Bootstrap and permutation tests of independence for point processes

Mélisande Albert, Yann Bouret, Magalie Fromont, Patricia Reynaud-Bouret

Bootstrap and permutation tests of independence for point processes.

Mélisande Albert[*], Yann Bouret[†], Magalie Fromont[‡] and Patricia Reynaud-Bouret[§]

Abstract: Motivated by a neuroscience question about synchrony detection in spike train 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 randomization/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. A simulation study is performed to illustrate the derived theoretical results, and to compare the performance of our new tests with existing ones in the neuroscientific literature.

Mathematics Subject Classification: Primary: 62M07, 62F40, 62E20, 60G55, 60F05; secondary: 62P10

Keywords: Independence test, U-statistics, point processes, bootstrap, randomisation, permutation, neuroscience, spike train analysis.

1 Introduction

Inspired by neuroscience problems, the present work is devoted to 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 chi-square test of independence (see [43, 44]) to the modern test of [24] 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

[*]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.

them, of particular interest are the tests based on the randomization/permutation principle introduced by Fisher [19], and covered thereafter in the series of papers by Pitman [49, 50], Scheffe [60], Hoeffding [33] for instance, or bootstrap approaches derived from Efron's [17] "naive" one. Note that permutation and bootstrap-based tests have a long history of applications, of which independence tests are just a very small part (see for instance [18, 46, 55, 56] for some reviews, or [1, 36, 39, 38, 22] for more recent works). Focusing on independence tests, two families of permutation or bootstrap-based tests may be distinguished at least: the whole family of rank tests including the tests of Hotelling and Pabst [35], Kendall [37], Wolfowitz [65] or Hoeffding [31] on the one hand, the family of Kolmogorov-Smirnov type tests, like Blum, Kiefer, and Rosenblatt's [8], Romano's [56] or Van der Vaart and Wellner's [63] ones on the other hand. In order to describe the properties of these tests, let us recall and fix a few definitions, which are furthermore used throughout this article. Tests are said to be *non-parametric* if they are free from the underlying distribution of the observed variables. For any prescribed α in $(0, 1)$, tests are said to be *exactly of level* α if their first kind error rate is less than α whatever the number of observations. This is a non-asymptotic property. Tests are also said to be *asymptotically of size* α if their first kind error rate tends to α when the number of observations tends to infinity. Finally, they are said to be *consistent* against some alternative if, under this alternative, their second kind error rate tends to 0 or equivalently their power tends to 1, when the number of observations tends to infinity. In this respect, the above mentioned tests of independence are all non-parametric and asymptotically of the prescribed size. Moreover, the tests based on permutation are exactly of the desired level. Some of these tests are proved to be consistent against many alternatives, such as Hoeffding's [31] one and the family of Kolmogorov-Smirnov type tests.

Detecting dependence is also a fundamental old point in the neuroscientific literature (see e.g. [23]). The neuroscience problem we were initially interested in consists in detecting interactions between occurrences of action potentials on two different neurons simultaneously recorded on n independent trials, as described in [26]. 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 [61]). So, 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 (see [26] and the references therein), consist in binning the spike trains at first in order to deal with vectorial data with reduced dimension. However, it has been shown 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. It

is therefore more realistic and reasonable to model recordings of spike trains by finite point processes, and to use independence tests specifically dedicated to such point processes. Asymptotic tests of independence between point processes have already been introduced in [62], but in the particular case of homogeneous Poisson processes. Such a parametric framework is necessarily restrictive and even possibly inappropriate here, as the very existence of any precise underlying distribution for the point processes modelling spike train data is subject to broad debate (see [51, 52]). We thus focus on non-parametric tests of independence for point processes. In this spirit, particular bootstrap methods under the name of *trial-shuffling* have been proposed in [48, 47] for binned data with relatively small dimension, without proper mathematical justification. Besides the loss of information the binning data pre-processing involves, it appears that the test statistics chosen in these papers do not lead to tests of asymptotic prescribed size as shown by our simulation study.

We here propose to construct new non-parametric tests of independence between two point processes, from the observation of n independent copies of these point processes, with as few assumptions as possible on their underlying distributions. Our test statistics are based on U-statistics. The corresponding critical values are obtained from bootstrap or permutation approaches. It has been acknowledged that when both approaches are available, permutation should be preferred, since the corresponding tests are exactly of the desired level [18, p. 218]. Nevertheless, we investigate permutation and bootstrap together, as bootstrap methods - through *trial-shuffling* - are the usual references in neuroscience. Moreover, for specific U-statistics, the corresponding tests share the same properties: both are proved to be asymptotically of the prescribed size and consistent against any reasonable alternatives, despite the fact that different tools are used to obtain these results. Indeed in particular, the distance between the bootstrapped distribution and the initial distribution under independence is here directly studied for the bootstrap approach, unlike the permutation approach. Finally both procedures have good performance in practice when the sample size is moderate to small, as is often the case in neuroscience due to biological or economical reasons.

As U-statistics are now usual tools for non-parametric statistical inference, many works deal with the application of bootstrap or permutation type methods to U-statistics. From the original work of Arvesen [5] about the Jackknife of U-statistics, to the recent one of Leucht and Neumann [40], several papers [6, 12, 2, 15] have been devoted to the general problem of bootstrapping a U-statistic. The use of bootstrap or permutation of U-statistics is specially considered in testing problems (see [34, 13] for instance), in particular in dependence detection problems with the Kolmogorov-Smirnov type tests cited above (see [56, 63]).

But all those works exclusively focus on U-statistics of i.i.d. real valued random variables or vectors. Up to our knowledge, there is no previous work on the bootstrap or permutation of general U-statistics for i.i.d. pairs of point processes, as considered in the present paper. The main difficulty thus lies in the nature of the mathematical objects we handle here, that is point processes and their associated point measures which are random measures. The proofs of our results, although inspired by Romano's [54, 56] work and Hoeffding's [33] precursor results on the

permutation, are therefore more technical and complex on many aspects detailed in the sequel. In addition, we aim at obtaining the asymptotic distribution of the bootstrapped or permuted test statistics under independence, but also under dependence (see Theorem 3.1 and Theorem 4.1). As concerns the permutation approach, such a result is, as far as we know, new even for more classical settings than point processes. It thus partially solves a problem stated as open question in [63].

This paper is organized as follows.

We first present in Section 2 the testing problem, and introduce the main notations. Starting from existing works in neuroscience, we introduce our test statistics, based on general kernel-based U-statistics.

Section 3 is devoted to our bootstrap approach. Are given new general results about the consistency of the bootstrap for the considered U-statistics, expressed in terms of Wasserstein's metric as in [6]. The convergence is studied under independence as well as under dependence. The corresponding bootstrap independence tests are therefore shown to be asymptotically of the desired size, and consistent against any reasonable alternative. The impact of using Monte Carlo methods to approximate the bootstrap quantiles is also investigated in this section.

Section 4 is devoted to the permutation approach which leads, by nature, to non-parametric independence tests exactly of the desired level, and this, even when a Monte Carlo method is used to approximate the permutation quantiles. Are then given new general results about the consistency of the permutation approach when the kernel of the U-statistic has a specific form. These results are still expressed in terms of Wasserstein's metric. As a consequence the corresponding permutation independence tests are proved to satisfy the same asymptotic properties as the bootstrap ones under the null hypothesis as well as under the same alternatives.

As a comparison of the performance of our tests with existing ones in neuroscience, especially when the sample sizes are moderate or even small, a simulation study is presented in Section 5.

A conclusion is given in the last section.

Finally notice that all proofs and some additional technical results can be found in a supplementary material.

2 From neuroscience interpretations to general test statistics

2.1 The testing problem

Throughout this article we consider finite point processes defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and observed on $[0,1]$, i.e. random point processes on $[0,1]$ whose total number of points is almost surely finite (see [2] for instance). 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 consists of the countable subsets of $[0,1]$. It is equipped with a metric $d_{\mathcal{X}}$ that we introduce

in (3.3). This metric, issued from the Skorohod topology, makes $\mathcal{X}$ separable and allows to define accordingly borelian sets on $\mathcal{X}$ and by extension on $\mathcal{X}^2$ through the product metric.

The point measure dN_x associated with an element x of $\mathcal{X}$ is defined for all measurable real-valued function f by $\int_{[0,1]} f(u)dN_x(u) = \sum_{u\in x} f(u)$. 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 $(\Omega, \mathcal{A}, \mathbb{P})$.

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, with marginals P^1 and P^2 if $P(\mathcal{B}) = \mathbb{P}(X \in \mathcal{B})$, $P^1(\mathcal{B}^1) = \mathbb{P}(X^1 \in \mathcal{B}^1)$, and $P^2(\mathcal{B}^2) = \mathbb{P}(X^2 \in \mathcal{B}^2)$, for every borelian set $\mathcal{B}$ of $\mathcal{X}^2$, and all borelian sets $\mathcal{B}^1$, $\mathcal{B}^2$ of $\mathcal{X}$.

Given the observation of an i.i.d. sample $\mathbb{X}_n = (X_1, \ldots, X_n)$ from the same distribution P as X, with $X_i = (X_i^1, X_i^2)$ for every $i = 1 \ldots n$, we aim at testing (H_0) X^1 and X^2 are independent against (H_1) X^1 and X^2 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 $\mathbb{X}_n = (X_1, \ldots, X_n)$ models pairs of rescaled spike trains issued from two distinct and simultaneously recorded neurons during n trials, the main dependence feature that needs to be detected between both neurons corresponds to synchronization in time, referred to as coincidences [26]. 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.

In [62], the notion of coincidence count between two point processes X^1 and X^2 with delay δ ($\delta > 0$) is defined by

$$\varphi_\delta^{coinc}(X^1, X^2) = \int_{[0,1]^2} \mathbb{1}_{|u-v|\leq\delta} dN_{X^1}(u)dN_{X^2}(v) = \sum_{u\in X^1, v\in X^2} \mathbb{1}_{|u-v|\leq\delta}. \tag{2.1}$$

Notice that other coincidence count functions have been used in the neuroscience literature such as the binned coincidence count function (i.e. based on binned data) introduced in [25] or its shifted version [27] (see also [62] for explicit formulae). The interaction function introduced in [59] provides a further example of possible function used to detect dependence in neuroscience, which is of the form

$$\varphi^w(X^1, X^2) = \int_{[0,1]^2} w(u,v)dN_{X^1}(u)dN_{X^2}(v). \tag{2.2}$$

Under the assumption that both X^1 and X^2 are homogeneous Poisson processes, the independence test of [62] rejects (H_0) when a test statistic based on $\sum_{i=1}^n \varphi_\delta^{coinc}\left(X_i^1, X_i^2\right)$

is larger than a given critical value. This critical value is deduced from the asymptotic Gaussian distribution of the test statistic under (H_0). The test is proved to be asymptotically of the desired size, 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 [51, 52].

2.3 General U-statistics as independence test statistics

In the parametric homogeneous Poisson framework of [62], the expectation of $\varphi_\delta^{coinc}\left(X_i^1, X_i^2\right)$ 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 $\varphi_\delta^{coinc}\left(X_i^1, X_i^2\right)$ to lead to a test statistic with a centered asymptotic distribution under (H_0).

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 centering plug-in tool is not available. We propose to use instead a self-centering trick, which amounts, combined with a rescaling step, to considering the statistic

$$\frac{1}{n(n-1)} \sum_{i \neq i' \in \{1,\dots,n\}} \left(\varphi_\delta^{coinc}\left(X_i^1, X_i^2\right) - \varphi_\delta^{coinc}\left(X_i^1, X_{i'}^2\right)\right). \tag{2.3}$$

It is clear that the function φ_δ^{coinc} used in [62] suits the dependence feature the neuroscientists expect to detect in a spike train analysis. However, it is not necessarily the best choice for other kinds of dependence features to be detected in a general point processes analysis. Note furthermore that the statistic (2.3) can be written as a U-statistic of the i.i.d. sample $\mathbb{X}_n = (X_1, \dots, X_n)$ with a symmetric kernel, as defined by Hoeffding [30].

Let us therefore consider the general independence test statistics which are U-statistics of the form

$$U_{n,h}(\mathbb{X}_n) = \frac{1}{n(n-1)} \sum_{i \neq i' \in \{1,\dots,n\}} h\left(X_i, X_{i'}\right), \tag{2.4}$$

where $h : (\mathcal{X}^2)^2 \to \mathbb{R}$ is a symmetric kernel such that:

$(\mathcal{A}_{Cent})$	For all $n \geq 2$, $U_{n,h}(\mathbb{X}_n)$ is zero mean under (H_0), i.e. for X_1 and X_2, i.i.d. with distribution $P^1 \otimes P^2$ on $\mathcal{X}^2$, $\mathbb{E}\left[h\left(X_1, X_2\right)\right] = 0$.

In the sequel, we call *Coincidence case* the case where $h = h_{\varphi_\delta^{coinc}}$, with

$$h_{\varphi_\delta^{coinc}}(x,y) = \frac{1}{2}\left(\varphi_\delta^{coinc}\left(x^1, x^2\right) + \varphi_\delta^{coinc}\left(y^1, y^2\right) - \varphi_\delta^{coinc}\left(x^1, y^2\right) - \varphi_\delta^{coinc}\left(y^1, x^2\right)\right), \tag{2.5}$$

so that $U_{n,h_{\varphi_\delta^{coinc}}}(\mathbb{X}_n)$ is equal to the statistic (2.3).

A more general choice, which of course includes the above *Coincidence case*, is obtained by replacing φ_δ^{coinc} by any generic integrable function φ. This is the *Linear case*. For any integrable function φ, the kernel h is then taken equal to h_φ, with

$$h_\varphi(x,y) = \frac{1}{2}\left(\varphi\left(x^1,x^2\right) + \varphi\left(y^1,y^2\right) - \varphi\left(x^1,y^2\right) - \varphi\left(y^1,x^2\right)\right). \tag{2.6}$$

This example is of utmost importance in the present work since it provides a first proved case of consistency for the permutation approach under the null hypothesis as well as under the alternative (see Theorem 4.1). In this case, note that $(\mathcal{A}_{Cent})$ is straightforwardly satisfied, i.e. $U_{n,h_\varphi}(\mathbb{X}_n)$ is zero mean under (H_0). Note furthermore that $U_{n,h_\varphi}(\mathbb{X}_n)$ is here an unbiased estimator of

$$\int\int \varphi\left(x^1,x^2\right)\left(dP(x^1,x^2) - dP^1(x^1)dP^2(x^2)\right),$$

without any assumption on the underlying point processes. This is therefore a reasonable independence test statistic. If X^1 and X^2 were finite dimensional variables with continuous distributions w.r.t. the Lebesgue measure, this test statistic would be closely related to generalized Kolmogorov-Smirnov tests of independence. For instance, the test statistics of Blum, Kiefer, and Rosenblatt [8], Romano [56], Van der Vaart and Wellner in [63] are equivalent to $\sqrt{n}\sup_{v^1\in\mathcal{V}^1,v^2\in\mathcal{V}^2}\left|U_{n,h_{\varphi_{(v^1,v^2)}}}(\mathbb{X}_n)\right|$, 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 $\varphi_{(v^1,v^2)}(x^1,x^2) = v^1(x^1)v^2(x^2)$.

