}} \to 0$ and so result 3 in Corollary 2.3 can be used.

Remark 3.2. The results in this section directly extend the range of cases where valid inference for $\int_{\mathbb{R}} f_0^2(x)\, dx$ can be performed. Under case (B), inference is possible when the density is known to be irregular ($s \leq 1/4$): we believe this is the first example in the literature of a semi-parametric problem where inference is possible in spite of an infinite-dimensional non-smooth nuisance parameter. Under case (A), inference can also be performed when the density is known to be regular ($s > 1/4$) and the estimator is under-smoothed; because inference remains valid under the same rules when the bandwidth is optimal, we obtain inferential rules robust to hyper-parameter selection when the density is known to be regular. A similar motivation for robustness to "small bandwidth" can be found in Cattaneo, Crump, and Jansson (2014b) for a different smooth functional. In their problem, there is no counterpart to case (A) because their results are only valid when their nuisance parameter is highly regular.

Remark 3.3. The results in this section also contribute to the theory of U-statistics with n-dependent kernels from the inferential side. We not only recover the qualitative features exhibited previously in Hardle and Mammen (1993) and Cattaneo, Crump, and Jansson (2014b,a) but do so in a low-regularity setting. We construct, moreover, a leave-one-out variance estimator whose simple structure may be appropriately leveraged in similar problems.

3.1 A simple consistent leave-one-out variance estimator

We show in this subsection how to construct an estimator of $\operatorname{Var} U_n$ that remains consistent in the non-regular regime corresponding to either case (A) or case (B). As made clear in Remark 3.1, we should work for additional generality under Assumption NB1: this allows us to consider both regular and non-regular inference together. The consistency in the non-regular regime, however, is valid both under Assumption NB2 and case (B) and under Assumption NB1 and case (A).

The main idea behind our estimator is inspired by the variance estimator in Cattaneo, Crump, and Jansson (2014b). The main difference is that we carefully unpack all the terms in the expansion leading to a simpler estimator that could be applied to other inferential problems with second-order U-statistics with n-dependent kernels. The cost of this simpler estimator is a more complicated proof. Based on Equation (1.3), Equation (1.4), and the results in Lemma 2.1 and Lemma 2.2, it is natural to consider the estimators

$$\begin{aligned}\widetilde{\sigma_L^2} =& n^{-1} \sum_{i=1}^{n} \binom{n-1}{2}^{-1} \sum_{\substack{j=1 \\ j\neq i}}^{n-1} \sum_{\substack{k=j+1 \\ k\neq i}}^{n} k_n(X_i, X_j) k_n(X_i, X_k) \\ &- \left(\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_n(X_i, X_j) \right)^2\end{aligned}$$

and

$$\widetilde{\sigma_W^2} = h_n \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_n(X_i, X_j)^2.$$

Then, based on Equation (1.2), it is natural to consider the estimator of $\operatorname{Var} U_n$ given by

$$\widetilde{V}_n = 4n^{-1}\widetilde{\sigma_L^2} + \binom{n}{2}^{-1} h_n^{-1}\widetilde{\sigma_W^2}.$$

We show that this estimator is consistent for the asymptotic variance. Given the characterization of $\operatorname{Var} U_n$

$$\operatorname{Var} U_n = 4n^{-1}\Big(\sigma_L^2 + o(1)\Big) + \binom{n}{2}^{-1} h_n^{-1}\Big(\sigma_W^2 + o(1)\Big),$$

the proof of the consistency of $\widetilde{V}_n$ follows directly from the lemma stated below. The proof of this lemma is long and is relegated to Section A.3. of the Supplementary Material.

Lemma 3.1. *If $n^2 h_n \to \infty$, then:*

1. $$\widetilde{\sigma_L^2} = \sigma_L^2 + o_P(1);$$

2. $$\widetilde{\sigma_W^2} = \sigma_W^2 + o_P(1).$$

Proposition 3.1. *If $n^2 h_n \to \infty$, then*

$$\widetilde{V}_n^{-1/2}(U_n - \theta_0) \rightsquigarrow N(0, 1)$$

with

$$\widetilde{V}_n = 4n^{-1}\widetilde{\sigma_L^2} + \binom{n}{2}^{-1} h_n^{-1}\widetilde{\sigma_W^2}.$$

Proof. This follows directly from Lemma 3.1, the characterization of $\operatorname{Var} U_n$ in Equation (A.1), and an application of Slutsky's theorem in Corollary 2.3. □

Remark 3.4. The form of $\widetilde{\sigma_L^2}$ stems from the moment characterization

$$n\operatorname{Var} L_n = \mathbb{E}\,[u_n^1(X_1)^2] - \mathbb{E}\,[u_n^1(X_1)]^2.$$

The first term in the estimator $\widetilde{\sigma_L^2}$ is then a leave-one-out bias-corrected estimator of the plug-in estimator for $\mathbb{E}\,[u_n(X_1)^2]$. Indeed, we have that

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\Big((n-1)^{-1}\sum_{\substack{j=1\\ j\neq i}}^{n} k_n(X_i,X_j)\Big)^2 &= n^{-1}\sum_{i=1}^{n}(n-1)^{-2}\sum_{\substack{j=1\\ j\neq i}}^{n} k_n(X_i, X_j)^2 \\ &\quad + 2n^{-1}\sum_{i=1}^{n}(n-1)^{-2}\sum_{\substack{j=1\\ j\neq i}}^{n-1}\sum_{\substack{k=j+1\\ k\neq i}}^{n} k_n(X_i, X_j)k_n(X_i, X_k). \end{aligned}$$

The first term is the one contributing to σ_W^2 under $nh_n \to 0$, while the second term is the one contributing to σ_L^2 under $nh_n \to \infty$. This justifies the form of the first term in $\widetilde{\sigma_L^2}$. For completeness, we show that the first term

$$\widetilde{\sigma_{LW}^2} = n^{-1} \sum_{i=1}^{n} (n-1)^{-2} \sum_{j=1\, j\neq i}^{n} k_n(X_i, X_j)^2$$

is the one contributing to σ_W^2 when $nh_n \to 0$. To see this, note that $\mathbb{E}\left[n^{-1}\sum_{i=1}^n (n-1)^{-2}\sum_{j=1\, j\neq i}^n k_n(X_i,X_j)^2\right] = (n-1)^{-1}\mathbb{E}\left[k_n(X_i,X_j)^2\right]$. Then

$$\begin{aligned}
\mathbb{E}&\left[\left|nh_n n^{-1}\sum_{i=1}^{n}(n-1)^{-2}\sum_{j=1,\, j\neq i}^{n} k_n(X_i,X_j)^2 - nh_n(n-1)^{-1}\mathbb{E}\left[k_n(X_i,X_j)^2\right]\right|^2\right] \\
&= h_n^2 n^2 (n-1)^{-2}\mathrm{Var}\left(\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} k_n(X_i,X_j)^2\right) \\
&= O(n^{-2}h_n^{-1} + n^{-1}),
\end{aligned}$$

where the last equality follows from the bounds derived in Section C of the Supplementary Material. We conclude by L^2-convergence that

$$nh_n n^{-1}\sum_{i=1}^{n}(n-1)^{-2}\sum_{j=1\, j\neq i}^{n} k_n(X_i,X_j)^2 = \sigma_W^2 + o_P(1).$$

3.2 Inconsistency of the plug-in variance estimator

We show in this subsection that the plug-in variance estimator is inconsistent in the non-regular regime corresponding either to case (A) or to case (B). As in the previous section, we should work for additional generality under Assumption NB1 and its corresponding case (A), but the same remark applies for the validity of the result under Assumption NB2 and case (B) – see Remark 3.1.

Consider then the following plug-in estimators

$$\widehat{\sigma_L^2} = n^{-1}\sum_{i=1}^{n} \widehat{l}_{n,i}{}^2, \tag{3.1}$$

$$\widehat{\sigma_W^2} = h_n \binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \widehat{w}_{n,i,j}{}^2, \tag{3.2}$$

with

$$\widehat{u_n^0} = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_n(X_i, X_j), \tag{3.3}$$

$$\widehat{l}_{n,i} = (n-1)^{-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} k_n(X_i, X_j) - \widehat{u_n^0}, \tag{3.4}$$

$$\widehat{w}_{n,i,j} = k_n(X_i, X_j) - \widehat{l}_{n,i} - \widehat{l}_{n,j} + \widehat{u_n^0}. \tag{3.5}$$

We first collect in the following lemmas the limits in probability of the estimators $\widehat{\sigma_L^2}$ and $\widehat{\sigma_W^2}$ for a whole range of bandwidth sequence rates. These limits are then directly used to show consistency and inconsistency of different (rescaled) plug-in variance estimators. The proofs of these lemmas are relegated to Section A.4. of the Supplementary Material.

Lemma 3.2.

