

Bootstrap and permutation tests of independence for point processes

Mélanie Albert*, Yann Bouret†, Magalie Fromont‡ and Patricia Reynaud-Bouret§

June 5, 2014

Abstract: Motivated by a neuroscience question about synchrony detection in spike trains analysis, we deal with the independence testing problem for point processes. We introduce non-parametric test statistics, which are rescaled general U -statistics, whose corresponding critical values are constructed from bootstrap and randomisation or permutation approaches, making as few assumptions as possible on the underlying distribution of the point processes. We derive general consistency results for the bootstrap and for the permutation w.r.t. to Wasserstein's metric, which induce weak convergence as well as convergence of second order moments. The obtained bootstrap or permutation independence tests are thus proved to be asymptotically of the prescribed size, and to be consistent against any reasonable alternative, randomisation or permutation independence tests having the further advantage to be exactly (that is non-asymptotically) of the prescribed level, even when Monte Carlo methods are used to approximate the randomised quantiles.

Mathematics Subject Classification: 62M07, 62E20, 60G55, 60F05.

Keywords: Independence test, U -statistics, point processes, bootstrap, randomisation, permutation.

1 Introduction

Originally motivated by a request of neuroscientists about synchrony detection in spike trains studies, we focus in the present work on independence tests for point processes. The question of testing whether two random variables are independent is of course largely encountered in the statistical literature, as it is one of the central goals of data analysis. From the historical Pearson's (see [38, 39]) chi-square test of independence to the recent test of Gretton et al. [22, 20, 21] using kernel methods in the spirit of statistical learning, many non-parametric independence tests have been developed for real valued random variables or even random vectors. Among them, of particular interest are the tests based on the randomisation or permutation principle introduced by Fisher [16], and covered thereafter in the series of papers by Pitman [43, 44, 45], Scheffe [54], Hoeffding [31] for instance, or bootstrap approaches derived from Efron's [15] "naive" one. Two families of such permutation or bootstrap-based independence tests may be distinguished at least : the whole family of rank tests including the tests of Hotelling and Pabst [34], Kendall [35], Wolfowitz [62] or Hoeffding [30] on the one hand, the family of Kolmogorov-Smirnov type tests, like Blum, Kiefer, and Rosenblatt's [6], Romano's [50] or Van der Vaart and Wellner's [58] ones on the other hand. These tests are purely non-parametric that is they are completely free of the underlying distributions of the observed random variables or vectors. They are proved to achieve asymptotically the right desired size: the probability, under independence, that the independence hypothesis is rejected tends to a prescribed α in $]0, 1[$, as the size of the original samples of observations grows to infinity. Moreover, the tests based on permutation are also known to be exactly (non-asymptotically meaning) of the desired level, that is the probability, under independence, that the independence hypothesis is rejected is smaller than the prescribed α , for any positive size of samples. Some of these tests are also proved to be consistent against many alternatives, such as Hoeffding's [30] one and the family of Kolmogorov-Smirnov type tests. Detecting dependence is also a fundamental old point in the neuroscientific literature (see [19] e.g.), as understanding how different areas of the brain interact with respect to a given stimulus helps to understand how the

*Univ. Nice Sophia Antipolis, CNRS, LJAD, UMR 7351, 06100 Nice, France.

†Univ. Nice Sophia Antipolis, CNRS, LPMC, UMR 7336, 06100 Nice, France.

‡Univ. Européenne de Bretagne, Institut de Recherche Mathématique de Rennes (IRMAR), France.

§Univ. Nice Sophia Antipolis, CNRS, LJAD, UMR 7351, 06100 Nice, France.

brain works. The neuroscience problem we were initially interested in consists of detecting interactions between occurrences of action potentials on two different neurons, observed on n independent trials, as described in [25]. Indeed, it is nowadays possible to simultaneously record the time occurrences of action potentials, which are quick and brutal variations of the electric potential of the neuron membrane, for several neurons at once, via extracellular multielectrodes. Each recorded set of time occurrences of action potentials for each neuron is usually referred to as a spike train, the spikes being the time occurrences themselves. It is commonly accepted that these spikes are one of the main components of the brain activity (see [56]). Therefore, when observing two spike trains coming from two different neurons, one of the main elementary problem is to assess whether these two spike trains are independent or not. Unfortunately, even if the real recordings of spike trains are discretized in time, due to the record resolution, and thus belong to finite dimensional spaces, the dimension of these spaces is so huge (from ten thousand up to one million) that it is neither realistic nor reasonable to model them by finite dimensional variables, and to apply usual independence tests. Several methods, such as the classical Unitary Events method introduced by Grün [23], consists of binning the spike trains at first in order to deal with vectorial data with reduced dimension. But it has been shown (see [26] e.g.) that these dimension reduction methods involve an information loss of more than 60% in some cases, making this kind of preprocessing quite proscribed despite its simplicity of use.

Modelling the recordings of spike trains by point processes that are only almost surely finite without any a priori bound on the total number of points, and using, constructing if needed, independence tests specifically dedicated to such point processes then appear as realistic and reasonable solutions.

Another field of applications, where detecting dependence between point processes is crucial, is, for instance, genomics, where point processes represent positions of motifs on the DNA strand as described in [27], or positions of Transcription Regulatory Elements on the DNA strand as described in [10].

In the present work, we mainly focus on the application in neuroscience detailed above. Since the existence of any precise underlying distribution for the point processes modelling spike trains data is subject to broad debate, to construct model free independence tests for point processes is of utmost importance in this context at least (see [24] for more details). To this end, bootstrap methods were used in [42, 41, 61] for binned data with relatively small dimension, without any proper mathematical justification. Besides the loss of information the binning data pre-processing involves, it appears in fact that the test statistics chosen in these papers are not suitable for consistency of the bootstrap approach, due to an incorrect centring device leading to some considerable bias. Asymptotic independence tests were furthermore introduced in [57] for point processes, but limited to homogeneous Poisson processes. Such a Poisson assumption is necessarily restrictive and even possibly inappropriate, considering spike trains analyses (see [46]), as well as other fields of applications. Moreover, these tests, of an asymptotic nature, may suffer from a lack of performance when the number of trials is small or moderate.

We therefore propose to construct new non-parametric tests of independence between two point processes, from the observation of an i.i.d. sample with the same distribution as the considered pair of point processes, making as few assumptions as possible on this underlying distribution. Our test statistics, correctly re-centred and re-scaled, are based on general U -statistics. The corresponding critical values are obtained from bootstrap or permutation approaches, so that the final tests that are proved to be asymptotically (with respect to the sample size) of the desired size and consistent against many alternatives, also have good performance in practice when the sample size is moderate or even small, as is often the case in neuroscience or genomics for biological or economic reasons.

A huge number of papers deal with the bootstrap or permutation of U -statistics of i.i.d. real valued random variables or vectors. Among them, we can cite for instance [4, 8, 1, 13], or [36] devoted to the bootstrap in a general framework, [33] or [11] devoted to the permutation in various testing frameworks, and of course, the papers cited above (see [50, 58]), which are devoted to Kolmogorov-Smirnov type tests, and based on particular U -statistics in an independence testing framework.

To our knowledge, there is no previous work on the bootstrap or permutation of general U -statistics of i.i.d. pairs of point processes, as considered in the present work. The main difficulties lie here in the nature of the mathematical objects we handle, that is point processes and their associated point measures which are random measures on the one hand, in the general nature of the results we aim at on the other hand. The proofs of our results, although inspired by Romano's [49, 50] work and Hoeffding's [31] precursor results on the permutation, are therefore more technical and complex on many aspects, that we will point out.

This paper is organized as follows.

We first present the problem of testing, and introduce the main notations. Starting from existing works in neuroscience, we introduce the test statistics we propose to use, based on general kernel-based U -statistics,

before listing and discussing the main assumptions made on the kernels and the underlying point processes. These points constitute the content of Section 2.

Section 3 is devoted to our bootstrap approach, from its description to the theoretical asymptotic properties of the corresponding independence tests, through general results about the consistency of the bootstrap for the considered U -statistics, expressed in terms of Wasserstein's metric as in [4]. We thus states that our tests are asymptotically of the desired size, and that they are consistent against every reasonable alternative.

Section 4 is devoted to the randomisation or permutation approach. We also begin with a description of the approach, then, we give general results about its consistency when the kernel of the U -statistic has a particular common form. These results are still expressed in terms of Wasserstein's metric, and thus induce usual weak convergence as well as convergence of second order moments. They also lead to prove that the corresponding permutation independence tests satisfy the same asymptotic properties as the bootstrap ones, with the further advantage of being exactly (that is non-asymptotically) of the desired level, even when a Monte Carlo method is used to approximate the randomised quantiles. This last point is easily deduced from a very useful lemma of Romano and Wolf (see [51, Lemma 1]).

Finally notice that the proofs of the results are postponed in Section 5 at the end of the paper.

2 From the neuroscience interpretation to a general test statistic

2.1 The testing problem

We consider in this paper point processes defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and observed on $[0, 1]$. More precisely, we consider finite point processes on $[0, 1]$, that is random point processes on $[0, 1]$, for which the total number of points is almost surely finite (see [12] for instance for more details). Typically, in a neuroscience framework, such finite point processes may represent spike trains recorded on a given finite interval of time, and rescaled so that their values may be assumed to belong to $[0, 1]$. The set \mathcal{X} of all their possible values is equipped with a metric $d_{\mathcal{X}}$, that we introduce in (9). This metric, issued from the Skorohod topology, makes \mathcal{X} separable and allows to define accordingly borelian sets on \mathcal{X} , and by extension through the product metric, on \mathcal{X}^2 .

Let us recall that the point measure dN_x associated to an element x of \mathcal{X} is defined for all measurable real-valued function f by :

$$\int_{[0,1]} f(u) dN_x(u) = \sum_{u \in x} f(u),$$

so that, in particular, the total number of points of x , denoted by $\#x$, is equal to $\int_{[0,1]} dN_x(u)$.

Moreover, for a point process X defined on $(\Omega, \mathcal{A}, \mathbb{P})$ and observed on $[0, 1]$, $\int f(u) dN_X(u)$ becomes a real random variable, defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Now, a pair $X = (X^1, X^2)$ of finite point processes defined on $(\Omega, \mathcal{A}, \mathbb{P})$, and observed on $[0, 1]$, has joint distribution P if $P(\mathcal{B}) = \mathbb{P}(X = (X^1, X^2) \in \mathcal{B})$, for any borelian set \mathcal{B} of \mathcal{X}^2 , and marginal distributions P^1 and P^2 if $P^1(\mathcal{B}^1) = \mathbb{P}(X^1 \in \mathcal{B}^1)$, and $P^2(\mathcal{B}^2) = \mathbb{P}(X^2 \in \mathcal{B}^2)$ for any borelian sets \mathcal{B}^1 and \mathcal{B}^2 of \mathcal{X} .

Given the observation of an i.i.d. sample $\mathbb{X}_n = (X_1, \dots, X_n)$ from the same distribution P as X , with $X_i = (X_i^1, X_i^2)$ for every $i = 1 \dots n$, we aim at testing

$$(H_0) \ X^1 \text{ and } X^2 \text{ are independent} \quad \text{against} \quad (H_1) \ X^1 \text{ and } X^2 \text{ are not independent},$$

which can also be written as

$$(H_0) \ P = P^1 \otimes P^2 \quad \text{against} \quad (H_1) \ P \neq P^1 \otimes P^2.$$

2.2 Independence test based on coincidences in neuroscience

Considering that the i.i.d. sample $\mathbb{X}_n = (X_1, \dots, X_n)$ models pairs of rescaled spike trains recorded for two neurons during n trials, the main dependence feature that needs to be detected between both neurons corresponds to synchronizations in time, referred to as coincidences (see [25] for instance). More precisely, neuroscientists expect to detect if such coincidences occur significantly, that is more than what may be due to chance. They speak in this case of a detected synchrony, which is of the utmost importance to understand how a neural network behaves.

Formalising the approach of [26], Tuleau-Malot et al. [57] first introduced the notion of coincidence count between two point processes X^1 and X^2 with delay δ ($\delta > 0$), defined by :

$$\varphi_\delta^{coinc}(X^1, X^2) = \int_{[0,1]^2} \text{coinc}_\delta(u, v) dN_{X^1}(u) dN_{X^2}(v), \quad (1)$$

with $\text{coinc}_\delta(u, v) = \mathbb{1}_{\{|u-v| \leq \delta\}}$.

Thus, $\varphi_\delta^{coinc}(X^1, X^2) = \sum_{u \in X^1, v \in X^2} \mathbb{1}_{\{|u-v| \leq \delta\}}$ is equal to the number of pairs (u, v) such that: u is a point of X^1 , v is point of X^2 , and $|u - v| \leq \delta$. Then, under the assumption that both X^1 and X^2 are homogeneous Poisson processes, Tuleau-Malot et al. proposed an independence test of (H_0) against (H_1) , rejecting (H_0) when a test statistic based on $\sum_{i=1}^n \varphi_\delta^{coinc}(X_i^1, X_i^2)$ is larger than a given critical value. This critical value is deduced from the asymptotic Gaussian distribution of the test statistic, under the null hypothesis of independence. The test is proved to be asymptotically of the desired level, but only under the homogeneous Poisson processes assumption.

However, it is now well-known that this assumption, as well as many other model assumptions, fails to be satisfied in practice for spike trains (see [46]). To construct a test detecting synchrony, with as few assumptions as possible on the underlying processes, is therefore an interesting, and challenging purpose.

In the present work, we propose a new non-parametric independence test for point processes, which allows, without any Poisson or other specific model assumption on the underlying processes, to detect particular time dependence features such as synchrony, but also more general ones.

2.3 General non-degenerate U -statistics as independence test statistics

In the parametric homogeneous Poisson framework of [57], the expectation of $\varphi_\delta^{coinc}(X_i^1, X_i^2)$ has a simple expression as a function of δ and the intensities λ_1 and λ_2 of X^1 and X^2 . Since λ_1 and λ_2 can be easily estimated, an estimator of this expectation can thus be obtained using the plug-in principle, and subtracted from $\varphi_\delta^{coinc}(X_i^1, X_i^2)$ to lead to a test statistic with a centred asymptotic distribution under the null hypothesis of independence.

In the present non-parametric framework where we want to make as few assumptions as possible on the point processes X^1 and X^2 , such a centring plug-in tool is not available, and we propose to use instead a self-centring trick, which amounts to consider

$$\sum_{i \neq i' \in \{1, \dots, n\}} (\varphi_\delta^{coinc}(X_i^1, X_i^2) - \varphi_\delta^{coinc}(X_i^1, X_{i'}^2)).$$

Furthermore, it is clear that the function φ_δ^{coinc} used in [57] suits the dependence feature the neuroscientists expect to detect in a spike trains analysis. However, it is not necessarily the best choice for other kinds of dependence features to be detected in a general point processes analysis. So we introduce a more general interaction function defined by

$$\varphi_\delta^w(X^1, X^2) = \int_{[0,1]^2} w_\delta(u, v) dN_{X^1}(u) dN_{X^2}(v), \quad (2)$$

where w_δ is an integrable function possibly depending on a real-valued vector δ of parameters. Of course, the choice $w_\delta = \text{coinc}_\delta$ with $\delta > 0$ allows to get back the coincidence count function defined in (1), while the choice $w_{(j,k)}(u, v) = \psi_{j,k}(v - u)$, where $\psi_{j,k}$ is a rescaled (j) and translated (k) Haar mother wavelet, leads to the interaction function used in [53] for an interaction test with strong theoretical and practical properties in a specific Poisson framework.

Even more generally, we could choose any integrable function φ , whose aim would be to detect particular dependence features on the pair (X^1, X^2) for instance, and for which we introduce:

$$T_{n,\varphi} = \sum_{i \neq i' \in \{1, \dots, n\}} (\varphi(X_i^1, X_i^2) - \varphi(X_i^1, X_{i'}^2)). \quad (3)$$

Since $T_{n,\varphi}/(n(n-1))$ is an unbiased estimator of $\int \int \varphi(x^1, x^2) (dP(x^1, x^2) - dP^1(x^1)dP^2(x^2))$, it may be a reasonable independence test statistic, with expectation equal to 0 under (H_0) , and this without any assumption on the underlying point processes.

Notice that if X^1 and X^2 were finite dimensional variables with continuous distributions with respect to the Lebesgue measure, this test statistic would be closely related to generalized Kolmogorov-Smirnov type tests of

independence. For instance, the test statistics of Blum, Kiefer, and Rosenblatt [6], Romano [50], Van der Vaart and Wellner in [58] are equal to

$$n^{-3/2} \sup_{v^1 \in \mathcal{V}^1, v^2 \in \mathcal{V}^2} \left| T_{n, \varphi(v^1, v^2)} \right|,$$

where, respectively:

- $\mathcal{V}^1 = \mathcal{V}^2 = \mathbb{R}$ and $\varphi(v^1, v^2)(x^1, x^2) = \mathbb{1}_{[-\infty, v^1]}(x^1) \mathbb{1}_{[-\infty, v^2]}(x^2)$,
- \mathcal{V}^1 and \mathcal{V}^2 are countable V.-C. classes of subsets of \mathbb{R}^d , and $\varphi(v^1, v^2)(x^1, x^2) = \mathbb{1}_{\{v^1\}}(x^1) \mathbb{1}_{\{v^2\}}(x^2)$,
- \mathcal{V}^1 and \mathcal{V}^2 are well-chosen classes of real-valued functions and $\varphi(v^1, v^2)(x^1, x^2) = v^1(x^1) v^2(x^2)$.

Now, let us notice that setting

$$h_\varphi(x, y) = \frac{1}{2} \left(\varphi(x^1, x^2) + \varphi(y^1, y^2) - \varphi(x^1, y^2) - \varphi(y^1, x^2) \right), \quad (4)$$

for every $x = (x^1, x^2)$, $y = (y^1, y^2)$ in \mathcal{X}^2 , then $T_{n, \varphi}$ can also be written as

$$T_{n, \varphi} = \sum_{i \neq i' \in \{1, \dots, n\}} h_\varphi(X_i, X_{i'}), \text{ with } X_i = (X_i^1, X_i^2), X_{i'} = (X_{i'}^1, X_{i'}^2),$$

so $T_{n, \varphi}/(n(n-1))$ is a classical U -statistic with a symmetric kernel h_φ , as introduced by Hoeffding [29].

Therefore, we finally consider in the present work independence test statistics which are based on U -statistics of the form:

$$U_{n, h}(\mathbb{X}_n) = \frac{1}{n(n-1)} \sum_{i \neq i' \in \{1, \dots, n\}} h(X_i, X_{i'}), \quad (5)$$

for some kernel h of the form h_φ given in (4), or even for some general symmetric kernel $h : (\mathcal{X}^2)^2 \rightarrow \mathbb{R}$ such that $\mathbb{E}[h(X_i, X_{i'})] = 0$ for every $X_i, X_{i'}$ i.i.d. with distribution P on \mathcal{X}^2 satisfying $(H_0) P = P^1 \otimes P^2$.

Notice that assuming that h is symmetric is not restrictive since for any kernel h , $2U_{n, h} = U_{n, h_{sym}}$ where $h_{sym}(x, y) = h(x, y) + h(y, x)$ always defines a symmetric kernel.

Following the works of Romano [50] or Van der Vaart and Wellner [58], the tests we propose here are based on bootstrap and permutation approaches for the above general U -statistics.

Most of our theoretical results are true, whether the considered U -statistics $U_{n, h}(\mathbb{X}_n)$ are degenerate or not. However, for degenerate U -statistics, the interpretation of these results is very poor. So, we will focus on non-degenerate U -statistics, and therefore discuss below what degeneracy means about the underlying processes.

Since under (H_0) $U_{n, h}(\mathbb{X}_n)$ is assumed to have zero mean, it will be degenerate if and only if for X with distribution $P^1 \otimes P^2$ and for $P^1 \otimes P^2$ -almost every x in \mathcal{X}^2 ,

$$\mathbb{E}[h(x, X)] = 0.$$

In the particular case where $h = h_\varphi$, with h_φ defined by (4) and φ some integrable function, this amounts to state that for $X = (X^1, X^2)$ with distribution $P^1 \otimes P^2$ and for $P^1 \otimes P^2$ -almost every $x = (x^1, x^2)$,

$$\varphi(x^1, x^2) + \mathbb{E}[\varphi(X^1, X^2)] - \mathbb{E}[\varphi(x^1, X^2)] - \mathbb{E}[\varphi(X^2, x^1)] = 0.$$

This condition's fulfillment would however imply a very particular link between φ and the distribution of the bivariate point process X , which is unknown. Let us look more deeply at the case where $\varphi = \varphi_\delta^w$ given by (2). In this case, the above condition can be rewritten as follows: for $P^1 \otimes P^2$ -almost every $x = (x^1, x^2)$,

$$\begin{aligned} & \int w_\delta(u, v) dN_{x^1}(u) dN_{x^2}(v) + \mathbb{E} \left[\int w_\delta(u, v) dN_{X^1}(u) dN_{X^2}(v) \right] \\ & - \mathbb{E} \left[\int w_\delta(u, v) dN_{x^1}(u) dN_{X^2}(v) \right] - \mathbb{E} \left[\int w_\delta(u, v) dN_{X^1}(u) dN_{x^2}(v) \right] = 0. \quad (6) \end{aligned}$$

Therefore, if for instance $P^1(\{\emptyset\}) > 0$ and $P^2(\{\emptyset\}) > 0$, which is the case for Bernoulli processes (discretized processes), Poisson processes, Hawkes processes or more generally any processes with conditional intensities, this implies, by taking $x^1 = x^2 = \emptyset$, that

$$\mathbb{E} \left[\int w_\delta(u, v) dN_{X^1}(u) dN_{X^2}(v) \right] = 0. \quad (7)$$

If the function w_δ is nonnegative, then this finally gives that $\int w_\delta(u, v) dN_{X^1}(u) dN_{X^2}(v)$ is almost surely equal to 0. However, in practice, it is always possible to choose a function w_δ so that the observation of at least one of the $\int w_\delta(u, v) dN_{X^1}(u) dN_{X^2}(v)$'s is not equal to zero, and the U -statistic is not degenerate.

When considering spike trains analysis in neuroscience, notice first that assuming that the processes may be empty (but not empty almost surely) is an obvious assumption in practice. Indeed, there often exist trials (usually short) where, just by chance, no spikes have been detected for some trials while some spikes have been detected for other trials. Moreover, if $w_\delta = \text{coinc}_\delta$ as in (1), δ is chosen large enough so that coincidences are observed in practice, so (7) is not satisfied, and the U -statistic is non-degenerate.

However notice that when $w_\delta(u, v) = \psi_{j,k}(v - u)$, where $\delta = (j, k)$ and $\psi_{j,k}$ is a rescaled and translated Haar mother wavelet as in [53], when considering specific Poisson processes, the condition (7) is fulfilled, since $\int \psi_{j,k}(u) du = 0$ is required to make the test statistic in [53] a reasonable test statistics. But in this case, the U -statistic is still non-degenerate since the left-hand side of (6) has zero mean and positive variance, as soon as both Poisson processes have non zero intensities (see Section 3.5 for more details).

3 Bootstrap tests of independence

3.1 Description of the bootstrap approach and informal view of the results

Recall that we observe an i.i.d. sample $\mathbb{X}_n = (X_1, \dots, X_n)$ with $X_i = (X_i^1, X_i^2)$ of distribution P on \mathcal{X}^2 , whose marginals are respectively denoted P^1 and P^2 , and that the test statistics we consider are based on the U -statistics $U_{n,h}(\mathbb{X}_n)$ defined by (5). Since the distribution of these test statistics is not free from the unknown underlying marginal distributions P^1 and P^2 under the null hypothesis, we turn to a classical bootstrap approach, whose aim is to mimic the distribution of the test statistics under (H_0) , for large, but also moderate or small sample sizes. Since each $X_i = (X_i^1, X_i^2)$ is $P^1 \otimes P^2$ -distributed under (H_0) , the first and second coordinates of the elements of \mathbb{X}_n are resampled according to the corresponding marginal empirical distributions. More precisely, we consider, for $j = 1, 2$, the marginal sample distribution P_n^j of (X_1^j, \dots, X_n^j) given by

$$P_n^j = \frac{1}{n} \sum_{i=1}^n \delta_{X_i^j}. \quad (8)$$

Then the bootstrap sample $\mathbb{X}_n^* = (X_{n,1}^*, \dots, X_{n,n}^*)$, with $X_{n,i}^* = (X_{n,i}^{1*}, X_{n,i}^{2*})$, is defined as an n i.i.d. sample from the distribution $P_n^1 \otimes P_n^2$.

In particular, we prove (see Theorem 1) that, under appropriate assumptions, the conditional distribution of $\sqrt{n}U_{n,h}(\mathbb{X}_n^*)$ given \mathbb{X}_n is asymptotically close to the one of $\sqrt{n}U_{n,h}(\mathbb{X}_n^\perp)$, where \mathbb{X}_n^\perp is an i.i.d sample from the distribution $P^1 \otimes P^2$. Therefore, using the quantiles of the conditional distribution of $\sqrt{n}U_{n,h}(\mathbb{X}_n^*)$ given \mathbb{X}_n as critical values, we can build unilateral or bilateral tests of (H_0) against (H_1) with asymptotic size α . These tests also satisfy consistence properties against alternatives such that $\int h(x, x') dP(x) dP(x') \neq 0$ (see Theorem 2).

Before stating these results and to avoid any confusion, we introduce a few notations.

- For any functional $Z : (\mathcal{X}^2)^n \rightarrow \mathbb{R}$, $\mathcal{L}(Z, Q)$ denotes the distribution of the variable $Z(\mathbb{X}_n)$, where \mathbb{X}_n is an i.i.d. sample from the distribution Q on \mathcal{X}^2 .
- If the distribution $Q = Q(W)$ is random and depends on a random variable W , then $\mathcal{L}(Z, Q|W)$ is the conditional distribution of the variable $Z(\mathbb{X}_n)$, where \mathbb{X}_n is an i.i.d. sample from the distribution $Q = Q(W)$ on \mathcal{X}^2 , given W .
- " Q -a.s. in $(X_n)_n$ " at the end of a statement means that the statement only depends on the sequence $(X_n)_n$, where the X_n 's are i.i.d with distribution Q , and that there exists an event \mathcal{C} only depending on $(X_n)_n$ such that $\mathbb{P}(\mathcal{C}) = 1$, on which the statement is true.

- " $Q_n \Rightarrow Q$ " means that the sequence of distributions $(Q_n)_n$ converges towards Q in the weak sense, that is for any real valued, continuous and bounded function g , $\int g(z) dQ_n(z) \xrightarrow{n \rightarrow +\infty} \int g(z) dQ(z)$.
- Finally, as we often work conditionally on \mathbb{X}_n , we will denote by $\mathbb{E}^*[\cdot]$ the conditional expectation given the sample \mathbb{X}_n .

In particular, the conditional distribution of $\sqrt{n}U_{n,h}(\mathbb{X}_n^*)$ given \mathbb{X}_n is denoted by $\mathcal{L}(\sqrt{n}U_{n,h}, P_n^1 \otimes P_n^2 | \mathbb{X}_n)$.

3.2 Main assumptions

Since the random variables we deal with are not real-valued variables but point processes, the proofs of our results will need many assumptions, which may seem to be somewhat abstract. We therefore describe and discuss them in the present section.

Centring assumptions on the kernel h

First notice that, under (H_0) , the U -statistic introduced in (5) has zero mean if the kernel h satisfies the following assumption:

$$(\mathcal{A}_{Cent}) \quad \left| \begin{array}{l} \text{For } X_i \text{ and } X_{i'}, \text{ i.i.d. with distribution } P^1 \otimes P^2 \text{ on } \mathcal{X}^2, \\ \mathbb{E}[h(X_i, X_{i'})] = 0. \end{array} \right.$$

An empirical and resampled version of this assumption is furthermore stated as:

$$(\mathcal{A}_{Cent}^*) \quad \left| \begin{array}{l} \text{For } x_1 = (x_1^1, x_1^2), \dots, x_n = (x_n^1, x_n^2) \text{ in } \mathcal{X}^2, \\ \sum_{i_1, i_2, i'_1, i'_2=1}^n h\left((x_{i_1}^1, x_{i_2}^2), (x_{i'_1}^1, x_{i'_2}^2)\right) = 0. \end{array} \right.$$

Notice that both assumptions are fulfilled as soon as h is of the form h_φ given by (4), though it is not necessary. The kernel $h : ((x^1, x^2), (y^1, y^2)) \mapsto \#x^1 \cdot \#x^2 \cdot \#y^1 \cdot \#y^2 (\#x^1 - \#y^1)(\#x^2 - \#y^2)$ indeed satisfies both assumptions though it can not be written as an h_φ .

Moment assumptions on the kernel h

Since the metric we use in the following to prove the asymptotic validity of our bootstrap procedure is the \mathbb{L}^2 -Wasserstein distance, we will need moment assumptions on the kernel h . In particular, we will assume that the variance of $U_{n,h}(\mathbb{X}_n)$ exists, i.e.

$$(\mathcal{A}_{Mmt}) \quad \left| \begin{array}{l} \text{For } X_i \text{ and } X_{i'}, \text{ i.i.d. with distribution } P \text{ on } \mathcal{X}^2, \\ \mathbb{E}[h^2(X_i, X_{i'})] < \infty. \end{array} \right.$$

We will also need a resampled version of this assumption as:

$$(\mathcal{A}_{Mmt}^*) \quad \left| \begin{array}{l} \text{For } X_1, X_2, X_3, X_4 \text{ i.i.d. with distribution } P \text{ on } \mathcal{X}^2, i_1, i_2, i'_1, i'_2 \text{ in } \{1, 2, 3, 4\}, \\ \mathbb{E}\left[h^2\left((X_{i_1}^1, X_{i_2}^2), (X_{i'_1}^1, X_{i'_2}^2)\right)\right] < +\infty. \end{array} \right.$$

Notice that when (\mathcal{A}_{Mmt}^*) is satisfied, this implies that:

- (\mathcal{A}_{Mmt}) is satisfied (taking $i_1 = i_2, i'_1 = i'_2$, and $i'_1 \neq i_1$),
- for X_i with distribution P , $\mathbb{E}[h^2(X_i, X_i)] < +\infty$ (taking $i_1 = i_2 = i'_1 = i'_2$),
- for $X_i, X_{i'}$ i.i.d with distribution $P^1 \otimes P^2$, $\mathbb{E}[h^2(X_i, X_{i'})] < +\infty$ (taking i_1, i_2, i'_1, i'_2 all different).