Thus, up to our knowledge, the existing test statistics are based on functions φ of product type. However, as seen in Section 2.2, when dealing with point processes, natural functions φ are not of this type, as for instance φ_δ^{coinc}.

2.4 Non-degeneracy of the U-statistics under (H_0)

Following the works of Romano [56] or Van der Vaart and Wellner [63], the tests we propose here are based on bootstrap and permutation approaches for the above general U-statistics. Most of the assumptions on h depend on the chosen method (permutation or bootstrap) and are postponed to the corresponding section. However another assumption is common, besides $(\mathcal{A}_{Cent})$:

$(\mathcal{A}_{non-deg})$ | For all $n \geq 2$, $U_{n,h}(\mathbb{X}_n)$ is non-degenerate under (H_0), i.e. for all X_1 and X_2, i.i.d. with distribution $P^1 \otimes P^2$ on $\mathcal{X}^2$, $\mathrm{Var}\left(\mathbb{E}\left[h(X_1,X_2)|X_1\right]\right) \neq 0$.

This assumption is needed in all results with weak convergence to a Gaussian limit, as its variance has to be strictly positive (see e.g. Proposition 3.5 or Theorem 4.1). Since under (H_0), $U_{n,h}(\mathbb{X}_n)$ is assumed to have zero mean, it is degenerate

under (H_0) 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 *Linear case*, this condition implies a very particular link between φ and the distribution of the bivariate point process X, which is unknown. The following result gives some basic condition to fulfill $(\mathcal{A}_{non-deg})$ when φ is the coincidence count function.

Proposition 2.1. *If the empty set is charged by the marginals, i.e. if* $P^1(\{\emptyset\}) > 0$ *and* $P^2(\{\emptyset\}) > 0$ *and if* $\varphi_\delta^{coinc}(X_1, X_2)$ *(see (2.1)) is not almost surely null under* (H_0)*, then when* h *is given by (2.5),* $(\mathcal{A}_{non-deg})$ *is satisfied.*

The proof can be found in the supplementary material together with a more informal discussion on the *Linear case* with $\varphi = \varphi^w$ as given by (2.2).

As far as experimental neuronal data are concerned, assuming that the processes may be empty is an obvious assumption as there often exist trials (usually short) where, just by chance, no spikes have been detected. Moreover, practitioners usually choose δ large enough such that coincidences are observed in practice and therefore $\varphi_\delta^{coinc}(X_1, X_2)$ is not almost surely null. Hence in practice, the non-degeneracy assumption is always satisfied in the *Coincidence case*.

Throughout this article, $(X_i)_i$ denotes a sequence of i.i.d. pairs of point processes, with $X_i = (X_i^1, X_i^2)$ of distribution P on $\mathcal{X}^2$, whose marginals are P^1 and P^2 on $\mathcal{X}$. For $n \geq 2$, let $\mathbb{X}_n = (X_1, \ldots, X_n)$ and $U_{n,h}(\mathbb{X}_n)$ as in (2.4), with a fixed measurable symmetric kernel h satisfying $(\mathcal{A}_{Cent})$. To shorten mathematical expression, $U_n(\mathbb{X}_n)$ refers from now on to $U_{n,h}(\mathbb{X}_n)$.

3 Bootstrap tests of independence

Since the distribution of the test statistic $U_n(\mathbb{X}_n)$ 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, which aims at mimicking the distribution of the test statistic under (H_0), for large, but also moderate or small sample sizes.

In order to describe this bootstrap approach, and to state our results without any confusion, we give below additional notations, and then discuss the main assumptions needed in these results.

3.1 Additional notations

Bootstrap formalism Under (H_0), each $X_i = (X_i^1, X_i^2)$ is $P^1 \otimes P^2$- distributed. Hence, our bootstrap approach consists in resampling the first and second coordinates of the elements of $\mathbb{X}_n$ according to the corresponding marginal empirical distributions P_n^j given by

$$\text{for } j = 1, 2, \quad P_n^j = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i^j}. \tag{3.1}$$

The bootstrap sample $\mathbb{X}_n^* = \left(X_{n,1}^*, \ldots, X_{n,n}^*\right)$, with $X_{n,i}^* = (X_{n,i}^{*1}, X_{n,i}^{*2})$, is then defined as an n i.i.d. sample from the distribution $P_n^1 \otimes P_n^2$, and the bootstrap version of $U_n(\mathbb{X}_n)$ is given by $U_n(\mathbb{X}_n^*)$.

Convergence formalism The bootstrap distribution of interest is the conditional distribution of $\sqrt{n}U_n(\mathbb{X}_n^*)$ given $\mathbb{X}_n$, to be compared with the initial distribution of $\sqrt{n}U_n(\mathbb{X}_n)$ under (H_0). In order to state properly our convergence results, we therefore introduce the following formalism.

- For any functional $Z : (\mathcal{X}^2)^n \to \mathbb{R}$, $\mathcal{L}(Z, Q)$ denotes the distribution of $Z(\mathbb{Y}_n)$, where $\mathbb{Y}_n$ is an i.i.d. sample from the distribution Q on $\mathcal{X}^2$. In particular, the distribution of $\sqrt{n}U_n(\mathbb{X}_n)$ under (H_0) is denoted by $\mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)$.
- If the distribution $Q = Q(W)$ depends on a random variable W, $\mathcal{L}(Z, Q|\, W)$ is the conditional distribution of $Z(\mathbb{Y}_n)$, $\mathbb{Y}_n$ being an i.i.d. sample from the distribution $Q = Q(W)$ on $\mathcal{X}^2$, given W.

  In particular, the conditional distribution of $\sqrt{n}U_n(\mathbb{X}_n^*)$ given $\mathbb{X}_n$ is denoted by $\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 \middle|\, \mathbb{X}_n\right)$.
- " Q-a.s. in $(X_i)_i$" at the end of a statement means that the statement only depends on the sequence $(X_i)_i$, where the X_i's are i.i.d with distribution Q, and that there exists an event $\mathcal{C}$ only depending on $(X_i)_i$ such that $\mathbb{P}(\mathcal{C}) = 1$, on which the statement is true. Here Q is usually equal to P.
- "$Q_n \underset{n\to+\infty}{\Longrightarrow} Q$" 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, $\int g(z)dQ_n(z) \to_{n\to+\infty} \int g(z)dQ(z)$.
- Finally, as we often work conditionally on $\mathbb{X}_n$, we denote by $\mathbb{E}^*[\cdot]$ the conditional expectation given the sample $\mathbb{X}_n$.

As explained above, the bootstrap distribution of interest is the conditional distribution $\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 \middle|\, \mathbb{X}_n\right)$, which will be proved to be asymptotically close to $\mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)$. Following the historical paper by Bickel et Freedman [6], the closeness between these two distributions, which are both distributions on $\mathbb{R}$, is here measured via the classical $\mathbb{L}^2$-Wasserstein's metric.

The $\mathbb{L}^2$-Wasserstein's metric d_2 (also called Mallows' metric) is defined by

$$d_2^2(Q, Q') = \inf\left\{\mathbb{E}\left[(Z - Z')^2\right],\ (Z, Z') \text{ with marginals } Q \text{ and } Q'\right\}, \tag{3.2}$$

for all the distributions Q, Q' with finite second order moments. Recall that convergence w.r.t. d_2 is equivalent to both weak convergence and convergence of second order moments.

3.2 Main assumptions

The random variables we deal with are not real-valued variables but point processes, so the assumptions needed in our results may be difficult to interpret in this setting. We therefore devote this whole section to their description and discussion.

In addition to Assumption $(\mathcal{A}_{Cent})$, we need its following empirical version:

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

Notice that this assumption, as well as $(\mathcal{A}_{Cent})$, is fulfilled in the *Linear case* where h is of the form h_φ given by (2.6), but $(\mathcal{A}^*_{Cent})$ does not imply that h is of the form h_φ (see the supplementary material for a counterexample).

Moment assumptions Due to the $\mathbb{L}^2$-Wasserstein metric used here to study the consistency of the bootstrap approach, moment assumptions are required. In particular, the variance of $U_n(\mathbb{X}_n)$ should exist, i.e.

$$(\mathcal{A}_{Mmt}) \quad \left| \begin{array}{l} \text{For } X_1 \text{ and } X_2 \text{, i.i.d. with distribution } P \text{ on } \mathcal{X}^2, \\ \mathbb{E}\left[h^2\left(X_1, X_2\right)\right] < +\infty, \end{array} \right.$$

and more generally we need:

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

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 $X \sim P$, $\mathbb{E}\left[h^2\left(X, X\right)\right] < +\infty$ (taking $i_1 = i_2 = i_1' = i_2'$),
- for X_1, X_2 i.i.d with distribution $P^1 \otimes P^2$, $\mathbb{E}\left[h^2\left(X_1, X_2\right)\right] < +\infty$ (taking i_1, i_2, i_1', i_2' all different).

A sufficient condition for $(\mathcal{A}^*_{Mmt})$ and $(\mathcal{A}_{Mmt})$ to be satisfied is that there exist positive constants α_1, α_2, C such that for every $x = (x^1, x^2), y = (y^1, y^2)$ in $\mathcal{X}^2$, $|h(x,y)| \leq C\left((\#x^1)^{\alpha_1} + (\#y^1)^{\alpha_1}\right)\left((\#x^2)^{\alpha_2} + (\#y^2)^{\alpha_2}\right)$, with $\mathbb{E}[(\#X^1)^{4\alpha_1}] < +\infty$ and $\mathbb{E}[(\#X^2)^{4\alpha_2}] < +\infty$.

In the *Linear case* where h is of the form h_φ given by (2.6), a possible sufficient condition is that there exist some positive constants α_1, α_2, and 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 $\mathbb{E}[(\#X^2)^{4\alpha_2}] < +\infty$. In particular, in the *Coincidence case*, the coincidence count function φ_δ^{coinc} satisfies: for every x^1, x^2 in $\mathcal{X}$, $|\varphi_\delta^{coinc}(x^1, x^2)| \leq (\#x^1)(\#x^2)$. So, $(\mathcal{A}^*_{Mmt})$ and $(\mathcal{A}_{Mmt})$ are satisfied as soon as $\mathbb{E}[(\#X^1)^4] < +\infty$ and $\mathbb{E}[(\#X^2)^4] < +\infty$.

Such moment bounds for the total number of points of the processes are in fact satisfied by many kinds of point processes: discretized point processes at resolution $0 < r < 1$ (see [62] for a definition), which have at most $1/r$ points, Poisson

processes, whose total number of points obeys a Poisson distribution having exponential moments of any order, and point processes with bounded conditional intensities, which can be constructed by thinning homogeneous Poisson processes (see [42]). Similar moment bounds can also be obtained (see [28]) for linear stationary Hawkes processes with positive interaction functions that are classical models in spike train analysis (see e.g. [45, 62]). This finally may be extended to point processes whose conditional intensities are upper bounded by intensities of linear stationary Hawkes processes with positive interaction functions, by thinning arguments. This includes more general Hawkes processes (see [10]) and in particular Hawkes processes used to model inhibition in spike train analysis (see [28, 62, 53] or [52]).

Continuity of the kernel The set $\mathcal{X}$ can be embedded in the space $\mathcal{D}$ of càdlàg functions on $[0,1]$ through the identification

$$N : x \in \mathcal{X} \mapsto \left(N_x : t \mapsto \int_0^1 \mathbb{1}_{u \le t} dN_x(u) \right) \in \mathcal{D}.$$

Notice that the quantity N_x is actually the counting process associated with x (see [11] for instance): at time t, $N_x(t)$ is the number of points of x less than t. Now consider the uniform Skorohod topology on $\mathcal{D}$ (see [7]), associated with the metric $d_{\mathcal{D}}$ defined by

$$d_{\mathcal{D}}(f,g) = \inf \left\{ \varepsilon > 0 \ ; \ \exists \lambda \in \Lambda, \left\{ \begin{array}{l} \sup_{t\in[0,1]} |\lambda(t) - t| \le \varepsilon, \\ \sup_{t\in[0,1]} |f(\lambda(t)) - g(t)| \le \varepsilon \end{array} \right. \right\},$$

where Λ is the set of strictly increasing, continuous mappings of $[0,1]$ onto itself. Notice that here, λ represents a uniformly small deformation of the time scale. Thanks to the identification N above, $\mathcal{X}$ can then be endowed with the topology induced by $d_{\mathcal{X}}$ defined on $\mathcal{X}$ by

$$d_{\mathcal{X}}(x,x') = d_{\mathcal{D}}(N(x), N(x')) \quad \text{for every } x,\, x' \text{ in } \mathcal{X}. \tag{3.3}$$

As an illustration, if x and x' are in $\mathcal{X}$, for ε in $(0,1)$, $d_{\mathcal{X}}(x,x') \le \varepsilon$ implies that x and x' have the same cardinality, and for k in $\{1,\dots,\#x\}$, the k^{th} point of x is at distance less than ε from the k^{th} point of x'. Since $(\mathcal{D}, d_{\mathcal{D}})$ is a separable metric space, so are $(\mathcal{X}, d_{\mathcal{X}})$, $\left(\mathcal{X}^2, d_{\mathcal{X}^2}\right)$, where $d_{\mathcal{X}^2}$ is the product metric defined from $d_{\mathcal{X}}$ (see [3, p 32]), and $\left(\mathcal{X}^2 \times \mathcal{X}^2, d\right)$, where d, the product metric defined from $d_{\mathcal{X}^2}$, is given by

$$d\big((x,y),(x',y')\big) = \sup \left\{ \sup_{j=1,2} \left\{ d_{\mathcal{X}}(x^j, x'^j) \right\}, \sup_{j=1,2} \left\{ d_{\mathcal{X}}(y^j, y'^j) \right\} \right\}, \tag{3.4}$$

for every $x = (x^1, x^2)$, $y = (y^1, y^2)$, $x' = (x'^1, x'^2)$, $y' = (y'^1, y'^2)$ in $\mathcal{X}^2$.
The kernel h chosen to define the U-statistic $U_n(\mathbb{X}_n)$ in (2.4) should satisfy:

$(\mathcal{A}_{Cont})$ | There exists a subset $\mathcal{C}$ of $\mathcal{X}^2 \times \mathcal{X}^2$, such that
(i) h is continuous on $\mathcal{C}$ for the topology induced by d,
(ii) $(P^1 \otimes P^2)^{\otimes 2}(\mathcal{C}) = 1$.

Here are some examples in the *Linear case* for which $(\mathcal{A}_{Cont})$ holds.