1. If $nh_n \to \infty$, then

$$\widehat{\sigma_L^2} = \sigma_L^2 + o_P(1).$$

2. If $nh_n \to C \in (0, \infty)$, then

$$n^{-1}\widehat{\sigma_L^2} = n^{-1}\Big(\sigma_L^2 + \frac{2}{C}\sigma_W^2 + o_P(1)\Big).$$

3. If $n^2 h_n \to \infty$ and $nh_n \to 0$, then

$$n^{-1}\widehat{\sigma_L^2} = \binom{n}{2}^{-1} h_n^{-1}\Big(\sigma_W^2 + o_P(1)\Big)$$

Lemma 3.3. *If $n^2 h_n \to \infty$, then*

$$\widehat{\sigma_W^2} = \sigma_W^2 + o_P(1).$$

Recall again the characterization of $\operatorname{Var} U_n$ given by

$$\operatorname{Var} U_n = 4n^{-1}\Big(\sigma_L^2 + o(1)\Big) + \binom{n}{2}^{-1} h_n^{-1}\Big(\sigma_W^2 + o(1)\Big).$$

Then it follows directly from the two previous lemmas that the plug-in estimator

$$\hat{V}_{n,p} = 4n^{-1}\widehat{\sigma_L^2} + \binom{n}{2}^{-1} h_n^{-1}\widehat{\sigma_W^2}$$

is consistent when the linear term dominates but inconsistent otherwise. However, it is possible to construct a rescaled version

$$\hat{V}_{n,r} = 4n^{-1}\widehat{\sigma_L^2} - 3\binom{n}{2}^{-1} h_n^{-1}\widehat{\sigma_W^2}$$

which is directly seen to be consistent in all cases. These results are summarized in the next propositions.

Proposition 3.2. *If $nh_n \longrightarrow 0$ and $n^2 h_n \longrightarrow \infty$, then*

$$\hat{V}_{n,p} - 3\mathrm{Var}\, U_n = o_p(1)$$

with

$$\hat{V}_{n,p} = 4n^{-1}\widehat{\sigma_L^2} + \binom{n}{2}^{-1} h_n^{-1}\widehat{\sigma_W^2}.$$

Proof. This follows directly from Lemma 3.2 and Lemma 3.3 and the characterization of $\mathrm{Var}\, U_n$ in Equation (A.1). □

Proposition 3.3. *If $n^2 h_n \longrightarrow \infty$, then*

$$\hat{V}_{n,r}^{-1/2}(U_n - \theta_0) \rightsquigarrow N(0,1)$$

with

$$\hat{V}_{n,r} = 4n^{-1}\widehat{\sigma_L^2} - 3\binom{n}{2}^{-1} h_n^{-1}\widehat{\sigma_W^2}.$$

Proof. This follows directly from Lemma 3.2 and Lemma 3.3, the characterization of $\mathrm{Var}\, U_n$ in Equation (A.1), and an application of Slutsky's theorem in Corollary 2.3. □

Remark 3.5. The estimator introduced in Section 3.1 is directly seen to correct the deficiency of the plug-in variance estimator. There exists another method to restore consistency of the plug-in variance estimator, which is also considered in Cattaneo, Crump, and Jansson (2014b). It consists in estimating the variance with a bandwidth sequence H_n converging at a different rate than the bandwidth sequence h_n used to estimate θ_0. The validity of the method follows directly from Lemma 3.2 by taking H_n in the estimation of $\widehat{\sigma_L^2}$ such that $nH_n \longrightarrow \infty$. In this case, the plug-in estimator with double bandwidth sequences is directly seen to be consistent without rescaling.

3.3 Inconsistency of the non-parametric bootstrap

We finally show that the non-parametric bootstrap fails to reproduce the underlying distribution in the non-regular regime corresponding to either case (A) or case (B). We should again work under Assumption NB1 and the associated non-regular case (A), but the same remark for the validity under Assumption NB2 and case (B) still holds – see Remark 3.1.

Remark 3.6. The bootstrap failure for our problem is directly reminiscent of those reported in Hardle and Mammen (1993) and Cattaneo, Crump, and Jansson (2014a). The same underlying mechanism explains the failure in spite of asymptotic normality – it can already be seen from the previous results where we had to "manually" rescale the variance of the quadratic term to obtain a consistent estimator of $\mathrm{Var}\, U_n$. Sensibly similar issues and solutions were already reported on jackknife estimate of variance for U-statistics – see Efron and Stein (1981).

Let $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ be an i.i.d. sample. Then take $\mathcal{X}_n^* = \{X_1^*, X_2^*, \ldots, X_n^*\}$ an i.i.d. sample from the empirical distribution $\mathbb{P}_n$ based on $\mathcal{X}_n$. Equivalently, $\mathcal{X}^*$ can be obtained by uniformly sampling n times from $\mathcal{X}_n$ with replacement. Denote by $P^*, \mathbb{E}^*, \mathrm{Var}^*, \mathrm{cov}^*$, the probability, expectation, variance, and covariance taken with respect to the empirical distribution conditional on $\mathcal{X}_n$. We introduce the bootstrap analogue to the estimator previously introduced

$$U_n^* = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{h_n} K\left(\frac{X_i^* - X_j^*}{h_n}\right) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_n(X_i^*, X_j^*).$$

The statistics U_n^* is the same second-order U-statistics with n-dependent kernel k_n as U_n but computed over the random sample $\mathcal{X}_n^*$ instead of $\mathcal{X}_n$. Note that conditional on $\mathcal{X}_n$, the empirical distribution is a discrete (non-random) distribution, namely, multinomial with uniform weights $1/n$. It follows that the statistics U_n^* admits a Hoeffding decomposition with respect to the empirical distribution conditional on $\mathcal{X}_n$. In virtue of Lemma 1.1, we have

$$U_n^* = \mathbb{E}^*[U_n^*] + 2L_n^* + W_n^*$$

with

$$L_n^* = \frac{1}{n} \sum_{i=1}^{n} \left[u_n^{1*}(X_i^*) - u_n^{0*}\right]$$
$$W_n^* = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[u_n^{2*}(X_i^*, X_j^*) - u_n^{1*}(X_i^*) - u_n^{1*}(X_j^*) + u_n^{0*}\right]$$

and

$$u_n^{0*} = \mathbb{E}^*[U_n^*]$$
$$u_n^{1*}(X_i^*) = \mathbb{E}^*[k_n(X_i^*, X_j^*)|X_i^*]$$
$$u_n^{2*}(X_i^*, X_j^*) = k_n(X_i^*, X_j^*)$$

where $i \neq j$ are any two indexes.

By Lemma 1.2, we have

$$\mathrm{Var}^* U_n^* = \frac{4}{n} \mathrm{Var}^*(u_n^{1*}(X_1)) + \frac{2}{n(n-1)} \mathrm{Var}^*(u_n^{2*}(X_1, X_2) - u_n^{1*}(X_1) - u_n^{1*}(X_2)).$$

Moreover, for the same reason as in the proof of Lemma 1.2, we also have

$$\mathrm{Var}^* U_n^* = \frac{4}{n} \mathrm{Var}^*(u_n^{1*}(X_1)) + \frac{2}{n(n-1)} \left[\mathrm{Var}^*(u_n^{2*}(X_1, X_2)) - \mathrm{Var}^*(u_n^{1*}(X_1))\right].$$

Then to compute $\mathrm{Var}^* U_n^*$, we make use of the multinomial representation of the empirical measure conditional on the observed sample. Similar computations have been used repeatedly when bootstrapping U-statistics (be they with standard kernels or n-dependent kernels), see, for instance, Dehling and Mikosch (1994) or Cattaneo, Crump, and Jansson (2014a). In particular, note that $u_n^{1*}(X_i^*)$ can be rewritten as

$$u_n^{1*}(X_i^*) = \mathbb{E}_{\Xi}[k_n(\xi_i(X_1,\ldots,X_n),\xi_j(X_1,\ldots,X_n))|\xi_i, X_1, X_2,\ldots,X_n]$$

where Ξ is the multimomial distribution with uniform weights and $\xi_1,\xi_2,\ldots,\xi_n$ is an i.i.d. sample from this distribution, and so it follows that

$$u_n^{1*}(X_i^*) = \frac{1}{n}\sum_{j=1}^{n} k_n(X_i^*, X_j).$$

Then we have

$$u_n^{0*} = \mathbb{E}^*[u_n^{1*}(X_i^*)] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} k_n(X_i, X_j) = \frac{n-1}{n}U_n.$$

Additional moment calculations then lead to the following result, which yields inconsistency of the bootstrap variance whenever linearity subsides. The proof can be found in Section A.5. of the Supplementary material.

Proposition 3.4. *If $n^2 h_n \to \infty$, then*

$$\mathrm{Var}^*\, U_n^* - 4n^{-1}\sigma_L^2 - 3\binom{n}{2}^{-1} h_n^{-1}\sigma_W^2 = o_P(1).$$

In particular, if $nh_n \to 0$, then

$$\mathrm{Var}^*\, U_n^* - 3\mathrm{Var}\, U_n = o_P(1).$$

This proves the inconsistency of the bootstrap variance. However, inconsistency of the bootstrap variance is not generally sufficient for inconsistency of the bootstrap distribution. To see that it holds in this case, suppose by contradiction that $\binom{n}{2}^{1/2} h_n^{1/2}(U_n^* - \frac{n-1}{n}U_n) \rightsquigarrow N(0,\sigma_W^2)$ in probability. Then the fact that $\mathrm{Var}^*\, U_n^* - 4n^{-1}\sigma_L^2 - 3\binom{n}{2}^{-1} h_n^{-1}\sigma_W^2 = o_P(1)$ is enough to ensure uniform integrability and convergence of second moments, as in Lemma 2.1. in Kato (2011) which extends Theorem 4.5.2. in Chung (2001) to conditional distributions. Then $\lim_{n\to\infty}\binom{n}{2} h_n \mathrm{Var}^*\, U_n^* = \sigma_W^2$ in probability, a contradiction.