A sufficient condition for (\mathcal{A}_{Mmt}^*) (and consequently (\mathcal{A}_{Mmt})) to be satisfied is that there exist some positive constants α_1, α_2, C such that for every $x = (x^1, x^2), y = (y^1, y^2)$ in \mathcal{X}^2 ,

$$|h(x, y)| \leq C((\#x^1)^{\alpha_1} + (\#y^1)^{\alpha_1})(\#x^2)^{\alpha_2} + (\#y^2)^{\alpha_2},$$

with $\mathbb{E}[(\#X^1)^{4\alpha_1}] < +\infty$ and $\mathbb{E}[(\#X^2)^{4\alpha_2}] < +\infty$.

In the particular case where h is of the form h_φ given by (4), a possible sufficient condition is that there exist some positive constants α_1, α_2, C such that for every x^1, x^2 in \mathcal{X} , $|\varphi(x^1, x^2)| \leq C(\#x^1)^{\alpha_1}(\#x^2)^{\alpha_2}$, with $\mathbb{E}[(\#X^1)^{4\alpha_1}] < +\infty$ and $\mathbb{E}[(\#X^2)^{4\alpha_2}] < +\infty$.

Notice that the coincidence count function φ_δ^{coinc} defined by (1), satisfies, for every x^1, x^2 in \mathcal{X} ,

$$|\varphi_\delta^{coinc}(x^1, x^2)| \leq (\#x^1)(\#x^2).$$

Therefore, (\mathcal{A}_{Mmt}^*) and (\mathcal{A}_{Mmt}) are satisfied as soon as $\mathbb{E}[(\#X^1)^4] < +\infty$ and $\mathbb{E}[(\#X^2)^4] < +\infty$.

Moment bounds for the total number of points of the processes, such as the ones stated above, are in fact satisfied in many situations. Since any discretized point process at resolution $0 < r < 1$ (see [57] for a definition) has at most $1/r$ points, such moment bounds are obviously satisfied in the case where the considered point processes are discretized ones. Moreover, since any Poisson process has a total number of points obeying a Poisson distribution, which has exponential moment of any order, such moment bounds are also obviously satisfied in the case of Poisson processes, as well as in the case of point processes with bounded conditional intensities, which can be constructed by thinning of homogeneous Poisson processes on \mathbb{R} [37]. Similar moment bounds can also be obtained (see [28]) for linear stationary Hawkes processes with positive interaction functions, that are quite classical models in spike trains analysis (see for instance [40, 57]). This finally may be extended to point processes whose conditional intensities is upper bounded by intensities of linear stationary Hawkes processes with positive interaction functions, by thinning arguments again, which includes more general Hawkes processes (see [7]) and in particular Hawkes processes used to model inhibition in spike trains analysis (see [28, 57, 48] or [47]).

Continuity of the kernel h

Let us recall that \mathcal{X} is the set of the possible values of finite point processes defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and observed on $[0, 1]$, and notice that it can be identified with the space \mathcal{D} of càdlàg functions on $[0, 1]$ through the identification $I : x \in \mathcal{X} \mapsto \left(t \mapsto \int_0^1 \mathbb{1}_{\{u \leq t\}} dx_u\right) \in \mathcal{D}$. Now, considering the uniform Skorohod topology on \mathcal{D} , associated with the metric $d_{\mathcal{D}}$ defined by (see [5] e.g.):

$$d_{\mathcal{D}}(f, f') = \inf \left\{ \varepsilon > 0 \mid \exists \lambda \in \Lambda, \sup_{t \in [0, 1]} |\lambda(t) - t| \leq \varepsilon, \sup_{t \in [0, 1]} |f(\lambda(t)) - f'(t)| \leq \varepsilon \right\},$$

where Λ denotes the class of strictly increasing, continuous mappings of $[0, 1]$ onto itself, \mathcal{X} can be endowed with the topology induced by the metric $d_{\mathcal{X}}$ defined on \mathcal{X} by:

$$d_{\mathcal{X}}(x, x') = d_{\mathcal{D}}(I(x), I(x')) \quad \text{for every } x, x' \text{ in } \mathcal{X}. \quad (9)$$

Since $(\mathcal{D}, d_{\mathcal{D}})$ is a separable metric space, so are the spaces $(\mathcal{X}, d_{\mathcal{X}})$, $(\mathcal{X}^2, d_{\mathcal{X}^2})$, where $d_{\mathcal{X}^2}$ is the usual product metric defined from $d_{\mathcal{X}}$ (see [14, p32] e.g.), and $(\mathcal{X}^2 \times \mathcal{X}^2, d)$, where d is the product metric defined from $d_{\mathcal{X}}$, finally given by:

$$d((x, y), (x', y')) = \sup \left\{ \sup_{j=1,2} \left\{ d_{\mathcal{X}}(x^j, x'^j) \right\}, \sup_{j=1,2} \left\{ d_{\mathcal{X}}(y^j, y'^j) \right\} \right\}, \quad (10)$$

for every $x = (x^1, x^2)$, $y = (y^1, y^2)$, $x' = (x'^1, x'^2)$, $y' = (y'^1, y'^2)$ in \mathcal{X}^2 .

We assume that the kernel h defining the U -statistic $U_{n,h}(\mathbb{X}_n)$ in (5) satisfies:

$$(\mathcal{A}_{Cont}) \quad \left\{ \begin{array}{l} \text{There exists a subset } \mathcal{C} \text{ of } \mathcal{X}^2 \times \mathcal{X}^2, \text{ such that} \\ \text{(i) } h \text{ is continuous in every } (x, y) \text{ in } \mathcal{C} \text{ for the topology induced by the metric } d \text{ defined by (10),} \\ \text{(ii) } (P^1 \otimes P^2)^{\otimes 2}(\mathcal{C}) = 1. \end{array} \right.$$

Notice that in the particular case where h is of the form $h_{\varphi_\delta^w}$ defined by (2) and (4), with a continuous function $w_\delta : [0, 1]^2 \rightarrow \mathbb{R}$, then (\mathcal{A}_{Cont}) is necessarily fulfilled.

Considering the coincidence count kernel $h_{\varphi_\delta^{coinc}}$, defined by (1) and (4), since the function $(u, v) \in [0, 1]^2 \mapsto \mathbb{1}_{\{|u-v| \leq \delta\}}$ is not continuous, it is not so clear. However, one has the following result whose proof is given in Section 5.14.

Proposition 1. *The coincidence count kernel $h_{\varphi_\delta^{coinc}}$ defined on $\mathcal{X}^2 \times \mathcal{X}^2$ by (1) and (4) is continuous w.r.t. the topology induced by d , defined by (10), on the set*

$$\mathcal{C}_\delta = \left\{ ((x^1, x^2), (y^1, y^2)) \in \mathcal{X}^2 \times \mathcal{X}^2 \mid (\{x^1\} \cup \{y^1\}) \cap (\{x^2 \pm \delta\} \cup \{y^2 \pm \delta\}) = \emptyset \right\}. \quad (11)$$

As suggested in [57], when dealing with discretized point processes at resolution r in spike trains analysis, the right choice for δ is $dr + r/2$ for some integer d . Thus, in this case, the condition (ii) in (\mathcal{A}_{Cont}) is satisfied by the set \mathcal{C}_δ of (11), hence (\mathcal{A}_{Cont}) is satisfied by $h_{\varphi_\delta^{coinc}}$. Furthermore, when dealing with point processes that have conditional intensities, so that they may be constructed by thinning Poisson processes (see [37] e.g.), the probability $(P^1 \otimes P^2)^{\otimes 2}$ of \mathcal{C}_δ in (11) is larger than the probability that $X \cap (X' \pm \delta) = \emptyset$, when X and X' are two independent Poisson processes. Thus, (\mathcal{A}_{Cont}) is also satisfied by $h_{\varphi_\delta^{coinc}}$ in this case.

3.3 Consistency of the bootstrap approach

As in the historical paper by Bickel et Freedman [4], the closeness between $\mathcal{L}(\sqrt{n}U_{n,h}, P_n^1 \otimes P_n^2 | \mathbb{X}_n)$ and $\mathcal{L}(\sqrt{n}U_{n,h}, P^1 \otimes P^2)$, that are both distributions on \mathbb{R} , is here measured via the classical \mathbb{L}^2 -Wasserstein's metric d_2 (also called Mallows' metric) on the space Γ_2 of distributions Q on \mathbb{R} such that $\int z^2 dQ(z) < \infty$. Recall that the metric d_2 is defined by:

$$d_2^2(Q, Q') = \inf \{ \mathbb{E}[(Z - Z')^2] \mid (Z, Z') \text{ has marginal distributions } Q \text{ and } Q' \}, \quad (12)$$

for every Q, Q' in Γ_2 , and that convergence w.r.t. d_2 is equivalent to both weak convergence and convergence of second order moments.

The validity of the bootstrap approach described above for our independence tests is mainly due to the following consistency result.

Theorem 1. *Let $(X_n)_n$ be a sequence of i.i.d. pairs of point processes with distribution P on \mathcal{X}^2 and with marginals P^1 and P^2 on \mathcal{X} . For every $n \geq 2$, let $\mathbb{X}_n = (X_1, \dots, X_n)$, let P_n^j for $j = 1, 2$ be the empirical marginal distributions defined by (8) and let $U_{n,h}$ be defined by (5), with a measurable symmetric kernel h on $\mathcal{X}^2 \times \mathcal{X}^2$. Assume that (\mathcal{A}_{Cent}) , (\mathcal{A}_{Cent}^*) , (\mathcal{A}_{Mmt}^*) and (\mathcal{A}_{Cont}) are all satisfied. Then,*

$$d_2(\mathcal{L}(\sqrt{n}U_{n,h}, P_n^1 \otimes P_n^2 | \mathbb{X}_n), \mathcal{L}(\sqrt{n}U_{n,h}, P^1 \otimes P^2)) \xrightarrow{n \rightarrow +\infty} 0 \quad P\text{-a.s. in } (X_n)_n.$$

The proof follows similar arguments to the ones of [4] for the bootstrap of the mean, or to [13] and [36] for the bootstrap of U -statistics. The difficulty of the present proof lies in the original data being here finite point processes instead of real-valued random variables. The main point is therefore to transpose finite point processes in the separable Skorohod metric space $(\mathcal{D}, d_{\mathcal{D}})$, where weak convergence of sample probability distributions is available (see [60]).

This theorem derives from the following two propositions which may be useful in various frameworks. The first one states a non-asymptotic result, while the second one gives rather natural results of convergence.

Proposition 2. *Under assumptions (\mathcal{A}_{Cent}) , (\mathcal{A}_{Cent}^*) , (\mathcal{A}_{Mmt}) , and in the notation of Theorem 1, there exists some constant $C > 0$ such that for every integer $n \geq 2$,*

$$d_2^2(\mathcal{L}(\sqrt{n}U_{n,h}, P_n^1 \otimes P_n^2 | \mathbb{X}_n), \mathcal{L}(\sqrt{n}U_{n,h}, P^1 \otimes P^2)) \leq C \inf \left\{ \mathbb{E}^* \left[(h(Y_{n,1}^*, Y_{n,2}^*) - h(Y_1, Y_2))^2 \right] \right. \\ \left. Y_{n,1}^* \sim P_n^1 \otimes P_n^2, Y_1 \sim P^1 \otimes P^2, \text{ and } (Y_{n,2}^*, Y_2) \text{ is an independent copy of } (Y_{n,1}^*, Y_1) \right\}.$$

Proposition 3. *Let $(X_n)_n$ be a sequence of i.i.d. pairs of point processes with distribution P on \mathcal{X}^2 and with marginals P^1 and P^2 on \mathcal{X} . Let $\mathbb{X}_n = (X_1, \dots, X_n)$ for every $n \geq 2$, let h be a measurable symmetric kernel on $\mathcal{X}^2 \times \mathcal{X}^2$, and let $U_{n,h}(\mathbb{X}_n)$ be the U -statistic defined by (5).*

- *If $\mathbb{E}[|h(X, X')|] < +\infty$, for X, X' i.i.d with distribution P , then one has*

$$U_{n,h}(\mathbb{X}_n) \xrightarrow{n \rightarrow +\infty} \mathbb{E}[h(X, X')] = \int h(x, x') dP(x) dP(x') \quad P\text{-a.s. in } (X_n)_n. \quad (13)$$

- *Under (\mathcal{A}_{Mmt}^*) , one moreover obtains*

$$\frac{1}{n^4} \sum_{i,j,k,l=1}^n h^2((X_i^1, X_j^2), (X_k^1, X_l^2)) \xrightarrow{n \rightarrow +\infty} \mathbb{E}[h^2((X_1^1, X_2^2), (X_3^1, X_4^2))] \quad P\text{-a.s. in } (X_n)_n. \quad (14)$$

3.4 Convergence of cumulative distribution functions and quantiles

As usual, $\mathcal{N}(m, v)$ stands for the Gaussian distribution with mean m and variance v , $\Phi_{m,v}$ for its cumulative distribution function (c.d.f.) and $\Phi_{m,v}^{-1}$ for the corresponding quantile function. From the results of Rubin and Vitale [52] generalising Hoeffding's [29] Central Limit Theorem for nondegenerate U -statistics when the X_i 's are random vectors, we can also easily deduce the following Central Limit Theorem for $U_{n,h}(\mathbb{X}_n)$.

Proposition 4. *Let $(X_n)_n$ be a sequence of i.i.d. pairs of point processes with distribution $P^1 \otimes P^2$ on \mathcal{X}^2 . Let h be a measurable symmetric kernel satisfying (\mathcal{A}_{Cent}) and (\mathcal{A}_{Mmt}) and for every $n \geq 2$, let $U_{n,h}$ be the U -statistic defined by (5). Then, if $U_{n,h}$ is nondegenerate,*

$$\mathcal{L}(\sqrt{n}U_{n,h}, P^1 \otimes P^2) \Rightarrow \mathcal{N}(0, \sigma_{h,P^1 \otimes P^2}^2),$$

with

$$\sigma_{h,P^1 \otimes P^2}^2 = 4\text{Var}(\mathbb{E}[h(X, X') | X]) = 4\text{Cov}(h(X, X'), h(X, X'')) > 0, \quad (15)$$

X, X', X'' being i.i.d. pairs of point processes with distribution $P^1 \otimes P^2$ on \mathcal{X}^2 .

Moreover, since $\mathbb{E}[(\sqrt{n}U_{n,h}(\mathbb{X}_n))^2] \rightarrow \sigma_{h,P^1 \otimes P^2}^2$, the following also holds

$$d_2(\mathcal{L}(\sqrt{n}U_{n,h}, P^1 \otimes P^2), \mathcal{N}(0, \sigma_{h,P^1 \otimes P^2}^2)) \xrightarrow{n \rightarrow \infty} 0.$$

Comment. The above asymptotic normality result may lead to a rather simple test, of the desired asymptotic size. Indeed, by Slutsky's lemma and the law of large numbers for U -statistics of order 3, if

$$S_{n,h}(\mathbb{X}_n) = \sqrt{n}U_{n,h}(\mathbb{X}_n)/\hat{\sigma}_h,$$

with

$$\hat{\sigma}_h^2 = \frac{4}{n(n-1)(n-2)} \sum_{i,j,k \in \{1, \dots, n\}, \# \{i,j,k\} = 3} h(X_i, X_j)h(X_i, X_k),$$

then under (H_0) , $S_{n,h}(\mathbb{X}_n)$ converges in distribution towards $\mathcal{N}(0, 1)$.

Therefore, the test that rejects (H_0) when $|S_{n,h}(\mathbb{X}_n)| \geq \Phi_{0,1}^{-1}(1 - \alpha/2)$ is of asymptotic size α . One can also easily prove that it is consistent (i.e. of asymptotic power 1) against any alternative P such that $\mathbb{E}[h(X, X')] \neq 0$, for X, X' i.i.d. with distribution P and satisfying (\mathcal{A}_{Mmt}) .

Such tests, which are purely asymptotic, may of course suffer from a lack of power when the sample size n is small or even moderate, which is typically the case for the application in neuroscience described in Section 2, since the number of trials can never be very large for biological reasons (from few tens up to few hundreds at best).

This is the reason why we turn here to bootstrap and permutation approaches developed by Romano [50] in another framework, which are known to better fit for small or moderate sample sizes.

Notice that the asymptotic normality of $\sqrt{n}U_{n,h}(\mathbb{X}_n)$, stated in Proposition 4, has not been used in the proof of Theorem 1. Since it notably means that $\sqrt{n}U_{n,h}(\mathbb{X}_n)$ weakly converges to a distribution with continuous c.d.f., we can however use it to prove the convergence of the conditional c.d.f. or quantiles of the considered bootstrap distributions in the following Corollary.

Corollary 1. *Let $(X_n)_n$ be a sequence of i.i.d. pairs of point processes with distribution P on \mathcal{X}^2 and with marginals P^1 and P^2 on \mathcal{X} . For every $n \geq 2$, let $\mathbb{X}_n = (X_1, \dots, X_n)$, and in the notation of Theorem 1, let \mathbb{X}_n^* be a bootstrap sample defined as an i.i.d sample from the distribution $P_n^1 \otimes P_n^2$. Let \mathbb{X}_n^\perp be another i.i.d. sample with distribution $P^1 \otimes P^2$ on \mathcal{X}^2 , with size n . Then under the same assumptions as in Theorem 1, when $U_{n,h}(\mathbb{X}_n)$ is non-degenerate under (H_0) , then*

$$\sup_{z \in \mathbb{R}} |\mathbb{P}(\sqrt{n}U_{n,h}(\mathbb{X}_n^*) \leq z | \mathbb{X}_n) - \mathbb{P}(\sqrt{n}U_{n,h}(\mathbb{X}_n^\perp) \leq z)| \xrightarrow{n \rightarrow +\infty} 0 \quad P\text{-a.s. in } (X_n)_n. \quad (16)$$

If moreover, for every $\eta \in (0, 1)$, $q_{n,h,\eta}^*(\mathbb{X}_n)$ and $q_{n,h,\eta}^\perp$ respectively denote the conditional η -quantile of $\sqrt{n}U_{n,h}(\mathbb{X}_n^*)$ given \mathbb{X}_n and the η -quantile of $\sqrt{n}U_{n,h}(\mathbb{X}_n^\perp)$, then

$$q_{n,h,\eta}^*(\mathbb{X}_n) \xrightarrow{n \rightarrow +\infty} \Phi_{0, \sigma_{h,P^1 \otimes P^2}^2}^{-1}(\eta) \quad \text{and} \quad |q_{n,h,\eta}^*(\mathbb{X}_n) - q_{n,h,\eta}^\perp| \xrightarrow{n \rightarrow +\infty} 0 \quad P\text{-a.s. in } (X_n)_n, \quad (17)$$

where $\sigma_{h,P^1 \otimes P^2}^2$ is defined as in (15).

3.5 Asymptotic properties of the bootstrap tests

Starting from Corollary 1 and with the same notation, for any fixed α in $(0, 1)$, we introduce the following tests:

$$\begin{cases} \Delta_{h,\alpha}^{*+}(\mathbb{X}_n) &= \mathbb{1}_{\{\sqrt{n}U_{n,h}(\mathbb{X}_n) > q_{n,h,1-\alpha}^*(\mathbb{X}_n)\}}, \\ \Delta_{h,\alpha}^{*-}(\mathbb{X}_n) &= \mathbb{1}_{\{\sqrt{n}U_{n,h}(\mathbb{X}_n) < q_{n,h,\alpha}^*(\mathbb{X}_n)\}}, \\ \Delta_{h,\alpha}^*(\mathbb{X}_n) &= \Delta_{h,\alpha/2}^{*+}(\mathbb{X}_n) + \Delta_{h,\alpha/2}^{*-}(\mathbb{X}_n). \end{cases} \quad (18)$$

Note that $q_{n,h,\eta}^*(\mathbb{X}_n)$ is random, depending on \mathbb{X}_n and that it may be exactly computed by considering all possible n^{2n} bootstrap samples. The algorithmic complexity of such an exact computation is usually so large, that a Monte Carlo approximation of the bootstrap quantiles, based on resampling from the original data \mathbb{X}_n , is preferred in practice. This Monte Carlo step is considered in the next section.

From Corollary 1, we deduce that, under non very restrictive assumptions, $\Delta_{h,\alpha}^{*+}$, $\Delta_{h,\alpha}^{*-}$ and $\Delta_{h,\alpha}^*$ are asymptotically of size α , and also consistent against particular alternatives.

Theorem 2. *Let $\alpha \in (0, 1)$. Let $(X_n)_n$ be a sequence of i.i.d. pairs of point processes with distribution P on \mathcal{X}^2 and with marginals P^1 and P^2 on \mathcal{X} . For every $n \geq 2$, let $\mathbb{X}_n = (X_1, \dots, X_n)$ and let Δ^* be one of the three tests defined in (18). Assume that (\mathcal{A}_{Cent}) , (\mathcal{A}_{Cent}^*) , (\mathcal{A}_{Mnt}^*) and (\mathcal{A}_{Cont}) are satisfied and that $U_{n,h}(\mathbb{X}_n)$ is non-degenerate under (H_0) . Then,*

- Δ^* is asymptotically of size α , that is: if $P = P^1 \otimes P^2$, then $\mathbb{P}(\Delta^*(\mathbb{X}_n) = 1) \xrightarrow{n \rightarrow +\infty} \alpha$.
- If $\Delta^* = \Delta_{h,\alpha}^{*+}$, then Δ^* is consistent against any alternative P such that $\int h(x, x') dP(x) dP(x') > 0$, that is: for such P , $\mathbb{P}(\Delta^*(\mathbb{X}_n) = 1) \xrightarrow{n \rightarrow +\infty} 1$.
- If $\Delta^* = \Delta_{h,\alpha}^{*-}$, then Δ^* is consistent against any alternative P such that $\int h(x, x') dP(x) dP(x') < 0$.
- If $\Delta^* = \Delta_{h,\alpha}^*$, then Δ^* is consistent against any alternative P such that $\int h(x, x') dP(x) dP(x') \neq 0$.

Notice that in the case where $h = h_\varphi$ with an integrable function φ (see (4)),

$$\int h(x, x') dP(x) dP(x') = \int \varphi(x^1, x^2) [dP(x^1, x^2) - dP^1(x^1) dP^2(x^2)].$$

This means that the bilateral test $\Delta_{h,\alpha}^*$ is consistent for any alternative such that $\int \varphi(x^1, x^2) dP(x^1, x^2)$ is different from what is expected under (H_0) , i.e. $\int \varphi(x^1, x^2) dP^1(x^1) dP^2(x^2)$. When $\varphi = \varphi_\delta^w$ defined by (2), this amounts to state that

$$\beta_{w,\delta} = \int w_\delta(u, v) (\mathbb{E}[dN_{X^1}(u) dN_{X^2}(v)] - \mathbb{E}[dN_{X^1}(u)] \mathbb{E}[dN_{X^2}(v)]) \neq 0.$$

Under the specific Poisson assumptions of Sansonnet and Tuleau-Malot [53], if $w_\delta(u, v) = \psi_{j,k}(v - u)$ where $\delta = (j, k)$ and $\psi_{j,k}$ is a rescaled and translated Haar mother wavelet, $\beta_{w,\delta}$ is linked to the coefficient in the Haar basis of the so-called interaction function, which measures the dependence between both processes X^1 and X^2 . Working non asymptotically, one of the main result of [53] states, after reformulation in the present setting, that if $\beta_{w,\delta}$ is larger than an explicit lower bound, then the Type II error is less than a prescribed $\beta \in (0, 1)$. Theorem 3.7 thus generalizes their result to a set-up with much less reductive assumptions on the underlying stochastic models, but in an asymptotic way.

3.6 Bootstrap tests with Monte Carlo approximation

As seen above the tests defined in (18) have satisfactory theoretical properties, but involve an exact computation of the conditional quantiles $q_{n,h,\eta}^*(\mathbb{X}_n)$. Though such a computation is possible, it is not often reasonable in practice even when the sample size n is moderate, since computing $U_{n,h}(\mathbb{X}_n^{\pi_n})$ itself may be complex from an algorithmic point of view, from some particular choices for h . Therefore, we choose to approximate the conditional quantiles $q_{n,\varphi,\eta}^*(\mathbb{X}_n)$, as usual, by a Monte Carlo method.

We prove that, even if such a Monte Carlo method is used, the resulting tests have the same asymptotic properties as the ones defined in (18), as the number of bootstrap samples grows to infinity.

Let $B \geq 1$ be a chosen number of iterations for the Monte Carlo method, and $(\mathbb{X}_n^{*1}, \dots, \mathbb{X}_n^{*B})$ be B independent bootstrap samples from \mathbb{X}_n , that is B i.i.d. random variables from the distribution $P_n^1 \otimes P_n^2$.

For each \mathbf{b} in $\{1, \dots, B\}$, we introduce $U^{*\mathbf{b}} = U_{n,h}(\mathbb{X}_n^{*\mathbf{b}})$. The order statistic associated with (U^{*1}, \dots, U^{*B}) is now denoted as usual by $(U^{*(1)}, \dots, U^{*(B)})$.

Given α in $(0, 1)$, we can now introduce the Monte Carlo-based bootstrap tests defined by:

$$\begin{cases} \Delta_{B,h,\alpha}^{*+}(\mathbb{X}_n) = \mathbb{1}_{\{U_{n,h}(\mathbb{X}_n) > U^{*(\lceil (1-\alpha)B \rceil)}\}}, \\ \Delta_{B,h,\alpha}^{*-}(\mathbb{X}_n) = \mathbb{1}_{\{U_{n,h}(\mathbb{X}_n) < U^{*(\lfloor \alpha B \rfloor + 1)}\}}, \\ \Delta_{B,h,\alpha}^*(\mathbb{X}_n) = \Delta_{B,h,\alpha/2}^{*+}(\mathbb{X}_n) + \Delta_{B,h,\alpha/2}^{*-}(\mathbb{X}_n). \end{cases} \quad (19)$$

Proposition 5. *Let $\alpha \in (0, 1)$, and let $(X_n)_n$ be a sequence of i.i.d. pairs of point processes with distribution P on \mathcal{X}^2 , with marginals P^1 and P^2 on \mathcal{X} . For every $n \geq 2$, let $\mathbb{X}_n = (X_1, \dots, X_n)$, and for $B \geq 1$, let Δ_B^* be one of the tests defined in (19). Under the same assumptions as in Theorem 2, if $B_n \xrightarrow{n \rightarrow \infty} +\infty$, then*

- $\Delta_{B_n}^*$ is asymptotically of size α , that is: if $P = P^1 \otimes P^2$, then $\mathbb{P}(\Delta_{B_n}^*(\mathbb{X}_n) = 1) \xrightarrow{n \rightarrow +\infty} \alpha$.
- If $\Delta_{B_n}^* = \Delta_{B_n,h,\alpha}^{*+}$, then $\Delta_{B_n}^*$ is consistent against any alternative P such that $\int h(x, x') dP(x) dP(x') > 0$, that is: for such P , $\mathbb{P}(\Delta_{B_n}^*(\mathbb{X}_n) = 1) \xrightarrow{n \rightarrow +\infty} 1$.
- If $\Delta_{B_n}^* = \Delta_{B_n,h,\alpha}^{*-}$, then $\Delta_{B_n}^*$ is consistent against any alternative P such that $\int h(x, x') dP(x) dP(x') < 0$.
- If $\Delta_{B_n}^* = \Delta_{B_n,h,\alpha}^*$, then $\Delta_{B_n}^*$ is consistent against any alternative P such that $\int h(x, x') dP(x) dP(x') \neq 0$.

3.7 Back to spike trains analysis in neuroscience

Let us focus on the case where h is of the form $h_{\varphi_\delta^{\text{coinc}}}$, defined by (4) and (1), which originally motivated the study. The assumptions of Theorem 3.7 are at least fulfilled in those three generic cases that are of major importance in neuroscience:

1. The processes X^1 and X^2 are discretized at resolution r and satisfy $P^1(\{\emptyset\})P^2(\{\emptyset\}) > 0$, like in particular Bernoulli processes. In this case, a natural choice for δ is $\delta = dr + r/2$ for some integer d .
2. The processes X^1 and X^2 have bounded conditional intensities, like in particular Poisson processes.
3. The processes X^1 and X^2 have conditional intensities bounded by the intensity of linear stationary Hawkes processes with positive interaction, like general Hawkes processes used in particular to model inhibition (see [28] e.g.).

In any case, δ should be chosen large enough to ensure that $\varphi_\delta^{\text{coinc}}(X^1, X^2)$ is not a.s. null, but it is always possible to choose such a δ a posteriori, by looking at the coincidence count on the observed trials.