Proposition 3.1. *Let $w : [0,1]^2 \to \mathbb{R}$ be a continuous integrable function. Then the kernel h_{φ^w} defined on $\mathcal{X}^2 \times \mathcal{X}^2$ by (2.2) and (2.6) is continuous w.r.t. the topology induced by d, defined by (3.4).*

The above result does not apply to $h_{\varphi_\delta^{coinc}}$ but the following one holds.

Proposition 3.2. *The coincidence count kernel $h_{\varphi_\delta^{coinc}}$ defined on $\mathcal{X}^2 \times \mathcal{X}^2$ by (2.1) and (2.6) is continuous w.r.t. the topology induced by d, on*

$$\mathcal{C}_\delta = \big\{ \big((x^1,x^2),(y^1,y^2)\big) \in \mathcal{X}^2 \times \mathcal{X}^2 \ ; \\ \big(\{x^1\} \cup \{y^1\}\big) \cap \big(\{x^2 \pm \delta\} \cup \{y^2 \pm \delta\}\big) = \emptyset \big\}. \quad (3.5)$$

As suggested in [62], when dealing with discretized point processes at resolution r, the right choice for δ is $kr + r/2$ for an integer k, so $(P^1 \otimes P^2)^{\otimes 2}(\mathcal{C}_\delta) = 1$, and $h_{\varphi_\delta^{coinc}}$ satisfies $(\mathcal{A}_{Cont})$. Furthermore, when dealing with independent point processes with conditional intensities, those processes may be constructed by thinning two independent Poisson processes X and X'. Hence, in this case, the probability $(P^1 \otimes P^2)^{\otimes 2}$ of $\mathcal{C}_\delta$ in (3.5) is larger than $\mathbb{P}\left(X \cap (X' \pm \delta) = \emptyset\right)$, whose value is 1. So when dealing with point processes with conditional intensities, $h_{\varphi_\delta^{coinc}}$ also satisfies $(\mathcal{A}_{Cont})$.

3.3 Consistency of the bootstrap approach

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

Theorem 3.1. *For every $n \geq 2$, let P_n^j for $j = 1,2$ be the empirical marginal distributions defined by (3.1). Then, under $(\mathcal{A}_{Cent})$, $(\mathcal{A}^*_{Cent})$, $(\mathcal{A}^*_{Mmt})$ and $(\mathcal{A}_{Cont})$,*

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 \middle| \mathbb{X}_n\right), \mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)\right) \underset{n\to+\infty}{\longrightarrow} 0, \ P\text{-a.s. in } (X_i)_i.$$

The proof follows similar arguments to the ones of [6] for the bootstrap of the mean, or to [15] and [40] for the bootstrap of U-statistics. The main novel point here consists in using the identification (3.4) and the properties of the separable Skorohod metric space $(\mathcal{D}, d_\mathcal{D})$, where weak convergence of sample probability distributions is available (see [7]). This theorem derives in fact 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 3.3. *Under $(\mathcal{A}_{Cent})$, $(\mathcal{A}^*_{Cent})$, $(\mathcal{A}^*_{Mmt})$, with the notation of Theorem 3.1, there exists an absolute constant $C > 0$ such that for $n \geq 2$,*

$$d_2^2\left(\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 \middle| \mathbb{X}_n\right), \mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)\right) \\ \leq C \inf\Big\{\mathbb{E}^*\Big[\big(h\left(Y^*_{n,a}, Y^*_{n,b}\right) - h\left(Y_a, Y_b\right)\big)^2\Big], Y^*_{n,a} \sim P_n^1 \otimes P_n^2, \ Y_a \sim P^1 \otimes P^2, \\ \text{and } (Y^*_{n,b}, Y_b) \text{ is an independent copy of } (Y^*_{n,a}, Y_a)\Big\}.$$

Comment. In the above proposition, the infimum is taken over all the possible distributions of $(Y^*_{n,a}, Y_a)$ having the correct marginals, $(Y^*_{n,b}, Y_b)$ being just an independent copy of $(Y^*_{n,a}, Y_a)$. In particular, $Y^*_{n,a}$ is not necessarily independent of Y_a.

Proposition 3.4. *If* $\mathbb{E}\left[|h(X_1, X_2)|\right] < +\infty$, *then*

$$U_n(\mathbb{X}_n) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\left[h(X_1, X_2)\right] = \int h(x, x')dP(x)dP(x'), P\text{-a.s. in } (X_i)_i. \tag{3.6}$$

Under $(\mathcal{A}^*_{Mmt})$, *one moreover obtains that* P*-a.s. in* $(X_i)_i$,

$$\frac{1}{n^4}\sum_{i,j,k,l=1}^{n} h^2\left(\left(X_i^1, X_j^2\right), \left(X_k^1, X_l^2\right)\right) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\left[h^2\left(\left(X_1^1, X_2^2\right), \left(X_3^1, X_4^2\right)\right)\right].$$

3.4 Convergence of cumulative distribution functions (c.d.f.) and quantiles

As usual, $\mathcal{N}(m, v)$ stands for the Gaussian distribution with mean m and variance v, $\Phi_{m,v}$ for its c.d.f. and $\Phi^{-1}_{m,v}$ for its quantile function. From the results of Rubin and Vitale [58] generalizing Hoeffding's [30] Central Limit Theorem for non-degenerate U-statistics when the X_i's are random vectors, we deduce the following Central Limit Theorem for $U_n(\mathbb{X}_n)$.

Proposition 3.5. *Assume that* h *satisfies* $(\mathcal{A}_{non-deg})$, $(\mathcal{A}_{Cent})$ *and* $(\mathcal{A}_{Mmt})$. *Let* $\sigma^2_{P^1\otimes P^2}$ *be defined by*

$$\sigma^2_{P^1\otimes P^2} = 4Var\left(\mathbb{E}\left[h\left(X_1, X_2\right)|X_1\right]\right), \tag{3.7}$$

when X_1 *and* X_2 *are* $P^1 \otimes P^2$*-distributed. Then*

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right), \mathcal{N}(0, \sigma^2_{P^1\otimes P^2})\right) \underset{n\to+\infty}{\longrightarrow} 0.$$

Comments.

(i) Notice that $(\mathcal{A}_{non-deg})$ is equivalent to $\sigma^2_{P^1\otimes P^2} > 0$. In the case where $(\mathcal{A}_{non-deg})$ does not hold, i.e. if $\sigma^2_{P^1\otimes P^2} = 0$, the quantity $\sqrt{n}U_n(\mathbb{X}_n)$ tends in probability towards 0. In this case, Theorem 3.1 implies that the two distributions $\mathcal{L}\left(\sqrt{n}U_n, P^1_n \otimes P^2_n | \mathbb{X}_n\right)$ and $\mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)$ are not only close, but that they are actually both tending to the Dirac mass in 0. Indeed, degenerate U-statistics of order 2 have a faster rate of convergence than $\sqrt{n}$ (see [3] for instance for explicit limit theorems). So in this degenerate case, one could not use $\sqrt{n}U_n(\mathbb{X}_n)$ as a test statistic anymore (without changing the normalization). But as mentioned above, $(\mathcal{A}_{non-deg})$ is usually satisfied in practice (see Section 2.4 for the *Coincidence case*).

(ii) The asymptotic normality result of Proposition 3.5 may lead to a rather simple test, asymptotically of the prescribed size. Let

$$S_n = \frac{\sqrt{n}U_n(\mathbb{X}_n)}{\hat{\sigma}}, \tag{3.8}$$

with

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

By Slutsky's lemma and the law of large numbers for U-statistics of order 3, under (H_0), S_n converges in distribution to $\mathcal{N}(0,1)$. The test rejecting (H_0) when $|S_n| \geq \Phi_{0,1}^{-1}(1-\alpha/2)$ is therefore asymptotically of size α. It is also consistent against any reasonable alternative P, satisfying $(\mathcal{A}_{Mmt})$ and such that $\mathbb{E}[h(X, X')] \neq 0$, for X, X' i.i.d. with distribution P. Such a purely asymptotic test 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 for biological reasons (from few tens up to few hundreds at best). Though the bootstrap approach is mainly justified by asymptotic arguments, the simulation study presented in Section 5 shows its efficiency in a non-asymptotic context, with respect to this simpler test.

As Proposition 3.5 implies that the limit distribution of $\sqrt{n}U_n(\mathbb{X}_n)$ has a continuous c.d.f., the convergence of the conditional c.d.f. or quantiles of the considered bootstrap distributions holds. Note that these conditional bootstrap distributions are discrete, so the corresponding quantile functions are to be understood as the generalized inverses of the cumulative distribution functions.

Corollary 3.1. *For $n \geq 2$, with the notation of Theorem 3.1, let $\mathbb{X}_n^*$ be a bootstrap sample, i.e. an i.i.d n-sample from the distribution $P_n^1 \otimes P_n^2$. Let $\mathbb{X}_n^{\perp\!\!\!\perp}$ be another i.i.d. n-sample from the distribution $P^1 \otimes P^2$ on $\mathcal{X}^2$. Under $(\mathcal{A}_{non-deg})$ and the assumptions of Theorem 3.1,*

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}\left(\sqrt{n}U_n\left(\mathbb{X}_n^*\right) \leq z \middle| \mathbb{X}_n\right) - \mathbb{P}\left(\sqrt{n}U_n(\mathbb{X}_n^{\perp\!\!\!\perp}) \leq z\right) \right| \underset{n \to +\infty}{\longrightarrow} 0, \ P\text{-a.s. in } (X_i)_i.$$

If moreover, for η in $(0,1)$, $q_{\eta,n}^*(\mathbb{X}_n)$ denotes the conditional η-quantile of $\sqrt{n}U_n(\mathbb{X}_n^*)$ given $\mathbb{X}_n$ and $q_{\eta,n}^{\perp\!\!\!\perp}$ denotes the η-quantile of $\sqrt{n}U_n(\mathbb{X}_n^{\perp\!\!\!\perp})$,

$$|q_{\eta,n}^*(\mathbb{X}_n) - q_{\eta,n}^{\perp\!\!\!\perp}| \underset{n \to +\infty}{\longrightarrow} 0, \ P\text{-a.s. in } (X_i)_i. \tag{3.9}$$

3.5 Asymptotic properties of the bootstrap tests

We are interested in the asymptotic behavior of sequences of tests all based on test statistics of the form $U_n(\mathbb{X}_n)$. The bootstrap approach, whose consistency is studied above, allows to define bootstrap-based critical values for these tests. Since the permutation approach studied in Section 4 is based on the same test statistics, but with critical values obtained by permutation, we introduce here a condensed and common formalism for the upper-, lower- and two-tailed tests considered in this work, taking into account that the only change in our two considered approaches concerns the critical values. This will help to state our results in the shortest manner.

Let α be fixed in $(0,1)$, and q be a sequence of upper and lower critical values:

$$q = \left(q^+_{\alpha,n}(\mathbb{X}_n), q^-_{\alpha,n}(\mathbb{X}_n)\right)_{n\geq 2}.$$

From this sequence of q, let us now define the family $\Gamma(q)$ of three sequences of tests $\Delta^+ = (\Delta^+_{\alpha,n})_{n\geq 2}$, $\Delta^- = (\Delta^-_{\alpha,n})_{n\geq 2}$, and $\Delta^{+/-} = (\Delta^{+/-}_{\alpha,n})_{n\geq 2}$, where

$$\left\{\begin{array}{lcll} \Delta^+_{\alpha,n}(\mathbb{X}_n) & = & \mathbb{1}_{\sqrt{n}U_n(\mathbb{X}_n) > q^+_{\alpha,n}(\mathbb{X}_n)} & \text{(upper-tailed test),} \\ \Delta^-_{\alpha,n}(\mathbb{X}_n) & = & \mathbb{1}_{\sqrt{n}U_n(\mathbb{X}_n) < q^-_{\alpha,n}(\mathbb{X}_n)} & \text{(lower-tailed test),} \\ \Delta^{+/-}_{\alpha,n}(\mathbb{X}_n) & = & \max\left(\Delta^+_{\alpha/2,n}(\mathbb{X}_n), \Delta^-_{\alpha/2,n}(\mathbb{X}_n)\right) & \text{(two-tailed test),} \end{array}\right. \tag{3.10}$$

the last test being implicitly defined by the corresponding choices in $\alpha/2$.

Of course, q, $\Gamma(q)$, as well as Δ^+, Δ^- and $\Delta^{+/-}$, depend on the choice of α, but since α is fixed at the beginning, to keep the notation as simple as possible, this dependence is, like the one in h, omitted in the notations.

Depending on the choice of q, the classical asymptotic properties that can be expected to be satisfied by $\Gamma(q)$ are $(\mathcal{P}_{size})$ and $(\mathcal{P}_{consist.})$ defined by

$(\mathcal{P}_{size})$ | Each sequence $\Delta = (\Delta_{\alpha,n})_{n\geq 2}$ in $\Gamma(q)$ is asymptotically of size α, i.e. if $P = P^1 \otimes P^2$, $\mathbb{P}\left(\Delta_{\alpha,n}(\mathbb{X}_n) = 1\right) \to_{n\to+\infty} \alpha$,

$(\mathcal{P}_{consist.})$ | Each sequence $\Delta = (\Delta_{\alpha,n})_{n\geq 2}$ in $\Gamma(q)$ is consistent, i.e. $\mathbb{P}\left(\Delta_{\alpha,n}(\mathbb{X}_n) = 1\right) \to_{n\to+\infty} 1$, for every P such that
- $\int h(x,x')dP(x)dP(x') > 0$ if $\Delta = \Delta^+$,
- $\int h(x,x')dP(x)dP(x') < 0$ if $\Delta = \Delta^-$,
- $\int h(x,x')dP(x)dP(x') \neq 0$ if $\Delta = \Delta^{+/-}$.

Following Corollary 3.1, our bootstrap tests family is defined from (3.10) by $\Gamma(q^*)$, with

$$q^* = \left(q^*_{1-\alpha,n}(\mathbb{X}_n), q^*_{\alpha,n}(\mathbb{X}_n)\right)_{n\geq 2}. \tag{3.11}$$

Theorem 3.2. *Let $\Gamma(q^*)$ be the family of tests defined by (3.10) and (3.11). If $(\mathcal{A}_{Cent})$, $(\mathcal{A}_{non-deg})$, $(\mathcal{A}^*_{Cent})$, $(\mathcal{A}^*_{Mmt})$ and $(\mathcal{A}_{Cont})$ hold, then $\Gamma(q^*)$ satisfies both $(\mathcal{P}_{size})$ and $(\mathcal{P}_{consist.})$.*

Comments. In the *Linear case* where h is equal to h_φ defined by (2.6),

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

This means that under the assumptions of Theorem 3.2, the two-tailed test of $\Gamma(q^*)$ is consistent against any alternative such that $\int \varphi(x^1,x^2)dP(x^1,x^2)$ differs from what is expected under (H_0), i.e. $\int \varphi(x^1,x^2)dP^1(x^1)dP^2(x^2)$.