Remark 3.7. Given the nature of bootstrap failure in this problem, there is a number of natural potential candidates to restore consistency. The first ones are those in Cattaneo, Crump, and Jansson (2014a), respectively subsampling and bootstrapping the studentized statistics for the consistent variance estimator. A second set of solutions is based on recentering the kernel of the U-statistics in the bootstrap world, or equivalently adjusting the random sampling weights – see Arcones and Gine (1992) and Dehling and Mikosch (1994). Other reweighing solutions based on a martingale representation of the estimator in the spirit of Otsu and Rai (2017), extending the wild bootstrap in Hardle and Mammen (1993), can also be investigated. A last set of candidates is based on the smoothed bootstrap where resampling is not based on a (conditional) discrete distribution, but a continuous one. Assessing the validity of these bootstrap methods is left for future research.

References

Arcones, M. A., and E. Gine (1992): "On the Bootstrap of U and V Statistics," *The Annals of Statistics*, pp. 655–674.

Bickel, P. J., and Y. Ritov (1988): "Estimating Integrated Squared Density Derivatives: Sharp Best Order of Convergence Estimates," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 381–393.

Cattaneo, M. D., R. K. Crump, and M. Jansson (2014a): "Bootstrapping Density-Weighted Average Derivatives," *Econometric Theory*, 30(6), 1135–1164.

——— (2014b): "Small Bandwidth Asymptotics for Density-Weighted Average Derivatives," *Econometric Theory*, 30(1), 176–200.

Chung, K. (2001): *A Course in Probability Theory*. Elsevier Science.

de Jong, P. (1987): "A Central Limit Theorem for Generalized Quadratic Forms," *Probability Theory and Related Fields*, 75(2), 261–277.

Dehling, H., and T. Mikosch (1994): "Random Quadratic Forms and the Bootstrap for U-Statistics," *Journal of Multivariate Analysis*, 51(2), 392–413.

Efromovich, S., and M. Low (1996): "On Optimal Adaptive Estimation of a Quadratic Functional," *The Annals of Statistics*, 24(3), 1106–1125.

Efron, B., and C. Stein (1981): "The Jackknife Estimate of Variance," *The Annals of Statistics*, pp. 586–596.

Eubank, R., and S. Wang (1999): "A Central Limit Theorem for the Sum of Generalized Linear and Quadratic Forms," *Statistics: A Journal of Theoretical and Applied Statistics*, 33(1), 85–91.

Folland, G. B. (1999): *Real Analysis: Modern Techniques and their Applications*, vol. 40. John Wiley & Sons.

Giné, E., and R. Nickl (2008): "A Simple Adaptive Estimator of the Integrated Square of a Density," *Bernoulli*, pp. 47–61.

Hall, P. (1984): "Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators," *Journal of multivariate analysis*, 14(1), 1–16.

Hall, P., and J. S. Marron (1987): "Estimation of Integrated Squared Density Derivatives," *Statistics & Probability Letters*, 6(2), 109–115.

Hardle, W., and E. Mammen (1993): "Comparing Nonparametric versus Parametric Regression Fits," *The Annals of Statistics*, pp. 1926–1947.

Hoeffding, W. (1948): "A Class of Statistics with Asymptotically Normal Distribution," *The Annals of Mathematical Statistics*, pp. 293–325.

KATO, K. (2011): "A Note on Moment Convergence of Bootstrap M-Estimators," *Statistics & Decisions*, 28(1), 51–61.

KENNEDY, E. H., S. BALAKRISHNAN, AND L. WASSERMAN (2020): "Discussion of" On nearly assumption-free tests of nominal confidence interval coverage for causal parameters estimated by machine learning"," *Statistical Science*, 35(3), 540–544.

LAURENT, B. (1996): "Efficient Estimation of Integral Functionals of a Density," *The Annals of Statistics*, 24(2), 659–681.

——— (2005): "Adaptive Estimation of a Quadratic Functional of a Density by Model Selection," *ESAIM: Probability and Statistics*, 9, 1–18.

LEE, A. J. (1990): *U-statistics: Theory and Practice*. CRC Press.

LEONENKO, N. N., L. PRONZATO, AND V. SAVANI (2008): "A Class of Renyi Information Estimators for Multidimensional Densities," *Annals of Statistics*, 36(5), 2153–2182.

LIU, L., R. MUKHERJEE, AND J. M. ROBINS (2020): "On Nearly Assumption-Free Tests of Nominal Confidence Interval Coverage for Causal Parameters Estimated by Machine Learning," *Statistical Science*, 35(3), 518–539.

OTSU, T., AND Y. RAI (2017): "Bootstrap Inference of Matching Estimators for Average Treatment Effects," *Journal of the American Statistical Association*, 112(520), 1720–1732.

ROBINS, J., AND A. VAN DER VAART (2006): "Adaptive Nonparametric Confidence Sets," *The Annals of Statistics*, pp. 229–253.

ROBINS, J. M., L. LI, E. TCHETGEN TCHETGEN, AND A. VAN DER VAART (2016): "Asymptotic Normality of Quadratic Estimators," *Stochastic Processes and their Applications*, 126(12), 3733–3759.

Supplementary Material for "Integrated square of a density: asymptotic normality and non-regular semi-parametric inference"

Paul Delatte

The Supplementary Material to "Integrated square of a density: asymptotic normality and non-regular semi-parametric inference" contains four sections. The first section (A) contains most of the proofs of the results found in the main text. The second section (B) states and proves a result based on a density argument extending L^1-continuity. The third section (C) contains additional bounds on the moments of some U-statistics that appear in Section A. The last section (D) reproduces and comments on the bias result in Giné and Nickl (2008).

A Proofs

A.1 Proof of the moment bounds of U_n

Proof of Lemma 2.3. 1. By change of variable, we have

$$\mathbb{E}\left[|u_n^1(X_i)|^q\right] = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} K_h(u) f_0(x-u)\, du \right|^q f_0(x)\, dx.$$

Since $f \in L^q$, we conclude using the mollification theorem and continuity of the norm.

2. By change of variable and Fubini's theorem, we have

$$\mathbb{E}\left[|u_n^2(X_i, X_j)|^q\right] = \frac{1}{h^{q-1}} \int_{\mathbb{R}} |K(u)|^q \int_{\mathbb{R}} f_0(x-uh) f_0(x)\, dx\, du,$$

and since $K \in L^q$, we conclude by using Lemma B.1 and dominated convergence.

3. By Holder's inequality, we have

$$\mathbb{E}\left[|u_n^2(X_i, X_j)|^r |u_n^1(X_i)|^q\right] \leq \left(\mathbb{E}\left[|u_n^2(X_i, X_j)^{2r}\right] \mathbb{E}\left[|u_n^1(X_i)|^{2q}\right]\right)^{1/2}$$

and we conclude directly by using (1.) and (2.).

4. By change of variable and Fubini's theorem, we have

$$\begin{aligned}
&\mathbb{E}\left[|u_n^2(X_i, X_j)|^r |u_n^2(X_i, X_k)|^q\right] \\
&\qquad = \int_{\mathbb{R}} \int_{\mathbb{R}} |K_h(x-y)|^r f_0(y)\, dy \int_{\mathbb{R}} |K_h(x-z)|^q f_0(z)\, dz f_0(x)\, dx \\
&\qquad = \frac{1}{h^{r+q-2}} \int_{\mathbb{R}} \int_{\mathbb{R}} |K(u)|^r |K(v)|^q \int_{\mathbb{R}} f_0(x-uh) f_0(x-wh) f_0(x)\, dx\, du\, dv
\end{aligned}$$

and we conclude by using the extension of Lemma B.1 and dominated convergence. □

Proof of Lemma 2.4. 5. It suffices to show that

$$\mathbb{E}\left[|u_n^2(X_i, X_j)|^r |u_n^2(X_i, X_k)|^q |u_n^2(X_j, X_k)|^t\right] = O(h^{-r-q-t+2}).$$

The result then obtains by induction on l, using either independence, result (4.), or this result. By using the same change of variable as in (4.), we directly obtain that

$$\begin{aligned}
&\mathbb{E}\left[|u_n^2(X_i, X_j)|^r |u_n^2(X_i, X_k)|^q |u_n^2(X_j, X_k)|^t\right] \\
&= \frac{1}{h^{r+q-2}} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{h^t} |K(u-v)|^t |K(u)|^r |K(v)|^q \int_{\mathbb{R}} f_0(x-uh) f_0(x-wh) f_0(x)\, dx\, du\, dv \\
&\leq \frac{||K||_\infty}{h^{r+q+t-2}} \int_{\mathbb{R}} \int_{\mathbb{R}} |K(u)|^r |K(v)|^q \int_{\mathbb{R}} f_0(x-uh) f_0(x-wh) f_0(x)\, dx\, du\, dv
\end{aligned}$$

and we can conclude as in (4.).