Theorem means here that $\Delta_{h_{\varphi_\delta^{\text{coinc}}, \alpha}}^*$ is asymptotically of power 1 for any alternative P such that

$$\int \mathbb{1}_{\{|v-u| \leq \delta\}} \mathbb{E}[dN_{X^1}(u) dN_{X^2}(v)] \neq \int \mathbb{1}_{\{|v-u| \leq \delta\}} \mathbb{E}[dN_{X^1}(u)] \mathbb{E}[dN_{X^2}(v)].$$

Notice that one cannot find such a δ if heuristically, the repartition of the delays $|v - u|$ between points of X^1 and X^2 is the same under (H_0) and under (H_1) . This also means for neuroscientists, that the cross-correlogram¹ shows no different behavior between the dependent and independent case and this even if one has access to an infinite number of trials. Hence this heuristically means one cannot find such a δ if the dependence cannot be measured in terms of delay between points. Though this is quite not likely to happen for classical spike trains, the question of the choice of δ remains an open question for model free independence test.

¹histogram of the delays, that is classically represented in neuroscience as the first description of the data

4 Permutation tests of independence

4.1 Description of the permutation approach and overview of the results

We still observe an i.i.d sample $\mathbb{X}_n = (X_1, \dots, X_n)$ with $X_i = (X_i^1, X_i^2)$ of distribution P on \mathcal{X}^2 , whose marginals are respectively denoted by P^1 and P^2 .

We consider here the U -statistic $U_{n,h}(\mathbb{X}_n)$ defined by (5), in the particular case where h equals h_φ for some integrable function φ , as defined in (4). The permutation approach we propose to use consists of randomly permuting the second coordinates of the observed pairs of point processes. More precisely, if Π_n denotes a random permutation of $\{1, \dots, n\}$, the corresponding permuted sample is defined by

$$\mathbb{X}_n^{\Pi_n} = \left(X_1^{\Pi_n}, \dots, X_n^{\Pi_n} \right) \text{ with } X_i^{\Pi_n} = \left(X_i^1, X_{\Pi_n(i)}^2 \right), \quad (20)$$

and we denote by P_n^* the conditional distribution of $\mathbb{X}_n^{\Pi_n}$ given \mathbb{X}_n .

Like for the bootstrap, the idea of the randomization or permutation principle is to mimic the distribution of the test statistics, assuming that (H_0) is satisfied.

We actually prove (see Theorem 3 and Theorem 4) that under appropriate assumptions, the conditional distribution of $\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n})$ given \mathbb{X}_n is asymptotically close to the distribution of $\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^\perp)$, where \mathbb{X}_n^\perp is an i.i.d. sample from the distribution $P^1 \otimes P^2$. Thus, we can propose new permutation tests of independence of the desired asymptotic size, using the conditional quantiles of $\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n})$ given \mathbb{X}_n as critical values, and prove that these tests are consistent against any reasonable alternative. Following the statement of our bootstrap results in Section 3, we still express the closeness in distributions between the permuted and original statistics in terms of Wasserstein's metric, which distinguishes, to our knowledge, our results from previous ones in the permutation tests scene.

At this stage, since the permutation independence tests satisfy the same asymptotic properties as the bootstrap ones, but with much more computation difficulties to prove them, one might wonder whether the introduction of the permutation tests is of genuine interest. Though the bootstrap approaches are known to perform well when moderate or even small sample sizes are considered in practice, there are very few non-asymptotic theoretical results giving evidence of it, since such results are based on concentration inequalities that are not always accessible in complex models (see [2, 3, 18] in testing frameworks, [17] in a classification framework).

One of the main advantages of the permutation approach we use here lies in the resulting tests being exactly of the desired level, even when a Monte Carlo method is used to approximate the critical values. Such non-asymptotic results, stated in Proposition 7 and Proposition 8, can be proved without any sophisticated concentration inequality-type tool, only thanks to [51, Lemma 1] and the following Proposition, at the heart of the permutation principle.

Proposition 6. *Let π_n be a given deterministic permutation of $\{1, \dots, n\}$, and Π_n be a uniformly distributed random permutation of $\{1, \dots, n\}$, independent of the observed sample $\mathbb{X}_n = (X_1, \dots, X_n)$. If $P = P^1 \otimes P^2$, that is under (H_0) , then the permuted samples $\mathbb{X}_n^{\pi_n}$ and $\mathbb{X}_n^{\Pi_n}$ defined in (20) both have the same distribution as the original sample \mathbb{X}_n .*

4.2 Consistency of the permutation approach

Since we only consider in this section U -statistics based on symmetric kernels of the form h_φ , as defined in (4), we notice that the centring assumption (\mathcal{A}_{Cent}) is satisfied. Therefore, we will only need here the following moment assumption on φ :

$$(\mathcal{A}_{\varphi, Mnt}) \quad \left| \begin{array}{l} \text{For } X \text{ with distribution } P \text{ or } P^1 \otimes P^2 \text{ on } \mathcal{X}^2, \\ \mathbb{E} [\varphi^4(X^1, X^2)] < \infty. \end{array} \right.$$

Theorem 3. *Let $(X_n)_n$ be a sequence of i.i.d. pairs of point processes with distribution P on \mathcal{X}^2 , with marginals P^1 and P^2 . We consider an integrable function φ satisfying $(\mathcal{A}_{\varphi, Mnt})$, h_φ and U_{n,h_φ} defined by (4) and (5). For every $n \geq 2$, let $\mathbb{X}_n = (X_1, \dots, X_n)$, let Π_n be a uniformly distributed random permutation independent of $(X_n)_n$, and $\mathbb{X}_n^{\Pi_n}$ be the corresponding permuted sample defined by (20). In the notation of Section 3,*

$$d_2 \left(\mathcal{L} \left(\sqrt{n}U_{n,h_\varphi}, P_n^* | \mathbb{X}_n \right), \mathcal{N} \left(0, \sigma_{h_\varphi, P^1 \otimes P^2}^2 \right) \right) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0, \quad (21)$$

where $\xrightarrow{\mathbb{P}}$ stands for the usual convergence in \mathbb{P} -probability.

Comment. Notice that we do not need the continuity assumption for the kernel h_φ that we had in the bootstrap approach, but the existence of a fourth order moment for this kernel, at the price however that the convergence for Wasserstein's metric occurs in probability and not almost surely as for the bootstrap.

We also wish to emphasize that the above result, whose proof is based on asymptotic normality results for martingale difference arrays, goes further than the ones of Romano [50] for instance, since we have a convergence result which is satisfied whether the independence hypothesis is satisfied or not.

From Theorem 3, we deduce the following Corollary, which will be, combined with Theorem 4, a key point to prove that we can use the conditional quantiles of $\mathcal{L}(\sqrt{n}U_{n,h_\varphi}, P_n^* | \mathbb{X}_n)$ as critical values to obtain permutation tests of independence of the desired asymptotic size.

Corollary 2. *Let $(X_n)_n$ be a sequence of i.i.d. pairs of point processes with distribution P on \mathcal{X}^2 , with marginals P^1 and P^2 . For every $n \geq 2$, let $\mathbb{X}_n = (X_1, \dots, X_n)$, and Π_n be a uniformly distributed random permutation independent of the $(X_n)_n$. For η in $(0, 1)$, let $q_{n,\varphi,\eta}^*(\mathbb{X}_n)$ denote the η -quantile of $\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n})$ given \mathbb{X}_n that is $\mathcal{L}(\sqrt{n}U_{n,h_\varphi}, P_n^* | \mathbb{X}_n)$. Then, under the same assumptions as in Theorem 3,*

$$q_{n,\varphi,\eta}^*(\mathbb{X}_n) \xrightarrow[n \rightarrow +\infty]{P} \Phi_{0,\sigma_{h,P^1 \otimes P^2}}^{-1}(\eta). \quad (22)$$

4.3 Asymptotic properties of the permutation tests

In the notation of Corollary 2, for any fixed α in $(0, 1)$, we introduce the following tests:

$$\begin{cases} \Delta_{\varphi,\alpha}^{*+}(\mathbb{X}_n) &= \mathbb{1}_{\{\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n) > q_{n,\varphi,1-\alpha}^*(\mathbb{X}_n)\}}, \\ \Delta_{\varphi,\alpha}^{*-}(\mathbb{X}_n) &= \mathbb{1}_{\{\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n) < q_{n,\varphi,\alpha}^*(\mathbb{X}_n)\}}, \\ \Delta_{\varphi,\alpha}^*(\mathbb{X}_n) &= \Delta_{\varphi,\alpha/2}^{*+}(\mathbb{X}_n) + \Delta_{\varphi,\alpha/2}^{*-}(\mathbb{X}_n). \end{cases} \quad (23)$$

Notice that like for the bootstrap approach, the critical values used in the tests defined by (23) are random, depending on the observed sample \mathbb{X}_n .

Recall that Π_n is a uniformly distributed random permutation on the set \mathfrak{S}_n of all the permutations of $\{1, \dots, n\}$ and independent of \mathbb{X}_n . Hence, given \mathbb{X}_n , the conditional distribution of $\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n})$ is discrete and takes the values $\{\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\pi_n})\}_{\pi_n \in \mathfrak{S}_n}$. Let $\sqrt{n}U_{n,h_\varphi}^{(1)}(\mathbb{X}_n) \leq \dots \leq \sqrt{n}U_{n,h_\varphi}^{(n!)}(\mathbb{X}_n)$ be these ordered values. Then, $q_{n,\varphi,\eta}^*(\mathbb{X}_n) = \sqrt{n}U_{n,h_\varphi}^{(\lceil n! \eta \rceil)}(\mathbb{X}_n)$, and it is therefore possible to compute it exactly.

Theorem 4. *Let $\alpha \in (0, 1)$. Let $(X_n)_n$ be a sequence of i.i.d. pairs of point processes with distribution P on \mathcal{X}^2 and with marginals P^1 and P^2 on \mathcal{X} . For every $n \geq 2$, let $\mathbb{X}_n = (X_1, \dots, X_n)$, Δ^* be one of the permutation tests defined by (23), and assume that $(A_{\varphi,Mmt})$ is satisfied. Then,*

- Δ^* is asymptotically of size α , that is: if $P = P^1 \otimes P^2$, then $\mathbb{P}(\Delta^*(\mathbb{X}_n) = 1) \xrightarrow[n \rightarrow +\infty]{} \alpha$.
- If $\Delta^* = \Delta_{\varphi,\alpha}^{*+}$, then Δ^* is consistent against any alternative P such that $\int h_\varphi(x, x') dP(x) dP(x') > 0$, or equivalently $\int \varphi(x^1, x^2) (dP(x^1, x^2) - dP^1(x^1) dP^2(x^2)) > 0$ that is: for such P , $\mathbb{P}(\Delta^*(\mathbb{X}_n) = 1) \xrightarrow[n \rightarrow +\infty]{} 1$.
- If $\Delta^* = \Delta_{\varphi,\alpha}^{*-}$, then Δ^* is consistent against any alternative P such that $\int h_\varphi(x, x') dP(x) dP(x') < 0$.
- If $\Delta^* = \Delta_{\varphi,\alpha}^*$, then Δ^* is consistent against any alternative P such that $\int h_\varphi(x, x') dP(x) dP(x') \neq 0$.

4.4 Non-asymptotic properties of the permutation tests

As seen above, one of the advantages of the permutation approach, as compared with the bootstrap one, lies on the result of Proposition 6, which allows in fact to prove that the tests also satisfy non-asymptotic properties.

Considering for instance $\Delta^* = \Delta_{\varphi, \alpha}^{*+}$, then under (H_0) ,

$$\begin{aligned} \mathbb{P}(\Delta^*(\mathbb{X}_n) = 1) &= \mathbb{P}(\sqrt{n}U_{n, h_\varphi}(\mathbb{X}_n) > q_{n, \varphi, 1-\alpha}^*(\mathbb{X}_n)) \\ &= \frac{1}{n!} \sum_{\pi_n \in \mathfrak{S}_n} \mathbb{P}(\sqrt{n}U_{n, h_\varphi}(\mathbb{X}_n^{\pi_n}) > q_{n, \varphi, 1-\alpha}^*(\mathbb{X}_n^{\pi_n})) \end{aligned} \quad (24)$$

$$= \frac{1}{n!} \sum_{\pi_n \in \mathfrak{S}_n} \mathbb{P}(\sqrt{n}U_{n, h_\varphi}(\mathbb{X}_n^{\pi_n}) > q_{n, \varphi, 1-\alpha}^*(\mathbb{X}_n)), \quad (25)$$

where (24) comes from Proposition 6, and the equality $q_{n, \varphi, 1-\alpha}^*(\mathbb{X}_n^{\pi_n}) = q_{n, \varphi, 1-\alpha}^*(\mathbb{X}_n)$ in (25) is explained from the fact that the conditional distribution of $\sqrt{n}U_{n, h_\varphi}(\mathbb{X}_n^{\Pi_n \circ \pi_n})$ given \mathbb{X}_n is the same as the one of $\sqrt{n}U_{n, h_\varphi}(\mathbb{X}_n^{\Pi_n})$. Therefore,

$$\begin{aligned} \mathbb{P}(\Delta^*(\mathbb{X}_n) = 1) &= \sum_{\pi_n \in \mathfrak{S}_n} \mathbb{P}(\sqrt{n}U_{n, h_\varphi}(\mathbb{X}_n^{\Pi_n}) > q_{n, \varphi, 1-\alpha}^*(\mathbb{X}_n) | \Pi_n = \pi_n) \mathbb{P}(\Pi_n = \pi_n) \\ &= \mathbb{P}(\sqrt{n}U_{n, h_\varphi}(\mathbb{X}_n^{\Pi_n}) > q_{n, \varphi, 1-\alpha}^*(\mathbb{X}_n)) \\ &= \mathbb{E}[\mathbb{P}(\sqrt{n}U_{n, h_\varphi}(\mathbb{X}_n^{\Pi_n}) > q_{n, \varphi, 1-\alpha}^*(\mathbb{X}_n) | \mathbb{X}_n)] \\ &\leq \alpha. \end{aligned}$$

These arguments may be adapted to every permutation test defined in (23), leading to the following Proposition.

Proposition 7. *Let $\alpha \in (0, 1)$. Let $(X_n)_n$ be a sequence of i.i.d. pairs of point processes with distribution P on \mathcal{X}^2 and with marginals P^1 and P^2 on \mathcal{X} . For every $n \geq 2$, let $\mathbb{X}_n = (X_1, \dots, X_n)$, and Δ^* be one of the permutation tests defined by (23). Then Δ^* is exactly of level α , that is: if $P = P^1 \otimes P^2$, $\mathbb{P}(\Delta^*(\mathbb{X}_n) = 1) \leq \alpha$.*

4.5 Permutation tests with Monte Carlo approximation

The tests defined in (23) involve an exact computation of the conditional quantiles $q_{n, \varphi, \eta}^*(\mathbb{X}_n)$. Such a computation is possible by sorting the $n!$ values of $\{U_{n, h_\varphi}(\mathbb{X}_n^{\pi_n})\}_{\pi_n \in \mathfrak{S}_n}$, but, as for the bootstrap approach, it has not often a reasonable algorithmic complexity. Therefore, we also choose to approximate the conditional quantiles $q_{n, \varphi, \eta}^*(\mathbb{X}_n)$ by a Monte Carlo method, and we prove that the resulting tests have the same asymptotic properties as the ones defined in (23).

Moreover, it is interesting to point out that, even if they are based on a Monte Carlo approximation, these tests are still exactly of the desired level. This is proved using Proposition 6 and [51, Lemma 1] again.

Let $B \geq 1$ be a chosen number of iterations for the Monte Carlo method, and $(\Pi_n^1, \dots, \Pi_n^B)$ be a sample of i.i.d. random permutations uniformly distributed on \mathfrak{S}_n . For each \mathbf{b} in $\{1, \dots, B\}$, we introduce $U^{*\mathbf{b}} = U_{n, h_\varphi}(\mathbb{X}_n^{\Pi_n^{\mathbf{b}}})$, and we denote by $U^{*(B+1)}$ the statistic $U_{n, h_\varphi}(\mathbb{X}_n)$ computed on the original sample \mathbb{X}_n . The order statistic associated with $(U^{*1}, \dots, U^{*(B+1)})$ is now denoted as usual by $(U^{*(1)}, \dots, U^{*((B+1))})$.

Given α in $(0, 1)$, we can now introduce the Monte Carlo-based permutation tests defined by:

$$\begin{cases} \Delta_{B, \varphi, \alpha}^{*+}(\mathbb{X}_n) = \mathbb{1}_{\{U_{n, h_\varphi}(\mathbb{X}_n) > U^{*((1-\alpha)(B+1))}\}}, \\ \Delta_{B, \varphi, \alpha}^{*-}(\mathbb{X}_n) = \mathbb{1}_{\{U_{n, h_\varphi}(\mathbb{X}_n) < U^{*(\lfloor \alpha(B+1) \rfloor + 1)}\}}, \\ \Delta_{B, \varphi, \alpha}^*(\mathbb{X}_n) = \Delta_{B, \varphi, \alpha/2}^{*+}(\mathbb{X}_n) + \Delta_{B, \varphi, \alpha/2}^{*-}(\mathbb{X}_n). \end{cases} \quad (26)$$

Proposition 8. *Let $\alpha \in (0, 1)$, and let $(X_n)_n$ be a sequence of i.i.d. pairs of point processes with distribution P on \mathcal{X}^2 , with marginals P^1 and P^2 on \mathcal{X} . For every $n \geq 2$, let $\mathbb{X}_n = (X_1, \dots, X_n)$, and for $B \geq 1$, let $\Delta_{B, \varphi}^*$ be one of the tests defined in (26). If $B_n \xrightarrow[n \rightarrow \infty]{} +\infty$, if U_{n, h_φ} is non degenerate, and if $(\mathcal{A}_{\varphi, Mnt})$ is satisfied, then*

- $\Delta_{B_n, \varphi}^*$ is asymptotically of size α , that is: if $P = P^1 \otimes P^2$, then $\mathbb{P}(\Delta_{B_n, \varphi}^*(\mathbb{X}_n) = 1) \xrightarrow[n \rightarrow +\infty]{} \alpha$.
- If $\Delta_{B_n, \varphi}^* = \Delta_{B_n, \varphi, \alpha}^{*+}$, then $\Delta_{B_n, \varphi}^*$ is consistent against any alternative P such that $\int h_\varphi(x, x') dP(x) dP(x') > 0$, that is: for such P , $\mathbb{P}(\Delta_{B_n, \varphi}^*(\mathbb{X}_n) = 1) \xrightarrow[n \rightarrow +\infty]{} 1$.
- If $\Delta_{B_n, \varphi}^* = \Delta_{B_n, \varphi, \alpha}^{*-}$, then $\Delta_{B_n, \varphi}^*$ is consistent against any alternative P such that $\int h_\varphi(x, x') dP(x) dP(x') < 0$.

- If $\Delta_{B_n}^* = \Delta_{B_n, \varphi, \alpha}^*$, then $\Delta_{B_n}^*$ is consistent against any alternative P such that $\int h_\varphi(x, x') dP(x) dP(x') \neq 0$.

Proposition 9. Let $\alpha \in (0, 1)$, and let $(X_n)_n$ be a sequence of i.i.d. pairs of point processes with distribution P on \mathcal{X}^2 , with marginals P^1 and P^2 on \mathcal{X} . For every $n \geq 2$, let $\mathbb{X}_n = (X_1, \dots, X_n)$, and Δ_B^* be one of the tests defined in (26), for some $B \geq 1$. Then Δ_B^* is exactly of level α , that is: if $P = P^1 \otimes P^2$, $\mathbb{P}(\Delta_B^*(\mathbb{X}_n) = 1) \leq \alpha$.

5 Proofs

All along the proofs in this section, C and C' will denote universal positive constants, that may vary from one line to another one.

5.1 Proof of Proposition 2

For some integer $n \geq 2$, let $(Y_{n,i}^*, Y_i)_{1 \leq i \leq n}$ be an i.i.d. sample such that for every $i = 1 \dots n$, $Y_{n,i}^* \sim P_n^1 \otimes P_n^2$, and $Y_i \sim P^1 \otimes P^2$, and so that in particular:

$$\frac{\sqrt{n}}{n(n-1)} \sum_{i \neq i'} h(Y_{n,i}^*, Y_{n,i'}^*) \sim \mathcal{L}(\sqrt{n}U_n, P_n^1 \otimes P_n^2 | \mathbb{X}_n),$$

and

$$\frac{\sqrt{n}}{n(n-1)} \sum_{i \neq i'} h(Y_i, Y_{i'}) \sim \mathcal{L}(\sqrt{n}U_n, P^1 \otimes P^2).$$

From the definition of Wasserstein's metric d_2 , recalled in (12), we then deduce that:

$$d_2^2(\mathcal{L}(\sqrt{n}U_n, P_n^1 \otimes P_n^2 | \mathbb{X}_n), \mathcal{L}(\sqrt{n}U_n, P^1 \otimes P^2)) \leq \frac{1}{n(n-1)^2} \mathbb{E}^* \left[\left(\sum_{i \neq i'} (h(Y_{n,i}^*, Y_{n,i'}^*) - h(Y_i, Y_{i'})) \right)^2 \right], \quad (27)$$

where the upper bound is finite under (\mathcal{A}_{Mnt}^*) .

Introducing the quantities

$$\mathbb{E}_{(i,i',j,j')} = \mathbb{E}^* \left[\left(h(Y_{n,i}^*, Y_{n,i'}^*) - h(Y_i, Y_{i'}) \right) \left(h(Y_{n,j}^*, Y_{n,j'}^*) - h(Y_j, Y_{j'}) \right) \right],$$

for (i, i', j, j') in $\{1, 2, \dots, n\}^4$, and the sets

$$I_m = \left\{ (i, i', j, j') \in \{1, 2, \dots, n\}^4 \mid i \neq i', j \neq j', \# \{i, i', j, j'\} = m \right\},$$

for m in $\{2, 3, 4\}$, where $\# \{i, i', j, j'\}$ denotes the number of different elements in $\{i, i', j, j'\}$, we easily see that:

$$\mathbb{E}^* \left[\left(\sum_{i \neq i'} h(Y_{n,i}^*, Y_{n,i'}^*) - \sum_{i \neq i'} h(Y_i, Y_{i'}) \right)^2 \right] = \sum_{(i,i',j,j') \in I_4} \mathbb{E}_{(i,i',j,j')} + \sum_{(i,i',j,j') \in I_3} \mathbb{E}_{(i,i',j,j')} + \sum_{(i,i',j,j') \in I_2} \mathbb{E}_{(i,i',j,j')}. \quad (28)$$

Let us now upper bound each term of this sum separately.

1. If $(i, i', j, j') \in I_4$, then by independence,

$$\begin{aligned} \mathbb{E}_{(i,i',j,j')} &= \mathbb{E}^*[h(Y_{n,i}^*, Y_{n,i'}^*) - h(Y_i, Y_{i'})] \mathbb{E}^*[h(Y_{n,j}^*, Y_{n,j'}^*) - h(Y_j, Y_{j'})] \\ &= (\mathbb{E}^*[h(Y_{n,i}^*, Y_{n,i'}^*)] - \mathbb{E}[h(Y_i, Y_{i'})]) (\mathbb{E}^*[h(Y_{n,j}^*, Y_{n,j'}^*)] - \mathbb{E}[h(Y_j, Y_{j'})]). \end{aligned}$$

Yet, under assumptions (\mathcal{A}_{Cent}) and (\mathcal{A}_{Cent}^*) , $\mathbb{E}[h(Y_i, Y_{i'})] = \mathbb{E}^*[h(Y_{n,i}^*, Y_{n,i'}^*)] = 0$, so $\mathbb{E}_{(i,i',j,j')} = 0$.

2. If $(i, i', j, j') \in I_3$, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbb{E}_{(i, i', j, j')} &\leq \sqrt{\mathbb{E}^* \left[\left(h(Y_{n,i}^*, Y_{n,i'}^*) - h(Y_i, Y_{i'}) \right)^2 \right]} \sqrt{\mathbb{E}^* \left[\left(h(Y_{n,j}^*, Y_{n,j'}^*) - h(Y_j, Y_{j'}) \right)^2 \right]} \\ &= \mathbb{E}^* \left[\left(h(Y_{n,1}^*, Y_{n,2}^*) - h(Y_1, Y_2) \right)^2 \right]. \end{aligned}$$

3. If $(i, i', j, j') \in I_2$, we immediately obtain that

$$\mathbb{E}_{(i, i', j, j')} = \mathbb{E}^* \left[\left(h(Y_{n,1}^*, Y_{n,2}^*) - h(Y_1, Y_2) \right)^2 \right].$$

Since $\#I_3 = 4n(n-1)(n-2)$ and $\#I_2 = 2n(n-1)$, we thus obtain that

$$\sum_{\substack{i \neq i' \\ j \neq j'}} \mathbb{E}_{(i, i', j, j')} \leq 4n(n-1)^2 \mathbb{E}^* \left[\left(h(Y_{n,1}^*, Y_{n,2}^*) - h(Y_1, Y_2) \right)^2 \right]. \quad (29)$$

From (27), (28), (29), we finally derive that

$$d_2^2(\mathcal{L}(\sqrt{n}U_n, P_n^1 \otimes P_n^2 | \mathbb{X}_n), \mathcal{L}(\sqrt{n}U_n, P^1 \otimes P^2)) \leq 4 \mathbb{E}^* \left[\left(h(Y_{n,1}^*, Y_{n,2}^*) - h(Y_1, Y_2) \right)^2 \right],$$

where the i.i.d sample $(Y_{n,i}^*, Y_i)_{1 \leq 2}$ may be arbitrarily chosen such that for every i in $\{1, 2\}$, $Y_{n,i}^* \sim P_n^1 \otimes P_n^2$, $Y_i \sim P^1 \otimes P^2$. This ends the proof of Proposition 2.

5.2 Proof of Proposition 3

Let us first notice that (13) is a direct application of the strong law of large numbers for U -statistics, stated for instance in [32].

Next, let us notice that if

$$g_m(X_{i_1}, \dots, X_{i_m}) = \sum_{(i,j,k,l) \in I_{\{i_1, \dots, i_m\}}} h^2((X_i^1, X_j^2), (X_k^1, X_l^2)),$$

with $I_{\{i_1, \dots, i_m\}} = \left\{ (i, j, k, l) \in \{i_1, \dots, i_m\}^4 / \# \{i, j, k, l\} = m \right\}$, and m in $\{1, \dots, 4\}$, then

$$\frac{1}{n^4} \sum_{i,j,k,l=1}^n h^2((X_i^1, X_j^2), (X_k^1, X_l^2)) = \sum_{m=1}^4 \frac{1}{m!} \left(\frac{1}{n^4} \sum_{\substack{(i_1, \dots, i_m) \in \{1, \dots, n\}^m \\ i_1, \dots, i_m \text{ all different}}} g_m(X_{i_1}, \dots, X_{i_m}) \right).$$

Each of the four terms in the right hand side of the above decomposition being, up to a multiplicative factor, some classical U -statistic, and since under $(\mathcal{A}_{Mmt}^*, \mathbb{E}[|g_m(X_{i_1}, \dots, X_{i_m})|] < +\infty)$, we can now apply the strong law of large numbers for U -statistics proved by Hoeffding in [32] for instance. Therefore P -a.s. in $(X_n)_n$,

$$\frac{1}{n(n-1) \dots (n-m+1)} \sum_{\substack{(i_1, \dots, i_m) \in \{1, \dots, n\}^m \\ i_1, \dots, i_m \text{ all different}}} g_m(X_{i_1}, \dots, X_{i_m}) \xrightarrow{n \rightarrow +\infty} \mathbb{E}[g_m(X_1, \dots, X_m)].$$

In particular, for $m \in \{1, 2, 3\}$, P -a.s. in $(X_n)_n$,

$$\frac{1}{n^4} \sum_{\substack{(i_1, \dots, i_m) \in \{1, \dots, n\}^m \\ i_1, \dots, i_m \text{ all different}}} g_m(X_{i_1}, \dots, X_{i_m}) \xrightarrow{n \rightarrow +\infty} 0,$$

and

$$\frac{1}{n^4} \sum_{\substack{(i_1, \dots, i_4) \in \{1, \dots, n\}^4 \\ i_1, \dots, i_4 \text{ all different}}} g_4(X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4}) \xrightarrow{n \rightarrow +\infty} \mathbb{E}[g_4(X_1, X_2, X_3, X_4)].$$

Finally noticing that $\mathbb{E}[g_4(X_1, X_2, X_3, X_4)] = 4! \mathbb{E}[h^2((X_1^1, X_2^2), (X_3^1, X_4^2))]$, this concludes the proof.