(i) In particular, in the *Coincidence case* where h is equal to $h_{\varphi_\delta^{coinc}}$ defined by (2.5), the assumptions of Theorem 3.2 are fulfilled for instance if X^1 and X^2 are discretized at resolution r, with $\delta = kr + r/2$ for some integer k, or if X^1 and X^2 have bounded conditional intensities, with δ large enough so that $\varphi_\delta^{coinc}(X^1, X^2)$ is not a.s. null. Theorem 3.2 means in such cases that the corresponding two-tailed test is asymptotically of power 1, for any alternative P such that

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

Note that no δ ensuring this condition can be found 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). For neuroscientists, it means that the cross-correlogram (histogram of the delays, that is classically represented as the first description of the data) does not show different behaviors in the dependent and independent cases. This would only occur if the dependence could not be measured in terms of delay between points.

(ii) Furthermore, when φ is equal to φ^w defined by (2.2) with a continuous integrable function w (see Proposition 3.1), Theorem 3.2 means that the corresponding bilateral test is consistent against any alternative such that

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

For the function w chosen in [59] and under specific Poisson assumptions, β_w is linked to a 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 [59] states, after reformulation in the present setting, that if β_w is larger than an explicit lower bound, then the second kind error rate of the upper-tailed test is less than a prescribed β in $(0,1)$. Theorem 3.2 thus generalizes the result of [59] 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

The above family of bootstrap tests $\Gamma(q^*)$ involves an exact computation of the conditional quantiles $q^*_{\eta,n}(\mathbb{X}_n)$. Though such a computation is possible, it is not often reasonable in practice even when the sample size n is moderate (say $n \geq 15$), since computing $U_n(\mathbb{X}_n^*)$ itself may be complex from an algorithmic point of view, for particular choices of h. Therefore, the conditional quantiles $q^*_{\eta,n}(\mathbb{X}_n)$ are approximated, as usual, by a Monte Carlo method.

More precisely, for a chosen number $B \geq 1$ of iterations for the Monte Carlo method, let $\left(\mathbb{X}_n^{*1}, \ldots, \mathbb{X}_n^{*B}\right)$ be B independent bootstrap samples from $\mathbb{X}_n$, that is B i.i.d. random variables from the distribution $(P_n^1 \otimes P_n^2)^{\otimes n}$. For each b in $\{1, \ldots, B\}$, define $U^{*b} = U_n\left(\mathbb{X}_n^{*b}\right)$. The order statistic associated with $\left(U^{*1}, \ldots, U^{*B}\right)$ is denoted as usual by $\left(U^{*(1)}, \ldots, U^{*(B)}\right)$. Then, for any sequence $(B_n)_{n\geq 2}$ of positive integers, the family of Monte Carlo bootstrap tests is defined from (3.10) by

$\Gamma(q^*_{MC})$, with

$$q^*_{MC} = \left(\sqrt{n}U^{*(\lceil(1-\alpha)B_n\rceil)}, \sqrt{n}U^{*(\lfloor\alpha B_n\rfloor+1)}\right)_{n\geq 2}. \tag{3.12}$$

Proposition 3.6. *Let $\Gamma(q^*_{MC})$ be the family of tests defined by (3.10) and (3.12). If $B_n \to_{n\to+\infty} +\infty$, under the same assumptions as in Theorem 3.2, then $\Gamma(q^*_{MC})$ also satisfies both $(\mathcal{P}_{size})$ and $(\mathcal{P}_{consist.})$.*

4 Permutation tests of independence

4.1 Description of the permutation approach and review of its main known non-asymptotic properties

Permutation formalism The permutation approach we propose here consists in randomly permuting the second coordinates of the observed pairs of point processes. More precisely, if π_n is a permutation of $\{1,\dots,n\}$, one defines

$$\mathbb{X}_n^{\pi_n} = (X_1^{\pi_n},\dots,X_n^{\pi_n}) \text{ with } X_i^{\pi_n} = \left(X_i^1, X^2_{\pi_n(i)}\right), \tag{4.1}$$

and if Π_n is a random permutation taken uniformly at random in $\mathfrak{S}_n$ independently of $\mathbb{X}_n$, the corresponding permuted sample is defined by $\mathbb{X}_n^{\Pi_n}$. Let $P_n^\star$ be the conditional distribution of $\mathbb{X}_n^{\Pi_n}$ given $\mathbb{X}_n$.

For every $n \geq 2$ and η in $(0,1)$, let $q^\star_{\eta,n}(\mathbb{X}_n)$ denote the η-quantile of $\mathcal{L}\left(\sqrt{n}U_n, P_n^\star|\mathbb{X}_n\right)$. The corresponding sequence of upper and lower critical values is defined by

$$q^\star = \left(q^\star_{1-\alpha,n}\left(\mathbb{X}_n\right), q^\star_{\alpha,n}\left(\mathbb{X}_n\right)\right)_{n\geq 2}. \tag{4.2}$$

As for the bootstrap approach, the computation of $q^\star$ is usually not feasible for relatively small n and Monte-Carlo procedures come into play. Let $B \geq 1$ be a chosen number of iterations for the Monte Carlo method, and $\left(\Pi_n^1,\dots,\Pi_n^B\right)$ be a sample of i.i.d. random permutations uniformly distributed on $\mathfrak{S}_n$. For each b in $\{1,\dots,B\}$, we introduce $U^{\star b} = U_n\left(\mathbb{X}_n^{\Pi_n^b}\right)$, and we denote by $U^{\star B+1}$ the statistic $U_n\left(\mathbb{X}_n\right)$ computed on the original sample $\mathbb{X}_n$. The order statistic associated with $\left(U^{\star 1},\dots,U^{\star B+1}\right)$ is denoted as usual by $\left(U^{\star(1)},\dots,U^{\star(B+1)}\right)$. For a sequence $(B_n)_{n\geq 2}$ of positive integers, the sequence of upper and lower Monte Carlo permutation critical values is then defined by

$$q^\star_{MC} = \left(\sqrt{n}U^{\star(\lceil(1-\alpha)(B_n+1)\rceil)}, \sqrt{n}U^{\star(\lfloor\alpha(B_n+1)\rfloor+1)}\right)_{n\geq 2}. \tag{4.3}$$

Known non-asymptotic properties The main advantage of the permutation approach lies in the resulting tests being exactly of the desired level, even when the critical values are approximated by a Monte Carlo method. Such non-asymptotic results are well known (see for instance [57, Lemma 1] and [46]) and have

no known counterpart for the bootstrap approach, despite very few non-asymptotic results in more classical settings [4, 21, 20]. They are the main reason for usually preferring permutation to bootstrap when both methods are available (see [18]). We summarize and formalize these non-asymptotic properties in the next proposition.

Proposition 4.1. *Let α in $(0,1)$ and let $\Gamma(q^\star)$ be the family of permutation tests defined by (3.10) combined with (4.2). Then for every $n \geq 2$ and any test $\Delta_{\alpha,n}$ of $\Gamma(q^\star)$, $\Delta_{\alpha,n}$ is exactly of level α, i.e. if $P = P^1 \otimes P^2$, $\mathbb{P}(\Delta_{\alpha,n}(\mathbb{X}_n) = 1) \leq \alpha$.*

This also holds for the family of Monte Carlo permutation tests $\Gamma(q^\star_{MC})$, defined by (3.10) combined with (4.3), whatever the choice of the number of iterations $(B_n)_{n\geq 2}$.

4.2 Consistency of the permutation approach

In this section, we focus on the *Linear case* where h is of the form h_φ for some integrable function φ, as defined in (2.6). Indeed, it is the most general case for which we are able to prove a combinatorial Central Limit Theorem under any alternative as well as under the null hypothesis (Theorem 4.1). Hence in this section, U_n refers to U_{n,h_φ}. Notice that the centering assumption $(\mathcal{A}_{Cent})$ is then always satisfied by $U_n(\mathbb{X}_n)$. We here only need the following moment assumption:

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

Though we have no exact counterpart of Theorem 3.1 for our permutation approach, the following result combined with Proposition 3.5 gives a similar result.

Theorem 4.1. *For all $n \geq 2$, let $P_n^\star$ be the conditional distribution of a permuted sample given $\mathbb{X}_n$. In the* Linear case *where the kernel h is of the form (2.6) for an integrable function φ, under $(\mathcal{A}_{non-deg})$ and $(\mathcal{A}_{\varphi,Mmt})$ and with the notations of Section 3,*

$$d_2\left(\mathcal{L}\left(\sqrt{n}U_n, P_n^\star \middle| \mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0, \tag{4.4}$$

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

Comments. As pointed out above, unlike the bootstrap approach, the conditional permutation distribution of the test statistic is not here directly compared with the initial distribution of the test statistic under the null hypothesis. It is in fact compared with the Gaussian limit distribution of the test statistic under the null hypothesis, when the non-degeneracy assumption $(\mathcal{A}_{non-deg})$ holds. Moreover, the convergence occurs here in probability and not almost surely, but note that no continuity assumption for the kernel h_φ is used anymore. The price to pay is that the moment assumption is stronger than the one used for the bootstrap. This assumption, due to our choice to use an existing Central Limit Theorem for martingale difference arrays in the proof, is probably merely technical and maybe dispensable. Indeed, the result of Theorem 4.1 is close to asymptotic results for permutation known as *combinatorial Central Limit Theorems* [32, 46], where this kind of higher moment assumption can be replaced by some Lindeberg conditions

[41, 29]. However, all these existing results can only be applied directly in our case either when $(X_i)_i$ is deterministic or under the null hypothesis. Up to our knowledge, no combinatorial Central Limit Theorem has been proved for non deterministic non exchangeable variables, like here under any alternative.

The above result is therefore one of the newest results presented here and its scope is well beyond the only generalization to the point processes setting. Indeed, because it holds not only under (H_0) but also under (H_1), it goes further than any existing one for independence test statistics such as the ones of Romano [56]. Note also that the behavior under dependence of the permuted test statistic of Van der Vaart and Wellner was also left as an open question in [63].

The proof is presented in the supplementary material.

From Theorem 4.1, we deduce the following corollary.

Corollary 4.1. *Under the assumptions of Theorem 4.1 and with the notations of Proposition 3.5, for η in $(0,1)$,*

$$q^{\star}_{\eta,n}\left(\mathbb{X}_n\right) \xrightarrow[n\to+\infty]{\mathbb{P}} \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(\eta).$$

4.3 Asymptotic properties of the permutation tests

As for the bootstrap tests, we obtain the following results.

Theorem 4.2. *Let $\Gamma(q^{\star})$ be the family of permutation tests defined by (3.10) and (4.2). In the* Linear case*, if $(\mathcal{A}_{non-deg})$ and $(\mathcal{A}_{\varphi,Mmt})$ hold, then $\Gamma(q^{\star})$ satisfies both $(\mathcal{P}_{size})$ and $(\mathcal{P}_{consist.})$. Moreover, if $B_n \to_{n\to+\infty} +\infty$, the family of Monte Carlo permutation tests $\Gamma(q^{\star}_{MC})$, defined by (3.10) and (4.3), also satisfies both $(\mathcal{P}_{size})$ and $(\mathcal{P}_{consist.})$.*

5 Simulation study

In this section, we study our testing procedures from a practical point of view, by giving estimations of the size and the power for various underlying distributions that are coherent with real neuronal data. This allows to verify the usability of these new methods in practice, and to compare them to existing classical methods. A real data sets study and a more operational and complete method for neuroscientists derived from the present ones is the subject of an ongoing work. The programs have been optimized, parallelized in `C++` and interfaced with `R`. The code is available at `https://code.google.com/p/neuro-stat/`.

5.1 Presentation of the study

All along the study, h is taken equal to $h_{\varphi^{coinc}_{\delta}}$ (see (2.5)), where φ^{coinc}_{δ} is defined in (2.1) and $\alpha = 0.05$. We only present the results for upper-tailed tests, but an analogous study has been performed for lower-tailed tests with similar results. Five different testing procedures are compared.

5.1.1 Testing procedures

(CLT) Test based on the Central Limit Theorem for U-statistics (see Proposition 3.5) which rejects (H_0) when the test statistic S_n in (3.8) is larger than the $(1-\alpha)$-quantile of the standard normal distribution.

(B) Monte Carlo bootstrap upper-tailed test of $\Gamma(q^*_{MC})$ ((3.10) and (3.12)).

(P) Monte Carlo permutation upper-tailed test of $\Gamma(q^\star_{MC})$ ((3.10) and (4.3)).

(GA) Upper-tailed tests introduced in [62, Definition 3] under the notation $\Delta^+_{GAUE}(\alpha)$, based on a Gaussian approximation of the total number of coincidences.

(TS) Trial-shuffling test based on a Monte Carlo approximation of the p-value introduced in [47, equation (3)], but adapted to the present notion of coincidences. This test is the reference distribution-free method for neuroscientists. To be more precise, let $C(\mathbb{X}_n) = \sum_{i=1}^n \varphi_\delta^{coinc}\left(X_i^1, X_i^2\right)$ be the total number of coincidences of $\mathbb{X}_n$. The trial-shuffling method consists in uniformly drawing with replacement n i.i.d. pairs of indices $\{(i^*(k), j^*(k))\}_{1\leq k\leq n}$ in $\{(i,j), 1\leq i\neq j\leq n\}$, and considering the associated TS-sample $\mathbb{X}_n^{TS} = \left(\left(X^1_{i^*(k)}, X^2_{j^*(k)}\right)\right)_{1\leq k\leq n}$. The Monte Carlo p-value is defined by $\alpha_B^{TS} = \frac{1}{B}\sum_{b=1}^B \mathbb{1}_{C\left(\mathbb{X}_n^{TS,b}\right)\geq C(\mathbb{X}_n)}$, where $\left(\mathbb{X}_n^{TS,1}, \ldots, \mathbb{X}_n^{TS,B}\right)$ are B independent TS-samples, and the test rejects (H_0) if $\alpha_B^{TS} \leq \alpha$. This procedure is therefore close in spirit to our bootstrap procedure except that it is applied on a non-centered quantity under (H_0), namely $C(\mathbb{X}_n)$.

The number B of steps in the Monte Carlo methods is taken equal to 10000.

5.1.2 Simulated data

Various types of point processes are simulated here to check the distribution-free character of our approaches and to investigate their limits. Of course, each of the considered point processes satisfies the moment assumptions on the number of points so that the theorems in this article can be applied. From now on and to be coherent with the neuroscience application which originally motivated this work, the point processes are simulated on $[0, 0.1]$. Indeed the following experiments have been done to match neurophysiological parameters [62, 25] and the classical necessary window for detection is usually of duration 0.1 seconds.