6. This follows directly from Holder's inequality, independence, (5.) and (1.) as in (3.). □

Proof of Lemma 2.5. 1. We have $l(X_i) = u_n^1(X_i) - u_n^0 = u_n^1(X_i) - \mathbb{E}[u_n^1(X_i)]$, hence $\mathbb{E}[|l(X_i)|^q] \leq C(q)\mathbb{E}[|u_n^1(X_i)|^q]$, and we conclude by Lemma 2.3.

2. If $q = 1$, the result follows from the triangle inequality and Lemma 2.2. Suppose now $q \geq 2$. By the multinomial theorem and Lemma 2.3, the term that dominates asymptotically is $\mathbb{E}[|u_n^2(X_i, X_j)|^q] = O(h^{-q+1})$, since $\mathbb{E}[|u_n^2(X_i, X_j)|^{q-1}|u_n^1(X_i)|] = O(h^{-q+3/2})$ and all other terms are of lower order. This concludes the proof.

3. If $r = q = 1$, the result follows from the triangle inequality and Lemma 2.3. Suppose now w.l.o.g. that $r > 1$. By the multinomial theorem and Lemma 2.4 with $l = 3$, the only two terms that can dominate asymptotically are $\mathbb{E}[|u_n^2(X_i, X_j)|^r |u_n^2(X_i, X_k)|^q] = O(h^{-r-q+2})$ and $\mathbb{E}[|u_n^2(X_i, X_j)|^{r-1} |u_n^2(X_i, X_k)|^q |u_n^1(X_i)|] = O(h^{-r+1-q+\frac{3-1}{2}}) = O(h^{-r-q+2})$, since all the other terms are of lower order. This concludes the proof.

4. The same argument generalizes by induction on l, as there are always only two terms in the multinomial expansions that can dominate asymptotically. □

A.2 Proof of the weak limits

Proof of Proposition 2.1. For simplicity, we write $h = h_n$. To directly apply the results of de Jong (1987) and Eubank and Wang (1999), we recall and introduce some notations

$$\begin{aligned}
&l(X_i) = u_n^1(X_i) - u_n^0, \\
&w(X_i, X_j) = u_n^2(X_i, X_j) - u_n^1(X_i) - u_n^1(X_j) + u_n^0,
\end{aligned}$$

and

$$L_i = n^{-1/2} l(X_i) \qquad \text{and} \qquad W_{i,j} = \binom{n}{2}^{-1/2} h^{1/2} w(X_i, X_j)$$

for $1 \le i, j \le n$, $i \neq j$. From there, it follows that

$$\sqrt{n} L_n = \sum_{i=1}^{n} L_i =: L(n) \qquad \text{and} \qquad \sqrt{\binom{n}{2}} h W_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} W_{i,j} =: W(n).$$

In particular, $\operatorname{Var}(\sqrt{n} L_n) = O(1)$ and $\operatorname{Var}(\sqrt{\binom{n}{2}} h W_n) = O(1)$, so

$$\operatorname{Var}\left(\sqrt{n} L_n + \sqrt{\binom{n}{2}} h W_n\right) = O(1),$$

and so, in all the Lyapunov-type conditions, the normalizing variances can be taken to be 1. The conditions (1.3) to (1.6) in Eubank and Wang (1999) then rewrite as

$$\binom{n}{2}^{-1} h \max_{1 \le i \le n} \sum_{j=1}^{n} \operatorname{Var}(w(X_i, X_j)) \to 0, \tag{EW1.3}$$

$$\mathbb{E}\left[W(n)^4\right] / (\operatorname{Var} W(n))^2 \to 3, \tag{EW1.4}$$

$$n^{-2} \sum_{i=1}^{n} \mathbb{E}\left[l(X_i)^4\right] \to 0, \tag{EW1.5}$$

$$\binom{n}{2}^{-1} n^{-1} h \mathbb{E}\left[\left(\sum_{i=2}^{n} \sum_{j=1}^{i-1} \mathbb{E}\left[w(X_i, X_j) l(X_i) | X_1, \dots, X_{i-1}\right]\right)^2\right] \to 0. \tag{EW1.6}$$

The i.i.d. assumption allows us to considerably simplifies those expressions. In particular, the conditions above are equivalent to

$$n^{-1} h \operatorname{Var}(w(X_1, X_2)) \to 0, \tag{EW1.3bis}$$

$$\mathbb{E}\left[W(n)^4\right] / (\operatorname{Var} W(n))^2 \to 3, \tag{EW1.4}$$

$$n^{-1} \mathbb{E}\left[l(X_1)^4\right] \to 0, \tag{EW1.5bis}$$

$$n^{-1} h \operatorname{Var}\big(\mathbb{E}\left[w(X_2, X_1) l(X_2) | X_1\right]\big) \to 0. \tag{EW1.6bis}$$

The last equivalence follows from $\mathbb{E}[X^2] = \operatorname{Var}(X) + \mathbb{E}[X]^2$ and $\mathbb{E}[w(X_i, X_j) l(X_i)] = 0$. To handle (EW1.4), we make use of the expansion of $\mathbb{E}[W(n)^4]$ in Table 1 in de Jong (1987). In particular, using the notations of de Jong (1987), it follows from our normalization that (EW1.4) holds whenever the terms $G_{\mathrm{I}}, G_{\mathrm{II}}, G_{\mathrm{III}}, G_{\mathrm{IV}}$ tend to zero and the term G_{V} is asymptotically equivalent to $(\operatorname{Var}(W(n))^2/2$. Using the i.i.d. assumption and our notations, this reduces to the following conditions

$$n^{-2}h^2\mathbb{E}\left[w(X_1,X_2)^4\right] \to 0 \tag{dJ.G_{I}}$$

$$n^{-1}h^2\mathbb{E}\left[w(X_1,X_2)^2 w(X_1,X_3)^2\right] \to 0 \tag{dJ.G_{II}}$$

$$n^{-1}h^2\mathbb{E}\left[w(X_1,X_2)^2 w(X_1,X_3)w(X_3,X_2)\right] \to 0 \tag{dJ.G_{III}}$$

$$h^2\mathbb{E}\left[w(X_1,X_2)w(X_1,X_3)w(X_4,X_2)w(X_4,X_3)\right] \to 0 \tag{dJ.G_{IV}}$$

$$h\mathbb{E}\left[w(X_1,X_2)^2\right]/\mathrm{Var}(W(n)) \to 1 \tag{dJ.G_{V}}$$

where the last equivalence follows from $3\binom{n}{4}\binom{n}{2}^{-2} \sim 1/2$. We now use Lemma 2.5 to prove that all limits are as given. For (EW1.3bis), we have

$$\mathrm{Var}\, w(X_1,X_2) \leq E[w(X_1,X_2)^2] = O(h^{-1}),$$

hence the result. For (dJ.G_{I}), we have

$$E[w(X_1,X_2)^4] = O(h^{-3}),$$

and so the result follows since $n^2h \to \infty$. For (dJ.G_{II}),

$$\mathbb{E}\left[w(X_1,X_2)^2 w(X_1,X_3)^2\right] = O(h^{-2}),$$

so the result follows. For (dJ.G_{III}), we have

$$\mathbb{E}\left[w(X_1,X_2)^2 w(X_1,X_3)w(X_3,X_2)\right] = O(h^{-2}),$$

so the result follows. For (dJ.G_{IV}), we have

$$\mathbb{E}\left[w(X_1,X_2)w(X_1,X_3)w(X_4,X_2)w(X_4,X_3)\right] = O(h^{-3/2}),$$

so the result follows since $h \to 0$. For (dJ.G_{V}), the result follows immediately from Lemma 2.2 and Lemma 2.5. For (EW1.5bis), we have

$$\mathbb{E}\left[l(X_1)^4\right] = O(1),$$

so the result follows. For (EW1.6bis), we have

$$\mathrm{Var}(\mathbb{E}\left[w(X_2,X_1)l(X_2)|X_1\right]) \leq \mathbb{E}\left((\mathbb{E}\left[w(X_2,X_1)l(X_2)|X_1\right])^2\right).$$

By monotonicity, conditional Holder's inequality, and independence,

$$\mathrm{Var}(\mathbb{E}\left[w(X_2,X_1)l(X_2)|X_1\right]) \leq \mathbb{E}\left[w(X_2,X_1)\right](\mathbb{E}\left[l(X_2^2)\right])^{1/2} = O(1),$$

and so the result follows. This concludes the proof. □

Proof of Corollary 2.2. From Lemma 2.1 and Lemma 2.2 and Equation (1.2), we have

$$\operatorname{Var} U_n = 4n^{-1}\left(\sigma_L^2 + o(1)\right) + \binom{n}{2}^{-1} h_n^{-1}\left(\sigma_W^2 + o(1)\right). \tag{A.1}$$

In particular, $(\operatorname{Var} U_n)^{-1/2} = O(n^{1/2} \wedge nh_n^{1/2})$. By distinguishing three cases if necessary, the result then follows immediately from Equation (1.1), Slutsky's theorem, Proposition 2.1, and the normality of the marginals of bivariate normals. □

Proof of Corollary 2.3. 1. If $nh_n \to \infty$, then $(\operatorname{Var} U_n)^{-1/2} \sim (4\sigma_L^2)^{-1/2} n^{1/2}$. By Lemma D.1, $n^{1/2}(\mathbb{E}[U_n] - \theta_0) = O(n^{1/2}h^{2s})$ and so $n^{1/2}(\mathbb{E}[U_n] - \theta_0) = o(1)$ since $nh_n^{4s} \to 0$. The result then follows from Slutsky's theorem and Corollary 2.2.