5.3 Proof of Theorem 1

By Proposition 2, for all $n \geq 2$,

$$\begin{aligned} d_2\left(\mathcal{L}\left(\sqrt{n}U_{n,h}, P_n^1 \otimes P_n^2 | \mathbb{X}_n\right), \mathcal{L}\left(\sqrt{n}U_{n,h}, P^1 \otimes P^2\right)\right) \\ \leq C \inf_{\substack{(Y_{n,1}^*, Y_1), (Y_{n,2}^*, Y_2) \text{ i.i.d. /} \\ Y_{n,1}^*, Y_{n,2}^* \sim P_n^1 \otimes P_n^2, Y_1, Y_2 \sim P^1 \otimes P^2}} \mathbb{E}^* \left[\left(h(Y_{n,1}^*, Y_{n,2}^*) - h(Y_1, Y_2) \right)^2 \right]. \end{aligned}$$

Our goal is here to construct, for almost all ω in Ω , a sequence of random variables $(\bar{Y}_{n,\omega,1}^*)_{n \geq 1}$ such that for every $n \geq 1$, $\bar{Y}_{n,\omega,1}^* \sim P_{n,\omega}^1 \otimes P_{n,\omega}^2$, where $P_{n,\omega}^j = n^{-1} \sum_{i=1}^n \delta_{X_i^j(\omega)}$ ($j = 1, 2$) are the empirical measures corresponding to the realisation $\mathbb{X}_n(\omega)$, a random variable $\bar{Y}_{\omega,1} \sim P^1 \otimes P^2$, and $\left\{ (\bar{Y}_{n,\omega,2}^*)_{n \geq 1}, \bar{Y}_{\omega,2} \right\}$ an independent copy of $\left\{ (\bar{Y}_{n,\omega,1}^*)_{n \geq 1}, \bar{Y}_{\omega,1} \right\}$ on some probability space $(\Omega'_\omega, \mathcal{A}'_\omega, \mathbb{P}'_\omega)$ depending on ω such that

$$\mathbb{E}'_\omega \left[\left(h(\bar{Y}_{n,\omega,1}^*, \bar{Y}_{n,\omega,2}^*) - h(\bar{Y}_{\omega,1}, \bar{Y}_{\omega,2}) \right)^2 \right] \xrightarrow{n \rightarrow +\infty} 0, \quad (30)$$

where \mathbb{E}'_ω denotes the expectation corresponding to \mathbb{P}'_ω . From this, we shall directly deduce that, for almost all ω in Ω ,

$$\begin{aligned} \inf_{\substack{(Y_{n,1}^*, Y_1), (Y_{n,2}^*, Y_2) \text{ i.i.d. /} \\ Y_{n,1}^*, Y_{n,2}^* \sim P_{n,\omega}^1 \otimes P_{n,\omega}^2, Y_1, Y_2 \sim P^1 \otimes P^2}} \mathbb{E}^* \left[\left(h(Y_{n,1}^*, Y_{n,2}^*) - h(Y_1, Y_2) \right)^2 \right] (\omega) \\ \leq \mathbb{E}'_\omega \left[\left(h(\bar{Y}_{n,\omega,1}^*, \bar{Y}_{n,\omega,2}^*) - h(\bar{Y}_{\omega,1}, \bar{Y}_{\omega,2}) \right)^2 \right] \xrightarrow{n \rightarrow +\infty} 0, \end{aligned}$$

which will conclude the proof of the theorem.

Consider $(\Omega, \mathcal{A}, \mathbb{P})$ the probability space on which all the X_n 's are defined. In what follows, we can keep in mind that Ω represents the randomness in the original sequence $(X_n)_n$. Thus, a given ω in Ω represents a given realisation of $(X_n)_n$.

As a preliminary step, from Proposition 3, there exists some subset Ω_1 of Ω such that $\mathbb{P}(\Omega_1) = 1$ and for every ω in Ω_1 ,

$$\frac{1}{n^4} \sum_{i,j,k,l=1}^n h^2((X_i^1(\omega), X_j^2(\omega)), (X_k^1(\omega), X_l^2(\omega))) \xrightarrow{n \rightarrow +\infty} \mathbb{E} [h^2((X_1^1, X_2^2), (X_3^1, X_4^2))]. \quad (31)$$

Applying Theorem 3 in [60], since $(\mathcal{X}, d_{\mathcal{X}})$ defined by (9) is separable, P -a.s. in $(X_n)_n$, $P_n^1 \Rightarrow P^1$ and $P_n^2 \Rightarrow P^2$. We deduce that there exists some subset Ω_2 of Ω such that $\mathbb{P}(\Omega_2) = 1$ and for every ω in Ω_2 ,

$$P_{n,\omega}^1 \otimes P_{n,\omega}^2 \Rightarrow P^1 \otimes P^2, \quad (32)$$

Now, let us consider $\Omega_0 = \Omega_1 \cap \Omega_2$, and fix ω in Ω_0 .

Following the proof of Skorokhod's representation theorem in [14, Theorem 11.7.2 p.415], since $(\mathcal{X}^2, d_{\mathcal{X}^2})$, as defined from (9) in Section 3.2, is a separable space, it is possible to construct some probability space $(\Omega'_\omega, \mathcal{A}'_\omega, \mathbb{P}'_\omega)$, and some random variables $\bar{Y}_{n,\omega,1}^* : \Omega'_\omega \rightarrow \mathcal{X}^2$, $\bar{Y}_{n,\omega,2}^* : \Omega'_\omega \rightarrow \mathcal{X}^2$ with distribution $P_{n,\omega}^1 \otimes P_{n,\omega}^2$, and $\bar{Y}_{\omega,1} : \Omega'_\omega \rightarrow \mathcal{X}^2$, $\bar{Y}_{\omega,2} : \Omega'_\omega \rightarrow \mathcal{X}^2$ with distribution $P^1 \otimes P^2$ such that

$$\mathbb{P}'_\omega\text{-a.s.}, \quad \bar{Y}_{n,\omega,1}^* \xrightarrow{n \rightarrow +\infty} \bar{Y}_{\omega,1},$$

$$\mathbb{P}'_\omega\text{-a.s.}, \quad \bar{Y}_{n,\omega,2}^* \xrightarrow{n \rightarrow +\infty} \bar{Y}_{\omega,2},$$

$\left\{ (\bar{Y}_{n,\omega,1}^*)_{n \geq 1}, \bar{Y}_{\omega,1} \right\}$ and $\left\{ (\bar{Y}_{n,\omega,2}^*)_{n \geq 1}, \bar{Y}_{\omega,2} \right\}$ being independent, so that:

$$\mathbb{P}'_\omega\text{-a.s.}, \quad (\bar{Y}_{n,\omega,1}^*, \bar{Y}_{n,\omega,2}^*) \xrightarrow{n \rightarrow +\infty} (\bar{Y}_{\omega,1}, \bar{Y}_{\omega,2}), \quad (33)$$

with respect to the product metric d defined by (10).

But under (\mathcal{A}_{Cont}) , h is continuous on \mathcal{C} such that $\mathbb{P}'_\omega((\bar{Y}_{\omega,1}, \bar{Y}_{\omega,2}) \in \mathcal{C}) = (P^1 \otimes P^2)^{\otimes 2}(\mathcal{C}) = 1$, hence

$$\mathbb{P}'_{\omega}\text{-a.s.}, \quad h(\bar{Y}_{n,\omega,1}^*, \bar{Y}_{n,\omega,2}^*) \xrightarrow{n \rightarrow +\infty} h(\bar{Y}_{\omega,1}, \bar{Y}_{\omega,2}).$$

Since \mathbb{P}'_{ω} -a.s. convergence implies convergence in probability, in order to obtain the \mathbb{L}^2 -convergence in (30), according to Theorem 16.6 p. 165 of [55], we only need to prove that the sequence $(h^2(\bar{Y}_{n,\omega,1}^*, \bar{Y}_{n,\omega,2}^*))_{n \geq 1}$ is uniformly integrable, or that

$$\mathbb{E}'_{\omega}[h^2(\bar{Y}_{n,\omega,1}^*, \bar{Y}_{n,\omega,2}^*)] \xrightarrow{n \rightarrow +\infty} \mathbb{E}'_{\omega}[h^2(\bar{Y}_{\omega,1}, \bar{Y}_{\omega,2})].$$

Moreover by (31) we have

$$\begin{aligned} \mathbb{E}'_{\omega}[h^2(\bar{Y}_{n,\omega,1}^*, \bar{Y}_{n,\omega,2}^*)] &= \frac{1}{n^4} \sum_{i,j,k,l=1}^n h^2((X_i^1(\omega), X_j^2(\omega)), (X_k^1(\omega), X_l^2(\omega))) \\ &\xrightarrow{n \rightarrow +\infty} \mathbb{E}[h^2((X_1^1, X_2^2), (X_3^1, X_4^2))] = \mathbb{E}'_{\omega}[h^2(\bar{Y}_{\omega,1}, \bar{Y}_{\omega,2})]. \end{aligned}$$

We finally obtained (30) for any ω in Ω_0 , with $\mathbb{P}(\Omega_0) = 1$, which ends the proof.

5.4 Proof of Proposition 4

Let $(X_n)_n$ be a sequence of i.i.d pairs of point processes with distribution $P^1 \otimes P^2$ on \mathcal{X}^2 . Then, according to (\mathcal{A}_{Cent}) , for all $i \neq j$, $\mathbb{E}[h(X_i, X_j)] = 0$.

For a better readability, we introduce $\mathbb{E}[h|X_i] = \mathbb{E}[h(X_i, X)|X_i] = \mathbb{E}[h(X, X_i)|X_i]$ for some X with distribution $P^1 \otimes P^2$, and independent of X_i . By Hoeffding's decomposition for non degenerate U -statistics, we obtain

$$\begin{aligned} \sqrt{n}U_{n,h}(\mathbb{X}_n) &= \frac{2}{\sqrt{n}(n-1)} \sum_{i < j} h(X_i, X_j) \\ &= \frac{2}{\sqrt{n}(n-1)} \underbrace{\sum_{i < j} (\mathbb{E}[h|X_i] + \mathbb{E}[h|X_j])}_{T_n} + \frac{2}{\sqrt{n}(n-1)} \underbrace{\sum_{i < j} g(X_i, X_j)}_{M_n}, \end{aligned}$$

with $g(X_i, X_j) = h(X_i, X_j) - \mathbb{E}[h|X_i] - \mathbb{E}[h|X_j]$.

- Let us prove that

$$\frac{2}{\sqrt{n}(n-1)} M_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \quad (34)$$

We first notice that for $i < j$, $\mathbb{E}[g(X_i, X_j)] = 0$, $\mathbb{E}[g^2(X_i, X_j)] < +\infty$, and give a simpler expression for $\mathbb{E}[M_n^2] = \sum_{i < j} \sum_{k < l} \mathbb{E}[g(X_i, X_j)g(X_k, X_l)]$. If $\{i, j\} \cap \{k, l\} = \emptyset$, $i < j$, $k < l$, then $\mathbb{E}[g(X_i, X_j)g(X_k, X_l)] = (\mathbb{E}[g(X_i, X_j)])^2 = 0$. If $\#\{i, j\} \cap \{k, l\} = 1$, with for instance $k = i$, $j \neq l$, ($i < j$, $i < l$) (the other cases may be treated similarly), then

$$\mathbb{E}[g(X_i, X_j)g(X_i, X_l)] = \mathbb{E}[\mathbb{E}[g(X_i, X_j)|X_i] \mathbb{E}[g(X_i, X_l)|X_i]] = 0.$$

Finally, we obtain

$$\mathbb{E}[M_n^2] = \sum_{i < j} \mathbb{E}[g^2(X_i, X_j)] = \frac{n(n-1)}{2} \mathbb{E}[g^2(X_1, X_2)].$$

As a consequence, for $\varepsilon > 0$, Chebychev's inequality leads to

$$\begin{aligned} \mathbb{P}\left(\left|\frac{2}{\sqrt{n}(n-1)} M_n\right| > \varepsilon\right) &= \mathbb{P}\left(|M_n| > \sqrt{n}(n-1) \frac{\varepsilon}{2}\right) \\ &\leq \frac{4\mathbb{E}[M_n^2]}{n(n-1)^2 \varepsilon^2} = \frac{2\mathbb{E}[g^2(X_1, X_2)]}{(n-1)\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0. \end{aligned}$$

- Now, let us prove that

$$\frac{2}{\sqrt{n}(n-1)}T_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma_{h, P^1 \otimes P^2}^2). \quad (35)$$

Notice that $T_n = \sum_{i=1}^n (n-i)\mathbb{E}[h|X_i] + \sum_{j=1}^n (j-1)\mathbb{E}[h|X_j] = (n-1) \sum_{i=1}^n \mathbb{E}[h|X_i]$.
Thus,

$$\frac{2}{\sqrt{n}(n-1)}T_n = \frac{2}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}[h|X_i].$$

Since the $\mathbb{E}[h|X_i]$ are i.i.d, with $\mathbb{E}[\mathbb{E}[h|X_i]] = 0$ and $\text{Var}(\mathbb{E}[h|X_i]) = \sigma_{h, P^1 \otimes P^2}^2/4$, thanks to (\mathcal{A}_{Mmt}) , the Central Limit Theorem leads to (35).

Finally, combining (34) and (35), Slutsky's lemma ensures the convergence in distribution of $\sqrt{n}U_{n,h}(\mathbb{X}_n)$ towards $\mathcal{N}(0, \sigma_{h, P^1 \otimes P^2}^2)$.

Now, notice that

$$\mathbb{E} \left[(\sqrt{n}U_{n,h}(\mathbb{X}_n))^2 \right] = \frac{1}{n(n-1)^2} \sum_{i \neq i'} \sum_{j \neq j'} \mathbb{E}[h(X_i, X_{i'})h(X_j, X_{j'})].$$

Let us consider all the cases where $i \neq i'$ and $j \neq j'$.

If $\#\{i, i', j, j'\} = 4$, then by independence, $\mathbb{E}[h(X_i, X_{i'})h(X_j, X_{j'})] = \mathbb{E}[h(X_i, X_{i'})] \mathbb{E}[h(X_j, X_{j'})] = 0$.

If $\#\{i, i', j, j'\} = 3$, then, by symmetry of h , $\mathbb{E}[h(X_i, X_{i'})h(X_j, X_{j'})] = \mathbb{E}[h(X_1, X_2)h(X_1, X_2)] = \sigma_{h, P^1 \otimes P^2}^2/4$.

Finally, if $\#\{i, i', j, j'\} = 2$, $\mathbb{E}[h(X_i, X_{i'})h(X_j, X_{j'})] = \mathbb{E}[(h(X_1, X_2))^2]$.

Therefore,

$$\mathbb{E} \left[(\sqrt{n}U_{n,h}(\mathbb{X}_n))^2 \right] = \frac{1}{n(n-1)^2} \left(4n(n-1)(n-2) \frac{\sigma_{h, P^1 \otimes P^2}^2}{4} + 2n(n-1) \mathbb{E}[(h(X_1, X_2))^2] \right) \xrightarrow[n \rightarrow \infty]{} \sigma_{h, P^1 \otimes P^2}^2,$$

which ends the proof of Proposition 4.

5.5 Proof of Corollary 1

By Proposition 4, we have that

$$\mathcal{L}(\sqrt{n}U_{n,h}, P^1 \otimes P^2) \Rightarrow \mathcal{N}(0, \sigma_{h, P^1 \otimes P^2}^2), \quad (36)$$

where $\mathcal{N}(0, \sigma_{h, P^1 \otimes P^2}^2)$ has a continuous c.d.f. Therefore, by application of [59, Lemma 2.11],

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}(\sqrt{n}U_{n,h}(\mathbb{X}_n^\perp) \leq z) - \Phi_{0, \sigma_{h, P^1 \otimes P^2}^2}(z) \right| \xrightarrow[n \rightarrow +\infty]{} 0.$$

Furthermore, since convergence w.r.t the d_2 distance implies weak convergence, Theorem 1 combined with (36) leads to

$$\mathcal{L}(\sqrt{n}U_{n,h}, P_n^1 \otimes P_n^2 | \mathbb{X}_n) \Rightarrow \mathcal{N}(0, \sigma_{h, P^1 \otimes P^2}^2) \quad P\text{-a.s. in } (X_n)_n, \quad (37)$$

so

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}(\sqrt{n}U_{n,h}(\mathbb{X}_n^*) \leq z | \mathbb{X}_n) - \Phi_{0, \sigma_{h, P^1 \otimes P^2}^2}(z) \right| \xrightarrow[n \rightarrow +\infty]{} 0 \quad P\text{-a.s. in } (X_n)_n. \quad (38)$$

As a consequence,

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}(\sqrt{n}U_{n,h}(\mathbb{X}_n^*) \leq z | \mathbb{X}_n) - \mathbb{P}(\sqrt{n}U_{n,h}(\mathbb{X}_n^\perp) \leq z) \right| \xrightarrow[n \rightarrow +\infty]{} 0 \quad P\text{-a.s. in } (X_n)_n.$$

As for the second part of the result, since $\Phi_{0, \sigma_{h, P^1 \otimes P^2}^2}^{-1}$ is continuous on $[0, 1]$, it is sufficient to apply [59, Lemma 21.2] to both (36) and (37), on the event where (37) holds, to obtain that $q_{n,h,\eta}^*(\mathbb{X}_n)$ and $q_{n,h,\eta}^\perp$ both converge towards $\Phi_{0, \sigma_{h, P^1 \otimes P^2}^2}^{-1}(\eta)$ on the same event. This ends the proof.

5.6 Proof of Theorem 2

Let us focus on the case $\Delta = \Delta_{h,\alpha}^{*+}$, the proof in the other cases being similar. Given α in $(0, 1)$, we deduce from Corollary 1 that

$$q_{n,h,1-\alpha}^*(\mathbb{X}_n) \xrightarrow[n \rightarrow +\infty]{} \Phi_{0,\sigma_{h,P^1 \otimes P^2}}^{-1}(1-\alpha) \quad P\text{-a.s. in } (X_n)_n. \quad (39)$$

Then, from Proposition 4 and Slutsky's lemma, under (H_0) , $(\sqrt{n}U_{n,h}(\mathbb{X}_n), q_{n,h,1-\alpha}^*(\mathbb{X}_n))$ converges in distribution towards $(Z, \Phi_{0,\sigma_{h,P^1 \otimes P^2}}^{-1}(1-\alpha))$, where $Z \sim \mathcal{N}(0, \sigma_{h,P^1 \otimes P^2}^2)$. Therefore, under (H_0) ,

$$\mathbb{P}(\sqrt{n}U_{n,h}(\mathbb{X}_n) > q_{n,h,1-\alpha}^*(\mathbb{X}_n)) \xrightarrow[n \rightarrow +\infty]{} \alpha.$$

Under (H_1) , by Proposition 3,

$$U_{n,h}(\mathbb{X}_n) \xrightarrow[n \rightarrow +\infty]{} \int h(x, x') dP(x) dP(x') > 0 \quad P\text{-a.s. in } (X_n)_n.$$

Furthermore, from (39), we deduce that

$$q_{n,h,1-\alpha}^*(\mathbb{X}_n) / \sqrt{n} \xrightarrow[n \rightarrow +\infty]{} 0 \quad P\text{-a.s. in } (X_n)_n,$$

as a consequence, there exists an integer n_0 large enough such that for every $n \geq n_0$,

$$\mathbb{P}(\sqrt{n}U_{n,h}(\mathbb{X}_n) \leq q_{n,h,1-\alpha}^*(\mathbb{X}_n)) \leq \mathbb{P}\left(\frac{1}{2} \int h(x, x') dP(x) dP(x') \leq q_{n,h,1-\alpha}^*(\mathbb{X}_n) / \sqrt{n}\right) \xrightarrow[n \rightarrow +\infty]{} 0,$$

since $\int h(x, x') dP(x) dP(x') > 0$.

5.7 Proof of Proposition 5

Let us focus again on the case $\Delta_{B_n}^* = \Delta_{B_n,h,\alpha}^{*+}$, the other cases being similar. First notice that $\Delta_{B_n}^*(\mathbb{X}_n) = \mathbb{1}_{\{\sqrt{n}U_{n,h}(\mathbb{X}_n) > \sqrt{n}U^*(\lceil (1-\alpha)B_n \rceil)\}}$.

Let us introduce the c.d.f F_{n,\mathbb{X}_n}^* of $\mathcal{L}(\sqrt{n}U_{n,h,\varphi}, P_n^1 \otimes P_n^2 | \mathbb{X}_n)$ defined by $F_{n,\mathbb{X}_n}^*(z) = \mathbb{P}(\sqrt{n}U_{n,h}(\mathbb{X}_n^*) \leq z | \mathbb{X}_n)$, and let $F_{n,\mathbb{X}_n}^{*,B_n}$ be the empirical c.d.f of $\mathcal{L}(\sqrt{n}U_{n,h,\varphi}, P_n^1 \otimes P_n^2 | \mathbb{X}_n)$ associated with the bootstrap sample $(\mathbb{X}_n^{*1}, \dots, \mathbb{X}_n^{*B_n})$, that is

$$\forall z \in \mathbb{R}, \quad F_{n,\mathbb{X}_n}^{*,B_n}(z) = \frac{1}{B_n} \sum_{b=1}^{B_n} \mathbb{1}_{\{\sqrt{n}U_{n,h,\varphi}(\mathbb{X}_n^{*b}) \leq z\}}.$$

Then, using the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality (see [59] p.268 for instance), we derive that for all $n \geq 2$, all possible realisation \mathbf{x}_n of \mathbb{X}_n , and for all $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{z \in \mathbb{R}} |F_{n,\mathbf{x}_n}^{*,B_n}(z) - F_{n,\mathbf{x}_n}^*(z)| > \varepsilon\right) \leq 2e^{-2B_n\varepsilon^2}.$$

Therefore, we obtain that

$$\begin{aligned} \mathbb{P}\left(\sup_{z \in \mathbb{R}} |F_{n,\mathbb{X}_n}^{*,B_n}(z) - F_{n,\mathbb{X}_n}^*(z)| > \varepsilon\right) &= \mathbb{E}\left[\mathbb{P}\left(\sup_{z \in \mathbb{R}} |F_{n,\mathbb{X}_n}^{*,B_n}(z) - F_{n,\mathbb{X}_n}^*(z)| > \varepsilon \middle| \mathbb{X}_n\right)\right] \\ &\leq 2e^{-2B_n\varepsilon^2} \xrightarrow[n \rightarrow \infty]{} 0, \end{aligned}$$

that is exactly

$$\sup_{z \in \mathbb{R}} |F_{n,\mathbb{X}_n}^{*,B_n}(z) - F_{n,\mathbb{X}_n}^*(z)| \xrightarrow[n \rightarrow \infty]{} 0. \quad (40)$$

Moreover, by (38), $\sup_{z \in \mathbb{R}} |F_{n,\mathbb{X}_n}^*(z) - \Phi_{0,\sigma_{h,P^1 \otimes P^2}}^{-1}(z)| \xrightarrow[n \rightarrow +\infty]{} 0 \quad P\text{-a.s. in } (X_n)_n$.

Therefore, we obtain that

$$\sup_{z \in \mathbb{R}} |F_{n,\mathbb{X}_n}^{*,B_n}(z) - \Phi_{0,\sigma_{h,P^1 \otimes P^2}}^{-1}(z)| \xrightarrow[n \rightarrow \infty]{} 0. \quad (41)$$

We finish the proof using the similar arguments as in [59, Lemma 21.2], combined with a subsequence argument since the convergence in (41) occurs in probability, and not almost surely. Let ϕ_0 be an extraction. Then, by (41), there exists an extraction ϕ_1 , and some $\Omega_0 \subset \Omega$ such that $\mathbb{P}(\Omega_0) = 1$, and for every ω in Ω_0 ,

$$\sup_{z \in \mathbb{R}} \left| F_{\phi_1 \circ \phi_0(n), \mathbb{X}_{\phi_1 \circ \phi_0(n)}}^{*B_{\phi_1 \circ \phi_0(n)}}(\omega)(z) - \Phi_{0, \sigma_{h, P^1 \otimes P^2}^2}(z) \right| \xrightarrow{n \rightarrow \infty} 0.$$

From now on, fix ω in Ω_0 . Notice that this comes down to fix a realisation of \mathbb{X}_n , and a realisation of $(\mathbb{X}_n^{*1}, \dots, \mathbb{X}_n^{*B_n})$, and in particular, $F_{n, \mathbb{X}_n}^{*B_n}(\omega)$ is deterministic.

Consider $V \sim \mathcal{N}(0, 1)$. Then, we obtain that $F_{\phi_1 \circ \phi_0(n), \mathbb{X}_{\phi_1 \circ \phi_0(n)}}^{*B_{\phi_1 \circ \phi_0(n)}}(\omega)(V) \xrightarrow[n \rightarrow \infty]{a.s.} \Phi_{0, \sigma_{h, P^1 \otimes P^2}^2}(V)$. In particular, for all η in $(0, 1)$,

$$\begin{aligned} \Phi_{0,1} \left(\left(F_{\phi_1 \circ \phi_0(n), \mathbb{X}_{\phi_1 \circ \phi_0(n)}}^{*B_{\phi_1 \circ \phi_0(n)}}(\omega) \right)^{-1}(\eta) \right) &= \mathbb{P} \left(F_{\phi_1 \circ \phi_0(n), \mathbb{X}_{\phi_1 \circ \phi_0(n)}}^{*B_{\phi_1 \circ \phi_0(n)}}(\omega)(V) < \eta \right) \\ &\xrightarrow{n \rightarrow \infty} \mathbb{P} \left(\Phi_{0, \sigma_{h, P^1 \otimes P^2}^2}(V) < \eta \right) \\ &= \Phi_{0,1} \left(\left(\Phi_{0, \sigma_{h, P^1 \otimes P^2}^2} \right)^{-1}(\eta) \right). \end{aligned}$$

Finally, as $\Phi_{0,1}$ is a one-to-one function and $\Phi_{0,1}^{-1}$ is continuous, we obtain that

$$\sqrt{\phi_1 \circ \phi_0(n)} U^*([\eta(B_{\phi_1 \circ \phi_0(n)})]) (\omega) = \left(F_{\phi_1 \circ \phi_0(n), \mathbb{X}_{\phi_1 \circ \phi_0(n)}}^{*B_{\phi_1 \circ \phi_0(n)}}(\omega) \right)^{-1}(\eta) \xrightarrow{n \rightarrow \infty} \Phi_{0, \sigma_{h, P^1 \otimes P^2}^2}^{-1}(\eta),$$

and this for all ω in Ω_0 , and any initial extraction ϕ_0 . Therefore, we obtain

$$\sqrt{n} U^*([\eta B_n]) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \Phi_{0, \sigma_{h, P^1 \otimes P^2}^2}^{-1}(\eta),$$

and the proof is ended exactly as the proof of Theorem 2.

5.8 Proof of Proposition 6

As the sigma-algebra considered on $(\mathcal{X}^2)^{\otimes n}$ is generated by the family containing all products $\prod_{i=1}^n (A_i^1 \times A_i^2)$, with A_i^1 and A_i^2 Borel sets of $(\mathcal{X}, d_{\mathcal{X}})$, we just need to verify that for all $A = \prod_{i=1}^n (A_i^1 \times A_i^2)$ of this form, $\mathbb{P}(\mathbb{X}_n^{\pi_n} \in A) = \mathbb{P}(\mathbb{X}_n \in A)$, and $\mathbb{P}(\mathbb{X}_n^{\Pi_n} \in A) = \mathbb{P}(\mathbb{X}_n \in A)$. We first have that under (H_0) ,

$$\begin{aligned} \mathbb{P}(\mathbb{X}_n^{\pi_n} \in A) &= \mathbb{P} \left(\bigcap_{i=1}^n \left\{ X_i \in A_i^1 \times A_{\pi_n^{-1}(i)}^2 \right\} \right) \\ &= \prod_{i=1}^n \left[P^1(A_i^1) P^2(A_{\pi_n^{-1}(i)}^2) \right] \\ &= \prod_{i=1}^n \left[P^1(A_i^1) P^2(A_i^2) \right] \\ &= \mathbb{P} \left(\bigcap_{i=1}^n \left\{ X_i \in A_i^1 \times A_i^2 \right\} \right) \\ &= \mathbb{P}(\mathbb{X}_n \in A). \end{aligned}$$

Now consider a random permutation Π_n , uniformly distributed on (\mathfrak{S}_n) , and independent of \mathbb{X}_n . Then

$$\begin{aligned} \mathbb{P}(\mathbb{X}_n^{\Pi_n} \in A) &= \sum_{\pi_n \in \mathfrak{S}_n} \mathbb{P}(\mathbb{X}_n^{\pi_n} \in A) \mathbb{P}(\Pi_n = \pi_n) \\ &= \frac{1}{n!} \sum_{\pi_n \in \mathfrak{S}_n} \mathbb{P}(\mathbb{X}_n \in A) \\ &= \mathbb{P}(\mathbb{X}_n \in A), \end{aligned}$$

which ends the proof.

5.9 Proof of Theorem 3

Sketch of proof. Let d_{BL} denote the bounded Lipschitz metric, which metrizes the weak convergence (see [14, Prop. 11.3.2 and Th. 11.3.3] for instance), and defined by

$$d_{BL}(\mu, \nu) = \sup_{f \in BL, \|f\|_{BL} \leq 1} \left| \int_{\mathbb{R}} f (d\mu - d\nu) \right|,$$

where BL denotes the set of bounded Lipschitz function on \mathbb{R} , and $\|f\|_{BL} = \|f\|_{\infty} + \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}$. In the following, when Z_n is real-valued random variable depending on \mathbb{X}_n and Π_n , we will denote by $\mathcal{L}(Z_n | \mathbb{X}_n)$ the conditional distribution of Z_n given \mathbb{X}_n with respect to the Lebesgue measure.

The first step of the proof consists of decomposing $\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n})$ in three terms:

$$\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}) = \frac{n}{n-1} \left(M_n^{\Pi_n}(\mathbb{X}_n) + \frac{R_n^{\Pi_n}(\mathbb{X}_n)}{\sqrt{n}} - \frac{T_n(\mathbb{X}_n)}{\sqrt{n}} \right),$$

where $M_n^{\Pi_n}(\mathbb{X}_n)$ is associated with a martingale difference array, and:

$$\mathbb{E} \left[\left| \frac{R_n^{\Pi_n}(\mathbb{X}_n)}{\sqrt{n}} \right| \middle| \mathbb{X}_n \right] \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \text{ and } \frac{T_n(\mathbb{X}_n)}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

So,

$$d_{BL} \left(\mathcal{L}(\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}) | \mathbb{X}_n), \mathcal{L} \left(\frac{n}{n-1} M_n^{\Pi_n}(\mathbb{X}_n) \middle| \mathbb{X}_n \right) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

The second step will be to prove, from the Cramér-Wold device and the Central Limit Theorem for martingales, that

$$d_{BL} \left(\mathcal{L}(M_n^{\Pi_n}(\mathbb{X}_n) | \mathbb{X}_n), \mathcal{N}(0, \sigma_{h_\varphi, P^1 \otimes P^2}^2) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

so that

$$d_{BL} \left(\mathcal{L}(\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}) | \mathbb{X}_n), \mathcal{N}(0, \sigma_{h_\varphi, P^1 \otimes P^2}^2) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

The third step will be to obtain the convergence:

$$\mathbb{E} \left[(\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}))^2 \middle| \mathbb{X}_n \right] \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sigma_{h_\varphi, P^1 \otimes P^2}^2,$$

so that finally

$$d_2 \left(\mathcal{L}(\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}) | \mathbb{X}_n), \mathcal{N}(0, \sigma_{h_\varphi, P^1 \otimes P^2}^2) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

First step: decomposition of $\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n})$.