Estimation of the size The three data sets simulated under (H_0) consist of i.i.d. samples of pairs of independent point processes. For simplicity, both processes have the same distribution, though this is not required.

Exp. A Homogeneous Poisson processes on $[0, 0.1]$ with intensity $\lambda = 60$.

Exp. B Inhomogeneous Poisson processes with intensity $f_\lambda : t \in [0, 0.1] \mapsto \lambda t$ and $\lambda = 60$.

Exp. C Hawkes Processes as detailed in [62] i.e. point processes with conditional intensity $\lambda(t) = \max\left(0, \mu - \int_0^t \nu \, \mathbb{1}_{]0,r]}(t-s)\, dN_X(s)\right)$, for t in $[0, 0.1]$, with spontaneous intensity $\mu = 60$, refractory period $r = 0.001$, and $\nu > \mu$ such that for all point T in X and t in $]T, T+r]$, $\lambda(t) = 0$. This choice of ν prevents two points to occur at a distance less than the refractory period r to reflect typical neuronal behavior. This model is also sometimes called Poisson process with dead time.

Study of the power The three data sets simulated under (H_1) are such that the number of coincidences is larger than expected under (H_0). The models (injection or Hawkes) are classical in neuroscience and already used in [62, 26].

Exp. D Homogeneous injection model. $X^1 = X^1_{ind} \cup X_{com}$ and $X^2 = X^2_{ind} \cup X_{com}$, with X^1_{ind} and X^2_{ind} being two independent homogeneous Poisson processes with intensity $\lambda_{ind} = 54$, X_{com} being a common homogeneous Poisson process with intensity $\lambda_{com} = 6$.

Exp. E Inhomogeneous injection model. Similar to **Exp. D**, except that both X^1_{ind} and X^2_{ind} are independent inhomogeneous Poisson processes with intensity $f_{\lambda_{ind}}$ (see **Exp. B**) and $\lambda_{ind} = 54$.

Exp. F Dependent bivariate Hawkes processes. The coordinates X^1 and X^2 of a same pair respectively have the conditional intensities:
$\lambda^1(t) = \max\left\{0, \mu - \int_0^t \nu \mathbb{1}_{]0,r]}(t-s)\, dN_{X^1}(s) + \int_0^t \eta \mathbb{1}_{]0,u]}(t-s)\, dN_{X^2}(s)\right\}$,
$\lambda^2(t) = \max\left\{0, \mu - \int_0^t \nu \mathbb{1}_{]0,r]}(t-s)\, dN_{X^2}(s) + \int_0^t \eta \mathbb{1}_{]0,u]}(t-s)\, dN_{X^1}(s)\right\}$, with the spontaneous intensity $\mu = 54$, the interaction intensity $\eta = 6$ in the period designated by $u = 0.005$ and the refractory period designated by $r = 0.001$ with $\nu \gg \mu + \eta u$ such that once again, $\lambda^j(t)$ is null on each $]T, T+r]$, for T in X^j. We arbitrarily took $\nu = 50(2\mu + \eta)$.

5.2 Results

Varying number of trials n In Figure 1, the delay is fixed at $\delta = 0.01$ and the number n of trials varies in $\{10, 20, 50, 100\}$. Note that when the number of trials is too small ($n = 10$), the estimated variance in (`CLT`) is sometimes negative, therefore, the test cannot be implemented.

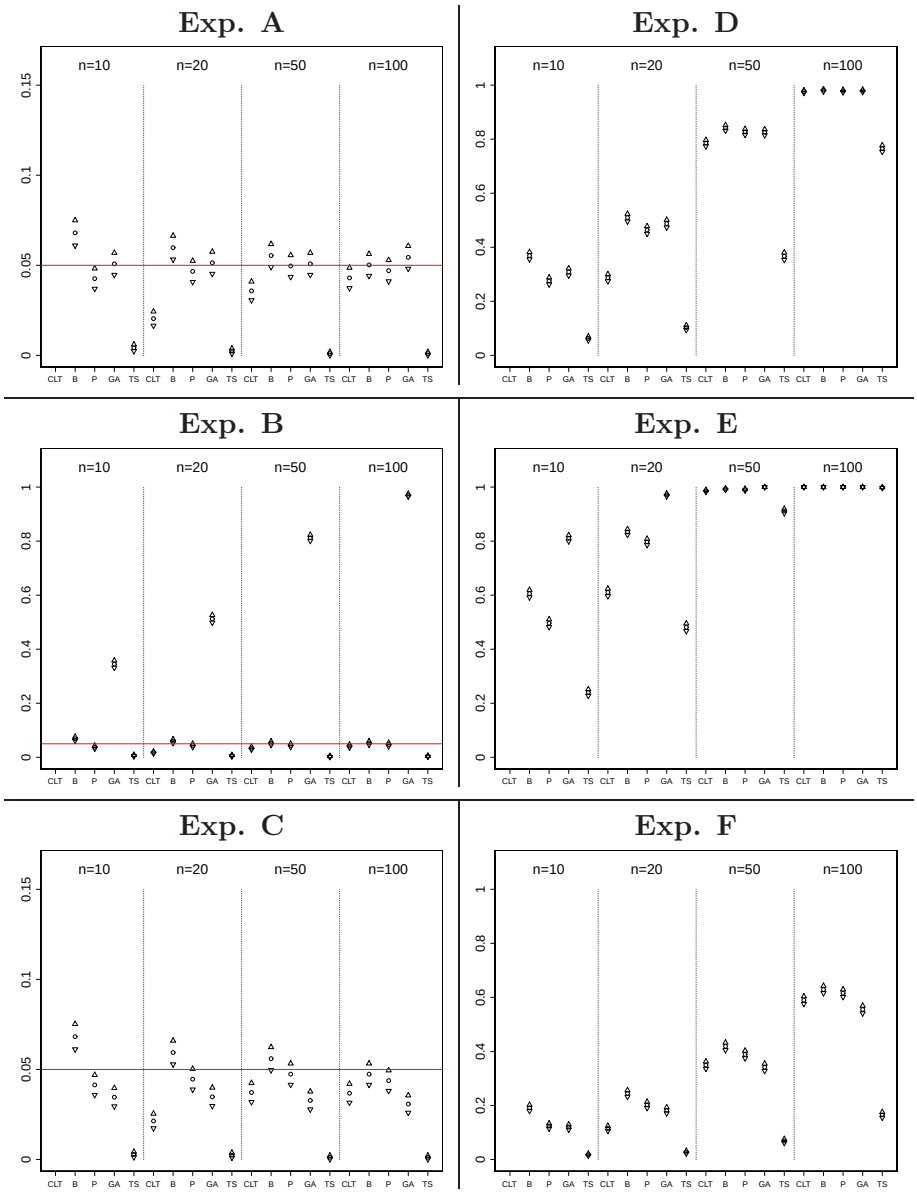

Figure 1 – Estimated sizes and powers for different number of trials n, all the tests being performed with a level $\alpha = 0.05$. The circles represent the percentage of rejection on 5000 simulations for each different method, the triangles represent the corresponding endpoints of a 95% confidence interval. The corresponding experiments are described in Section 5.1.2.

The left hand side of Figure 1 corresponds to estimated sizes. On the one hand, one can see in the case of homogeneous Poisson processes (**Exp. A**) and in the case of refractory Hawkes processes (**Exp. C**) that the methods (`CLT`), (`B`), (`P`) and (`GA`) are quite equivalent, but the size (first kind error rate) seems less controlled in the bootstrap approach (`B`) especially for small numbers of trials. Yet, one can see the convergence of the size of the bootstrap test towards α as the number of trials goes to infinity, which illustrates Proposition 3.6. Note that the (`CLT`) test also has a well controlled size even if it cannot be used for very small n. On the other hand, in the case of inhomogeneous Poisson processes (**Exp. B**), one can see that the (`GA`) test has a huge size and is thus inadequate here. Indeed it is based on the strong assumption that the data are homogeneous Poisson processes though they are in fact strongly nonstationary. The test tends thus to reject the independence null hypothesis even when the data are independent. Finally, in the three considered cases, the (`TS`) approach has a very small size, and is thus too conservative as one can see in the power study. We guess, from our experiments, that this lack of performance is mainly due to the fact that the (`TS`) approach is applied here on a non-centered quantity under (H_0), namely $C(\mathbb{X}_n)$.

The right hand side of Figure 1 corresponds to estimated powers, which increase as n grows. This is in line with the consistency of the tests. Now, as it could be expected when looking at its estimated sizes, for the (`TS`) approach, the estimated powers are distinctly lower than the ones for the other methods, which confirms its conservative behavior. The other approaches are more similar in **Exp. D** or **Exp. F** though (`B`) clearly seems to outperform all tests, but at the price of a less controlled size. Note that in the inhomogeneous case (**Exp. E**), (`GA`) seems to have the best power, but this time, at the price of a totally uncontrolled size.

This part of the simulation study illustrates the convergences of the size and the power of the bootstrap and permutation tests introduced here. The permutation approach seems to actually guarantee the best control of the size as expected, as compared with the bootstrap approach. Nevertheless both approaches are quite effective for any considered kind of point processes and any sample size, unlike the (`GA`) test which has very restrictive assumptions. The reference method (`TS`) for neuroscientists is clearly too conservative. Moreover, the (`CLT`) test seems to have also satisfying results, but with a slower convergence than the (`B`) and (`P`) ones. This seems to illustrate that the conditional bootstrap and permutation distributions give better approximations of the original one under independence than a simple Central Limit Theorem. This phenomenon is well-known as the second order accuracy of the bootstrap in more classical frameworks.

Varying delay δ We now investigate the impact of the choice for the delay δ by making δ vary in $\{0.001, 0.005, 0.01, 0.02\}$ for a fixed number of trials $n = 50$. The results for the sizes being similar to the previous study, only the estimated powers are presented in Figure 2.

On the top row of Figure 2, the same process is injected in both coordinates: the coincidences are exact in the sense that they have no delay. Therefore, the best choice for the delay parameter δ is the smallest possible value: the obtained

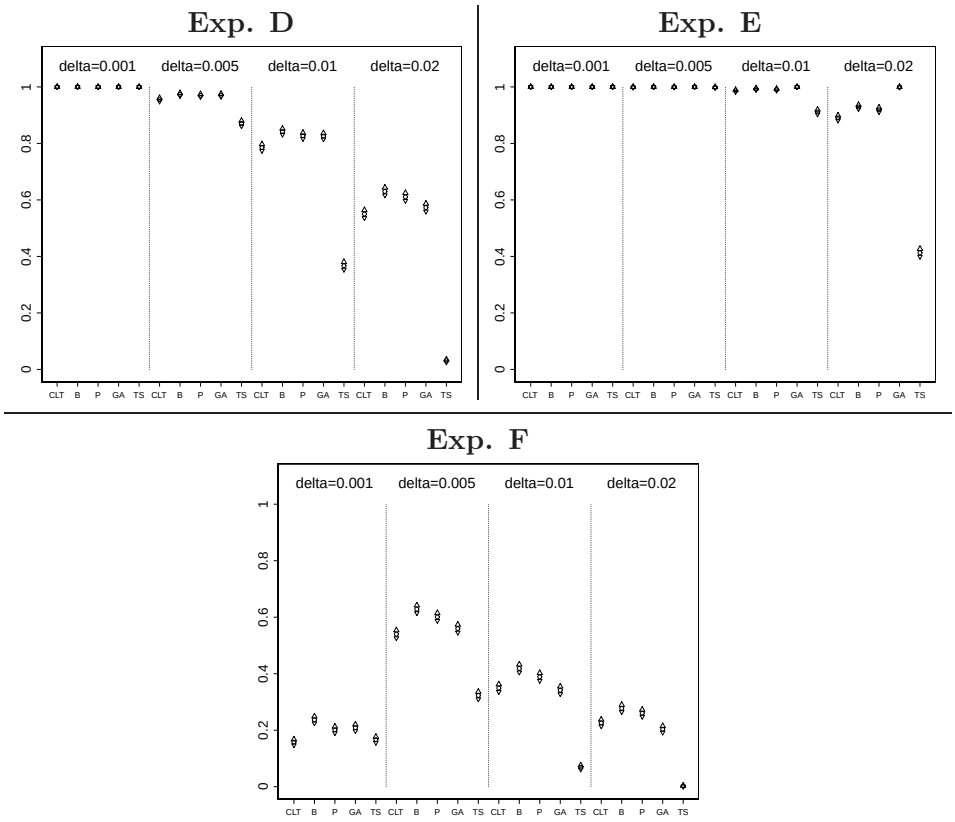

Figure 2 – Estimated powers for different δ. Same convention as in Figure 1.

power is 1 for very small δ's (e.g. $\delta = 0.001$) and then decreases as δ increases. On the contrary on the bottom row, it can be noticed that the highest power is for $\delta = 0.005$ which is the exact length of the interaction period u. Once again, the (`TS`) method performs poorly, as does the (`CLT`) method. The three other methods seem to be quite equivalent except in the inhomogeneous case (**Exp. E**) where the (`GA`) method has a power always equal to 1, but at the price of an uncontrolled size.

6 Conclusion

In the present article, we have proposed non-parametric independence tests between point processes based on U-statistics. The proposed critical values are obtained either by bootstrap or permutation methods. We have shown that both methods share the same asymptotic properties under the null hypothesis as well as under the alternative. From a theoretical point of view, the main asymptotic results (Theorem

3.1 and Theorem 4.1) have almost the same flavor. However, there are additional assumptions in the permutation case which make the bootstrap results more general (despite the additional continuity assumption, which is very mild). From a more concrete point of view, it is acknowledged (see e.g. [18]) that permutation should be preferred because of its very general non-asymptotic properties (see Proposition 4.1). This is confirmed by the experimental study, where clearly permutation leads to a better first kind error rate control. However, both approaches perform much better than a naive procedure, based on a basic application of a Central Limit Theorem, when the number of observation is small. They also outperform existing procedures of the neuroscience literature, namely [62], which assumes the point processes to be homogeneous Poisson processes and the trial-shuffling procedures [48, 47], which are biased bootstrap variants applied on a non-centered quantity.

One of the main open question with respect to the existing literature is whether our results can be translated to test statistics of the form $\sup_h U_{n,h}$. There are obviously two main obstacles. The first one lies in the nature of the observed random variables, that are point processes, and the fact that controlling a supremum of this type corresponds to a control of the whole U-process. This difficulty can probably be overcome, since the asymptotic Gaussian behavior of similar statistics has already been proved in very general spaces under (H_0) for product type kernels (see [9]). The study of such behavior under (H_1) is surely much more complex. The second obstacle comes from a more practical aspect. In neurosciences, and in the particular case of coincidence count, the use of $\sup_\delta U_{n,h_{\varphi_\delta^{coinc}}}$ leads to the following fundamental problems. On the one hand, it is not clear that one could easily compute such a statistic if the space where the supremum is taken is too large, typically $[0,1]$. On the other (more important) hand, neuroscientists are especially interested in the value of δ which leads to a rejection, since it actually provides the delay of interaction (see also Section 5). In this respect, an ongoing work involves multiple testing aspects, which seem to answer this issue.