2. If $nh_n \to C \in (0, \infty)$, then $(\operatorname{Var} U_n)^{-1/2} \sim (4\sigma_L^2 + \frac{2}{C}\sigma_W^2)^{-1/2} n^{1/2}$. As in (1.), since $nh_n^{4s} \to 0$, $n^{1/2}(\mathbb{E}[U_n] - \theta_0) = o(1)$. The result follows again from Slutsky's theorem and Corollary 2.2.

3. If $nh_n \to 0$, then $(\operatorname{Var} U_n)^{-1/2} \sim (\sigma_W^2)^{-1/2} h^{1/2} \binom{n}{2}^{1/2}$. By Lemma D.1, $h^{1/2}\binom{n}{2}^{1/2}(\mathbb{E}[U_n] - \theta_0) = O(nh^{2s+1/2})$ and so $n^{1/2}(\mathbb{E}[U_n] - \theta_0) = o(1)$ since $nh^{2s+1/2} \to 0$. The result follows again from Slutsky's theorem and Corollary 2.2. □

Proof of Corollary 2.4. 1. If $s > 1/4$, then $nh_n = Cn^{\frac{4s-1}{4s+1}} \to \infty$ and $nh^{4s} = Cn^{\frac{-4s+1}{4s+1}} \to 0$.

2. If $s = 1/4$, then $nh_n \to C \in (0, \infty)$.

3. If $s < 1/4$, then $nh_n = Cn^{\frac{4s-1}{4s+1}} \to 0$.

The results then follow directly from Corollary 2.2 as in the proof of Corollary 2.3. □

A.3 Proofs of the consistency of the simple variance estimator

Proof of Lemma 3.1. 1. We start with $\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} k_n(X_i, X_j)$. It is seen from an i.i.d. argument that

$$\mathbb{E}\left[\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} k_n(X_i, X_j)\right] = \mathbb{E}[k_n(X_1, X_2)].$$

Then, again by i.i.d. and Lemma 2.3,

$$\begin{aligned}\mathbb{E}\left[\left|\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} k_n(X_i, X_j) - \mathbb{E}[k_n(X_1, X_2)]\right|^2\right] &= \operatorname{Var}\left(\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} k_n(X_i, X_j)\right) \\ &= \operatorname{Var} U_n = O(n^{-1} + n^{-2}h^{-1}).\end{aligned}$$

By L^2-convergence, it follows that

$$\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} k_n(X_i, X_j) = \int_{\mathbb{R}} f_0(x)\,dx + o_P(1).$$

We now consider $n^{-1}\sum_{i=1}^{n}\binom{n-1}{2}^{-1}\sum_{j=1,\,j\neq i}^{n-1}\sum_{k=j+1,\,k\neq i}^{n}k_n(X_i,X_j)k_n(X_i,X_k)$. It is seen from an i.i.d. argument that

$$\mathbb{E}\left[n^{-1}\sum_{i=1}^{n}\binom{n-1}{2}^{-1}\sum_{\substack{j=1,\\ j\neq i}}^{n-1}\sum_{\substack{k=j+1,\\ k\neq i}}^{n}k_n(X_i,X_j)k_n(X_i,X_k)\right]=\mathbb{E}\left[k_n(X_1,X_2)k_n(X_1,X_3)\right].$$

Note then by i.i.d. and the law of iterated expectation that

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[k_n(X_1,X_2)|X_1\right]^2\right]&=\mathbb{E}\left[\mathbb{E}\left[k_n(X_1,X_2)|X_1\right]\mathbb{E}\left[k_n(X_1,X_3)|X_1\right]\right]\\&=\mathbb{E}\left[\mathbb{E}\left[k_n(X_1,X_2)k_n(X_1,X_3)|X_1\right]\right]\\&=\mathbb{E}\left[k_n(X_1,X_2)k_n(X_1,X_3)\right].\end{aligned}$$

Then

$$\begin{aligned}&\mathbb{E}\left[\left|n^{-1}\sum_{i=1}^{n}\binom{n-1}{2}^{-1}\sum_{\substack{j=1,\\ j\neq i}}^{n-1}\sum_{\substack{k=j+1,\\ k\neq i}}^{n}k_n(X_i,X_j)k_n(X_i,X_k)-\mathbb{E}\left[\mathbb{E}\left[k_n(X_1,X_2)|X_1\right]^2\right]\right|^2\right]\\&=\mathrm{Var}\left(n^{-1}\sum_{i=1}^{n}\binom{n-1}{2}^{-1}\sum_{j=1,\,j\neq i}^{n-1}\sum_{k=j+1,\,k\neq i}^{n}k_n(X_i,X_j)k_n(X_i,X_k)\right)\\&=n^{-1}\mathrm{Var}\left(\binom{n-1}{2}^{-1}\sum_{j=2}^{n-1}\sum_{k=j+1}^{n}k_n(X_1,X_j)k_n(X_1,X_k)\right)\\&\quad+n^{-2}n(n-1)\binom{n-1}{2}^{-2}\mathrm{cov}\left(\sum_{i=2}^{n-1}\sum_{j=i+1}^{n}k_n(X_1,X_i)k_n(X_1,X_j),\sum_{\substack{k=1\\ k\neq 2}}^{n-1}\sum_{\substack{l=k+1\\ l\neq 2}}^{n}k_n(X_2,X_k)k_n(X_2,X_l)\right)\\&=n^{-1}R_1+2n^{-2}\binom{n-1}{2}^{-1}R_2,\end{aligned}$$

with

$$R_1=\mathrm{Var}\left(\binom{n-1}{2}^{-1}\sum_{j=2}^{n-1}\sum_{k=j+1}^{n}k_n(X_1,X_j)k_n(X_1,X_k)\right),$$

and

$$R_2=\mathrm{cov}\left(\sum_{i=2}^{n-1}\sum_{j=i+1}^{n}k_n(X_1,X_i)k_n(X_1,X_j),\sum_{\substack{k=1\\ k\neq 2}}^{n-1}\sum_{\substack{l=k+1\\ l\neq 2}}^{n}k_n(X_2,X_k)k_n(X_2,X_l)\right),$$

where the penultimate equality follows from expanding the variance around the sum and identical distributions across i. In Lemma C.3, it is shown that

$$R_2=O(n^3+n^2h_n^{-1}+nh_n^{-2}).$$

For R_1, note that by the law of total variance, we have

$$R_1 = A + B$$

where

$$A = \mathbb{E}\left[\mathrm{Var}\left(\binom{n-1}{2}^{-1}\sum_{j=2}^{n-1}\sum_{k=j+1}^{n} k_n(X_1,X_j)k_n(X_1,X_k)\middle|X_1\right)\right],$$
$$B = \mathrm{Var}\left(\binom{n-1}{2}^{-1}\sum_{j=2}^{n-1}\sum_{k=j+1}^{n} \mathbb{E}\left[k_n(X_1,X_j)k_n(X_1,X_k)|X_1\right]\right).$$

Note that conditional on X_1, the term within the variance in A is a second-order U-statistics with kernel $k_n(X_1,X_2)k_n(X_1,X_3)$, hence it admits a Hoeffding decomposition and its variance can be bounded by standard argument. The quadratic term can be shown to be $O(n^{-2}h_n^{-2})$. The linear term can be shown to be $O(n^{-1})$. This is proved in Lemma C.1. Now, we analyze the term B. We have

$$\begin{aligned}
B &= \mathrm{Var}\left(\binom{n-1}{2}^{-1}\sum_{j=2}^{n-1}\sum_{k=j+1}^{n} \mathbb{E}\left[k_n(X_1,X_j)k_n(X_1,X_k)|X_1\right]\right)\\
&= \mathrm{Var}\left(\binom{n-1}{2}^{-1}\sum_{j=2}^{n-1}\sum_{k=j+1}^{n} \mathbb{E}\left[k_n(X_1,X_2)|X_1\right]^2\right)\\
&\le \mathbb{E}\left[\mathbb{E}\left[k_n(X_1,X_2)|X_1\right]^4\right] = O(1),
\end{aligned}$$

where the last equality follows from independence, the law of iterated expectation, and Lemma 2.4. It follows that

$$R_1 = O(1 + n^{-2}h_n^{-2} + n^{-1}),$$

and so

$$n^{-1}R_1 + 2n^{-2}\binom{n-1}{2}^{-1}R_2 = O(n^{-3}h_n^{-2} + n^{-1} + n^{-2} + n^{-2}h_n^{-1})$$

By L^2-convergence, it follows that, whenever $nh_n \to C \in (0,\infty]$, we have

$$n^{-1}\sum_{i=1}^{n}\binom{n-1}{2}^{-1}\sum_{j=1,\, j\neq i}^{n-1}\sum_{k=j+1,\, k\neq i}^{n} k_n(X_i,X_j)k_n(X_i,X_k) = \int_{\mathbb{R}} f_0(x)^3\,dx + o_P(1).$$

If $nh_n \to 0$, the same argument shows that

$$nh_n n^{-1}\sum_{i=1}^{n}\binom{n-1}{2}^{-1}\sum_{j=1,\, j\neq i}^{n-1}\sum_{k=j+1,\, k\neq i}^{n} k_n(X_i,X_j)k_n(X_i,X_k) = o_P(1).$$

A double application of the continuous mapping theorem then yields the result.