It is obvious that by the definition (4) of h_φ ,

$$U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}) = \frac{1}{n-1} U_n^{\Pi_n}, \tag{42}$$

where

$$\begin{aligned} U_n^{\Pi_n} &= \sum_{i=1}^n \varphi(X_i^1, X_{\Pi_n(i)}^2) - \frac{1}{n} \sum_{i,j=1}^n \varphi(X_i^1, X_j^2) \\ &= \sum_{i=1}^n \varphi(X_i^1, X_{\Pi_n(i)}^2) - \frac{1}{n} \sum_{i,j=1}^n \mathbb{E}[\varphi(X_i^1, X_j^2) | X_i^1] - \frac{1}{n} \sum_{i,j=1}^n \mathbb{E}[\varphi(X_i^1, X_j^2) | X_j^2] + \frac{1}{n} \sum_{i,j=1}^n \mathbb{E}[\varphi(X_i^1, X_j^2)] \\ &\quad - \frac{1}{n} \sum_{i,j=1}^n (\varphi(X_i^1, X_j^2) - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_i^1] - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_j^2] + \mathbb{E}[\varphi(X_i^1, X_j^2)]). \end{aligned}$$

On the one hand, if $\mathbb{E}_P[f]$ and $\mathbb{E}_{P^1 \otimes P^2}[f]$ respectively denote $\mathbb{E}[f(X_i^1, X_i^2)]$, and $\mathbb{E}[f(X_i^1, X_j^2)]$ with $j \neq i$, for any integrable function f , then

$$\begin{aligned} \frac{1}{n} \sum_{i,j=1}^n \mathbb{E}[\varphi(X_i^1, X_j^2)] &= \sum_{i,j=1}^n \mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{E}[\varphi(X_i^1, X_j^2)] - \sum_{i,j=1}^n \left(\mathbb{1}_{\{\Pi_n(i)=j\}} - \frac{1}{n} \right) \mathbb{E}[\varphi(X_i^1, X_j^2)] \\ &= \sum_{i,j=1}^n \mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{E}[\varphi(X_i^1, X_j^2)] - (\mathbb{E}_P[\varphi] - \mathbb{E}_{P^1 \otimes P^2}[\varphi]) \sum_{i=1}^n \left(\mathbb{1}_{\{\Pi_n(i)=i\}} - \frac{1}{n} \right). \end{aligned}$$

On the other hand,

$$\begin{aligned} \frac{1}{n} \sum_{i,j=1}^n \mathbb{E}[\varphi(X_i^1, X_j^2) | X_i^1] &= \sum_{i,j=1}^n \mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{E}[\varphi(X_i^1, X_j^2) | X_i^1] - \sum_{i,j=1}^n \left(\mathbb{1}_{\{\Pi_n(i)=j\}} - \frac{1}{n} \right) \mathbb{E}[\varphi(X_i^1, X_j^2) | X_i^1] \\ &= \sum_{i,j=1}^n \mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{E}[\varphi(X_i^1, X_j^2) | X_i^1] \\ &\quad - \sum_{i=1}^n \left(\mathbb{1}_{\{\Pi_n(i)=i\}} - \frac{1}{n} \right) (\mathbb{E}[\varphi(X_i^1, X_i^2) | X_i^1] - \mathbb{E}[\varphi(X_i^1, X^2) | X_i^1]), \end{aligned}$$

where $X = (X^1, X^2)$ is assumed to be P -distributed and independent of $(X_n)_n$, and in the same way,

$$\begin{aligned} \frac{1}{n} \sum_{i,j=1}^n \mathbb{E}[\varphi(X_i^1, X_j^2) | X_j^2] &= \sum_{i,j=1}^n \mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{E}[\varphi(X_i^1, X_j^2) | X_j^2] \\ &\quad - \sum_{j=1}^n \left(\mathbb{1}_{\{\Pi_n(j)=j\}} - \frac{1}{n} \right) (\mathbb{E}[\varphi(X_j^1, X_j^2) | X_j^2] - \mathbb{E}[\varphi(X^1, X_j^2) | X_j^2]). \end{aligned}$$

Therefore,

$$\begin{aligned} U_n^{\Pi_n} &= \sum_{i,j=1}^n \mathbb{1}_{\{\Pi_n(i)=j\}} (\varphi(X_i^1, X_j^2) - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_i^1] - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_j^2] + \mathbb{E}[\varphi(X_i^1, X_j^2)]) \\ &\quad + \sum_{i=1}^n \left(\mathbb{1}_{\{\Pi_n(i)=i\}} - \frac{1}{n} \right) \left(\mathbb{E}[\varphi(X_i^1, X_i^2) | X_i^1] + \mathbb{E}[\varphi(X_i^1, X_i^2) | X_i^2] \right. \\ &\quad \left. - \mathbb{E}[\varphi(X_i^1, X^2) | X_i^1] - \mathbb{E}[\varphi(X^1, X_i^2) | X_i^2] - \mathbb{E}_P[\varphi] + \mathbb{E}_{P^1 \otimes P^2}[\varphi] \right) \\ &\quad - \frac{1}{n} \sum_{i,j=1}^n (\varphi(X_i^1, X_j^2) - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_i^1] - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_j^2] + \mathbb{E}[\varphi(X_i^1, X_j^2)]). \end{aligned}$$

As a consequence,

$$\sqrt{n} U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}) = \frac{n}{n-1} \left(M_n^{\Pi_n}(\mathbb{X}_n) + \frac{R_n^{\Pi_n}(\mathbb{X}_n)}{\sqrt{n}} - \frac{T_n(\mathbb{X}_n)}{\sqrt{n}} \right), \quad (43)$$

with:

$$\begin{aligned} M_n^{\Pi_n}(\mathbb{X}_n) &= \frac{1}{\sqrt{n}} \sum_{i \neq j} \mathbb{1}_{\{\Pi_n(i)=j\}} (\varphi(X_i^1, X_j^2) - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_i^1] - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_j^2] + \mathbb{E}[\varphi(X_i^1, X_j^2)]), \\ R_n^{\Pi_n}(\mathbb{X}_n) &= \sum_{i=1}^n \left(\mathbb{1}_{\{\Pi_n(i)=i\}} - \frac{1}{n} \right) (\varphi(X_i^1, X_i^2) - \mathbb{E}[\varphi(X_i^1, X^2) | X_i^1] - \mathbb{E}[\varphi(X^1, X_i^2) | X_i^2] + \mathbb{E}_{P^1 \otimes P^2}[\varphi]), \\ T_n(\mathbb{X}_n) &= \frac{1}{n} \sum_{i \neq j} (\varphi(X_i^1, X_j^2) - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_i^1] - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_j^2] + \mathbb{E}[\varphi(X_i^1, X_j^2)]). \end{aligned}$$

Let us now prove that

$$d_{BL} \left(\mathcal{L}(\sqrt{n} U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}) | \mathbb{X}_n), \mathcal{L} \left(\frac{n}{n-1} M_n^{\Pi_n}(\mathbb{X}_n) \middle| \mathbb{X}_n \right) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

To do this, we first notice that for every function f in BL , and such that $\|f\|_{BL} \leq 1$,

$$\begin{aligned} \left| \mathbb{E} [f(\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n})) | \mathbb{X}_n] - \mathbb{E} \left[f \left(\frac{n}{n-1} M_n^{\Pi_n}(\mathbb{X}_n) \right) | \mathbb{X}_n \right] \right| &\leq \mathbb{E} \left[\left| \sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}) - \frac{n}{n-1} M_n^{\Pi_n}(\mathbb{X}_n) \right| | \mathbb{X}_n \right] \\ &\leq \frac{n}{n-1} \left(\mathbb{E} \left[\left| \frac{R_n^{\Pi_n}(\mathbb{X}_n)}{\sqrt{n}} \right| | \mathbb{X}_n \right] + \frac{|T_n(\mathbb{X}_n)|}{\sqrt{n}} \right), \end{aligned}$$

hence,

$$d_{BL} \left(\mathcal{L}(\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}) | \mathbb{X}_n), \mathcal{L} \left(\frac{n}{n-1} M_n^{\Pi_n}(\mathbb{X}_n) | \mathbb{X}_n \right) \right) \leq \frac{n}{n-1} \left(\mathbb{E} \left[\left| \frac{R_n^{\Pi_n}(\mathbb{X}_n)}{\sqrt{n}} \right| | \mathbb{X}_n \right] + \frac{|T_n(\mathbb{X}_n)|}{\sqrt{n}} \right). \quad (44)$$

Moreover, on the one hand, by Cauchy-Schwarz and Jensen's inequalities, and Π_n is independent of $(X_n)_n$,

$$\begin{aligned} \mathbb{E} \left[\left(\mathbb{E} \left[\left| \frac{R_n^{\Pi_n}(\mathbb{X}_n)}{\sqrt{n}} \right| | \mathbb{X}_n \right] \right)^2 \right] &\leq \frac{1}{n} \mathbb{E} \left[(R_n^{\Pi_n}(\mathbb{X}_n))^2 \right] \\ &\leq \frac{1}{n} \left(\sum_{i,j=1}^n \mathbb{E} \left[\left(\mathbb{1}_{\{\Pi_n(i)=i\}} - \frac{1}{n} \right) \left(\mathbb{1}_{\{\Pi_n(j)=j\}} - \frac{1}{n} \right) \right] \times \right. \\ &\quad \mathbb{E} \left[\left(\varphi(X_i^1, X_i^2) - \mathbb{E}[\varphi(X_i^1, X_i^2) | X_i^1] - \mathbb{E}[\varphi(X_i^1, X_i^2) | X_i^2] + \mathbb{E}_{P^1 \otimes P^2}[\varphi] \right) \times \right. \\ &\quad \left. \left. \left(\varphi(X_j^1, X_j^2) - \mathbb{E}[\varphi(X_j^1, X_j^2) | X_j^1] - \mathbb{E}[\varphi(X_j^1, X_j^2) | X_j^2] + \mathbb{E}_{P^1 \otimes P^2}[\varphi] \right) \right] \right] \right) \\ &\leq \frac{C}{n} (\mathbb{E}_P[\varphi^2] + \mathbb{E}_{P^1 \otimes P^2}[\varphi^2]) \sum_{i,j=1}^n \left(\mathbb{E}[\mathbb{1}_{\{\Pi_n(i)=i\}} \mathbb{1}_{\{\Pi_n(j)=j\}}] - \frac{1}{n^2} \right) \\ &\leq \frac{C}{n} (\mathbb{E}_P[\varphi^2] + \mathbb{E}_{P^1 \otimes P^2}[\varphi^2]) \left(\sum_{i=1}^n \left(\frac{1}{n} - \frac{1}{n^2} \right) + \sum_{i \neq j} \left(\frac{1}{n(n-1)} - \frac{1}{n^2} \right) \right) \\ &\leq \frac{C}{n} (\mathbb{E}_P[\varphi^2] + \mathbb{E}_{P^1 \otimes P^2}[\varphi^2]) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Therefore, from Markov's inequality, we deduce that

$$\mathbb{E} \left[\left| \frac{R_n^{\Pi_n}(\mathbb{X}_n)}{\sqrt{n}} \right| | \mathbb{X}_n \right] \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

On the other hand, denoting by $C_{i,j}$ the centred random variable

$$C_{i,j} = \varphi(X_i^1, X_j^2) - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_i^1] - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_j^2] + \mathbb{E}[\varphi(X_i^1, X_j^2)],$$

then

$$T_n(\mathbb{X}_n) = \frac{1}{n} \sum_{i \neq j} C_{i,j}.$$

Hence,

$$\mathbb{E} \left[\left(\frac{T_n(\mathbb{X}_n)}{\sqrt{n}} \right)^2 \right] = \frac{1}{n^3} \sum_{i \neq j} \sum_{k \neq l} \mathbb{E}[C_{i,j} C_{k,l}].$$

Notice that for $i \neq j$, $\mathbb{E}[C_{i,j} | X_i] = \mathbb{E}[C_{i,j} | X_j] = 0$.

If $\# \{i, j, k, l\} = 4$, then $\mathbb{E}[C_{i,j} C_{k,l}] = (\mathbb{E}[C_{i,j}])^2 = 0$.

If i, j, l are all different, then

$$\begin{aligned} \mathbb{E}[C_{i,j} C_{i,l}] &= \mathbb{E}[\mathbb{E}[C_{i,j} C_{i,l} | X_i, X_l]] \\ &= \mathbb{E}[\mathbb{E}[C_{i,j} | X_i] C_{i,l}] \\ &= 0. \end{aligned}$$

In the same way, for i, j, k all different, then $\mathbb{E}[C_{i,j}C_{k,i}] = 0$.

For $i \neq j$, then

$$\mathbb{E}[C_{i,j}^2] = \sigma_{h_\varphi, P^1 \otimes P^2}^2, \text{ and } \mathbb{E}[C_{i,j}C_{j,i}] \leq \sigma_{h_\varphi, P^1 \otimes P^2}^2, \quad (45)$$

by the Cauchy-Schwarz inequality.

Combining the above computations, we obtain that

$$\mathbb{E} \left[\left(\frac{T_n(\mathbb{X}_n)}{\sqrt{n}} \right)^2 \right] \leq 2 \frac{n(n-1)}{n^3} \sigma_{h_\varphi, P^1 \otimes P^2}^2 \xrightarrow{n \rightarrow \infty} 0,$$

therefore,

$$\frac{T_n(\mathbb{X}_n)}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

Finally, from (44), we derive that:

$$d_{BL} \left(\mathcal{L}(\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}) | \mathbb{X}_n), \mathcal{L}\left(\frac{n}{n-1}M_n^{\Pi_n}(\mathbb{X}_n) | \mathbb{X}_n\right) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0. \quad (46)$$

Second step: asymptotic normality of $M_n^{\Pi_n}(\mathbb{X}_n)$ given \mathbb{X}_n , in probability.

Recall that

$$M_n^{\Pi_n}(\mathbb{X}_n) = \frac{1}{\sqrt{n}} \sum_{i \neq j} \mathbb{1}_{\{\Pi_n(i)=j\}} (\varphi(X_i^1, X_j^2) - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_i^1] - \mathbb{E}[\varphi(X_i^1, X_j^2) | X_j^2] + \mathbb{E}[\varphi(X_i^1, X_j^2)]),$$

that is, with the above notation,

$$M_n^{\Pi_n}(\mathbb{X}_n) = \frac{1}{\sqrt{n}} \sum_{i \neq j} \mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j} = \frac{1}{\sqrt{n}} \sum_{i=2}^n \sum_{j=1}^{i-1} (\mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j} + \mathbb{1}_{\{\Pi_n(j)=i\}} C_{j,i}).$$

Let Π'_n be another uniformly distributed random permutation with values in \mathfrak{S}_n , independent of Π_n and \mathbb{X}_n , and

$$M_n^{\Pi'_n}(\mathbb{X}_n) = \frac{1}{\sqrt{n}} \sum_{i \neq j} \mathbb{1}_{\{\Pi'_n(i)=j\}} C_{i,j} = \frac{1}{\sqrt{n}} \sum_{i=2}^n \sum_{j=1}^{i-1} (\mathbb{1}_{\{\Pi'_n(i)=j\}} C_{i,j} + \mathbb{1}_{\{\Pi'_n(j)=i\}} C_{j,i}).$$

Lemma 1. *Considering the above notation,*

$$\mathcal{L} \left(\left(M_n^{\Pi_n}(\mathbb{X}_n), M_n^{\Pi'_n}(\mathbb{X}_n) \right)' \right) \Rightarrow \mathcal{N}_2 \left(0, \begin{pmatrix} \sigma_{h_\varphi, P^1 \otimes P^2}^2 & 0 \\ 0 & \sigma_{h_\varphi, P^1 \otimes P^2}^2 \end{pmatrix} \right),$$

where $\mathcal{N}_2(M, V)$ denotes the 2-dimensional Gaussian distribution with expectation M and variance-covariance matrix V .

Proof. According to the Cramér-Wold device, given a, b in \mathbb{R} , we aim at proving that

$$\mathcal{L} \left(aM_n^{\Pi_n}(\mathbb{X}_n) + bM_n^{\Pi'_n}(\mathbb{X}_n) \right) \Rightarrow \mathcal{N} \left(0, (a^2 + b^2) \sigma_{h_\varphi, P^1 \otimes P^2}^2 \right).$$

In order to deal with simpler mathematical expressions, we introduce below a few notations more.

- For any fixed integer $n \geq 2$, $\mathcal{F}_{n,k}$ denotes the filtration $\sigma(\Pi_n, \Pi'_n, X_1, X_2, \dots, X_k)$.
- Let

$$Y_{n,i} = \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} (\mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j} + \mathbb{1}_{\{\Pi_n(j)=i\}} C_{j,i}),$$

$$Y'_{n,i} = \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} (\mathbb{1}_{\{\Pi'_n(i)=j\}} C_{i,j} + \mathbb{1}_{\{\Pi'_n(j)=i\}} C_{j,i}),$$

$$M_{n,k} = \sum_{i=2}^k Y_{n,i} \text{ and } M'_{n,k} = \sum_{i=2}^k Y'_{n,i}, \text{ so that } M_n^{\Pi_n}(\mathbb{X}_n) = M_{n,n} \text{ and } M_n^{\Pi'_n}(\mathbb{X}_n) = M'_{n,n}.$$

Let us first prove that for a fixed integer $n \geq 2$, $(aM_{n,k} + bM'_{n,k}, \mathcal{F}_{n,k})_{2 \leq k \leq n}$ is a martingale difference array. Note that for $2 \leq i \leq n$,

$$\begin{aligned} \mathbb{E}[Y_{n,i} | \mathcal{F}_{n,i-1}] &= \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} \mathbb{E}[\mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j} + \mathbb{1}_{\{\Pi_n(j)=i\}} C_{j,i} | \mathcal{F}_{n,i-1}] \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} (\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{E}[C_{i,j} | X_j] + \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{E}[C_{j,i} | X_j]) \\ &= 0. \end{aligned}$$

In the same way, we have that $\mathbb{E}[Y'_{n,i} | \mathcal{F}_{n,i-1}] = 0$, so

$$\mathbb{E}[aM_{n,k} + bM'_{n,k} | \mathcal{F}_{n,k-1}] = aM_{n,k-1} + bM'_{n,k-1}.$$

We now use the following result which is commonly attributed to Brown [9].

Theorem 5. *Let $(X_{n,k})_{k \in \{1, \dots, p_n\}, n \in \mathbb{N}^*}$ be a martingale difference array, i.e. such that there exists an array of σ -algebra $(\mathcal{F}_{n,k})_{k \in \{1, \dots, p_n\}, n \in \mathbb{N}^*}$ that is increasing w.r.t. k such that for all $k = 1, \dots, p_n$,*

$$\mathbb{E}[X_{n,k} | \mathcal{F}_{n,k-1}] = 0.$$

Let $A_n = \sum_{k=1}^{p_n} \mathbb{E}[X_{n,k}^2 | \mathcal{F}_{n,k-1}]$. Assume that

- $A_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sigma^2 > 0$.
- For all $\varepsilon > 0$, $\sum_{k=1}^{p_n} \mathbb{E}[X_{n,k}^2 \mathbb{1}_{\{|X_{n,k}| > \varepsilon\}}] \rightarrow_{n \rightarrow \infty} 0$

Then $Z_n = \sum_{k=1}^{p_n} X_{n,k}$ converges in distribution towards $\mathcal{N}(0, \sigma^2)$.

From Theorem 5, we deduce that if:

$$\begin{aligned} (i) \quad & \sum_{i=2}^n \mathbb{E}[(aY_{n,i} + bY'_{n,i})^2 | \mathcal{F}_{n,i-1}] \xrightarrow[n \rightarrow \infty]{\mathbb{P}} (a^2 + b^2) \sigma_{h_\varphi, P^1 \otimes P^2}^2, \\ (ii) \quad & \sum_{i=2}^n \mathbb{E}[(aY_{n,i} + bY'_{n,i})^2 \mathbb{1}_{\{|aY_{n,i} + bY'_{n,i}| > \varepsilon\}}] \rightarrow 0 \text{ for any } \varepsilon > 0, \end{aligned}$$

then

$$\mathcal{L}(aM_n^{\Pi_n}(\mathbb{X}_n) + bM_n^{\Pi'_n}(\mathbb{X}_n)) \Rightarrow \mathcal{N}(0, (a^2 + b^2) \sigma_{h_\varphi, P^1 \otimes P^2}^2).$$

Let us now check that both (i) and (ii) are satisfied.

Assumption (i). Noticing that

$$\sum_{i=2}^n \mathbb{E}[(aY_{n,i} + bY'_{n,i})^2 | \mathcal{F}_{n,i-1}] = (a^2 + b^2) \sum_{i=2}^n \mathbb{E}[Y_{n,i}^2 | \mathcal{F}_{n,i-1}] + 2ab \sum_{i=2}^n \mathbb{E}[Y_{n,i} Y'_{n,i} | \mathcal{F}_{n,i-1}], \quad (47)$$

the proof of (i) can be decomposed into two points.

The first point consists of deriving that

$$\sum_{i=2}^n \mathbb{E}[Y_{n,i}^2] \xrightarrow[n \rightarrow \infty]{} \sigma_{h_\varphi, P^1 \otimes P^2}^2, \quad \text{and} \quad \mathbb{V}\text{ar}\left(\sum_{i=2}^n \mathbb{E}[Y_{n,i}^2 | \mathcal{F}_{n,i-1}]\right) \xrightarrow[n \rightarrow \infty]{} 0,$$

which leads, thanks to Chebychev's inequality, to

$$\sum_{i=2}^n \mathbb{E}[Y_{n,i}^2 | \mathcal{F}_{n,i-1}] \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sigma_{h_\varphi, P^1 \otimes P^2}^2.$$

The second point consists of proving that

$$\mathbb{E} \left[\left(\sum_{i=2}^n \mathbb{E} [Y_{n,i} Y'_{n,i} | \mathcal{F}_{n,i-1}] \right)^2 \right] \xrightarrow{n \rightarrow \infty} 0,$$

so

$$\sum_{i=2}^n \mathbb{E} [Y_{n,i} Y'_{n,i} | \mathcal{F}_{n,i-1}] \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

First point.

On the one hand,

$$\sum_{i=2}^n \mathbb{E} [Y_{n,i}^2] = \frac{1}{n} \sum_{i=2}^n \sum_{j,k=1}^{i-1} \mathbb{E} [(\mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j} + \mathbb{1}_{\{\Pi_n(j)=i\}} C_{j,i}) (\mathbb{1}_{\{\Pi_n(i)=k\}} C_{i,k} + \mathbb{1}_{\{\Pi_n(k)=i\}} C_{k,i})].$$

Furthermore, if $1 \leq j \neq k \leq i-1$,

$$\begin{aligned} & \mathbb{E} [(\mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j} + \mathbb{1}_{\{\Pi_n(j)=i\}} C_{j,i}) (\mathbb{1}_{\{\Pi_n(i)=k\}} C_{i,k} + \mathbb{1}_{\{\Pi_n(k)=i\}} C_{k,i})] \\ &= \mathbb{E} [\mathbb{E} [(\mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j} + \mathbb{1}_{\{\Pi_n(j)=i\}} C_{j,i}) (\mathbb{1}_{\{\Pi_n(i)=k\}} C_{i,k} + \mathbb{1}_{\{\Pi_n(k)=i\}} C_{k,i}) | X_i, X_j, \Pi_n]] \\ &= \mathbb{E} [(\mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j} + \mathbb{1}_{\{\Pi_n(j)=i\}} C_{j,i}) (\mathbb{1}_{\{\Pi_n(i)=k\}} \mathbb{E} [C_{i,k} | X_i] + \mathbb{1}_{\{\Pi_n(k)=i\}} \mathbb{E} [C_{k,i} | X_i])] \\ &= 0. \end{aligned}$$

Thus,

$$\begin{aligned} \sum_{i=2}^n \mathbb{E} [Y_{n,i}^2] &= \frac{1}{n} \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E} [(\mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j} + \mathbb{1}_{\{\Pi_n(j)=i\}} C_{j,i})^2] \\ &= \frac{1}{n} \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j}^2 + \mathbb{1}_{\{\Pi_n(j)=i\}} C_{j,i}^2 + 2 \mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(j)=i\}} C_{i,j} C_{j,i}] \\ &= \frac{1}{n} \sum_{i=2}^n \sum_{j=1}^{i-1} \left(\frac{2}{n} \mathbb{E} [C_{i,j}^2] + \frac{2}{n(n-1)} \mathbb{E} [C_{i,j} C_{j,i}] \right) \\ &= \frac{2}{n^2} \sum_{i=2}^n (i-1) \left(\mathbb{E} [C_{1,2}^2] + \frac{1}{n-1} \mathbb{E} [C_{1,2} C_{2,1}] \right) \end{aligned}$$

so

$$\sum_{i=2}^n \mathbb{E} [Y_{n,i}^2] = \frac{n-1}{n} \mathbb{E} [C_{1,2}^2] + \frac{1}{n} \mathbb{E} [C_{1,2} C_{2,1}]. \quad (48)$$

From (45), we finally derive that

$$\sum_{i=2}^n \mathbb{E} [Y_{n,i}^2] \xrightarrow{n \rightarrow \infty} \sigma_{h_\varphi, P^1 \otimes P^2}^2. \quad (49)$$

On the other hand, we have that:

$$\begin{aligned}
\mathbb{E}[Y_{n,i}^2 | \mathcal{F}_{n,i-1}] &= \frac{1}{n} \sum_{j,k=1}^{i-1} \mathbb{E}[(\mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j} + \mathbb{1}_{\{\Pi_n(j)=i\}} C_{j,i}) (\mathbb{1}_{\{\Pi_n(i)=k\}} C_{i,k} + \mathbb{1}_{\{\Pi_n(k)=i\}} C_{k,i}) | \mathcal{F}_{n,i-1}] \\
&= \frac{1}{n} \sum_{j=1}^{i-1} \mathbb{E}[\mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j}^2 + 2\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(j)=i\}} C_{i,j} C_{j,i} + \mathbb{1}_{\{\Pi_n(j)=i\}} C_{j,i}^2 | \mathcal{F}_{n,i-1}] \\
&\quad + \frac{1}{n} \sum_{1 \leq j \neq k \leq i-1} \mathbb{E}[\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=i\}} C_{i,j} C_{k,i} + \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{1}_{\{\Pi_n(i)=k\}} C_{j,i} C_{i,k} | \mathcal{F}_{n,i-1}] \\
&= \frac{1}{n} \sum_{j=1}^{i-1} \mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{E}[C_{i,j}^2 | X_j] + \frac{1}{n} \sum_{j=1}^{i-1} \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{E}[C_{j,i}^2 | X_j] \\
&\quad + \frac{2}{n} \sum_{j=1}^{i-1} \mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{E}[C_{i,j} C_{j,i} | X_j] \\
&\quad + \frac{2}{n} \sum_{1 \leq j \neq k \leq i-1} \mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=i\}} \mathbb{E}[C_{i,j} C_{k,i} | X_j, X_k].
\end{aligned}$$

Then,

$$\sum_{i=2}^n (\mathbb{E}[Y_{n,i}^2 | \mathcal{F}_{n,i-1}] - \mathbb{E}[Y_{n,i}^2]) = A_{n,1} + A_{n,2} + 2A_{n,3} + 2A_{n,4},$$

with

$$\begin{aligned}
A_{n,1} &= \frac{1}{n} \sum_{1 \leq j < i \leq n} \left(\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{E}[C_{i,j}^2 | X_j] - \frac{1}{n} \mathbb{E}[C_{i,j}^2] \right), \\
A_{n,2} &= \frac{1}{n} \sum_{1 \leq j < i \leq n} \left(\mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{E}[C_{j,i}^2 | X_j] - \frac{1}{n} \mathbb{E}[C_{j,i}^2] \right), \\
A_{n,3} &= \frac{1}{n} \sum_{1 \leq j < i \leq n} \left(\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{E}[C_{i,j} C_{j,i} | X_j] - \frac{1}{n(n-1)} \mathbb{E}[C_{i,j} C_{j,i}] \right), \\
A_{n,4} &= \frac{1}{n} \sum_{1 \leq j \neq k < i \leq n} (\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=i\}} \mathbb{E}[C_{i,j} C_{k,i} | X_j, X_k]).
\end{aligned}$$

Thus,

$$\begin{aligned}
\text{Var} \left(\sum_{i=2}^n (\mathbb{E}[Y_{n,i}^2 | \mathcal{F}_{n,i-1}]) \right) &= \mathbb{E}[(A_{n,1} + A_{n,2} + 2A_{n,3} + 2A_{n,4})^2] \\
&\leq 4(\mathbb{E}[A_{n,1}^2] + \mathbb{E}[A_{n,2}^2] + 4\mathbb{E}[A_{n,3}^2] + 4\mathbb{E}[A_{n,4}^2]). \tag{50}
\end{aligned}$$

Let us now control each term of the above right-hand side.