Acknowledgments

We are grateful to both referees, whose discussions and comments allowed us to improve the present article. We thank F. Grammont for fruitful discussions. This work was granted access to the HPC and visualization resources of "Centre de Calcul Interactif" hosted by "Université Nice Sophia Antipolis". This research was 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.* The PhD grant of M. Albert is funded by the PACA french region.

Supplement

Supplement: Technical Results and Proofs of "Bootstrap and Permutation tests of independence for Point Processes"

This Supplement consists of all the proofs. It also contains some additional results about non-degeneracy and the empirical centering assumption. It can be found at the end of this article.

References

[1] J. Antoch and M. Hušková. Permutation tests in change point analysis. *Statist. Probab. Lett.*, 53(1):37–46, 2001.

[2] M. A. Arcones and E. Giné. On the bootstrap of U and V statistics. *Ann. Statist.*, 20(2):655–674, 1992.

[3] Miguel A. Arcones and Evarist Giné. Limit theorems for U-processes. *Ann. Probab.*, 21(3):1494–1542, 1993.

[4] S. Arlot, G. Blanchard, and E. Roquain. Some nonasymptotic results on resampling in high dimension, I: Confidence regions and II: Multiple tests. *Ann. Statist.*, 38(1):51–82, 83–99, 2010.

[5] James N Arvesen. Jackknifing U-statistics. *The Annals of Mathematical Statistics*, 40(6):2076–2100, 1969.

[6] P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9(6):1196–1217, 1981.

[7] P. Billingsley. *Convergence of probability measures*. Wiley-Interscience, 2009.

[8] J. R. Blum, J. Kiefer, and M. Rosenblatt. Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, 32(2):485–498, 1961.

[9] S. Bouzebda. General tests of independence based on empirical processes indexed by functions. *Statistical Methodology*, 21:59–87, 2014.

[10] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *Ann. Probab.*, 24(3):1563–1588, 1996.

[11] Pierre Brémaud. *Point processes and queues*. Springer-Verlag, New York-Berlin, 1981. Martingale dynamics, Springer Series in Statistics.

[12] J. Bretagnolle. Lois limites du bootstrap de certaines fonctionnelles. *Ann. Inst. H. Poincaré Probab. Statist.*, 19(3):281–296, 1983.

[13] E. Chung and J. P. Romano. Exact and asymptotically robust permutation tests. *Ann. Statist.*, 41(2):484–507, 2013.

[14] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I.* Probability and its Applications. Springer-Verlag, New York, 2003.

[15] H. Dehling and T. Mikosch. Random quadratic forms and the bootstrap for U-statistics. *J. Multivariate Anal.*, 51(2):392–413, 1994.

[16] R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2002.

[17] B. Efron. Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1):1–26, 1979.

[18] B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1993.

[19] R. A. Fisher. *The design of experiments*. 1935.

[20] M. Fromont. Model selection by bootstrap penalization for classification. *Mach. Learn.*, 66(2-3):165–207, 2007.

[21] M. Fromont, B. Laurent, and P. Reynaud-Bouret. The two-sample problem for poisson processes: Adaptive tests with a nonasymptotic wild bootstrap approach. *Ann. Statist.*, 41(3):1431–1461, 2013.

[22] Magalie Fromont, Béatrice Laurent, Matthieu Lerasle, and Patricia Reynaud-Bouret. Kernels based tests with non-asymptotic bootstrap approaches for two-sample problems. *Journal of Machine Learning Research: Workshop and Conference Proceedings, COLT 2012*, 23:23.1–23.22, 2012.

[23] G. L. Gerstein and D. H. Perkel. Simultaneous recorded trains of action potentials: analysis and functional interpretation. *Science*, 164:828–830, 1969.

[24] A. Gretton and L. Györfi. Consistent nonparametric tests of independence. *J.M.L.R.*, 11:1391–1423, 2010.

[25] S. Grün. *Unitary joint-events in multiple-neuron spiking activity: Detection, significance and interpretation.* PhD thesis, Thun: Verlag Harri Deutsch, 1996.

[26] S. Grün, M. Diesmann, and A. M. Aertsen. *Analysis of parallel spike trains*, chapter Unitary Events analysis. Springer Series in Computational Neuroscience, 2010.

[27] S. Grün, M. Diesmann, F. Grammont, A. Riehle, and A. M. Aertsen. Detecting unitary events without discretization of time. *J. Neuro. Meth.*, 93:67–79, 1999.

[28] N. R. Hansen, P. Reynaud-Bouret, and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *To appear in Bernoulli*, 2014.

[29] S.T. Ho and L. H. Y. Chen. An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probability*, 6(2):231–249, 1978.

[30] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.*, 19(3):293–325, 1948.

[31] W. Hoeffding. A non-parametric test of independence. *Ann. Math. Statist.*, 19(4):546–557, 1948.

[32] W. Hoeffding. A combinatorial central limit theorem. *Ann. Math. Statistics*, 22:558–566, 1951.

[33] W. Hoeffding. The large-sample power of tests based on permutation of the observations. *Ann. Math. Statist.*, 23(2):169–192, 1952.

[34] L. Horváth and M. Hušková. Testing for changes using permutations of U-statistics. *J. Statist. Plann. Inference*, 128(2):351–371, 2005.

[35] H. Hotelling and M. R. Pabst. Rank correlation and tests of significance involving no assumption of normality. *Ann. Math. Statist.*, 7(1):29–43, 1936.

[36] Arnold Janssen and Thorsten Pauls. How do bootstrap and permutation tests work? *Annals of statistics*, 31(3):768–806, 2003.

[37] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1–2):81–93, 1938.

[38] C. Kirch and J. Steinebach. Permutation principles for the change analysis of stochastic processes under strong invariance. *J. Comput. Appl. Math.*, 186(1):64–88, 2006.

[39] L. Klebanov, A. Gordon, Y. Xiao, H. Land, and A. Yakovlev. A permutation test motivated by microarray data analysis. *Comput. Statist. Data Anal.*, 50(12):3619–3628, 2006.

[40] A. Leucht and M. H. Neumann. Consistency of general bootstrap methods for degenerate U-type and V-type statistics. *J. Multivariate Anal.*, 100(8):1622–1633, 2009.

[41] M. Motoo. On the Hoeffding's combinatorial central limit theorem. *Ann. Inst. Statist. Math. Tokyo*, 8:145–154, 1957.

[42] Y. Ogata. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.

[43] 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.

[44] K. Pearson. On the probability that two independent distributions of frequency are really samples from the same population. *Biometrika*, 8:250–254, 1911.

[45] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter. How structure determines correlations in neuronal networks. *PLoS Computational Biology*, 7(5):e1002059, 2012.

[46] F. Pesarin and L. Salmaso. *Permutation tests for complex data: Theory, Applications and Software*. Series in Probability and Statistics. Wiley, 2010.

[47] 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.

[48] G. Pipa and S. Grün. Non-parametric significance estimation of joint-spike events by shuffling and resampling. *Neurocomputing*, 52–54:31–37, 2003.

[49] E. J. Pitman. Significance tests which may be applied to samples from any populations. *Suppl. J. R. Stat. Soc.*, 4:117–130, 225–232, 1937.

[50] E. J. Pitman. Significance tests which may be applied to samples from any populations. The analysis of variance test. *Biometrika*, 29:322–335, 1938.

[51] C. Pouzat and A. Chaffiol. Automatic spike train analysis and report generation. an implementation with R, R2HTML and STAR. *J. Neurosci. Meth.*, 2009.

[52] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *J. Math. Neuro.*, 4(3), 2014.

[53] 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.

[54] J. P. Romano. Bootstrap and randomization tests of some nonparametric hypotheses. Technical Report 270, Dept. Statistics, Stanford Univ., 1987.

[55] J. P. Romano. A bootstrap revival of some nonparametric distance tests. *J. Amer. Statist. Assoc.*, 83(403):698–708, 1988.

[56] J. P. Romano. Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.*, 17(1):141–159, 1989.

[57] J. P. Romano and M. Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc.*, 100(469):94–108, 2005.

[58] H. Rubin and R. A. Vitale. Asymptotic distribution of symmetric statistics. *Ann. Statist.*, pages 165–170, 1980.

[59] L. Sansonnet and C. Tuleau-Malot. A model of Poissonian interactions and detection of dependence. *to appear in Stat. Comput.*, 2014.

[60] H. Scheffe. Statistical inference in the non-parametric case. *Ann. Math. Statist.*, 14(4):305–332, 1943.

[61] W. Singer. Synchronization of cortical activity and its putative role in information processing and learning. *Annu. Rev. Physio.*, 55:349–374, 1993.

[62] C. Tuleau-Malot, A. Rouis, F. Grammont, and P. Reynaud-Bouret. Multiple tests based on a Gaussian approximation of the Unitary Events method. *Neural Computation*, 26(7), 2014.

[63] A. W. Van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, 1996.

[64] V. S. Varadarajan. On the convergence of sample probability distributions. *Sankhyà*, 19(1–2):23–26, 1958.

[65] J. Wolfowitz. Additive partition functions and a class of statistical hypotheses. *Ann. Math. Statist.*, 13(3):247–279, 2042.

Technical Results and Proofs of "Bootstrap and permutation tests of independence for point processes"

The references of Equations, Theorems, Propositions, etc, that use only numbers such as (3.1) for instance, refer to the main article *Bootstrap and permutation tests of independence for point processes*.

A Complete Proofs

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

A.1 Proof of Proposition 2.1

We focus on the *Coincidence case*. According to the comment following the definition of $(\mathcal{A}_{non-deg})$, $U_n(\mathbb{X}_n)$ is non-degenerate under (H_0) if one can find some borelian set $\mathcal{B}$ of $\mathcal{X}^2$ such that $P^1 \otimes P^2(\mathcal{B}) > 0$ and such that for all x in $\mathcal{B}$, $\mathbb{E}\left[h_{\varphi_\delta^{coinc}}(x, X)\right] \neq 0$, where X has distribution $P^1 \otimes P^2$.

Consider $\mathcal{B} = \{(\emptyset, \emptyset)\}$. Then $P^1 \otimes P^2(\mathcal{B}) = P^1(\emptyset) P^2(\emptyset) > 0$.
Moreover, as $\varphi_\delta^{coinc}(\cdot, \emptyset)$ and $\varphi_\delta^{coinc}(\emptyset, \cdot)$ are both the zero function, under (H_0),

$$\mathbb{E}\left[h_{\varphi_\delta^{coinc}}((\emptyset, \emptyset), X)\right] = \frac{1}{2}\mathbb{E}\left[\varphi_\delta^{coinc}\left(X^1, X^2\right)\right] > 0,$$

as $\varphi_\delta^{coinc}\left(X^1, X^2\right)$ is non-negative and not almost surely null under (H_0).

See also appendix B.1 for further results on the non-degeneracy of the U-statistic in more general cases.

A.2 Proof of Proposition 3.1

Consider $w : [0,1]^2 \to \mathbb{R}$ a continuous integrable function. Let us prove that $h = h_{\varphi^w}$ given by (2.2) and (2.6) is continuous for the topology induced by d (see (3.4)). Recall that for $x_1 = (x_1^1, x_1^2)$ and $x_2 = (x_2^1, x_2^2)$ in $(\mathcal{X}^2)$,

$$h_{\varphi^w}(x_1, x_2) = \frac{1}{2}\left(\varphi^w(x_1^1, x_1^2) + \varphi^w(x_2^1, x_2^2) - \varphi^w(x_1^1, x_2^2) - \varphi^w(x_2^1, x_1^2)\right).$$

The first step is to show that for each i, j in $\{1, 2\}$, the projection defined by

$$p_{i,j} : \begin{pmatrix} ((\mathcal{X}^2)^2, d) & \longrightarrow & (\mathcal{X}^2, d_{\mathcal{X}^2}) \\ \left((x_1^1, x_1^2), (x_2^1, x_2^2)\right) & \longmapsto & (x_i^1, x_j^2) \end{pmatrix},$$

is continuous. Let $\mathbf{x} = \left(\left(x_1^2, x_1^2\right), \left(x_2^1, x_2^2\right)\right)$ and $\mathbf{x}' = \left(\left(x'^1_1, {}'^2_1\right), \left(x'^1_2, x'^2_2\right)\right)$ in $(\mathcal{X}^2)^2$. Then,

$$d_{\mathcal{X}^2}(p_{i,j}(\mathbf{x}), p_{i,j}(\mathbf{x}')) = d_{\mathcal{X}^2}\left(\left(x_i^1, x_j^2\right), \left(x'^1_i, x'^2_j\right)\right) \leq d(\mathbf{x}, \mathbf{x}')\,.$$

Hence, $p_{i,j}$ is 1-Lipschitz and therefore continuous.

The second step is to show that if w is continuous on $\left([0,1]^2, \|\cdot\|_\infty\right)$, with $\|(u,v) - (u',v')\|_\infty = \max\{|u-u'|, |v-v'|\}$, then φ^w is also continuous.
Let $\varepsilon > 0$ and for z in $\mathcal{X}$, recall that N_z is the counting process associated with z, defined by

$$N_z(t) = \int_0^1 \mathbb{1}_{u \leq t} dN_z(u).$$

First notice that, w being continuous on the compact set $[0,1]^2$, w is uniformly continuous. Thus one can find some η in $(0,1)$ such that, for all (u,v), (u',v') in $[0,1]^2$,

$$\|(u,v) - (u',v')\|_\infty \leq \eta \quad \text{implies} \quad |w(u,v) - w(u',v')| \leq \varepsilon. \tag{A.1}$$

Consider such η.
Let $\{x_n\}_{n\geq 0}$ be a sequence in $\mathcal{X}^2$ such that $d_{\mathcal{X}^2}(x_n, x_0) \underset{n\to+\infty}{\longrightarrow} 0$ and let us show that $\varphi^w(x_n) \underset{n\to+\infty}{\longrightarrow} \varphi^w(x_0)$. There exists n_0 in $\mathbb{N}$ such that for all $n \geq n_0$, $d_{\mathcal{X}^2}(x_n, x_0) \leq \eta$. Then, for such n, by definition of $d_{\mathcal{X}^2}$, we have that $d_{\mathcal{D}}\left(N_{x_n^1}, N_{x_0^1}\right) \leq \eta$ and $d_{\mathcal{D}}\left(N_{x_n^2}, N_{x_0^2}\right) \leq \eta$. Thus, by definition of $d_{\mathcal{D}}$,

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

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

In particular, as η is chosen strictly smaller than 1 and as the $N_{x_n^j}$'s ($n \geq 0$, $j = 1, 2$) are counting processes with values in $\mathbb{N}$, $(1\text{-}ii)$ implies that $\forall t \in [0,1]$, $N_{x_n^1}(t) = N_{x_0^1}\left(\lambda_n^1(t)\right)$ and thus,

$$u_0 \in x_0^1 \quad \Leftrightarrow \quad u_n = \lambda_n^1(u_0) \in x_n^1.$$

Similarly, $(2\text{-}ii)$ implies that

$$v_0 \in x_0^2 \quad \Leftrightarrow \quad v_n = \lambda_n^2(v_0) \in x_n^2.$$

Therefore,

$$\begin{aligned}\varphi^w(x_n) &= \iint_{[0,1]^2} w(u,v)dN_{x_n^1}(u)dN_{x_n^2}(v)\\ &= \sum_{(u_n,v_n)\in x_n^1\times x_n^2} w(u_n,v_n)\\ &= \sum_{(u_0,v_0)\in x_0^1\times x_0^2} w\left(\lambda_n^1(u_0),\lambda_n^2(v_0)\right).\end{aligned}$$

Hence,

$$|\varphi^w(x_n)-\varphi^w(x_0)| \leq \sum_{(u_0,v_0)\in x_0^1\times x_0^2} \left|w\left(\lambda_n^1(u_0),\lambda_n^2(v_0)\right)-w\left(u_0,v_0\right)\right|.$$

Yet, by (1-i) and (2-i), for each (u_0,v_0) in $x_0^1\times x_0^2$, we have

$$\|\left(\lambda_n^1(u_0),\lambda_n^2(v_0)\right)-\left(u_0,v_0\right)\|_\infty \leq \eta,$$

and thus, applying (A.1), we obtain

$$|\varphi^w(x_n)-\varphi^w(x_0)| \leq \#x_0^1\#x_0^2\varepsilon,$$

and this for all $n\geq n_0$, which ends the proof of Proposition 3.1.