2. The proof operates with similar arguments. Note first that

$$\mathbb{E}\left[h_n\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}k_n(X_i,X_j)^2\right]=h_n\mathbb{E}\left[k_n(X_i,X_j)^2\right].$$

Then

$$\mathbb{E}\left[\left|h_n\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}k_n(X_i,X_j)^2-h_n\mathbb{E}\left[k_n(X_i,X_j)^2\right]\right|^2\right]=h_n^2\mathrm{Var}\left(\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}k_n(X_i,X_j)^2\right).$$

Note that $\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}k_n(X_i,X_j)^2$ is a second-order U-statistics with kernel $k_n(X_i,X_j)^2$. By Hoeffding decomposition and density arguments, its variance can be bounded in the same way as the variance of U_n. As proved in Lemma C.2, we can show that

$$\mathrm{Var}\left(\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}k_n(X_i,X_j)^2\right)=O(n^{-2}h_n^{-3})+O(n^{-1}h_n^{-2}).$$

We then have

$$\mathbb{E}\left[\left|h_n\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}k_n(X_i,X_j)^2-h_n\mathbb{E}\left[k_n(X_i,X_j)^2\right]\right|^2\right]=O(n^{-2}h_n^{-1}+n^{-1}).$$

By L^2-convergence, it follows that

$$h_n\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}k_n(X_i,X_j)^2=\sigma_W^2+o_P(1).$$

□

A.4 Proofs of the inconsistency of the plug-in variance estimator

Proof. We start by expanding $\widehat{\sigma_L^2}$. We have

$$\begin{aligned}\widehat{\sigma_L^2}&=n^{-1}\sum_{i=1}^{n}\left((n-1)^{-1}\sum_{\substack{j=1\\j\neq i}}^{n}k_n(X_i,X_j)-\widehat{u_n^0}\right)^2\\&=n^{-1}(n-1)^{-2}\sum_{i=1}^{n}\left(\sum_{\substack{j=1\\j\neq i}}^{n}k_n(X_i,X_j)\right)^2-\widehat{u_n^0}\binom{n}{2}^{-1}\sum_{\substack{j=1\\j\neq i}}^{n}k_n(X_i,X_j)+\widehat{u_n^0}^2\\&=n^{-1}(n-1)^{-2}\sum_{i=1}^{n}\left(\sum_{\substack{j=1\\j\neq i}}^{n}k_n(X_i,X_j)\right)^2-\widehat{u_n^0}^2.\end{aligned}$$

We now expand the first term on the right-end side

$$
\begin{aligned}
&n^{-1}(n-1)^{-2}\sum_{i=1}^{n}\Big(\sum_{\substack{j=1\\j\neq i}}^{n}k_n(X_i,X_j)\Big)^2\\
&\quad= n^{-1}(n-1)^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}k_n(X_i,X_j)^2+2n^{-1}(n-1)^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n-1}\sum_{\substack{k=j+1\\k\neq i}}^{n}k_n(X_i,X_j)k_n(X_i,X_k)\\
&\quad=\widetilde{\sigma^2_{LW}}+(n-2)(n-1)^{-1}\widetilde{\sigma^2_{L}}.
\end{aligned}
$$

The result then follows directly from Lemma 3.1 and Remark 3.4. □

Proof. We start by expanding $\widehat{\sigma^2_W}$. We have

$$
\begin{aligned}
\widehat{\sigma^2_W}&=h_n\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\Big[k_n(X_i,X_j)^2+(\widehat{l}_{n,i}+\widehat{l}_{n,j})^2+\widehat{u^0_n}^2\\
&\qquad-2k_n(X_i,X_j)(\widehat{l}_{n,i}+\widehat{l}_{n,j})+2k_n(X_i,X_j)\widehat{u^0_n}-2(\widehat{l}_{n,i}+\widehat{l}_{n,j})\widehat{u^0_n}\Big]\\
&=h_n\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\Big[k_n(X_i,X_j)^2+\widehat{l}_{n,i}{}^2+\widehat{l}_{n,j}{}^2+3\widehat{u^0_n}^2\\
&\qquad+2(n-1)^{-2}\sum_{\substack{k=1\\k\neq i}}^{n}\sum_{\substack{l=1\\l\neq j}}^{n}k_n(X_i,X_k)k_n(X_j,X_l)\\
&\qquad-2(n-1)^{-1}k_n(X_i,X_j)\Big(\sum_{\substack{k=1\\k\neq i}}^{n}k_n(X_i,X_k)+\sum_{\substack{l=1\\l\neq j}}^{n}k_n(X_j,X_l)\Big)\\
&\qquad-2k_n(X_i,X_j)\widehat{u^0_n}-4(\widehat{l}_{n,i}+\widehat{l}_{n,j})\widehat{u^0_n}\Big]
\end{aligned}
$$

The only terms that are not directly covered by the previous results are the summands with cross-terms, namely

$$
k_n(X_i,X_j)\Big(\sum_{\substack{k=1\\k\neq i}}^{n}k_n(X_i,X_k)+\sum_{\substack{l=1\\l\neq j}}^{n}k_n(X_j,X_l)\Big)
$$

and

$$
\sum_{\substack{k=1\\k\neq i}}^{n}\sum_{\substack{l=1\\l\neq j}}^{n}k_n(X_i,X_k)k_n(X_j,X_l).
$$

We show that all these terms are $o_p(1)$ by using previous results. We first have

$$
\Big|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}k_n(X_i,X_j)\sum_{\substack{k=1\\k\neq i}}^{n}k_n(X_i,X_k)\Big|\leq\sum_{i=1}^{n}\Big(\sum_{\substack{j=1\\j\neq i}}^{n}|k_n(X_i,X_j)|\Big)^2
$$

and

$$\left|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} k_n(X_i,X_j)\sum_{\substack{l=1\\l\neq j}}^{n} k_n(X_j,X_l)\right| \leq \sum_{i=1}^{n}\Big(\sum_{\substack{j=1\\j\neq i}}^{n}|k_n(X_i,X_j)|\Big)^2.$$

Similarly, we have

$$\left|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n}\sum_{\substack{l=1\\l\neq j}}^{n} k_n(X_i,X_k)k_n(X_j,X_l)\right| \leq 3\sum_{i=1}^{n}\Big(\sum_{\substack{j=1\\j\neq i}}^{n}|k_n(X_i,X_j)|\Big)^2.$$

From the proof of Lemma 3.2, we directly get that

$$(n-1)^{-1}h_n\binom{n}{2}^{-1}\sum_{i=1}^{n}\Big(\sum_{\substack{j=1\\j\neq i}}^{n}|k_n(X_i,X_j)|\Big)^2 = o_P(1).$$

Then by Lemma 3.1 and Lemma 3.2, we get that all terms in the expansion are $o_P(1)$, except

$$\widetilde{\sigma_W^2} = h_n\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} k_n(X_i,X_j)^2 = \sigma_W^2 + o_P(1).$$

This concludes the proof. □

A.5 Proof of the inconsistency of the non-parametric bootstrap

Proof of Proposition 3.4. From the multinomial representation, we have

$$\begin{aligned}
\operatorname{Var}^*(u_n^{1*}(X_i^*)) &= \mathbb{E}^*[(u_n^{1*}(X_i^*))^2] - (\mathbb{E}^*[u_n^{1*}(X_i^*)])^2 \\
&= \frac{1}{n^3}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} k_n(X_i,X_j)\right)^2 - \left(\frac{n-1}{n}\right)^2 U_n^2 \\
&= \left(\frac{n-1}{n}\right)^2\widehat{\sigma_L^2}
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{Var}^*(u_n^{2*}(X_i^*,X_j^*)) &= \mathbb{E}^*[k_n(X_i^*,X_j^*)^2] - (\mathbb{E}^*[k_n(X_i^*,X_j^*)])^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} k_n(X_i,X_j)^2 - \left(\frac{n-1}{n}\right)^2 U_n^2 \\
&= h_n^{-1}\frac{n-1}{n}\widetilde{\sigma_W^2} - \left(\frac{n-1}{n}\right)^2 U_n^2.
\end{aligned}$$