Convergence of $\mathbb{E}[A_{n,1}^2]$ and $\mathbb{E}[A_{n,2}^2]$.

$$\begin{aligned}
\mathbb{E}[A_{n,1}^2] &= \frac{1}{n^2} \sum_{1 \leq j < i \leq n} \sum_{1 \leq l < k \leq n} \mathbb{E} \left[\left(\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{E}[C_{i,j}^2 | X_j] - \frac{1}{n} \mathbb{E}[C_{i,j}^2] \right) \left(\mathbb{1}_{\{\Pi_n(k)=l\}} \mathbb{E}[C_{k,l}^2 | X_l] - \frac{1}{n} \mathbb{E}[C_{k,l}^2] \right) \right] \\
&= \frac{1}{n^2} \sum_{1 \leq j < i \leq n} \sum_{1 \leq l < k \leq n} \left(\mathbb{E}[\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] \mathbb{E}[\mathbb{E}[C_{i,j}^2 | X_j] \mathbb{E}[C_{k,l}^2 | X_l]] - \frac{1}{n^2} (\mathbb{E}[C_{i,j}^2])^2 \right).
\end{aligned}$$

Let us now consider all the cases where $1 \leq j < i \leq n$, and $1 \leq l < k \leq n$.

If $i = k$ and $j = l$, then

$$\mathbb{E}[\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] \mathbb{E}[\mathbb{E}[C_{i,j}^2 | X_j] \mathbb{E}[C_{k,l}^2 | X_l]] = \frac{1}{n} \mathbb{E}[(\mathbb{E}[C_{i,j}^2 | X_j])^2].$$

If $i = k$ and $j \neq l$, or if $i \neq k$ and $j = l$, then

$$\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] \mathbb{E} [\mathbb{E} [C_{i,j}^2 | X_j] \mathbb{E} [C_{k,l}^2 | X_l]] = 0.$$

If $i \neq k$ and $j \neq l$, then

$$\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] \mathbb{E} [\mathbb{E} [C_{i,j}^2 | X_j] \mathbb{E} [C_{k,l}^2 | X_l]] = \frac{1}{n(n-1)} (\mathbb{E} [C_{i,j}^2])^2.$$

By combining these results, from (45) and under the assumption $(\mathcal{A}_{\varphi, Mmt})$, we obtain that:

$$\mathbb{E} [A_{n,1}^2] \leq \frac{n-1}{2n^2} \left(\mathbb{E} [(\mathbb{E} [C_{i,j}^2 | X_j])^2] - \frac{\sigma_{h_\varphi, P^1 \otimes P^2}^4}{n} \right) + Cn^2 \left(\frac{1}{n(n-1)} - \frac{1}{n^2} \right) \sigma_{h_\varphi, P^1 \otimes P^2}^4 \xrightarrow{n \rightarrow \infty} 0.$$

One can prove in the same way that

$$\mathbb{E} [A_{n,2}^2] \xrightarrow{n \rightarrow \infty} 0.$$

Convergence of $\mathbb{E} [A_{n,3}^2]$.

$$\begin{aligned} \mathbb{E} [A_{n,3}^2] &= \frac{1}{n^2} \sum_{1 \leq j < i \leq n} \sum_{1 \leq l < k \leq n} \mathbb{E} \left[\left(\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{E} [C_{i,j} C_{j,i} | X_j] - \frac{1}{n(n-1)} \mathbb{E} [C_{i,j} C_{j,i}] \right) \times \right. \\ &\quad \left. \left(\mathbb{1}_{\{\Pi_n(k)=l\}} \mathbb{1}_{\{\Pi_n(l)=k\}} \mathbb{E} [C_{k,l} C_{l,k} | X_l] - \frac{1}{n(n-1)} \mathbb{E} [C_{k,l} C_{l,k}] \right) \right] \\ &= \frac{1}{n^2} \sum_{1 \leq j < i \leq n} \sum_{1 \leq l < k \leq n} \left(\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{1}_{\{\Pi_n(k)=l\}} \mathbb{1}_{\{\Pi_n(l)=k\}}] \mathbb{E} [\mathbb{E} [C_{i,j} C_{j,i} | X_j] \mathbb{E} [C_{k,l} C_{l,k} | X_l]] \right. \\ &\quad \left. - \frac{1}{n^2(n-1)^2} (\mathbb{E} [C_{i,j} C_{j,i}])^2 \right) \\ &= \frac{1}{n^2} \sum_{1 \leq j < i \leq n} \sum_{1 \leq l < k \leq n} \mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{1}_{\{\Pi_n(k)=l\}} \mathbb{1}_{\{\Pi_n(l)=k\}}] \mathbb{E} [\mathbb{E} [C_{i,j} C_{j,i} | X_j] \mathbb{E} [C_{k,l} C_{l,k} | X_l]] \\ &\quad - \frac{1}{4n^2} (\mathbb{E} [C_{1,2} C_{2,1}])^2. \end{aligned}$$

Let us again consider all the cases where $1 \leq j < i \leq n$, and $1 \leq l < k \leq n$.

If $i = k$ and $j = l$, then

$$\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{1}_{\{\Pi_n(k)=l\}} \mathbb{1}_{\{\Pi_n(l)=k\}}] \mathbb{E} [\mathbb{E} [C_{i,j} C_{j,i} | X_j] \mathbb{E} [C_{k,l} C_{l,k} | X_l]] = \frac{1}{n(n-1)} \mathbb{E} [(\mathbb{E} [C_{i,j} C_{j,i} | X_j])^2].$$

If $i = k$ and $j \neq l$, or if $i \neq k$ and $j = l$, then

$$\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{1}_{\{\Pi_n(k)=l\}} \mathbb{1}_{\{\Pi_n(l)=k\}}] \mathbb{E} [\mathbb{E} [C_{i,j} C_{j,i} | X_j] \mathbb{E} [C_{k,l} C_{l,k} | X_l]] = 0.$$

If $i \neq k$ and $j \neq l$, then

$$\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{1}_{\{\Pi_n(k)=l\}} \mathbb{1}_{\{\Pi_n(l)=k\}}] \mathbb{E} [\mathbb{E} [C_{i,j} C_{j,i} | X_j] \mathbb{E} [C_{k,l} C_{l,k} | X_l]] = \frac{(\mathbb{E} [C_{1,2} C_{2,1}])^2}{n(n-1)(n-2)(n-3)}.$$

Thus, under $(\mathcal{A}_{\varphi, Mmt})$, we finally have that:

$$\mathbb{E} [A_{n,3}^2] \leq \frac{1}{2n^2} \mathbb{E} [(\mathbb{E} [C_{1,2} C_{2,1} | X_1])^2] + C \frac{n (\mathbb{E} [C_{1,2} C_{2,1}])^2}{(n-1)(n-2)(n-3)} \xrightarrow{n \rightarrow \infty} 0.$$

Convergence of $\mathbb{E} [A_{n,4}^2]$.

$$\begin{aligned} \mathbb{E} [A_{n,4}^2] &= \frac{1}{n^2} \sum_{1 \leq j \neq k < i \leq n} \sum_{1 \leq p \neq q < l \leq n} \mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=i\}} \mathbb{1}_{\{\Pi_n(l)=p\}} \mathbb{1}_{\{\Pi_n(q)=l\}}] \times \\ &\quad \mathbb{E} [\mathbb{E} [C_{i,j} C_{k,i} | X_j, X_k] \mathbb{E} [C_{l,p} C_{q,l} | X_p, X_q]]. \end{aligned}$$

Let us now consider all the cases where $1 \leq j \neq k < i \leq n$, and $1 \leq p \neq q < l \leq n$.

If $\#\{j, k, p, q\} \geq 3$, there exists at least one element in $\{j, k, p, q\}$, j for instance (the other cases are studied in the same way), which differs from the other ones. Then,

$$\begin{aligned} \mathbb{E} [\mathbb{E} [C_{i,j} C_{k,i} | X_j, X_k] \mathbb{E} [C_{l,p} C_{q,l} | X_p, X_q]] &= \mathbb{E} [\mathbb{E} [\mathbb{E} [C_{i,j} C_{k,i} | X_j, X_k] \mathbb{E} [C_{l,p} C_{q,l} | X_p, X_q] | X_k, X_p, X_q]] \\ &= \mathbb{E} [\mathbb{E} [C_{i,j} C_{k,i} | X_k] \mathbb{E} [C_{l,p} C_{q,l} | X_p, X_q]] \\ &= \mathbb{E} [\mathbb{E} [\mathbb{E} [C_{i,j} C_{k,i} | X_i, X_k] | X_k] \mathbb{E} [C_{l,p} C_{q,l} | X_p, X_q]] \\ &= \mathbb{E} [\mathbb{E} [C_{k,i} \mathbb{E} [C_{i,j} | X_i] | X_k] \mathbb{E} [C_{l,p} C_{q,l} | X_p, X_q]]. \end{aligned}$$

Since $\mathbb{E} [C_{i,j} | X_i] = 0$, this leads to:

$$\mathbb{E} [\mathbb{E} [C_{i,j} C_{k,i} | X_j, X_k] \mathbb{E} [C_{l,p} C_{q,l} | X_p, X_q]] = 0. \quad (51)$$

If $j = p$, $k = q$, and $i = l$, then,

$$\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=i\}} \mathbb{1}_{\{\Pi_n(l)=p\}} \mathbb{1}_{\{\Pi_n(q)=l\}}] = \frac{1}{n(n-1)},$$

and

$$\begin{aligned} \mathbb{E} [\mathbb{E} [C_{i,j} C_{k,i} | X_j, X_k] \mathbb{E} [C_{l,p} C_{q,l} | X_p, X_q]] &= \mathbb{E} [(\mathbb{E} [C_{i,j} C_{k,i} | X_j, X_k])^2] \\ &= \mathbb{E} [(\mathbb{E} [C_{3,1} C_{2,3} | X_1, X_2])^2] \\ &< +\infty \quad \text{under } (\mathcal{A}_{\varphi, Mmt}). \end{aligned}$$

If $j = p$, $k = q$, and $i \neq l$, then $\mathbb{1}_{\{\Pi_n(k)=i\}} \mathbb{1}_{\{\Pi_n(q)=l\}} = 0$, so

$$\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=i\}} \mathbb{1}_{\{\Pi_n(l)=p\}} \mathbb{1}_{\{\Pi_n(q)=l\}}] = 0.$$

If $j = q$, $k = p$, and $i = l$, then $\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(l)=p\}} = 0$, so

$$\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=i\}} \mathbb{1}_{\{\Pi_n(l)=p\}} \mathbb{1}_{\{\Pi_n(q)=l\}}] = 0.$$

If $j = q$, $k = p$, and $i \neq l$, then

$$\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=i\}} \mathbb{1}_{\{\Pi_n(l)=p\}} \mathbb{1}_{\{\Pi_n(q)=l\}}] = \frac{(n-4)!}{n!},$$

and

$$\begin{aligned} \mathbb{E} [\mathbb{E} [C_{i,j} C_{k,i} | X_j, X_k] \mathbb{E} [C_{l,p} C_{q,l} | X_p, X_q]] &= \mathbb{E} [\mathbb{E} [C_{i,j} C_{k,i} | X_j, X_k] \mathbb{E} [C_{l,k} C_{j,l} | X_j, X_k]] \\ &= \mathbb{E} [\mathbb{E} [C_{i,j} C_{k,i} C_{l,k} C_{j,l} | X_j, X_k]] \\ &= \mathbb{E} [C_{1,4} C_{4,2} C_{3,1} C_{2,3}] \\ &< +\infty \quad \text{under } (\mathcal{A}_{\varphi, Mmt}). \end{aligned}$$

By combining these results, we obtain that:

$$\mathbb{E} [A_{n,4}^2] \leq C \frac{n^3}{n^2} \frac{\mathbb{E} [(\mathbb{E} [C_{3,1} C_{2,3} | X_1, X_2])^2]}{n(n-1)} + C' \frac{n^4}{n^2} \frac{(n-4)!}{n!} \mathbb{E} [C_{1,4} C_{4,2} C_{3,1} C_{2,3}] \xrightarrow{n \rightarrow \infty} 0.$$

From (50), and the above results of convergence towards 0 for $\mathbb{E} [A_{n,1}^2]$, $\mathbb{E} [A_{n,2}^2]$, $\mathbb{E} [A_{n,3}^2]$, and $\mathbb{E} [A_{n,4}^2]$, we firstly derive that:

$$\text{Var} \left(\sum_{i=2}^n (\mathbb{E} [Y_{n,i}^2 | \mathcal{F}_{n,i-1}]) \right) \xrightarrow{n \rightarrow \infty} 0.$$

Second point.

Notice that

$$\begin{aligned}\mathbb{E} [Y_{n,i} Y'_{n,i} | \mathcal{F}_{n,i-1}] &= \frac{1}{n} \sum_{j,k=1}^{i-1} \mathbb{E} [(\mathbb{1}_{\{\Pi_n(i)=j\}} C_{i,j} + \mathbb{1}_{\{\Pi_n(j)=i\}} C_{j,i}) (\mathbb{1}_{\{\Pi'_n(i)=k\}} C_{i,k} + \mathbb{1}_{\{\Pi'_n(k)=i\}} C_{k,i}) | \mathcal{F}_{n,i-1}] \\ &= B_{n,1} + B_{n,2} + B_{n,3} + B_{n,4},\end{aligned}$$

with

$$\begin{aligned}B_{n,1} &= \frac{1}{n} \sum_{1 \leq j < i \leq n} (\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi'_n(i)=j\}} \mathbb{E} [C_{i,j}^2 | X_j]), \\ B_{n,2} &= \frac{1}{n} \sum_{1 \leq j < i \leq n} (\mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{1}_{\{\Pi'_n(j)=i\}} \mathbb{E} [C_{j,i}^2 | X_j]), \\ B_{n,3} &= \frac{1}{n} \sum_{1 \leq j < i \leq n} ((\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi'_n(j)=i\}} + \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{1}_{\{\Pi'_n(i)=j\}}) \mathbb{E} [C_{i,j} C_{j,i} | X_j]), \\ B_{n,4} &= \frac{1}{n} \sum_{1 \leq j \neq k < i \leq n} \left(\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi'_n(i)=k\}} \mathbb{E} [C_{i,j} C_{i,k} | X_j, X_k] + \mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi'_n(k)=i\}} \mathbb{E} [C_{i,j} C_{k,i} | X_j, X_k] \right. \\ &\quad \left. + \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{1}_{\{\Pi'_n(i)=k\}} \mathbb{E} [C_{j,i} C_{i,k} | X_j, X_k] + \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{1}_{\{\Pi'_n(k)=i\}} \mathbb{E} [C_{j,i} C_{k,i} | X_j, X_k] \right).\end{aligned}$$

Thus,

$$\mathbb{E} \left[\left(\sum_{i=2}^n \mathbb{E} [Y_{n,i} Y'_{n,i} | \mathcal{F}_{n,i-1}] \right)^2 \right] \leq 4 (\mathbb{E} [B_{n,1}^2] + \mathbb{E} [B_{n,2}^2] + \mathbb{E} [B_{n,3}^2] + \mathbb{E} [B_{n,4}^2]). \quad (52)$$

Convergence of $\mathbb{E} [B_{n,1}^2]$ and $\mathbb{E} [B_{n,2}^2]$.

$$\begin{aligned}\mathbb{E} [B_{n,1}^2] &= \frac{1}{n^2} \sum_{1 \leq j < i \leq n} \sum_{1 \leq l < k \leq n} \mathbb{E} \left[(\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi'_n(i)=j\}} \mathbb{E} [C_{i,j}^2 | X_j]) (\mathbb{1}_{\{\Pi_n(k)=l\}} \mathbb{1}_{\{\Pi'_n(k)=l\}} \mathbb{E} [C_{k,l}^2 | X_l]) \right] \\ &= \frac{1}{n^2} \sum_{1 \leq j < i \leq n} \sum_{1 \leq l < k \leq n} \mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] \mathbb{E} [\mathbb{1}_{\{\Pi'_n(i)=j\}} \mathbb{1}_{\{\Pi'_n(k)=l\}}] \mathbb{E} [\mathbb{E} [C_{i,j}^2 | X_j] \mathbb{E} [C_{k,l}^2 | X_l]] \\ &\leq \frac{1}{n^3} \sum_{1 \leq j < i \leq n} \sum_{1 \leq l < k \leq n} \mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] \mathbb{E} [\mathbb{E} [C_{i,j}^2 | X_j] \mathbb{E} [C_{k,l}^2 | X_l]].\end{aligned}$$

Then, with the same computations as for the convergence of $\mathbb{E} [A_{n,1}^2]$ above, we prove that:

$$\mathbb{E} [B_{n,1}^2] \leq \frac{n-1}{2n^3} \mathbb{E} [(\mathbb{E} [C_{1,2}^2 | X_2])^2] + C \frac{\sigma_{h_\varphi, P^1 \otimes P^2}^4}{n-1} \xrightarrow{n \rightarrow \infty} 0$$

In the same way, we also prove that:

$$\mathbb{E} [B_{n,2}^2] \xrightarrow{n \rightarrow \infty} 0$$

Convergence of $\mathbb{E} [B_{n,3}^2]$.

$$\begin{aligned}\mathbb{E} [B_{n,3}^2] &\leq \frac{4}{n^2} \sum_{1 \leq j < i \leq n} \sum_{1 \leq l < k \leq n} \mathbb{E} \left[(\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi'_n(j)=i\}} \mathbb{E} [C_{i,j} C_{j,i} | X_j]) (\mathbb{1}_{\{\Pi_n(k)=l\}} \mathbb{1}_{\{\Pi'_n(l)=k\}} \mathbb{E} [C_{k,l} C_{l,k} | X_l]) \right] \\ &\leq \frac{4}{n^2} \sum_{1 \leq j < i \leq n} \sum_{1 \leq l < k \leq n} \mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] \mathbb{E} [\mathbb{1}_{\{\Pi'_n(j)=i\}} \mathbb{1}_{\{\Pi'_n(l)=k\}}] \mathbb{E} [\mathbb{E} [C_{i,j} C_{j,i} | X_j] \mathbb{E} [C_{k,l} C_{l,k} | X_l]].\end{aligned}$$

Now, with similar computations as for the convergence of $\mathbb{E}[A_{n,1}^2]$ above again, we prove that:

$$\mathbb{E}[B_{n,3}^2] \leq 2 \frac{n-1}{n^3} \mathbb{E}[(\mathbb{E}[C_{1,2}C_{2,1}|X_2])^2] + C \frac{(\mathbb{E}[C_{1,2}C_{2,1}])^2}{n-1} \xrightarrow{n \rightarrow \infty} 0.$$

Convergence of $\mathbb{E}[B_{n,4}^2]$.

Setting

$$B_{n,4,1} = \frac{1}{n} \sum_{1 \leq j \neq k < i \leq n} \mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi'_n(i)=k\}} \mathbb{E}[C_{i,j}C_{i,k}|X_j, X_k],$$

$$B_{n,4,2} = \frac{1}{n} \sum_{1 \leq j \neq k < i \leq n} \mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi'_n(k)=i\}} \mathbb{E}[C_{i,j}C_{k,i}|X_j, X_k],$$

$$B_{n,4,3} = \frac{1}{n} \sum_{1 \leq j \neq k < i \leq n} \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{1}_{\{\Pi'_n(i)=k\}} \mathbb{E}[C_{j,i}C_{i,k}|X_j, X_k],$$

$$B_{n,4,4} = \frac{1}{n} \sum_{1 \leq j \neq k < i \leq n} \mathbb{1}_{\{\Pi_n(j)=i\}} \mathbb{1}_{\{\Pi'_n(k)=i\}} \mathbb{E}[C_{j,i}C_{k,i}|X_j, X_k],$$

then $B_{n,4} = B_{n,4,1} + B_{n,4,2} + B_{n,4,3} + B_{n,4,4}$.

$$\begin{aligned} \mathbb{E}[B_{n,4,1}^2] &= \frac{1}{n^2} \sum_{1 \leq j \neq k < i \leq n} \sum_{1 \leq p \neq q < l \leq n} \mathbb{E}[\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(l)=p\}}] \mathbb{E}[\mathbb{1}_{\{\Pi'_n(i)=k\}} \mathbb{1}_{\{\Pi'_n(l)=q\}}] \times \\ &\quad \mathbb{E}[\mathbb{E}[C_{i,j}C_{i,k}|X_j, X_k] \mathbb{E}[C_{l,p}C_{l,q}|X_p, X_q]] . \end{aligned}$$

Let us consider all the cases where $1 \leq j \neq k < i \leq n$ and $1 \leq p \neq q < l \leq n$.

If $\#\{j, k, p, q\} = 3$, using a similar argument as in (51), we obtain that:

$$\mathbb{E}[\mathbb{E}[C_{i,j}C_{i,k}|X_j, X_k] \mathbb{E}[C_{l,p}C_{l,q}|X_p, X_q]] = 0.$$

If $j = p$, $k = q$, and $i = l$, then,

$$\mathbb{E}[\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(l)=p\}}] \mathbb{E}[\mathbb{1}_{\{\Pi'_n(i)=k\}} \mathbb{1}_{\{\Pi'_n(l)=q\}}]] = \frac{1}{n^2},$$

and

$$\begin{aligned} \mathbb{E}[\mathbb{E}[C_{i,j}C_{i,k}|X_j, X_k] \mathbb{E}[C_{l,p}C_{l,q}|X_p, X_q]] &= \mathbb{E}[(\mathbb{E}[C_{i,j}C_{i,k}|X_j, X_k])^2] \\ &= \mathbb{E}[(\mathbb{E}[C_{3,1}C_{3,2}|X_1, X_2])^2] \\ &< +\infty \quad \text{under } (\mathcal{A}_{\varphi, Mmt}). \end{aligned}$$

If $j = p$, $k = q$, and $i \neq l$, or if $j = q$, $k = p$, and $i = l$, then $\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(l)=p\}} = 0$, so

$$\mathbb{E}[\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(l)=p\}}] \mathbb{E}[\mathbb{1}_{\{\Pi'_n(i)=k\}} \mathbb{1}_{\{\Pi'_n(l)=q\}}]] = 0.$$

If $j = q$, $k = p$, and $i \neq l$, then

$$\mathbb{E}[\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(l)=p\}}] \mathbb{E}[\mathbb{1}_{\{\Pi'_n(i)=k\}} \mathbb{1}_{\{\Pi'_n(l)=q\}}]] = \frac{1}{n^2(n-1)^2},$$

and

$$\begin{aligned} \mathbb{E}[\mathbb{E}[C_{i,j}C_{i,k}|X_j, X_k] \mathbb{E}[C_{l,p}C_{l,q}|X_p, X_q]] &= \mathbb{E}[\mathbb{E}[C_{i,j}C_{i,k}|X_j, X_k] \mathbb{E}[C_{l,k}C_{l,j}|X_j, X_k]] \\ &= \mathbb{E}[\mathbb{E}[C_{i,j}C_{i,k}C_{l,k}C_{l,j}|X_j, X_k]] \\ &= \mathbb{E}[C_{1,4}C_{1,3}C_{2,3}C_{2,4}] \\ &< +\infty \quad \text{under } (\mathcal{A}_{\varphi, Mmt}). \end{aligned}$$

By combining these results, we obtain that:

$$\mathbb{E} [B_{n,4,1}^2] \leq C \frac{\mathbb{E} \left[(\mathbb{E} [C_{3,1} C_{3,2} | X_1, X_2])^2 \right]}{n} + C' \frac{\mathbb{E} [C_{1,4} C_{1,3} C_{2,3} C_{2,4}]}{(n-1)^2} \xrightarrow{n \rightarrow \infty} 0.$$

Following the same lines of proof, we furthermore obtain that $\mathbb{E} [B_{n,4,2}^2]$, $\mathbb{E} [B_{n,4,3}^2]$, and $\mathbb{E} [B_{n,4,4}^2]$ also converge towards 0.

As a consequence,

$$\mathbb{E} [B_{n,4}^2] \xrightarrow{n \rightarrow \infty} 0.$$

From (52), and the convergence towards 0 for $\mathbb{E} [B_{n,1}^2]$, $\mathbb{E} [B_{n,2}^2]$, $\mathbb{E} [B_{n,3}^2]$, and $\mathbb{E} [B_{n,4}^2]$, we derive that:

$$\mathbb{E} \left[\left(\sum_{i=2}^n \mathbb{E} [Y_{n,i} Y'_{n,i} | \mathcal{F}_{n,i-1}] \right)^2 \right] \xrightarrow{n \rightarrow \infty} 0,$$

which finally allows to conclude that assumption (i) is satisfied.

Assumption (ii). Given $\varepsilon > 0$, let us prove that

$$\sum_{i=2}^n \mathbb{E} \left[(aY_{n,i} + bY'_{n,i})^2 \mathbb{1}_{\{|aY_{n,i} + bY'_{n,i}| > \varepsilon\}} \right] \xrightarrow{n \rightarrow \infty} 0.$$

Since $\mathbb{1}_{\{|aY_{n,i} + bY'_{n,i}| > \varepsilon\}} \leq \left(\frac{aY_{n,i} + bY'_{n,i}}{\varepsilon} \right)^2$,

$$\begin{aligned} \sum_{i=2}^n \mathbb{E} \left[(aY_{n,i} + bY'_{n,i})^2 \mathbb{1}_{\{|aY_{n,i} + bY'_{n,i}| > \varepsilon\}} \right] &\leq \frac{1}{\varepsilon^2} \sum_{i=2}^n \mathbb{E} \left[(aY_{n,i} + bY'_{n,i})^4 \right] \\ &\leq \frac{2^3}{\varepsilon^2} \sum_{i=2}^n \left(a^4 \mathbb{E} [Y_{n,i}^4] + b^4 \mathbb{E} [Y'_{n,i}{}^4] \right) \\ &\leq \frac{2^3(a^4 + b^4)}{\varepsilon^2} \sum_{i=2}^n \mathbb{E} [Y_{n,i}^4]. \end{aligned}$$

But it is easy to see that

$$Y_{n,i} = \frac{1}{\sqrt{n}} \left(C_{i, \Pi_n(i)} \mathbb{1}_{\{\Pi_n(i) < i\}} + C_{\Pi_n^{-1}(i), i} \mathbb{1}_{\{\Pi_n^{-1}(i) < i\}} \right).$$

Hence,

$$\begin{aligned} \mathbb{E} [Y_{n,i}^4] &\leq \frac{2^3}{n^2} \mathbb{E} \left[C_{i, \Pi_n(i)}^4 \mathbb{1}_{\{\Pi_n(i) < i\}} + C_{\Pi_n^{-1}(i), i}^4 \mathbb{1}_{\{\Pi_n^{-1}(i) < i\}} \right] \\ &\leq \frac{2^3}{n^2} \sum_{j=1}^{i-1} \left(\mathbb{E} [C_{i,j}^4 \mathbb{1}_{\{\Pi_n(i)=j\}}] + \mathbb{E} [C_{j,i}^4 \mathbb{1}_{\{\Pi_n^{-1}(i)=j\}}] \right) \\ &\leq \frac{2^4}{n^2} \mathbb{E} [C_{1,2}^4]. \end{aligned}$$

We thus obtain that:

$$\sum_{i=2}^n \mathbb{E} \left[(aY_{n,i} + bY'_{n,i})^2 \mathbb{1}_{\{|aY_{n,i} + bY'_{n,i}| > \varepsilon\}} \right] \leq \frac{2^7(a^4 + b^4)}{\varepsilon^2 n} \mathbb{E} [C_{1,2}^4],$$

where the right-hand side tends to 0 as soon as $\mathbb{E} [C_{1,2}^4] < +\infty$.

This last condition is ensured by $(\mathcal{A}_{\varphi, Mmt})$, which allows to check that assumption (ii) is also checked, and that

$$\mathcal{L} \left(aM_n^{\Pi_n}(\mathbb{X}_n) + bM_n^{\Pi'_n}(\mathbb{X}_n) \right) \Rightarrow \mathcal{N} \left(0, (a^2 + b^2) \sigma_{h_\varphi, P^1 \otimes P^2}^2 \right).$$

This ends the proof of Lemma 1.

□

Recall that we aim at proving that

$$d_{BL} \left(\mathcal{L} \left(M_n^{\Pi_n} (\mathbb{X}_n) | \mathbb{X}_n \right), \mathcal{N} \left(0, \sigma_{h_\varphi, P^1 \otimes P^2}^2 \right) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

From Lemma 1, we deduce that for every t in \mathbb{R} ,

$$\begin{cases} \mathbb{P} \left(M_n^{\Pi_n} (\mathbb{X}_n) \leq t \right) \xrightarrow[n \rightarrow +\infty]{} \Phi_{0, \sigma_{h_\varphi, P^1 \otimes P^2}^2}^2(t), \\ \mathbb{P} \left(M_n^{\Pi_n} (\mathbb{X}_n) \leq t, M_n^{\Pi'_n} (\mathbb{X}_n) \leq t \right) \xrightarrow[n \rightarrow +\infty]{} \Phi_{0, \sigma_{h_\varphi, P^1 \otimes P^2}^2}^2(t). \end{cases}$$

Setting $M_n = M_n^{\Pi_n} (\mathbb{X}_n)$, and $M'_n = M_n^{\Pi'_n} (\mathbb{X}_n)$ for the sake of simplicity, this leads to

$$\begin{cases} \mathbb{E} \left[\mathbb{E} \left[\mathbb{1}_{\{M_n \leq t\}} | \mathbb{X}_n \right] \right] \xrightarrow[n \rightarrow +\infty]{} \Phi_{0, \sigma_{h_\varphi, P^1 \otimes P^2}^2}^2(t), \\ \mathbb{E} \left[\left(\mathbb{E} \left[\mathbb{1}_{\{M_n \leq t\}} | \mathbb{X}_n \right] \right)^2 \right] \xrightarrow[n \rightarrow +\infty]{} \Phi_{0, \sigma_{h_\varphi, P^1 \otimes P^2}^2}^2(t). \end{cases} \quad (53)$$

It is well-known (see [14, Th. 9.2.1] for instance), that in a separable metric space, convergence in probability is metrizable, and therefore is equivalent to almost-sure convergence of a sub-sequence of any initial subsequence. So, let us fix an initial extraction $\phi_0 : \mathbb{N} \rightarrow \mathbb{N}$, which defines a sub-sequence $(M_{\phi_0(n)})_{n \in \mathbb{N}}$ of $(M_n)_{n \in \mathbb{N}}$.