A.3 Proof of Proposition 3.2

Let us prove that in the *Coincidence case*, the kernel $h = h_{\varphi_\delta^{coinc}}$ given by (2.1) and (2.5) is continuous for the topology induced by the metric d (defined in (3.4)) in any (x_0,y_0) in $\mathcal{C}_\delta$ satisfying

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

As in the proof of Proposition 3.1, denote by N_z the counting process associated with z:

$$N_z(t)=\int_0^1 \mathbb{1}_{u\leq t}dN_z(u).$$

Consider a sequence $\{(x_n,y_n)\}_{n\in\mathbb{N}}$ of elements in $\mathcal{X}^2\times\mathcal{X}^2$, where $x_n=\left(x_n^1,x_n^2\right)$ and $y_n=\left(y_n^1,y_n^2\right)$ such that $d\left((x_n,y_n),(x_0,y_0)\right)\underset{n\to+\infty}{\longrightarrow}0$ and (x_0,y_0) belongs to $\mathcal{C}_\delta$.
We want to show that $|h\left(x_n,y_n\right)-h\left(x_0,y_0\right)|\underset{n\to+\infty}{\longrightarrow}0$.
Since (x_0,y_0) is in $\mathcal{C}_\delta$, for any t_0 in $\left\{x_0^2\pm\delta\right\}\cup\left\{y_0^2\pm\delta\right\}$, $t_0\notin x_0^1$, which means that $N_{x_0^1}$ is continuous in t_0 and therefore constant in a neighborhood:

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

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

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

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

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

Since $(x_0, y_0) \in \mathcal{C}_\delta \Leftrightarrow \left(\left\{x_0^2\right\} \cup \left\{y_0^2\right\}\right) \cap \left(\left\{x_0^1 \pm \delta\right\} \cup \left\{y_0^1 \pm \delta\right\}\right) = \emptyset$, one can construct $\eta_{x_0^2}$ and $\eta_{y_0^2}$ satisfying

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

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

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

$$\forall s \in \left\{x_0^1 \pm \delta\right\} \cup \left\{y_0^1 \pm \delta\right\}, \forall t \in [0,1], |t-s| \leq \eta \Rightarrow \left\{ \begin{array}{l} N_{x_0^2}(t) = N_{x_0^2}(s) \\ N_{y_0^2}(t) = N_{y_0^2}(s) \end{array} \right. . \quad \text{(A.3)}$$

As $d\left((x_n, y_n), (x_0, y_0)\right) \to_{n \to +\infty} 0$, there exists $n_0 \geq 0$ such that for $n \geq n_0$, $d\left((x_n, y_n), (x_0, y_0)\right) \leq \eta/4$. From the definition of d, we deduce that

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

and

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

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

By definition of h,

$$
\begin{aligned}
& h(x_n, y_n) - h(x_0, y_0) \qquad\qquad (A.4)\\
&= \frac{1}{2}\iint \mathbb{1}_{|u-v|\le\delta}\left\{dN_{x_n^1}dN_{x_n^2} + dN_{y_n^1}dN_{y_n^2} - dN_{x_n^1}dN_{y_n^2} - dN_{y_n^1}dN_{x_n^2}\right\}(u,v)\\
&\quad - \frac{1}{2}\iint \mathbb{1}_{|u-v|\le\delta}\left\{dN_{x_0^1}dN_{x_0^2} + dN_{y_0^1}dN_{y_0^2} - dN_{x_0^1}dN_{y_0^2} - dN_{y_0^1}dN_{x_0^2}\right\}(u,v)\\
&= \frac{1}{2}\iint \mathbb{1}_{|u-v|\le\delta}\Big(dN_{x_n^1}(u)\left(dN_{x_n^2} - dN_{x_0^2}\right)(v) + dN_{y_n^1}(u)\left(dN_{y_n^2} - dN_{y_0^2}\right)(v)\\
&\quad - dN_{x_n^1}(u)\left(dN_{y_n^2} - dN_{y_0^2}\right)(v) - dN_{y_n^1}(u)\left(dN_{x_n^2} - dN_{x_0^2}\right)(v)\\
&\quad + \left(dN_{x_n^1} - dN_{x_0^1}\right)(u)\; dN_{x_0^2}(v) + \left(dN_{y_n^1} - dN_{y_0^1}\right)(u)\; dN_{y_0^2}(v)\\
&\quad - \left(dN_{x_n^1} - dN_{x_0^1}\right)(u)\; dN_{y_0^2}(v) + \left(dN_{y_n^1} - dN_{y_0^1}\right)(u)\; dN_{x_0^2}(v)\Big).
\end{aligned}
$$

By symmetry of the problem, we just need to study the terms

$$A_n = \iint \mathbb{1}_{|u-v|\le\delta}\left(dN_{x_n^1} - dN_{x_0^1}\right)(u)\; dN_{x_0^2}(v),$$

and

$$B_n = \iint \mathbb{1}_{|u-v|\le\delta} dN_{x_n^1}(u)\left(dN_{x_n^2} - dN_{x_0^2}\right)(v).$$

Study of A_n

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

We have that

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

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

$$\left| \iint \mathbb{1}_{u<v-\delta} \left(dN_{x^1_n} - dN_{x^1_0} \right)(u)\ dN_{x^2_0}(v) \right| \leq \sum_{T \in x^2_0} \left| N^-_{x^1_n}(T-\delta) - N^-_{x^1_0}(T-\delta) \right|.$$

Therefore,

$$\begin{aligned} |A_n| \quad \leq \quad & \sum_{T \in x^2_0} \Bigg(\left| N_{x^1_n}(T+\delta) - N_{x^1_0}\left(\lambda^1_n(T+\delta)\right) \right| \\ & + \left| N_{x^1_0}\left(\lambda^1_n(T+\delta)\right) - N_{x^1_0}(T+\delta) \right| \\ & + \left| N^-_{x^1_n}(T-\delta) - N^-_{x^1_0}(T-\delta) \right| \Bigg). \end{aligned} \tag{A.5}$$

Let us study individually each term in the sum.
Fix T in x^2_0. By (1-ii),

$$\left| N_{x^1_n}(T+\delta) - N_{x^1_0}\left(\lambda^1_n(T+\delta)\right) \right| \leq \frac{\eta}{4} \leq \varepsilon. \tag{A.6}$$

From (1-i), one has $|\lambda^1_n(T+\delta) - (T+\delta)| \leq \frac{\eta}{2} \leq \eta$ which, with (A.2), implies

$$\left| N_{x^1_0}\left(\lambda^1_n(T+\delta)\right) - N_{x^1_0}(T+\delta) \right| = 0. \tag{A.7}$$

As $N^-_{x^1_n}(T-\delta) = \lim\limits_{\substack{u \to T-\delta \\ u<T-\delta}} N_{x^1_n}(u)$, there exists u_T in $[T-\delta-\eta/4, T-\delta[$ such that

$$\left| N^-_{x^1_n}(T-\delta) - N_{x^1_n}(u_T) \right| \leq \varepsilon,$$

so

$$\begin{aligned} \left| N^-_{x^1_n}(T-\delta) - N^-_{x^1_0}(T-\delta) \right| \leq \varepsilon & + \left| N_{x^1_n}(u_T) - N_{x^1_0}\left(\lambda^1_n(u_T)\right) \right| \\ & + \left| N_{x^1_0}\left(\lambda^1_n(u_T)\right) - N^-_{x^1_0}(T-\delta) \right|. \end{aligned} \tag{A.8}$$

From (1-ii), one has $\left| N_{x^1_n}(u_T) - N_{x^1_0}\left(\lambda^1_n(u_T)\right) \right| \leq \eta/4 \leq \varepsilon$.
Then, by continuity of $N_{x^1_0}$ at $T-\delta$, first remark that $N^-_{x^1_0}(T-\delta) = N_{x^1_0}(T-\delta)$.
Moreover, by (1-i) and construction of u_T,

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

hence, using (A.2), $\left| N_{x^1_0}\left(\lambda^1_n(u_T)\right) - N^-_{x^1_0}(T-\delta) \right| = 0$. So finally, (A.8) gives

$$\left| N^-_{x^1_n}(T-\delta) - N^-_{x^1_0}(T-\delta) \right| \leq 2\varepsilon. \tag{A.9}$$

Combining (A.5), (A.6), (A.7), and (A.9), we obtain that for any $n \geq n_0$:

$$|A_n| \leq 3\varepsilon \# x^2_0. \tag{A.10}$$

Study of B_n Recall that $B_n = \iint \mathbb{1}_{|u-v|\leq\delta} dN_{x_n^1}(u) \left(dN_{x_n^2} - dN_{x_0^2}\right)(v)$.
As for A_n, B_n is upper bounded by a sum of several terms, that we study separately.

$$B_n = \sum_{T\in x_n^1} \left(N_{x_n^2}(T+\delta) - N_{x_0^2}(T+\delta)\right) - \sum_{T\in x_n^1} \left(N^-_{x_n^2}(T-\delta) - N^-_{x_0^2}(T-\delta)\right).$$

So

$$B_n \leq |B_{n,1}| + |B_{n,2}| + |B_{n,3}| + |B_{n,4}|, \tag{A.11}$$

with

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

The control of B_n is quite similar to 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 (A.3) and (A.2) directly.

Control of $B_{n,1}$. Due to (2-*ii*), $\left|N_{x_n^2}(T+\delta) - N_{x_0^2}\left(\lambda_n^2(T+\delta)\right)\right| \leq \varepsilon$, so

$$|B_{n,1}| \leq \varepsilon \# x_n^1. \tag{A.12}$$

Control of $B_{n,2}$. One can easily see that

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

Fix now T in x_0^2.

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

As shown in (A.8), $\left|N^-_{x_n^1}(T-\delta) - N_{x_0^1}(T-\delta)\right| \leq 2\varepsilon$.
Furthermore, take v_T in $\left[\left(\lambda_n^2\right)^{-1}(T) - \delta - \eta/4, \left(\lambda_n^2\right)^{-1}(T) - \delta\right[$ such that

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

So,

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

By construction of v_T and λ_n^1 (see (1-ii)), $\left|N_{x_n^1}(v_T)-N_{x_0^1}\left(\lambda_n^1(v_T)\right)\right| \le \varepsilon$. Because of (A.2) which is true as

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

by (1-i), $\left|N_{x_0^1}\left(\lambda_n^1(v_T)\right)-N_{x_0^1}(T-\delta)\right| = 0$. Hence,

$$\left|N_{x_0^1}(T-\delta)-N^-_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}(T)-\delta\right)\right| \le 2\varepsilon.$$

Finally,

$$\left|N^-_{x_n^1}(T-\delta)-N^-_{x_n^1}\left(\left(\lambda_n^2\right)^{-1}(T)-\delta\right)\right| \le 4\varepsilon,$$

and

$$|B_{n,2}| \le 4\varepsilon \# x_0^2. \tag{A.13}$$

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

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

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

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

For each T in x_n^1, $\left|N^-_{x_n^2}(T-\delta)-N_{x_n^2}(T-\delta-\nu_n)\right| \le \varepsilon$ and $\left|N_{x_n^2}(T-\delta-\nu_n)-N_{x_0^2}\left(\lambda_n^2(T-\delta-\nu_n)\right)\right| \le \varepsilon$ by (2-ii). Therefore,

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

Now,

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

For each T in x_0^2,

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

where the last line comes from (1-ii), and (A.2).
We now find some w_T in $\left[\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n-\eta/4\;,\;\left(\lambda_n^2\right)^{-1}(T)+\delta+\nu_n\right[$ such that

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

so

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

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 (A.2), (1-$i$), and the construction of w_T,

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

and

$$\left|\left(\lambda_n^1(w_T)\right)-(T-\delta)\right|\leq\left|\left(\lambda_n^1(w_T)-w_T\right)\right|+\left|w_T-(T-\delta)\right|<\eta.$$

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

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

and

$$|B_{n,3}| \leq 2\varepsilon \# x_n^1 + 4\varepsilon \# x_0^2. \tag{A.14}$$

Control of $B_{n,4}$.

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

Let us fix T in x_0^2. We have

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

The first and the last terms are upper bounded by ε due to (1-*ii*). Furthermore, since $N_{x_0^1}\left(\lambda_n^1\left(\left(\lambda_n^2\right)^{-1}(T)+\delta\right)\right) = N_{x_0^1}(T+\delta) = N_{x_0^1}\left(\lambda_n^1(T+\delta)\right)$ by applying (A.2) and using (1-*i*) and (2-*i*),

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

So finally,

$$|B_{n,4}| \leq 2\varepsilon \# x_0^2. \tag{A.15}$$

Combining (A.11), (A.12), (A.13), (A.14), and (A.15), we can conclude that

$$|B_n| \leq 3\varepsilon \# x_n^1 + 10\varepsilon \# x_0^2. \tag{A.16}$$

We now just remark that $\left(\#x_n^1\right)_{n\geq n_0}$ 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, $\lambda_n^1(1) = 1$,

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

With (A.4), (A.10), and (A.16), this concludes the proof of Proposition 3.2.