It follows that

$$\operatorname{Var}^* U_n^* = \frac{4}{n}\left(\frac{n-1}{n}\right)^2\widehat{\sigma_L^2} + \binom{n}{2}^{-1}\left[h_n^{-1}\frac{n-1}{n}\widetilde{\sigma_W^2} - \left(\frac{n-1}{n}\right)^2 U_n^2 - 2\left(\frac{n-1}{n}\right)^2\widehat{\sigma_L^2}\right].$$

Then the result follows directly from Lemma 3.2 and Lemma 3.1. □

B A density argument

Lemma B.1. *Let $f \in L^1 \cap L^\infty$ and $a, b \in \mathbb{R}$. Then*

$$\lim_{h \to 0} \int |f(x+ah)f(x+bh) - f^2(x)|\, dx = 0.$$

Proof. We make use of a density argument. Suppose g is continuous and compactly supported. Then

$$\begin{aligned}\int |g(x+ah)g(x+bh) - g^2(x)|\, dx &\le \lambda(K) \sup_x |g(x+ah)g(x+bh) - g^2(x)| \\ &\to 0 \text{ as } h \to 0,\end{aligned} \tag{B.1}$$

since

$$|ab - cd| \le |b||a-c| + |c||b-d|$$

which yields

$$\begin{aligned}&\sup_x |g(x+ah)g(x+bh) - g^2(x)| \\ &\qquad \le \sup_x |g(x+ah)||g(x+bh) - g(x)| + \sup_x |g(x)||g(x+ah) - g(x)| \\ &\qquad \le c_1 \sup_x |g(x+bh) - g(x)| + c_2 \sup_x |g(x+ah) - g(x)| \\ &\qquad \to 0 \text{ as } h \to 0, \text{ by uniform continuity of } g.\end{aligned}$$

Now, by density of C_c in L^1, for $\varepsilon > 0$, there is $g \in C_c$ satisfying $\int |f(x) - g(x)|\, dx \le \varepsilon$. Then

$$\begin{aligned}&\int |f(x+ah)f(x+bh) - g(x+ah)g(x+bh)|\, dx \\ &\qquad \le \|f\|_\infty \int |f(x+ah) - g(x+ah)|\, dx + \|g\|_\infty \int f(x+bh) - g(x+bh)\, dx \\ &\qquad \le 2(\|f\|_\infty + \|g\|_\infty)\varepsilon\end{aligned}$$

by density and translation invariance. Moreover,

$$\begin{aligned}\int |f^2(x) - g^2(x)|\, dx &\le \|f\|_\infty \int |f(x) - g(x)|\, dx + \|g\|_\infty \int |f(x) - g(x)|\, dx \\ &\le 2(\|f\|_\infty + \|g\|_\infty)\varepsilon\end{aligned}$$

by density. Finally, by (B.1),

$$\int |g(x+ah)g(x+bh) - g^2(x)| \le \varepsilon$$

for h sufficiently close to 0. It follows by the triangular inequality and the previous inequalities that

$$\begin{aligned}
\int |f(x+ah)f(x+bh) - f^2(x)|\, dx& \\
&\leq \int |f(x+ah)f(x+bh) - g(x+ah)g(x+bh)|\, dx \\
&+ \int |g(x+ah)g(x+bh) - g^2(x)| + \int |f^2(x) - g^2(x)| \\
&\leq (1 + 4(\|f\|_\infty + \|g\|_\infty))\varepsilon.
\end{aligned}$$

□

The same argument can be extended by extending the absolute inequality for higher order products, e.g., for $k = 3$,

$$|abc - def| \leq |b||c||a-d| + |c||d||b-e| + |d||e||c-f|.$$

By the exact same density argument, we obtain the following extension.

Corollary B.1. *Let $f \in L^1 \cap L^\infty$ and $A \subset \mathbb{R}$ a finite subset. Then*

$$\lim_{h\to 0} \int |\Pi_{i\in A} f(x+ih) - f^{\#A}(x)|\, dx = 0.$$

C Additional bounds on the variance of some U-statistics

When proving L^2 convergence for estimators of the variance, a number of other second-order U-statistics appear whose variance need to be bounded. In this section, we collect some of these bounds. We also bound the covariance between two related U-statistics.

First, consider

$$A_1 = \binom{n-1}{2}^{-1} \sum_{j=2}^{n-1} \sum_{k=j+1}^{n} k_n(X_1, X_j) k_n(X_1, X_k).$$

Conditional on X_1, this is a second-order U-statistics with kernel $k_n(X_1, X_j)k_n(X_1, X_k)$. By the Hoeffding decomposition and the general variance bounds for second-order U-statistics, we know that, a.s.,

$$\mathrm{Var}(A_1|X_1) \leq \kappa_1 (n^{-2}\mathbb{E}\,[k_n(X_1,X_2)^2 k_n(X_1,X_3)^2 | X_1] + n^{-1}\mathbb{E}\,[\mathbb{E}\,[k_n(X_1,X_2)k_n(X_1,X_3)|X_2]^2|X_1])$$

where $\kappa_1 > 0$, and so, by monotonicity and tower property of the expectation, we have

$$\mathbb{E}\,[\mathrm{Var}(A_1|X_1)] \leq \kappa_1 (n^{-2}\mathbb{E}\,[k_n(X_1,X_2)^2 k_n(X_1,X_3)^2] + n^{-1}\mathbb{E}\,[\mathbb{E}\,[k_n(X_1,X_2)k_n(X_1,X_3)|X_2]^2]).$$

Lemma C.1. *We have that*

$$\mathbb{E}\,[\mathrm{Var}(A_1|X_1)] = O(n^{-2}h_n^{-2} + n^{-1})$$

Proof. The bound for the quadratic term follows directly from Lemma 2.3, that is,

$$\mathbb{E}\left[k_n(X_1, X_2)^2 k_n(X_1, X_3)^2\right] = O(h_n^{-2}).$$

For the linear term, a subtle application of the mollification theorem in two dimensions can deliver the result. However, a simpler argument using the properties of conditional expectations is presented. Note that

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left[k_n(X_1, X_2)k_n(X_1, X_3)|X_2\right]^2\right] \\
&\qquad = \mathbb{E}\left[\mathbb{E}\left[k_n(X_1, X_2)k_n(X_1, X_3)|X_2\right]\mathbb{E}\left[k_n(X_4, X_2)k_n(X_4, X_5)|X_2\right]\right] \\
&\qquad = \mathbb{E}\left[\mathbb{E}\left[k_n(X_1, X_2)k_n(X_1, X_3)k_n(X_4, X_2)k_n(X_4, X_5)|X_2\right]\right] \\
&\qquad = \mathbb{E}\left[k_n(X_1, X_2)k_n(X_1, X_3)k_n(X_4, X_2)k_n(X_4, X_5)\right].
\end{aligned}$$

Then, by Lemma 2.4, it follows that

$$\mathbb{E}\left[\mathbb{E}\left[k_n(X_1, X_2)k_n(X_1, X_3)|X_2\right]^2\right] = O(1).$$

□

Consider now

$$A_2 = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_n(X_i, X_j)^2.$$

This is a second-order U-statistics with kernel $k_n(X_i, X_j)^2$. By the Hoeffding decomposition and the general variance bounds for second-order U-statistics, we know that

$$\operatorname{Var} A_2 \leq \kappa_2(n^{-2}\mathbb{E}\left[k_n(X_1, X_2)^4\right] + n^{-1}\mathbb{E}\left[\mathbb{E}\left[k_n(X_1, X_2)^2|X_2\right]^2\right])$$

for some $\kappa_2 > 0$.

Lemma C.2. *We have that*

$$\operatorname{Var} A_2 = O(n^{-2}h_n^{-3} + n^{-1}h_n^{-2}).$$

Proof. The bound for the quadratic term follows directly from Lemma 2.3, that is,

$$\mathbb{E}\left[k_n(X_1, X_2)^4\right] = O(h_n^{-3}).$$

For the linear term, we use the properties of the conditional expectation. We have

$$\mathbb{E}\left[\mathbb{E}\left[k_n(X_1, X_2)^2|X_2\right]^2\right] = \mathbb{E}\left[k_n(X_1, X_2)^2 k_n(X_3, X_2)^2\right]$$

and again by Lemma 2.3, we have

$$\mathbb{E}\left[\mathbb{E}\left[k_n(X_1, X_2)^2|X_2\right]^2\right] = O(h_n^{-2}).$$

□

Now, consider the covariance

$$R_2 = \operatorname{cov}\left(\sum_{i=2}^{n-1}\sum_{j=i+1}^{n} k_n(X_1,X_i)k_n(X_1,X_j), \sum_{\substack{k=1\\k\neq 2}}^{n-1}\sum_{\substack{l=k+1\\l\neq 2}}^{n} k_n(X_2,X_k)k_n(X_2,X_l)\right).$$

Modulo the scaling factors, this is the covariance between two U-statistics (with random kernels anchored at X_1 and X_2, respectively). By finitude of all moments, a rough bound is directly given by $O(n^4 h_n^{-2})$. However, because of the i.i.d. assumption, many of summands are 0, namely all those such that $1, i, j$ are all different from $2, k, l$. This allows us to drastically refine the bound. We have that