Let us denote by $(q_m)_{m \in \mathbb{N}}$ a sequence such that $\{q_m, m \in \mathbb{N}\} = \mathbb{Q}$. For any m in \mathbb{N} , from (53), we derive that

$$\begin{cases} \mathbb{E} \left[\mathbb{E} \left[\mathbb{1}_{\{M_{\phi_0(n)} \leq q_m\}} | \mathbb{X}_{\phi_0(n)} \right] \right] \xrightarrow[n \rightarrow +\infty]{} \Phi_{0, \sigma_{h_\varphi, P^1 \otimes P^2}^2}^2(q_m), \\ \mathbb{E} \left[\left(\mathbb{E} \left[\mathbb{1}_{\{M_{\phi_0(n)} \leq q_m\}} | \mathbb{X}_{\phi_0(n)} \right] \right)^2 \right] \xrightarrow[n \rightarrow +\infty]{} \Phi_{0, \sigma_{h_\varphi, P^1 \otimes P^2}^2}^2(q_m), \end{cases}$$

which leads (by Chebychev's inequality) to

$$\mathbb{E} \left[\mathbb{1}_{\{M_{\phi_0(n)} \leq q_m\}} | \mathbb{X}_{\phi_0(n)} \right] \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \Phi_{0, \sigma_{h_\varphi, P^1 \otimes P^2}^2}^2(q_m). \quad (54)$$

Therefore, there exists an extraction ϕ_1 and a subset Ω_1 of Ω such that: $\mathbb{P}(\Omega_1) = 1$, and for every $\omega \in \Omega_1$,

$$\mathbb{E} \left[\mathbb{1}_{\{M_{\phi_1 \circ \phi_0(n)} \leq q_1\}} | \mathbb{X}_{\phi_1 \circ \phi_0(n)} \right] (\omega) \xrightarrow[n \rightarrow \infty]{} \Phi_{0, \sigma_{h_\varphi, P^1 \otimes P^2}^2}^2(q_1).$$

Now, let $m \geq 1$ for which there exists an extraction ϕ_m and a subset Ω_m of Ω such that: $\mathbb{P}(\Omega_m) = 1$, and for every $\omega \in \Omega_m$,

$$\mathbb{E} \left[\mathbb{1}_{\{M_{\phi_m \circ \phi_{m-1} \circ \dots \circ \phi_0(n)} \leq q_m\}} | \mathbb{X}_{\phi_m \circ \phi_{m-1} \circ \dots \circ \phi_0(n)} \right] (\omega) \xrightarrow[n \rightarrow \infty]{} \Phi_{0, \sigma_{h_\varphi, P^1 \otimes P^2}^2}^2(q_m).$$

Then, from (54), there also exist an extraction ϕ_{m+1} and a subset Ω_{m+1} of Ω such that: $\mathbb{P}(\Omega_{m+1}) = 1$, and for every $\omega \in \Omega_{m+1}$,

$$\mathbb{E} \left[\mathbb{1}_{\{M_{\phi_{m+1} \circ \phi_m \circ \phi_{m-1} \circ \dots \circ \phi_0(n)} \leq q_{m+1}\}} | \mathbb{X}_{\phi_{m+1} \circ \phi_m \circ \phi_{m-1} \circ \dots \circ \phi_0(n)} \right] (\omega) \xrightarrow[n \rightarrow \infty]{} \Phi_{0, \sigma_{h_\varphi, P^1 \otimes P^2}^2}^2(q_{m+1}).$$

Setting $\tilde{\Omega} = \bigcap_{m \in \mathbb{N}} \Omega_m$ and for every n in \mathbb{N} , $\tilde{\phi}(n) = \phi_n \circ \dots \circ \phi_2 \circ \phi_1(n)$, then $\mathbb{P}(\tilde{\Omega}) = 1$, and for every ω in $\tilde{\Omega}$, and m in \mathbb{N} ,

$$\mathbb{E} \left[\mathbb{1}_{\{M_{\tilde{\phi} \circ \phi_0(n)} \leq q_m\}} | \mathbb{X}_{\tilde{\phi} \circ \phi_0(n)} \right] (\omega) \xrightarrow[n \rightarrow \infty]{} \Phi_{0, \sigma_{h_\varphi, P^1 \otimes P^2}^2}^2(q_m).$$

Since $\Phi_{0, \sigma_{h_\varphi, P^1 \otimes P^2}^2}^2$ is a continuous distribution function, this means that:

$$d_{BL} \left(\mathcal{L} \left(M_{\tilde{\phi} \circ \phi_0(n)} | \mathbb{X}_{\tilde{\phi} \circ \phi_0(n)} \right), \mathcal{N} \left(0, \sigma_{h_\varphi, P^1 \otimes P^2}^2 \right) \right) \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

To conclude, we actually proved that

$$d_{BL} \left(\mathcal{L} \left(M_n^{\Pi_n} (\mathbb{X}_n) | \mathbb{X}_n \right), \mathcal{N} \left(0, \sigma_{h_\varphi, P^1 \otimes P^2}^2 \right) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

which, combined with (46), leads to:

$$d_{BL} \left(\mathcal{L} \left(\sqrt{n} U_{n, h_\varphi} (\mathbb{X}_n^{\Pi_n}) | \mathbb{X}_n \right), \mathcal{N} \left(0, \sigma_{h_\varphi, P^1 \otimes P^2}^2 \right) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

Third step: convergence of conditional second order moments.

Recall that from (42), $U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}) = \frac{1}{n-1}U_n^{\Pi_n}$, where

$$\begin{aligned} U_n^{\Pi_n} &= \sum_{i=1}^n \varphi(X_i^1, X_{\Pi_n(i)}^2) - \frac{1}{n} \sum_{i,j=1}^n \varphi(X_i^1, X_j^2) \\ &= \sum_{i,j=1}^n \left(\mathbb{1}_{\{\Pi_n(i)=j\}} - \frac{1}{n} \right) \varphi(X_i^1, X_j^2). \end{aligned}$$

Therefore,

$$\mathbb{E} \left[(\sqrt{n} U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n}))^2 \middle| \mathbb{X}_n \right] = \frac{n^2}{(n-1)^2} \left(\frac{1}{n} \mathbb{E} \left[(U_n^{\Pi_n})^2 \middle| \mathbb{X}_n \right] \right). \quad (55)$$

$$\begin{aligned} \frac{1}{n} \mathbb{E} \left[(U_n^{\Pi_n})^2 \middle| \mathbb{X}_n \right] &= \frac{1}{n} \sum_{i,j=1}^n \sum_{k,l=1}^n \mathbb{E} \left[\left(\mathbb{1}_{\{\Pi_n(i)=j\}} - \frac{1}{n} \right) \left(\mathbb{1}_{\{\Pi_n(k)=l\}} - \frac{1}{n} \right) \varphi(X_i^1, X_j^2) \varphi(X_k^1, X_l^2) \right] \\ &= \frac{1}{n} \sum_{i,j=1}^n \sum_{k,l=1}^n \left(\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] - \frac{1}{n^2} \right) \varphi(X_i^1, X_j^2) \varphi(X_k^1, X_l^2). \end{aligned}$$

Firstly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=4}} \left(\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] - \frac{1}{n^2} \right) \varphi(X_i^1, X_j^2) \varphi(X_k^1, X_l^2) = \frac{(n-2)(n-3)}{n^2} U_{n,1},$$

where

$$U_{n,1} = \frac{(n-4)!}{n!} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=4}} \varphi(X_i^1, X_j^2) \varphi(X_k^1, X_l^2),$$

is clearly a U -statistic of order 4. From the strong law of large numbers of Hoeffding [32], we thus have that

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=4}} \left(\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] - \frac{1}{n^2} \right) \varphi(X_i^1, X_j^2) \varphi(X_k^1, X_l^2) \xrightarrow[n \rightarrow \infty]{a.s.} (\mathbb{E} [\varphi(X_1^1, X_2^2)])^2.$$

Secondly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=3 \\ i=j, i=l, j=k, \text{ or } k=l}} \left(\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] - \frac{1}{n^2} \right) \varphi(X_i^1, X_j^2) \varphi(X_k^1, X_l^2) = \frac{2(n-2)}{n^2} U_{n,2},$$

where

$$U_{n,2} = \frac{(n-3)!}{n!} \sum_{\substack{i,k,l \in \{1,\dots,n\} \\ \#\{i,k,l\}=3}} (\varphi(X_i^1, X_i^2) \varphi(X_k^1, X_l^2) + \varphi(X_i^1, X_l^2) \varphi(X_k^1, X_i^2)),$$

is a U -statistic of order 3, so $\frac{2(n-2)}{n^2} U_{n,2} \xrightarrow[n \rightarrow \infty]{a.s.} 0$.

Thirdly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=3 \\ i=k, \text{ or } j=l}} \left(\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] - \frac{1}{n^2} \right) \varphi(X_i^1, X_j^2) \varphi(X_k^1, X_l^2) = -\frac{n(n-1)(n-2)}{n^3} U_{n,3},$$

where

$$U_{n,3} = \frac{(n-3)!}{n!} \sum_{\substack{i,k,l \in \{1,\dots,n\} \\ \#\{i,k,l\}=3}} (\varphi(X_i^1, X_k^2) \varphi(X_i^1, X_l^2) + \varphi(X_i^1, X_l^2) \varphi(X_k^1, X_l^2)),$$

is a U -statistic of order 3. So,

$$-\frac{n(n-1)(n-2)}{n^3}U_{n,3} \xrightarrow[n \rightarrow \infty]{a.s.} -\mathbb{E} \left[\left(\mathbb{E} [\varphi(X_1^1, X_2^2) | X_1] \right)^2 \right] - \mathbb{E} \left[\left(\mathbb{E} [\varphi(X_1^1, X_2^2) | X_2] \right)^2 \right].$$

Fourthly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=2 \\ i=j=k, i=j=l, \\ i=k=l, \text{ or } j=k=l}} \left(\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] - \frac{1}{n^2} \right) \varphi(X_i^1, X_j^2) \varphi(X_k^1, X_l^2) = -\frac{2(n-1)}{n^2} U_{n,4},$$

where

$$U_{n,4} = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} (\varphi(X_i^1, X_i^2) \varphi(X_i^1, X_j^2) + \varphi(X_i^1, X_i^2) \varphi(X_j^1, X_j^2)),$$

is a U -statistic of order 2, so $-\frac{2(n-1)}{n^2} U_{n,4} \xrightarrow[n \rightarrow \infty]{a.s.} 0$.

Fifthly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=2 \\ i=j \neq k=l, \text{ or } i=l \neq j=k}} \left(\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] - \frac{1}{n^2} \right) \varphi(X_i^1, X_j^2) \varphi(X_k^1, X_l^2) = \frac{1}{n^2} U_{n,5},$$

where

$$U_{n,5} = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} (\varphi(X_i^1, X_i^2) \varphi(X_j^1, X_j^2) + \varphi(X_i^1, X_j^2) \varphi(X_j^1, X_i^2)),$$

is a U -statistic of order 2, so $\frac{1}{n^2} U_{n,5} \xrightarrow[n \rightarrow \infty]{a.s.} 0$.

Sixthly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=2 \\ i=k \neq j=l}} \left(\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] - \frac{1}{n^2} \right) \varphi(X_i^1, X_j^2) \varphi(X_k^1, X_l^2) = \frac{(n-1)^2}{n^2} U_{n,6},$$

where

$$U_{n,6} = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \varphi^2(X_i^1, X_j^2),$$

is a U -statistic of order 2, so $\frac{(n-1)^2}{n^2} U_{n,6} \xrightarrow[n \rightarrow \infty]{a.s.} \mathbb{E} [\varphi^2(X_i^1, X_j^2)]$.

Seventhly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=1}} \left(\mathbb{E} [\mathbb{1}_{\{\Pi_n(i)=j\}} \mathbb{1}_{\{\Pi_n(k)=l\}}] - \frac{1}{n^2} \right) \varphi(X_i^1, X_j^2) \varphi(X_k^1, X_l^2) = \frac{n-1}{n^3} \sum_{i=1}^n \varphi(X_i^1, X_i^2),$$

which almost-surely tends to 0 thanks to the strong law of large numbers.

By combining all these results, we finally obtain that:

$$\frac{1}{n} \mathbb{E} \left[(U_n^{\Pi_n})^2 | \mathbb{X}_n \right] \xrightarrow[n \rightarrow \infty]{a.s.} \mathbb{E} [\varphi(X_1^1, X_2^2)]^2 - \mathbb{E} \left[\left(\mathbb{E} [\varphi(X_1^1, X_2^2) | X_1] \right)^2 \right] - \mathbb{E} \left[\left(\mathbb{E} [\varphi(X_1^1, X_2^2) | X_2] \right)^2 \right] + \mathbb{E} [\varphi^2(X_i^1, X_j^2)].$$

Noticing that

$$(\mathbb{E} [\varphi(X_1^1, X_2^2)])^2 - \mathbb{E} \left[\left(\mathbb{E} [\varphi(X_1^1, X_2^2) | X_1] \right)^2 \right] - \mathbb{E} \left[\left(\mathbb{E} [\varphi(X_1^1, X_2^2) | X_2] \right)^2 \right] + \mathbb{E} [\varphi^2(X_i^1, X_j^2)] = \sigma_{h_\varphi, P^1 \otimes P^2}^2,$$

from (55), we deduce that

$$\mathbb{E} \left[(\sqrt{n} U_{n, h_\varphi}(\mathbb{X}_n^{\Pi_n}))^2 | \mathbb{X}_n \right] \xrightarrow[n \rightarrow \infty]{a.s.} \sigma_{h_\varphi, P^1 \otimes P^2}^2.$$

Since

$$d_{BL} \left(\mathcal{L}(\sqrt{n} U_{n, h_\varphi}(\mathbb{X}_n^{\Pi_n}) | \mathbb{X}_n), \mathcal{N}(0, \sigma_{h_\varphi, P^1 \otimes P^2}^2) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

this allows to conclude that:

$$d_2 \left(\mathcal{L}(\sqrt{n} U_{n, h_\varphi}(\mathbb{X}_n^{\Pi_n}) | \mathbb{X}_n), \mathcal{N}(0, \sigma_{h_\varphi, P^1 \otimes P^2}^2) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

5.10 Proof of Corollary 2

Let us fix η in $(0, 1)$, and recall that we aim at proving that, under the same assumptions as in Theorem 3,

$$q_{n,\varphi,\eta}^*(\mathbb{X}_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \Phi_{0,\sigma_{h,P^1 \otimes P^2}^2}^{-1}(\eta),$$

where $q_{n,\varphi,\eta}^*(\mathbb{X}_n)$ denotes the η -quantile of $\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n^{\Pi_n})$ given \mathbb{X}_n .

Here, unlikely for the bootstrap approach, we only have in Theorem 3 a consistency result in probability. Thus, we use the same argument of sub-sequences as in the proof of the Theorem itself, in order to come down to an almost-sure convergence.

So let us consider an extraction $\phi_0 : \mathbb{N} \rightarrow \mathbb{N}$ which defines a sub-sequence.

Then, by Theorem 3, there exists an extraction ϕ_1 such that

$$d_{BL} \left(\mathcal{L} \left(\sqrt{\phi_1 \circ \phi_0(n)} U_{\phi_1 \circ \phi_0(n), h_\varphi} \left(\mathbb{X}_{\phi_1 \circ \phi_0(n)}^{\Pi_{\phi_1 \circ \phi_0(n)}} \right) \middle| \mathbb{X}_{\phi_1 \circ \phi_0(n)} \right), \mathcal{N} \left(0, \sigma_{h_\varphi, P^1 \otimes P^2}^2 \right) \right) \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (56)$$

In particular, applying [59, Lemma 21.2] on the event where the convergence above is true, we directly obtain that

$$q_{\phi_1 \circ \phi_0(n), \varphi, \eta}^*(\mathbb{X}_{\phi_1 \circ \phi_0(n)}) \xrightarrow[n \rightarrow +\infty]{a.s.} \Phi_{0,\sigma_{h_\varphi, P^1 \otimes P^2}^2}^{-1}(\eta),$$

which ends the proof.

5.11 Proof of Theorem 4

As for the bootstrap approach, we focus on the case $\Delta^* = \Delta_{\varphi, \alpha}^+(X_n)$, the proof in the other cases being similar. Given α in $(0, 1)$, we deduce from Corollary 2 that

$$q_{n,\varphi,1-\alpha}^*(\mathbb{X}_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \Phi_{0,\sigma_{h_\varphi, P^1 \otimes P^2}^2}^{-1}(1-\alpha) \quad (57)$$

Then, from Proposition 4 and Slutsky's lemma, under (H_0) , $(\sqrt{n}U_{n,h}(\mathbb{X}_n), q_{n,\varphi,1-\alpha}^*(\mathbb{X}_n))$ converges in distribution towards $(Z, \Phi_{0,\sigma_{h_\varphi, P^1 \otimes P^2}^2}^{-1}(1-\alpha))$, where $Z \sim \mathcal{N}(0, \sigma_{h_\varphi, P^1 \otimes P^2}^2)$. Therefore, under (H_0) ,

$$\mathbb{P} \left(\sqrt{n}U_{n,h_\varphi}(\mathbb{X}_n) > q_{n,\varphi,1-\alpha}^*(\mathbb{X}_n) \right) \xrightarrow[n \rightarrow +\infty]{} \alpha.$$

As for the consistency under (H_1) , we can argue exactly in the same way as for the bootstrap, to end the proof.

5.12 Proof of Proposition 8

In order to obtain the asymptotic results in Proposition 8, it is sufficient to prove that

$$\sqrt{n}U^*(\lceil (1-\alpha)(B_n+1) \rceil) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \Phi_{0,\sigma_{h,P^1 \otimes P^2}^2}^{-1}(1-\alpha). \quad (58)$$

Then, we conclude using exactly the same arguments as in the proof of Theorem 4.

To obtain (58), we follow the proof of [59, Lemma 21.2].

Let us introduce some more notations.

- Let F_{n,\mathbb{X}_n}^* be the conditional cumulative distribution function (c.d.f) of $\mathcal{L}(\sqrt{n}U_{n,h_\varphi}, P_n^* | \mathbb{X}_n)$, and let us first show that

$$\sup_{z \in \mathbb{R}} \left| F_{n,\mathbb{X}_n}^*(z) - \Phi_{0,\sigma_{h,P^1 \otimes P^2}^2}(z) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0. \quad (59)$$

As Theorem 3 provides only a convergence in probability, similar arguments of subsequences as in the proof of Corollary 2, have to be used. Let ϕ_0 be an initial extraction and ϕ_1 be the extraction such that (56) is satisfied. As convergence in the d_{BL} metric is equivalent to a weak convergence, and as the limit is continuous, by [59, Lemma 2.11] we obtain that $\sup_{z \in \mathbb{R}} \left| F_{\phi_1 \circ \phi_0(n), \mathbb{X}_{\phi_1 \circ \phi_0(n)}}^*(z) - \Phi_{0,\sigma_{h,P^1 \otimes P^2}^2}(z) \right| \xrightarrow[n \rightarrow \infty]{a.s.} 0$. This being true for any initial subsequence ϕ_0 , we obtain (59).

- Let $F_{n, \mathbb{X}_n}^{\star B_n}$ denote the empirical c.d.f of $\mathcal{L}(\sqrt{n}U_{n, h_\varphi}, P_n^\star | \mathbb{X}_n)$ associated to the sample $(\Pi_n^1, \dots, \Pi_n^{B_n})$, that is

$$\forall z \in \mathbb{R}, \quad F_{n, \mathbb{X}_n}^{\star B_n}(z) = \frac{1}{B_n} \sum_{b=1}^{B_n} \mathbb{1}_{\{\sqrt{n}U_{n, h_\varphi}(\mathbb{X}_n^{\Pi_n^b}) \leq z\}}.$$

Then, by the DKW inequality (see [59] p.268 for instance), we obtain as in the proof of Proposition 5,

$$\sup_{z \in \mathbb{R}} \left| F_{n, \mathbb{X}_n}^{\star B_n}(z) - F_{n, \mathbb{X}_n}^\star(z) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0. \quad (60)$$

- Finally, let

$$G_{n, \mathbb{X}_n}^{\star B_n}(z) = \frac{1}{B_n + 1} \sum_{b=1}^{B_n + 1} \mathbb{1}_{\{\sqrt{n}U^{\star b} \leq z\}}.$$

Since $G_{n, \mathbb{X}_n}^{\star B_n}(z) = \frac{1}{B_n + 1} \left(\mathbb{1}_{\{\sqrt{n}U_{n, h_\varphi}(\mathbb{X}_n) \leq z\}} + B_n F_{n, \mathbb{X}_n}^{\star B_n}(z) \right)$,

$$\sup_{z \in \mathbb{R}} \left| G_{n, \mathbb{X}_n}^{\star B_n}(z) - F_{n, \mathbb{X}_n}^{\star B_n}(z) \right| \leq \frac{2}{B_n + 1} \xrightarrow[n \rightarrow \infty]{a.s.} 0 \quad (61)$$

Finally, combining (59), (60) and (61), we obtain that

$$\sup_{z \in \mathbb{R}} \left| G_{n, \mathbb{X}_n}^{\star B_n}(z) - \Phi_{0, \sigma_{h, P^1 \otimes P^2}^2}(z) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0. \quad (62)$$

Since

$$\sqrt{n}U^{\star(\lceil(1-\alpha)(B_n+1)\rceil)} = \left(G_{n, \mathbb{X}_n}^{\star B_n} \right)^{-1}(1-\alpha),$$

we conclude in the same way as in the proof of Proposition 5.

5.13 Proof of Proposition 9

Under (H_0) , from Proposition 6, we deduce that $\mathbb{X}_n^{\Pi_n^1}, \dots, \mathbb{X}_n^{\Pi_n^B}, \mathbb{X}_n$, and hence also $U^{\star 1}, \dots, U^{\star(B+1)}$, are exchangeable real-valued random variables. Then,

$$\begin{aligned} \mathbb{P}\left(U^{\star(B+1)} > U^{\star(\lceil(1-\alpha)(B+1)\rceil)}\right) &= \mathbb{P}\left(\sum_{b=1}^{B+1} \mathbb{1}_{\{U^{\star b} < U^{\star(B+1)}\}} \geq (B+1) - \lfloor \alpha(B+1) \rfloor\right) \\ &= \mathbb{P}\left(\sum_{b=1}^{B+1} \mathbb{1}_{\{U^{\star b} \geq U^{\star(B+1)}\}} \leq \lfloor \alpha(B+1) \rfloor\right) \\ &= \mathbb{P}\left(\sum_{b=1}^{B+1} \mathbb{1}_{\{U^{\star b} \geq U^{\star(B+1)}\}} \leq \alpha(B+1)\right) \end{aligned}$$

By exchangeability of $U^{\star 1}, \dots, U^{\star(B+1)}$, applying Lemma 1 in [51], whose proof is given in [2], we finally obtain that:

$$\mathbb{P}\left(U^{\star(B+1)} > U^{\star(\lceil(1-\alpha)(B+1)\rceil)}\right) \leq \alpha.$$

5.14 Proof of Proposition 1

Let us prove that $h = h_{\varphi_\delta^{\text{c}o\text{inc}}}$ is continuous for the topology induced by d in any $(x_0, y_0) \in \mathcal{C}_\delta$ satisfying

$$(\{x_0^1\} \cup \{y_0^1\}) \cap (\{x_0^2 \pm \delta\} \cup \{y_0^2 \pm \delta\}) = \emptyset$$

Consider a sequence $\{(x_n, y_n)\}_{n \in \mathbb{N}}$ of elements of $\mathcal{X}^2 \times \mathcal{X}^2$, where $x_n = (x_n^1, x_n^2)$ and $y_n = (y_n^1, y_n^2)$ such that $d((x_n, y_n), (x_0, y_0)) \xrightarrow[n \rightarrow \infty]{} 0$ and (x_0, y_0) belongs to \mathcal{C}_δ .

We want to show that $|h(x_n, y_n) - h(x_0, y_0)| \xrightarrow[n \rightarrow \infty]{} 0$.

For an element z in \mathcal{X} , we denote by N_z the count function defined by:

$$N_z(t) = \int \mathbb{1}_{\{u \leq t\}} dN_z(u).$$

Since (x_0, y_0) is in \mathcal{C}_δ , for any t_0 in $\{x_0^2 \pm \delta\} \cup \{y_0^2 \pm \delta\}$, $t_0 \notin x_0^1$, which means that $N_{x_0^1}$ is continuous in t_0 :

$$\exists \eta_{t_0} > 0 / \forall t \in [0, 1], \quad |t - t_0| \leq \eta_{t_0} \quad \Rightarrow \quad N_{x_0^1}(t) = N_{x_0^1}(t_0).$$

As $\{x_0^2 \pm \delta\} \cup \{y_0^2 \pm \delta\}$ is finite, $\eta_{x_0^1} = \min_{t_0 \in \{x_0^2 \pm \delta\} \cup \{y_0^2 \pm \delta\}} \eta_{t_0} > 0$ is well defined, and satisfies

$$\forall u \in \{x_0^2 \pm \delta\} \cup \{y_0^2 \pm \delta\}, \quad \forall t \in [0, 1], \quad |t - u| \leq \eta_{x_0^1} \quad \Rightarrow \quad N_{x_0^1}(t) = N_{x_0^1}(u).$$

By the same argument using continuity of $N_{y_0^1}$ over $\{x_0^2 \pm \delta\} \cup \{y_0^2 \pm \delta\}$, we can find $\eta_{y_0^1} > 0$ such that

$$\forall u \in \{x_0^2 \pm \delta\} \cup \{y_0^2 \pm \delta\}, \quad \forall t \in [0, 1], \quad |t - u| \leq \eta_{y_0^1} \quad \Rightarrow \quad N_{y_0^1}(t) = N_{y_0^1}(u).$$

Since saying that (x_0, y_0) belongs to \mathcal{C}_δ is equivalent to saying that $(\{x_0^2\} \cup \{y_0^2\}) \cap (\{x_0^1 \pm \delta\} \cup \{y_0^1 \pm \delta\}) = \emptyset$, we can construct $\eta_{x_0^2}$ and $\eta_{y_0^2}$ satisfying

$$\forall u \in \{x_0^1 \pm \delta\} \cup \{y_0^1 \pm \delta\}, \quad \forall t \in [0, 1], \quad \begin{cases} |t - u| \leq \eta_{x_0^2} & \Rightarrow N_{x_0^2}(t) = N_{x_0^2}(u), \\ |t - u| \leq \eta_{y_0^2} & \Rightarrow N_{y_0^2}(t) = N_{y_0^2}(u). \end{cases}$$

Finally, if $\eta = \min \{\eta_{x_0^1}, \eta_{y_0^1}, \eta_{x_0^2}, \eta_{y_0^2}\} > 0$,

$$\forall s \in \{x_0^2 \pm \delta\} \cup \{y_0^2 \pm \delta\}, \quad \forall t \in [0, 1], \quad |t - s| \leq \eta \quad \Rightarrow \quad \begin{cases} N_{x_0^1}(t) = N_{x_0^1}(s), \\ N_{y_0^1}(t) = N_{y_0^1}(s), \end{cases} \quad (63)$$

and

$$\forall s \in \{x_0^1 \pm \delta\} \cup \{y_0^1 \pm \delta\}, \quad \forall t \in [0, 1], \quad |t - s| \leq \eta \quad \Rightarrow \quad \begin{cases} N_{x_0^2}(t) = N_{x_0^2}(s), \\ N_{y_0^2}(t) = N_{y_0^2}(s). \end{cases} \quad (64)$$

As $d((x_n, y_n), (x_0, y_0)) \xrightarrow{n \rightarrow \infty} 0$, there exists some $n_0 \geq 0$ such that for any n exceeding n_0 ,

$$d((x_n, y_n), (x_0, y_0)) \leq \frac{\eta}{4}.$$

From the definition of d , we deduce that:

$$\exists \lambda_n^1 \in \Lambda / \begin{cases} \sup_{t \in [0, 1]} |\lambda_n^1(t) - t| \leq \frac{\eta}{4}, & (1-i) \\ \sup_{t \in [0, 1]} |N_{x_n^1}(t) - N_{x_0^1}(\lambda_n^1(t))| \leq \frac{\eta}{4}, & (1-ii) \end{cases}$$

and

$$\exists \lambda_n^2 \in \Lambda / \begin{cases} \sup_{t \in [0, 1]} |\lambda_n^2(t) - t| \leq \frac{\eta}{4}, & (2-i) \\ \sup_{t \in [0, 1]} |N_{x_n^2}(t) - N_{x_0^2}(\lambda_n^2(t))| \leq \frac{\eta}{4}. & (2-ii) \end{cases}$$

Notice that similar results occur for y_n and y_0 , but there are not detailed here since we will not use them explicitly.