Lemma C.3. *We have that*

$$R_2 = O(n^3 + n^2 h_n^{-1} + n h_n^{-2}).$$

Proof. Note first that the following expansion holds

$$\begin{aligned}
R_2 = &\sum_{i=2}^{n-1}\sum_{j=i+1}^{n}\sum_{l=3}^{n} \operatorname{cov}(k_n(X_1,X_i)k_n(X_1,X_j), k_n(X_2,X_1)k_n(X_2,X_l)) \\
&+ \sum_{j=3}^{n}\sum_{\substack{k=1\\k\neq 2}}^{n-1}\sum_{\substack{l=k+1\\l\neq 2}}^{n} \operatorname{cov}(k_n(X_1,X_2)k_n(X_1,X_j), k_n(X_2,X_k)k_n(X_2,X_l)) \\
&+ \sum_{i=3}^{n-2}\sum_{j=i+1}^{n-1}\sum_{l=j+1}^{n} \operatorname{cov}(k_n(X_1,X_i)k_n(X_1,X_j), k_n(X_2,X_j)k_n(X_2,X_l)) \\
&+ \sum_{i=3}^{n-1}\sum_{j=i+1}^{n}\sum_{l=i+1}^{n} \operatorname{cov}(k_n(X_1,X_i)k_n(X_1,X_j), k_n(X_2,X_i)k_n(X_2,X_l)) \\
&+ \sum_{i=4}^{n-1}\sum_{j=i+1}^{n}\sum_{k=3}^{i-1} \operatorname{cov}(k_n(X_1,X_i)k_n(X_1,X_j), k_n(X_2,X_k)k_n(X_2,X_i)),
\end{aligned}$$

which we rewrite as

$$R_2 = S_1 + S_2 + S_3 + S_4 + S_5.$$

Note now that, for all $m \in \{1,2,3,4,5\}$,

$$S_m = O(n^3 + n^2 h_n^{-1} + n h_n^{-2}).$$

This follows from Lemma 2.4 and the fact that there are 4 different indexes in each summand, except when one of the free indexes is exactly equal to one of the other indexes. This is shown for

S_1 for illustration, but the same argument applies to the other terms. We have

$$
\begin{aligned}
S_1 = & \sum_{i=3}^{n-1} \sum_{j=i+1}^{n} \sum_{\substack{l=3 \\ l\neq i, l\neq j}}^{n} \operatorname{cov}(k_n(X_1, X_i)k_n(X_1, X_j), k_n(X_2, X_1)k_n(X_2, X_l)) \\
& + \sum_{i=3}^{n-1} \sum_{j=i+1}^{n} \operatorname{cov}(k_n(X_1, X_i)k_n(X_1, X_j), k_n(X_2, X_1)k_n(X_2, X_i)) \\
& + \sum_{i=3}^{n-1} \sum_{j=i+1}^{n} \operatorname{cov}(k_n(X_1, X_i)k_n(X_1, X_j), k_n(X_2, X_1)k_n(X_2, X_j)) \\
& + \sum_{j=3}^{n} \operatorname{cov}(k_n(X_1, X_2)k_n(X_1, X_j), k_n(X_2, X_1)k_n(X_2, X_j)) \\
& + \sum_{j=3}^{n} \sum_{\substack{l=3 \\ l\neq j}}^{n} \operatorname{cov}(k_n(X_1, X_2)k_n(X_1, X_j), k_n(X_2, X_1)k_n(X_2, X_l))
\end{aligned}
$$

By the i.i.d. assumption, each summand in the five terms of the expansion is equal, respectively, to

$$
\begin{aligned}
& \operatorname{cov}(k_n(X_1, X_3)k_n(X_1, X_4), k_n(X_2, X_1)k_n(X_2, X_5)) \\
& \operatorname{cov}(k_n(X_1, X_3)k_n(X_1, X_4), k_n(X_2, X_1)k_n(X_2, X_3)) \\
& \operatorname{cov}(k_n(X_1, X_3)k_n(X_1, X_4), k_n(X_2, X_1)k_n(X_2, X_4)) \\
& \operatorname{cov}(k_n(X_1, X_2)k_n(X_1, X_3), k_n(X_2, X_1)k_n(X_2, X_3)) \\
& \operatorname{cov}(k_n(X_1, X_2)k_n(X_1, X_3), k_n(X_2, X_1)k_n(X_2, X_4))
\end{aligned}
$$

From Lemma 2.4, it follows that the first term in the expansion is $O(n^3)$, the second, third, and fifth terms are $O(n^2 h_n^{-1})$, while the fourth one is $O(nh_n^{-2})$. □

D Bias results in Giné and Nickl (2008)

Lemma D.1 (Part 1 of Theorem 1 in Giné and Nickl (2008)). *If K satisfies Assumption K and f_0 satisfies Assumption D(s) with $s \in (0, 1/2]$. Then the bias of U_n satisfies*

$$
\mathbb{E}[U_n] - \theta_0 = O(h_n^{2s}).
$$

Proof. Write $K_{h_n}(x) = h_n^{-1} K_n(h_n^{-1}x)$. We have

$$\begin{aligned}
\mathbb{E}[U_n] - \theta_0 &= \int_{\mathbb{R}} \int_{\mathbb{R}} K_{h_n}(x-y) f_0(y)\, dy\, f_0(x)\, dx - \int_{\mathbb{R}} f_0(x) f_0(x)\, dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} K_{h_n}(x-y)(f_0(y) - f_0(x)) f_0(x)\, dy\, dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} K(u)(f_0(x - uh_n) - f_0(x)) f_0(x)\, du\, dx \\
&= \int_{\mathbb{R}} K(u) \left(\int_{\mathbb{R}} \bar{f}_0(uh_n - x) f_0(x)\, dx - \int_{\mathbb{R}} \bar{f}_0(0 - x) f_0(x)\, dx \right) du \\
&= \int_{\mathbb{R}} K(u)((\bar{f}_0 * f_0)(uh_n)) - (\bar{f}_0 * f_0)(0))\, du,
\end{aligned}$$

where $*$ denotes convolution and $\bar{f}_0(x) = f_0(-x)$. The second equality follows from the fact that the kernel integrates to one, the third from the change of variable $y = x - uh_n$, and the fourth from Fubini's theorem. Then by applying Lemma D.2 since $f_0 \in L^1$ as a probability density function, we get

$$\begin{aligned}
|\mathbb{E}[U_n] - \theta_0| &\leq \int_{\mathbb{R}} C \|f_0\|_{2,s}^2 |K(u)| |uh_n|^{2s}\, du \\
&= c_1 h_n^{2s}
\end{aligned}$$

where $c_1 = C\|f_0\|_{2,s}^2 \int_{\mathbb{R}} |K(u)||u|^{2s}\, du$ and $0 < C < \infty$ is a constant independent of f_0 and h_n. □

Lemma D.2 (Lemma 1 in Giné and Nickl (2008))**.** *If $f, g \in L^1$ satisfy Assumption D(s) with $0 < s \leq 1/2$, then for any $x \in \mathbb{R}$ and $t \neq 0$,*

$$\frac{|(f*g)(x+t) - (f*g)(x)|}{|t|^{2s}} \leq C\|f\|_{2,s}\|g\|_{2,s}$$

where $0 < C < \infty$ is a fixed constant independent of f, g, x and t.

Proof. Denote F the Fourier transform. Since $f, g \in L^1$ and g bounded, $f*g \in L^1$ and continuous, and since $f, g \in L^2$, we have $F(f*g) \in L^1$. We then have

$$\begin{aligned}
\frac{|(f*g)(x+t) - (f*g)(x)|}{|t|^{2s}} &\leq |t|^{-2s} \|F^{-1}F[(f*g)(\cdot + t) - (f*g)(\cdot)]\|_\infty \\
&\leq (2\pi)^{-1}|t|^{-2s} \|F[(f*g)(\cdot + t) - (f*g)(\cdot)]\|_1 \\
&= (2\pi)^{-1}|t|^{-2s} \int_{\mathbb{R}} |F(f*g)(u)(e^{-iut} - 1)|\, du \\
&= (2\pi)^{-1} \int_{\mathbb{R}} |Ff(u)||u|^s |Fg(u)||u|^s \frac{|(e^{-iut} - e^{-i0})|}{|u|^{2s}|t|^{2s}}\, du \\
&\leq C\|f\|_{2,s}\|g\|_{2,s}.
\end{aligned}$$

The first inequality follows from the definition of the L^∞ norm and the Fourier inversion theorem, the second from the inequality $\|f\|_\infty \leq \|Ff\|_1$ (which also follows from the Fourier inversion theorem). The first equality follows from the definition of the L^1 norm and the second from the convolution theorem. The last inequality follows from Hölder's inequality and the fact that $e^{-i(\cdot)}$ is bounded Lipschitz. □

Remark D.1. The assumption $s \leq 1/2$ is needed for c_1 to be finite under assumption K. This can be relaxed if $\int |K(u)||u|^{2s}\, du < \infty$. If the kernel K is non-negative (and so is a density function), then this condition is equivalent to the random variable with density K has finite $2s$ order moments. This is often the case for kernel of order 1, and so the result can be naturally generalized to $s \geq 1/2$.