By definition of h ,

$$\begin{aligned} h(x_n, y_n) - h(x_0, y_0) &= \frac{1}{2} \iint \mathbb{1}_{\{|u-v| \leq \delta\}} \{dN_{x_n^1} dN_{x_n^2} + dN_{y_n^1} dN_{y_n^2} - dN_{x_n^1} dN_{y_n^2} - dN_{y_n^1} dN_{x_n^2}\}(u, v) \\ &- \frac{1}{2} \iint \mathbb{1}_{\{|u-v| \leq \delta\}} \{dN_{x_0^1} dN_{x_0^2} + dN_{y_0^1} dN_{y_0^2} - dN_{x_0^1} dN_{y_0^2} - dN_{y_0^1} dN_{x_0^2}\}(u, v). \end{aligned}$$

Then,

$$\begin{aligned}
h(x_n, y_n) - h(x_0, y_0) &= \frac{1}{2} \iint \mathbb{1}_{\{|u-v| \leq \delta\}} \left(dN_{x_n^1}(u) \left(dN_{x_n^2} - dN_{x_0^2} \right)(v) + dN_{y_n^1}(u) \left(dN_{y_n^2} - dN_{y_0^2} \right)(v) \right. \\
&\quad - dN_{x_n^1}(u) \left(dN_{y_n^2} - dN_{y_0^2} \right)(v) - dN_{y_n^1}(u) \left(dN_{x_n^2} - dN_{x_0^2} \right)(v) \\
&\quad + \left(dN_{x_n^1} - dN_{x_0^1} \right)(u) dN_{x_0^2}(v) + \left(dN_{y_n^1} - dN_{y_0^1} \right)(u) dN_{y_0^2}(v) \\
&\quad \left. - \left(dN_{x_n^1} - dN_{x_0^1} \right)(u) dN_{y_0^2}(v) + \left(dN_{y_n^1} - dN_{y_0^1} \right)(u) dN_{x_0^2}(v) \right). \tag{65}
\end{aligned}$$

By symmetry of the problem, we just need to study the terms $A_n := \iint \mathbb{1}_{\{|u-v| \leq \delta\}} \left(dN_{x_n^1} - dN_{x_0^1} \right)(u) dN_{x_0^2}(v)$, and $B_n := \iint \mathbb{1}_{\{|u-v| \leq \delta\}} dN_{x_n^1}(u) \left(dN_{x_n^2} - dN_{x_0^2} \right)(v)$.

Study of A_n .

$$\begin{aligned}
A_n &= \iint \mathbb{1}_{\{|u-v| \leq \delta\}} \left(dN_{x_n^1} - dN_{x_0^1} \right)(u) dN_{x_0^2}(v) \\
&= \iint (\mathbb{1}_{\{u \leq v+\delta\}} - \mathbb{1}_{\{u < v-\delta\}}) \left(dN_{x_n^1} - dN_{x_0^1} \right)(u) dN_{x_0^2}(v) \\
&= \iint \mathbb{1}_{\{u \leq v+\delta\}} \left(dN_{x_n^1} - dN_{x_0^1} \right)(u) dN_{x_0^2}(v) - \iint \mathbb{1}_{\{u < v-\delta\}} \left(dN_{x_n^1} - dN_{x_0^1} \right)(u) dN_{x_0^2}(v)
\end{aligned}$$

We have that:

$$\begin{aligned}
\left| \iint \mathbb{1}_{\{u \leq v+\delta\}} \left(dN_{x_n^1} - dN_{x_0^1} \right)(u) dN_{x_0^2}(v) \right| &= \left| \int \left(N_{x_n^1}(v+\delta) - N_{x_0^1}(v+\delta) \right) dN_{x_0^2}(v) \right| \\
&\leq \sum_{T \in x_0^2} \left| N_{x_n^1}(T+\delta) - N_{x_0^1}(T+\delta) \right| \\
&\leq \sum_{T \in x_0^2} \left| N_{x_n^1}(T+\delta) - N_{x_0^1}(\lambda_n^1(T+\delta)) \right| \\
&\quad + \sum_{T \in x_0^2} \left| N_{x_0^1}(\lambda_n^1(T+\delta)) - N_{x_0^1}(T+\delta) \right|.
\end{aligned}$$

Now, using the notation $N_{x_i^1}^-(t) = \int \mathbb{1}_{\{u < t\}} dN_{x_i^1}(u)$,

$$\left| \iint \mathbb{1}_{\{u < v-\delta\}} \left(dN_{x_n^1} - dN_{x_0^1} \right)(u) dN_{x_0^2}(v) \right| \leq \sum_{T \in x_0^2} \left| N_{x_n^1}^-(T-\delta) - N_{x_0^1}^-(T-\delta) \right|.$$

Therefore,

$$\begin{aligned}
|A_n| &\leq \sum_{T \in x_0^2} \left(\left| N_{x_n^1}(T+\delta) - N_{x_0^1}(\lambda_n^1(T+\delta)) \right| + \left| N_{x_0^1}(\lambda_n^1(T+\delta)) - N_{x_0^1}(T+\delta) \right| \right. \\
&\quad \left. + \left| N_{x_n^1}^-(T-\delta) - N_{x_0^1}^-(T-\delta) \right| \right). \tag{66}
\end{aligned}$$

Let us study individually each term in the sum.

Fix T in x_0^2 . By (1-ii),

$$\left| N_{x_n^1}(T+\delta) - N_{x_0^1}(\lambda_n^1(T+\delta)) \right| \leq \frac{\eta}{4} \leq \varepsilon. \tag{67}$$

From (1-i), we derive that $|\lambda_n^1(T+\delta) - (T+\delta)| \leq \frac{\eta}{2} \leq \eta$ which, with (63), implies that

$$\left| N_{x_0^1}(\lambda_n^1(T+\delta)) - N_{x_0^1}(T+\delta) \right| = 0. \tag{68}$$

As $N_{x_n}^-(T - \delta) = \lim_{\substack{u \rightarrow T - \delta \\ u < T - \delta}} N_{x_n}^1(u)$, there exists $u_T \in [T - \delta - \eta/4, T - \delta[$ such that

$$\left| N_{x_n}^-(T - \delta) - N_{x_n}^1(u_T) \right| \leq \varepsilon,$$

so

$$\left| N_{x_n}^-(T - \delta) - N_{x_0}^-(T - \delta) \right| \leq \varepsilon + \left| N_{x_n}^1(u_T) - N_{x_0}^1(\lambda_n^1(u_T)) \right| + \left| N_{x_0}^1(\lambda_n^1(u_T)) - N_{x_0}^-(T - \delta) \right|. \quad (69)$$

From (1-ii), one has $\left| N_{x_n}^1(u_T) - N_{x_0}^1(\lambda_n^1(u_T)) \right| \leq \frac{\eta}{4} \leq \varepsilon$.

Then, by continuity of $N_{x_0}^1$ in $T - \delta$, first remark that $N_{x_0}^-(T - \delta) = N_{x_0}^1(T - \delta)$. Moreover, by (1-i) and construction of u_T ,

$$\left| \lambda_n^1(u_T) - (T - \delta) \right| \leq \left| \lambda_n^1(u_T) - u_T \right| + |u_T - (T - \delta)| \leq \frac{\eta}{4} + \frac{\eta}{4} < \eta,$$

hence, using (63), $\left| N_{x_0}^1(\lambda_n^1(u_T)) - N_{x_0}^-(T - \delta) \right| = 0$.

So finally, (69) gives

$$\left| N_{x_n}^-(T - \delta) - N_{x_0}^-(T - \delta) \right| \leq 2\varepsilon. \quad (70)$$

Combining (66), (67), (68), and (70), we obtain that:

$$|A_n| \leq 3\varepsilon \#x_0^2, \quad (71)$$

for any $n \geq n_0$.

Study of B_n . Recall that $B_n = \iint \mathbf{1}_{\{|u-v| \leq \delta\}} dN_{x_n}^1(u) \left(dN_{x_n}^2 - dN_{x_0}^2 \right) (v)$.

As for A_n , we can upper bound B_n by a sum of several terms, that can be studied separately.

$$\begin{aligned} B_n &= \iint \mathbf{1}_{\{u \leq v + \delta\}} dN_{x_n}^1(u) \left(dN_{x_n}^2 - dN_{x_0}^2 \right) (v) - \iint \mathbf{1}_{\{u < v - \delta\}} dN_{x_n}^1(u) \left(dN_{x_n}^2 - dN_{x_0}^2 \right) (v) \\ &= \sum_{T \in x_n^1} \left(N_{x_n}^2(T + \delta) - N_{x_0}^2(T + \delta) \right) - \sum_{T \in x_n^1} \left(N_{x_n}^-(T - \delta) - N_{x_0}^-(T - \delta) \right). \end{aligned}$$

So

$$B_n \leq |B_{n,1}| + |B_{n,2}| + |B_{n,3}| + |B_{n,4}|, \quad (72)$$

with:

$$\begin{aligned} B_{n,1} &= \sum_{T \in x_n^1} \left(N_{x_n}^2(T + \delta) - N_{x_0}^2(\lambda_n^2(T + \delta)) \right), \\ B_{n,2} &= \sum_{T \in x_n^1} \left(N_{x_0}^2(\lambda_n^2(T + \delta)) - N_{x_0}^2(T + \delta) \right), \\ B_{n,3} &= \sum_{T \in x_n^1} \left(N_{x_n}^-(T - \delta) - N_{x_0}^-(\lambda_n^2(T - \delta)) \right), \\ B_{n,4} &= \sum_{T \in x_n^1} \left(N_{x_0}^-(\lambda_n^2(T - \delta)) + N_{x_0}^-(T - \delta) \right). \end{aligned}$$

The control of B_n is quite similar as the one of A_n except that the sums are over $T \in x_n^1$ instead of $T \in x_0^1$, which prevents us to use (64) and (63) directly.

Control of $B_{n,1}$.

From (2-ii), we first deduce that $\left| N_{x_n}^2(T + \delta) - N_{x_0}^2(\lambda_n^2(T + \delta)) \right| \leq \varepsilon$, so

$$|B_{n,1}| \leq \varepsilon \#x_n^1. \quad (73)$$

Control of $B_{n,2}$.

One can easily see that:

$$\begin{aligned}
B_{n,2} &= \iint \mathbb{1}_{\{v \leq \lambda_n^2(u+\delta)\}} - \mathbb{1}_{\{v \leq u+\delta\}} dN_{x_0^2}(v) dN_{x_n^1}(u) \\
&= \iint \left(1 - \mathbb{1}_{\{u < (\lambda_n^2)^{-1}(v)-\delta\}}\right) - \left(1 - \mathbb{1}_{\{u < v-\delta\}}\right) dN_{x_0^2}(v) dN_{x_n^1}(u) \\
&= \sum_{T \in x_0^2} \left(N_{x_n^1}^-(T-\delta) - N_{x_n^1}^-((\lambda_n^2)^{-1}(T)-\delta) \right).
\end{aligned}$$

Fix now T in x_0^2 .

$$\left| N_{x_n^1}^-(T-\delta) - N_{x_n^1}^-((\lambda_n^2)^{-1}(T)-\delta) \right| \leq \left| N_{x_n^1}^-(T-\delta) - N_{x_0^1}(T-\delta) \right| + \left| N_{x_0^1}(T-\delta) - N_{x_n^1}^-((\lambda_n^2)^{-1}(T)-\delta) \right|.$$

As shown in (69), $\left| N_{x_n^1}^-(T-\delta) - N_{x_0^1}(T-\delta) \right| \leq 2\varepsilon$.

Furthermore, take $v_T \in \left[(\lambda_n^2)^{-1}(T) - \delta - \eta/4, (\lambda_n^2)^{-1}(T) - \delta \right]$ such that

$$\left| N_{x_n^1}^-((\lambda_n^2)^{-1}(T) - \delta) - N_{x_n^1}(v_T) \right| \leq \varepsilon.$$

So,

$$\left| N_{x_0^1}(T-\delta) - N_{x_n^1}^-((\lambda_n^2)^{-1}(T) - \delta) \right| \leq \varepsilon + \left| N_{x_n^1}(v_T) - N_{x_0^1}(\lambda_n^1(v_T)) \right| + \left| N_{x_0^1}(\lambda_n^1(v_T)) - N_{x_0^1}(T-\delta) \right|.$$

By construction of v_T and λ_n^1 (see (1-ii)), $\left| N_{x_n^1}(v_T) - N_{x_0^1}(\lambda_n^1(v_T)) \right| \leq \varepsilon$.

Because of (63) which is feasible as $\left| \lambda_n^1(v_T) - (T-\delta) \right| \leq \left| \lambda_n^1(v_T) - v_T \right| + |v_T - (T-\delta)| \leq \frac{\eta}{4} + \frac{\eta}{4} < \eta$ by (1-i) and the construction of v_T , $\left| N_{x_0^1}(\lambda_n^1(v_T)) - N_{x_0^1}(T-\delta) \right| = 0$.

Hence, $\left| N_{x_0^1}(T-\delta) - N_{x_n^1}^-((\lambda_n^2)^{-1}(T) - \delta) \right| \leq 2\varepsilon$.

Finally,

$$\left| N_{x_n^1}^-(T-\delta) - N_{x_n^1}^-((\lambda_n^2)^{-1}(T) - \delta) \right| \leq 4\varepsilon$$

and

$$|B_{n,2}| \leq 4\varepsilon \#x_0^2. \tag{74}$$

Control of $B_{n,3}$.

First, for all T in x_n^1 , we find some $\nu_{n,T} \in]0, \eta/4]$ such that

$$\forall u \in [T-\delta-\nu_{n,T}, T-\delta[, \quad \left| N_{x_n^2}^-(T-\delta) - N_{x_n^2}(u) \right| \leq \varepsilon.$$

Setting $\nu_n = \min_{T \in x_n^1} \nu_{n,T}$,

$$\begin{aligned}
|B_{n,3}| &\leq \sum_{T \in x_n^1} \left| N_{x_n^2}^-(T-\delta) - N_{x_n^2}(T-\delta-\nu_n) \right| + \sum_{T \in x_n^1} \left| N_{x_n^2}(T-\delta-\nu_n) - N_{x_0^2}(\lambda_n^2(T-\delta-\nu_n)) \right| \\
&\quad + \left| \sum_{T \in x_n^1} \left(N_{x_0^2}(\lambda_n^2(T-\delta-\nu_n)) - N_{x_0^2}(\lambda_n^2(T-\delta)) \right) \right|.
\end{aligned}$$

For each T , $\left| N_{x_n^2}^-(T-\delta) - N_{x_n^2}(T-\delta-\nu_n) \right| \leq \varepsilon$ and $\left| N_{x_n^2}(T-\delta-\nu_n) - N_{x_0^2}(\lambda_n^2(T-\delta-\nu_n)) \right| \leq \varepsilon$ by (2-ii). Therefore,

$$|B_{n,3}| \leq 2\varepsilon \#x_n^1 + \left| \sum_{T \in x_n^1} \left(N_{x_0^2}(\lambda_n^2(T-\delta-\nu_n)) - N_{x_0^2}(\lambda_n^2(T-\delta)) \right) \right|.$$

Now,

$$\begin{aligned}
& \sum_{T \in x_n^1} \left(N_{x_0^2} \left(\lambda_n^2 (T - \delta - \nu_n) \right) - N_{x_0^2}^- \left(\lambda_n^2 (T - \delta) \right) \right) \\
&= \iint \mathbf{1}_{\{v \leq \lambda_n^2(u - \delta - \nu_n)\}} - \mathbf{1}_{\{v < \lambda_n^2(u - \delta)\}} dN_{X_n^1}(u) dN_{X_0^2}(v) \\
&= \sum_{T \in x_0^2} \left(N_{x_n^1} \left((\lambda_n^2)^{-1}(T) + \delta \right) - N_{x_n^1}^- \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right).
\end{aligned}$$

For each T in x_0^2 ,

$$\begin{aligned}
& \left| N_{x_n^1} \left((\lambda_n^2)^{-1}(T) + \delta \right) - N_{x_n^1}^- \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right| \\
& \leq \left| N_{x_n^1} \left((\lambda_n^2)^{-1}(T) + \delta \right) - N_{x_0^1} \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta \right) \right) \right| \\
& + \left| N_{x_0^1} \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta \right) \right) - N_{x_0^1} \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right) \right| \\
& + \left| N_{x_0^1} \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right) - N_{x_n^1}^- \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right| \\
& \leq 2\varepsilon + \left| N_{x_0^1} \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right) - N_{x_n^1}^- \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right|,
\end{aligned}$$

where the last line comes from (1-ii), and (63).

We now find some $w_T \in \left[(\lambda_n^2)^{-1}(T) + \delta + \nu_n - \eta/4, (\lambda_n^2)^{-1}(T) + \delta + \nu_n \right]$ such that

$$\left| N_{x_n^1}^- \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) - N_{x_n^1}(w_T) \right| \leq \varepsilon,$$

so

$$\begin{aligned}
& \left| N_{x_0^1} \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right) - N_{x_n^1}^- \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right| \\
& \leq \left| N_{x_0^1} \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right) - N_{x_0^1} \left(\lambda_n^1(w_T) \right) \right| + \left| N_{x_0^1} \left(\lambda_n^1(w_T) \right) - N_{x_n^1}(w_T) \right| + \varepsilon.
\end{aligned}$$

From (1-ii), we deduce that $\left| N_{x_0^1} \left(\lambda_n^1(w_T) \right) - N_{x_n^1}(w_T) \right| \leq \varepsilon$.

Due to (63), (1-i), and the construction of w_T ,

$$\left| \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right) - (T - \delta) \right| \leq \frac{3\eta}{4} < \eta,$$

and

$$\begin{aligned}
\left| \left(\lambda_n^1(w_T) \right) - (T - \delta) \right| & \leq \left| \left(\lambda_n^1(w_T) - w_T \right) \right| + |w_T - (T - \delta)| \\
& \leq \frac{\eta}{4} + \frac{\eta}{4} \\
& < \eta,
\end{aligned}$$

so

$$\left| N_{x_0^1} \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right) - N_{x_0^1} \left(\lambda_n^1(w_T) \right) \right| = 0.$$

As a consequence,

$$\left| N_{x_n^1} \left((\lambda_n^2)^{-1}(T) + \delta \right) - N_{x_n^1}^- \left((\lambda_n^2)^{-1}(T) + \delta + \nu_n \right) \right| \leq 4\varepsilon,$$

and

$$|B_{n,3}| \leq 2\varepsilon \#x_n^1 + 4\varepsilon \#x_0^2. \tag{75}$$

Control of $B_{n,4}$.

$$\begin{aligned} B_{n,4} &= \iint (\mathbb{1}_{\{v < \lambda_n^2(u-\delta)\}} - \mathbb{1}_{\{v < u-\delta\}}) dN_{x_0^2}(v) dN_{x_n^1}(u) \\ &= \sum_{T \in x_0^2} \left(N_{x_n^1}(T + \delta) - N_{x_n^1} \left((\lambda_n^2)^{-1}(T) + \delta \right) \right) \end{aligned}$$

Let us fix T in x_0^2 . We have

$$\begin{aligned} \left| N_{x_n^1}(T + \delta) - N_{x_n^1} \left((\lambda_n^2)^{-1}(T) + \delta \right) \right| &\leq \left| N_{x_n^1}(T + \delta) - N_{x_0^1}(\lambda_n^1(T + \delta)) \right| \\ &\quad + \left| N_{x_0^1}(\lambda_n^1(T + \delta)) - N_{x_0^1} \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta \right) \right) \right| \\ &\quad + \left| N_{x_0^1} \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta \right) \right) - N_{x_n^1} \left((\lambda_n^2)^{-1}(T) + \delta \right) \right|. \end{aligned}$$

The first and the last terms are upper bounded by ε due to (1-ii).

Furthermore, since

$$N_{x_0^1} \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta \right) \right) = N_{x_0^1}(T + \delta) = N_{x_0^1}(\lambda_n^1(T + \delta)),$$

by applying (63) and using (1-i) and (2-i).

$$\left| N_{x_0^1}(\lambda_n^1(T + \delta)) - N_{x_0^1} \left(\lambda_n^1 \left((\lambda_n^2)^{-1}(T) + \delta \right) \right) \right| = 0.$$

So finally,

$$|B_{n,4}| \leq 2\varepsilon \#x_0^2. \quad (76)$$

Combining (72), (73), (74), (75), and (76), we can conclude that:

$$|B_n| \leq 3\varepsilon \#x_n^1 + 10\varepsilon \#x_0^2. \quad (77)$$

We now just need to remark that $(\#x_n^1)_{n \geq n_0}$ is bounded because it converges to $\#x_0^1$. Indeed, since $\#x_n^1 = N_{x_n^1}(1)$, $\#x_0^1 = N_{x_0^1}(1)$ and for every n , $\lambda_n^1(1) = 1$,

$$\begin{aligned} |\#x_n^1 - \#x_0^1| &= \left| N_{x_n^1}(1) - N_{x_0^1}(1) \right| \\ &= \left| N_{x_n^1}(1) - N_{x_0^1}(\lambda_n^1(1)) \right| \\ &\xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

With (65), (71), and (77), this concludes the proof of Proposition 1.

Acknowledgments

We are grateful to F. Grammont for fruitful discussions. This research is partly supported by the french Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration) and by the PEPS BMI 2012-2013 *Estimation of dependence graphs for thalamo-cortical neurons and multivariate Hawkes processes*.

References

- [1] M. A. Arcones and E. Giné. On the bootstrap of U and V statistics. *The Annals of Statistics*, 20(2):655–674, 1992.
- [2] S. Arlot, G. Blanchard, and E. Roquain. Some nonasymptotic results on resampling in high dimension, i: Confidence regions. *The Annals of Statistics*, pages 51–82, 2010.
- [3] S. Arlot, G. Blanchard, and E. Roquain. Some nonasymptotic results on resampling in high dimension, ii: Multiple tests. *The Annals of Statistics*, pages 83–99, 2010.

- [4] P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6):1196–1217, 1981.
- [5] P. Billingsley. *Convergence of probability measures*, volume 493. Wiley-Interscience, 2009.
- [6] J. R. Blum, J. Kiefer, and M. Rosenblatt. Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, 32(2):485–498, 1961.
- [7] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24(3):1563–1588, 1996.
- [8] J. Bretagnolle. Lois limites du bootstrap de certaines fonctionnelles. *Annales de l’institut Henri Poincaré (B) Probabilités et Statistiques*, 19(3):281–296, 1983.
- [9] B. M. Brown. Martingale central limit theorems. *The Annals of Mathematical Statistics*, pages 59–66, 1971.
- [10] L. Carstensen, A. Sandelin, O. Winther, and N. R. Hansen. Multivariate Hawkes process models of the occurrence of regulatory elements and an analysis of the pilot ENCODE regions. *BMC Bioinformatics*, 11(456), 2010.
- [11] E. Chung and J. P. Romano. Exact and asymptotically robust permutation tests. *The Annals of Statistics*, 41(2):484–507, 2013.
- [12] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I. Probability and its Applications* (New York). Springer-Verlag, New York, second edition, 2003. Elementary theory and methods.
- [13] H. Dehling and T. Mikosch. Random quadratic forms and the bootstrap for U-statistics. *Journal of Multivariate Analysis*, 51(2):392–413, 1994.
- [14] R. M. Dudley. *Real Analysis and Probability*. Number 74 in Cambridge studies in advanced mathematics. Cambridge University Press, 2nd edition, 2002.
- [15] B. Efron. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.
- [16] R. A. Fisher. *The design of experiments*. 1935.
- [17] M. Fromont. Model selection by bootstrap penalization for classification. *Machine Learning*, pages 165–207, 2007.
- [18] M. Fromont, B. Laurent, and P. Reynaud-Bouret. The two-sample problem for poisson processes: Adaptive tests with a nonasymptotic wild bootstrap approach. *The Annals of Statistics*, pages 1431–1461, 2013.
- [19] G. L. Gerstein and D. H. Perkel. Simultaneous recorded trains of action potentials: analysis and functional interpretation. *Science*, 164:828–830, 1969.
- [20] A. Gretton, K. Fukumizu, C. H. Teo, L. Song, B. Schölkopf, and A. J. Smola. A kernel statistical test of independence. *NIPS Proceedings*, 20:585–592, 2007.
- [21] A. Gretton and L. Györfi. Consistent nonparametric tests of independence. *The Journal of Machine Learning Research*, 11:1391–1423, 2010.
- [22] A. Gretton, R. Herbrich, A. J. Smola, O. Bousquet, and B. Schölkopf. Kernel methods for measuring independence. *The Journal of Machine Learning Research*, 6:2075–2129, 2005.
- [23] S. Grün. *Unitary joint-events in multiple-neuron spiking activity: Detection, significance and interpretation*. PhD thesis, Thun: Verlag Harri Deutsch, 1996.
- [24] S. Grün. Data-driven significance estimation for precise spike correlation. *Journal of Neurophysiology*, 101:1126–1140, 2009.
- [25] S. Grün, M. Diesmann, and A. M. Aertsen. *Analysis of parallel spike trains*, chapter Unitary Events analysis. Springer Series in Computational Neuroscience, 2010.
- [26] S. Grün, M. Diesmann, F. Grammont, A. Riehle, and A. M. Aertsen. Detecting unitary events without discretization of time. *Journal of Neuroscience Methods*, 93:67–79, 1999.

- [27] G. Gusto and S. Schbath. FADO: a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes' model. *Statistical Applications in Genetics and Molecular Biology*, 4:Art. 24, 28 pp. (electronic), 2005.
- [28] N. R. Hansen, P. Reynaud-Bouret, and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *To appear in Bernoulli*, 2014. <http://arxiv.org/abs/1208.0570>.
- [29] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948.
- [30] W. Hoeffding. A non-parametric test of independence. *The Annals of Mathematical Statistics*, 19(4):546–557, 1948.
- [31] W. Hoeffding. The large-sample power of tests based on permutation of the observations. *The Annals of Mathematical Statistics*, 23(2):169–192, 1952.
- [32] W. Hoeffding. The strong law of large numbers for U-statistics. Institute of Statistics, Mimeograph Series No. 302, 1961.
- [33] L. Horváth and M. Hušková. Testing for changes using permutations of U-statistics. *Journal of Statistical Planning and Inference*, 128(2):351–371, 2005.
- [34] H. Hotelling and M. R. Pabst. Rank correlation and tests of significance involving no assumption of normality. *The Annals of Mathematical Statistics*, 7(1):29–43, 1936.
- [35] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1–2):81–93, 1938.
- [36] A. Leucht and M. H. Neumann. Consistency of general bootstrap methods for degenerate U-type and V-type statistics. *Journal of Multivariate Analysis*, 100(8):1622–1633, 2009.
- [37] Y. Ogata. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.
- [38] K. Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900.
- [39] K. Pearson. On the probability that two independent distributions of frequency are really samples from the same population. *Biometrika*, 8:250–254, 1911.
- [40] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter. How structure determines correlations in neuronal networks. *PLoS Computational Biology*, 7(5):e1002059, 2012.
- [41] G. Pipa, M. Diesmann, and S. Grün. Significance of joint-spike events based on trial-shuffling by efficient combinatorial methods. *Complexity*, 8(4):1–8, 2003.
- [42] G. Pipa and S. Grün. Non-parametric significance estimation of joint-spike events by shuffling and resampling. *Neurocomputing*, 52–54:31–37, 2003.
- [43] E. J. Pitman. Significance tests which may be applied to samples from any populations. *Suppl. Jjournal of the Royal Sstatistical Ssociety*, 4:117–130, 1937.
- [44] E. J. Pitman. Significance tests which may be applied to samples from any populations: II. The correlation coefficient test. *Suppl. Jjournal of the Royal Sstatistical Ssociety*, 4:225–232, 1937.
- [45] E. J. Pitman. Significance tests which may be applied to samples from any populations: III. The analysis of variance test. *Biometrika*, 29:322–335, 1938.
- [46] C. Pouzat and A. Chaffiol. Automatic spike train analysis and report generation. an implementation with r, r2html and star. *Journal of Neuroscience Methods*, pages 119–144, 2009.
- [47] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. Technical report, <http://hal.archives-ouvertes.fr/hal-00789127>, 2013.
- [48] P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot. Inference of functional connectivity in neurosciences via Hawkes processes. In *1st IEEE Global Conference on Signal and Information Processing*, 2013, Austin Texas.

- [49] J. P. Romano. Bootstrap and randomization tests of some nonparametric hypotheses. Technical Report 270, Dept. Statistics, Stanford Univ., 1987.
- [50] J. P. Romano. Bootstrap and randomization tests of some nonparametric hypotheses. *The Annals of Statistics*, 17(1):141–159, 1989.
- [51] J. P. Romano and M. Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108, 2005.
- [52] H. Rubin and R. A. Vitale. Asymptotic distribution of symmetric statistics. *The Annals of Statistics*, pages 165–170, 1980.
- [53] L. Sansonnet and C. Tuleau-Malot. A model of Poissonian interactions and detection of dependence. Technical report, <http://hal.archives-ouvertes.fr/hal-00780598>, 2013.
- [54] H. Scheffe. Statistical inference in the non-parametric case. *The Annals of Mathematical Statistics*, 14(4):305–332, 1943.
- [55] R. L. Schilling. *Measures, integrals and martingales*, volume 13. Cambridge University Press, 2005.
- [56] W. Singer. Synchronization of cortical activity and its putative role in information processing and learning. *Annual Review of Physiology*, 55:349–374, 1993.
- [57] C. Tuleau-Malot, A. Rouis, F. Grammont, and P. Reynaud-Bouret. Multiple tests based on a Gaussian approximation of the Unitary Events method. Technical report, <http://hal.archives-ouvertes.fr/hal-00757323>, 2012.
- [58] A. W. Van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, 1996.
- [59] A.W. Van der Vaart. *Asymptotic statistics*. Cambridge series in Statistical and Probabilistic Mathematics, 1998.
- [60] V. S. Varadarajan. On the convergence of sample probability distributions. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 19(1–2):23–26, 1958.
- [61] V. Ventura. *Analysis of parallel spike trains*, chapter Bootstrap tests of hypotheses. Springer Series in Computational Neuroscience, 2010.
- [62] J. Wolfowitz. Additive partition functions and a class of statistical hypotheses. *The Annals of Mathematical Statistics*, 13(3):247–279, 2042.