A.4 Proof of Theorem 3.1

By Proposition 3.3, for all $n \geq 2$,

$$d_2\Big(\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 | \mathbb{X}_n\right), \mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)\Big) \leq C \inf_{\substack{\left(Y_{n,a}^*, Y_a\right), \left(Y_{n,b}^*, Y_b\right) \ i.i.d \ / \\ Y_{n,a}^*, Y_{n,b}^* \sim P_n^1 \otimes P_n^2, \ Y_a, Y_b \sim P^1 \otimes P^2}} \mathbb{E}^*\left[\left(h\left(Y_{n,a}^*, Y_{n,b}^*\right) - h\left(Y_a, Y_b\right)\right)^2\right].$$

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

$$\mathbb{E}'_\omega\left[\left(h\left(\bar{Y}_{n,\omega,a}^*, \bar{Y}_{n,\omega,b}^*\right) - h\left(\bar{Y}_{\omega,a}, \bar{Y}_{\omega,b}\right)\right)^2\right] \underset{n\to+\infty}{\longrightarrow} 0, \tag{A.17}$$

where $\mathbb{E}'_\omega$ denotes the expectation corresponding to $\mathbb{P}'_\omega$. Then from (A.17), we can conclude by noting that, for almost all ω in Ω,

$$\inf_{\substack{\left(Y_{n,a}^*, Y_a\right), \left(Y_{n,b}^*, Y_b\right) \ i.i.d \ / \\ Y_{n,a}^*, Y_{n,b}^* \sim P_{n,\omega}^1 \otimes P_{n,\omega}^2, \ Y_a, Y_b \sim P^1 \otimes P^2}} \mathbb{E}^*\left[\left(h\left(Y_{n,a}^*, Y_{n,b}^*\right) - h\left(Y_a, Y_b\right)\right)^2\right](\omega) \leq \mathbb{E}'_\omega\left[\left(h\left(\bar{Y}_{n,\omega,a}^*, \bar{Y}_{n,\omega,b}^*\right) - h\left(\bar{Y}_{\omega,a}, \bar{Y}_{\omega,b}\right)\right)^2\right] \underset{n\to+\infty}{\longrightarrow} 0.$$

To prove (A.17), consider $(\Omega, \mathcal{A}, \mathbb{P})$ the probability space on which all the X_i's are defined. In what follows, one can keep in mind that Ω represents the randomness in the original sequence $(X_i)_i$. Thus, a given ω in Ω represents a given realization of $(X_i)_i$.
As a preliminary step, from Proposition 3.4, there exists some subset Ω_1 of Ω such that $\mathbb{P}(\Omega_1) = 1$ and for every ω in Ω_1,

$$\frac{1}{n^4}\sum_{i,j,k,l=1}^n h^2\left(\left(X_i^1(\omega), X_j^2(\omega)\right), \left(X_k^1(\omega), X_l^2(\omega)\right)\right) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\left[h^2\left(\left(X_1^1, X_2^2\right), \left(X_3^1, X_4^2\right)\right)\right]. \tag{A.18}$$

Applying Theorem 3 in [7], since $(\mathcal{X}, d_{\mathcal{X}})$ defined by (3.3) is separable, P-a.s. in $(X_i)_i$, $P_n^1 \underset{n\to+\infty}{\Longrightarrow} P^1$ and $P_n^2 \underset{n\to+\infty}{\Longrightarrow} P^2$. Hence there exists some subset Ω_2 of Ω such that $\mathbb{P}(\Omega_2) = 1$ and for every ω in Ω_2,

$$P_{n,\omega}^1 \otimes P_{n,\omega}^2 \underset{n\to+\infty}{\Longrightarrow} P^1 \otimes P^2, \tag{A.19}$$

Now, consider $\Omega_0 = \Omega_1 \cap \Omega_2$, and fix ω in Ω_0.
Following the proof of Skorokhod's representation theorem in [3, Theorem 11.7.2, p. 415], since $(\mathcal{X}^2, d_{\mathcal{X}^2})$ is a separable space, it is possible to construct

- some probability space $(\Omega'_\omega, \mathcal{A}'_\omega, \mathbb{P}'_\omega)$,
- some random variables $\bar{Y}^*_{n,\omega,a} : \Omega'_\omega \to \mathcal{X}^2$, $\bar{Y}^*_{n,\omega,b} : \Omega'_\omega \to \mathcal{X}^2$ with distribution $P^1_{n,\omega} \otimes P^2_{n,\omega}$,
- $\bar{Y}_{\omega,a} : \Omega'_\omega \to \mathcal{X}^2$, $\bar{Y}_{\omega,b} : \Omega'_\omega \to \mathcal{X}^2$ with distribution $P^1 \otimes P^2$,

satisfying:

- $\mathbb{P}'_\omega$-a.s., $\bar{Y}^*_{n,\omega,a} \underset{n\to+\infty}{\longrightarrow} \bar{Y}_{\omega,a}$ and $\bar{Y}^*_{n,\omega,b} \underset{n\to+\infty}{\longrightarrow} \bar{Y}_{\omega,b}$,
- $\left\{ \left(\bar{Y}^*_{n,\omega,a}\right)_{n\geq 1}, \bar{Y}_{\omega,a} \right\}$ and $\left\{ \left(\bar{Y}^*_{n,\omega,b}\right)_{n\geq 1}, \bar{Y}_{\omega,b} \right\}$ are independent,

so that w.r.t. the metric d (see (3.4)),

$$\mathbb{P}'_\omega\text{-a.s.}, \quad \left(\bar{Y}^*_{n,\omega,a}, \bar{Y}^*_{n,\omega,b}\right) \underset{n\to+\infty}{\longrightarrow} \left(\bar{Y}_{\omega,a}, \bar{Y}_{\omega,b}\right). \tag{A.20}$$

But under $(\mathcal{A}_{Cont})$, h is continuous on a set $\mathcal{C}$ s. t. $\mathbb{P}'_\omega\left(\left(\bar{Y}_{\omega,a}, \bar{Y}_{\omega,b}\right) \in \mathcal{C}\right) = \left(P^1 \otimes P^2\right)^{\otimes 2}(\mathcal{C}) = 1$, hence

$$\mathbb{P}'_\omega\text{-a.s.}, \; h\left(\bar{Y}^*_{n,\omega,a}, \bar{Y}^*_{n,\omega,b}\right) \to_{n\to+\infty} h\left(\bar{Y}_{\omega,a}, \bar{Y}_{\omega,b}\right).$$

As $\mathbb{P}'_\omega$-a.s. convergence implies convergence in probability, to obtain (A.17), we only need to prove that the sequence $\left(h^2\left(\bar{Y}^*_{n,\omega,a}, \bar{Y}^*_{n,\omega,b}\right)\right)_{n\geq 1}$ is uniformly integrable, according to Theorem 16.6 p. 165 of [5]. We therefore conclude since (A.18) is equivalent to

$$\begin{aligned}\mathbb{E}'_\omega\left[h^2\left(\bar{Y}^*_{n,\omega,a}, \bar{Y}^*_{n,\omega,b}\right)\right] &= \frac{1}{n^4}\sum_{i,j,k,l=1}^{n} h^2\left((X^1_i(\omega), X^2_j(\omega)), (X^1_k(\omega), X^2_l(\omega))\right) \\ &\underset{n\to+\infty}{\longrightarrow} \mathbb{E}\left[h^2\left(\left(X^1_1, X^2_2\right), \left(X^1_3, X^2_4\right)\right)\right] = \mathbb{E}'_\omega\left[h^2\left(\bar{Y}_{\omega,a}, \bar{Y}_{\omega,b}\right)\right].\end{aligned}$$

(A.17) is thus obtained for any ω in Ω_0, with $\mathbb{P}(\Omega_0) = 1$. This ends the proof.

A.5 Proof of Proposition 3.3

Fix some integer $n \geq 2$ and recall that the P^j_n's, $(j = 1, 2)$, are the marginal empirical measures associated with $\mathbb{X}_n$.
Let $\left(Y^*_{n,i}, Y_i\right)_{1\leq i\leq n}$ be an i.i.d. sample such that for every $i = 1 \ldots n$, $Y^*_{n,i} \sim P^1_n \otimes P^2_n$, $Y_i \sim P^1 \otimes P^2$, and such that, from the definition of Wasserstein's metric d_2 recalled

in (3.2),

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

Notice that the upper bound is finite under $(\mathcal{A}_{Mmt}^*)$.
Introducing for (i,i',j,j') in $\{1,2,\ldots,n\}^4$, and m in $\{2,3,4\}$,

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

$$I_m = \Big\{(i,i',j,j') \in \{1,2,\ldots,n\}^4; i \neq i',\ j \neq j',\ \#\{i,i',j,j'\} = m\Big\},$$

where $\#\{i,i',j,j'\}$ denotes the number of different elements in $\{i,i',j,j'\}$, one has:

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

Let us now upper bound each term of this sum separately.
If (i,i',j,j') is in I_4, then by independence,

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

Under $(\mathcal{A}_{Cent})$ and $(\mathcal{A}_{Cent}^*)$, $\mathbb{E}\left[h(Y_i, Y_{i'})\right] = \mathbb{E}^*\big[h\big(Y_{n,i}^*, Y_{n,i'}^*\big)\big] = 0$, so
$\mathrm{E}_{(i,i',j,j')} = 0$.
If (i,i',j,j') is in I_3, by the Cauchy-Schwarz inequality,

$$\mathrm{E}_{(i,i',j,j')} \leq \mathbb{E}^*\left[\Big(h\big(Y_{n,a}^*, Y_{n,b}^*\big) - h(Y_a, Y_b)\Big)^2\right],$$

where $\left(Y_{n,a}^*, Y_a\right)$ and $\left(Y_{n,b}^*, Y_b\right)$ are independent copies of the $(Y_{n,i}, Y_i)$'s. If (i,i',j,j') is in I_2, then $\mathrm{E}_{(i,i',j,j')} = \mathbb{E}^*\left[\left(h\left(Y_{n,a}^*, Y_{n,b}^*\right) - h\left(Y_a, Y_b\right)\right)^2\right]$ is immediate.
But $\#I_3 = 4n(n-1)(n-2)$ and $\#I_2 = 2n(n-1)$, so

$$d_2^2\left(\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 \middle| \mathbb{X}_n\right), \mathcal{L}\left(\sqrt{n}U_n, P^1 \otimes P^2\right)\right)$$
$$\leq 4\mathbb{E}^*\left[\Big(h\big(Y_{n,a}^*, Y_{n,b}^*\big) - h\big(Y_a, Y_b\big)\Big)^2\right].$$

Since $\left(Y_{n,a}^*, Y_a\right)$ and $\left(Y_{n,b}^*, Y_b\right)$ may be arbitrarily chosen, Proposition 3.3 follows.

A.6 Proof of Proposition 3.4

Let us first notice that (3.6) is a direct application of the strong law of large numbers for U-statistics, proved by Hoeffding [4].

Next, for m in $\{1,\dots,4\}$, introduce

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

where $I_{\{i_1,\dots i_m\}}$ is the set $\left\{(i,j,k,l)\in\{i_1,\dots i_m\}^4;\#\{i,j,k,l\}=m\right\}$.
Then,

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

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

$$\frac{1}{n(n-1)\dots(n-m+1)}\sum_{(i_1,\dots,i_m)} g_m\left(X_{i_1},\dots,X_{i_m}\right)\underset{n\to+\infty}{\longrightarrow}\mathbb{E}\left[g_m\left(X_1,\dots,X_m\right)\right].$$

In particular, P-a.s. in $(X_i)_i$, $n^{-4}\sum_{(i_1,\dots,i_m)} g_m\left(X_{i_1},\dots,X_{i_m}\right)$ converges towards 0 for m in $\{1,2,3\}$, and towards $\mathbb{E}\left[g_4\left(X_1,X_2,X_3,X_4\right)\right]$ for $m=4$. Finally noticing that $\mathbb{E}\left[g_4\left(X_1,X_2,X_3,X_4\right)\right]=4!\mathbb{E}\left[h^2\left(\left(X_1^1,X_2^2\right),\left(X_3^1,X_4^2\right)\right)\right]$ allows to conclude.

A.7 Proof of Proposition 3.5

Let $(X_i)_i$ be a sequence of i.i.d pairs of point processes with distribution $P^1\otimes P^2$ on $\mathcal{X}^2$. According to $(\mathcal{A}_{Cent})$, for $i\neq j$, $\mathbb{E}\left[h(X_i,X_j)\right]=0$. For a better readability, we set $\mathbb{E}\left[h|X_i\right]=\mathbb{E}\left[h(X_i,X)|X_i\right]=\mathbb{E}\left[h(X,X_i)|X_i\right]$ 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 that

$$\sqrt{n}U_n(\mathbb{X}_n)=\frac{2}{\sqrt{n}(n-1)}\left(T_n+M_n\right),$$

where $T_n=\sum_{i<j}\left(\mathbb{E}\left[h|X_i\right]+\mathbb{E}\left[h|X_j\right]\right)$, and $M_n=\sum_{i<j}g(X_i,X_j)$, with $g(X_i,X_j)=h(X_i,X_j)-\mathbb{E}\left[h|X_i\right]-\mathbb{E}\left[h|X_j\right]$.
Firstly, we have that $\mathbb{E}\left[M_n^2\right]=\sum_{i<j}\sum_{k<l}\mathbb{E}\left[g(X_i,X_j)g(X_k,X_l)\right]$. But if $\{i,j\}\cap\{k,l\}=\emptyset$, $i<j$, $k<l$, $\mathbb{E}\left[g(X_i,X_j)g(X_k,X_l)\right]=\left(\mathbb{E}\left[g(X_i,X_j)\right]\right)^2=0$. 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

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

Therefore, $\mathbb{E}\left[M_n^2\right]=\sum_{i<j}\mathbb{E}\left[g^2(X_i,X_j)\right]=n(n-1)\mathbb{E}\left[g^2(X_1,X_2)\right]/2$, and since $\mathbb{E}\left[g^2(X_i,X_j)\right]<+\infty$, from Chebychev's inequality, we deduce that

$$\frac{2}{\sqrt{n}(n-1)}M_n\xrightarrow[n\to+\infty]{\mathbb{P}}0. \tag{A.21}$$

Secondly, we have that $T_n=(n-1)\sum_{i=1}^n\mathbb{E}\left[h|X_i\right]$. Since the $\mathbb{E}\left[h|X_i\right]$'s are i.i.d, with $\mathbb{E}\left[\mathbb{E}\left[h|X_i\right]\right]=0$ and $\mathbb{V}\mathrm{ar}\left(\mathbb{E}\left[h|X_i\right]\right)=\sigma^2_{P^1\otimes P^2}/4$, thanks to $(\mathcal{A}_{Mmt})$, the Central Limit Theorem leads to

$$\frac{2}{\sqrt{n}(n-1)}T_n\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right). \tag{A.22}$$

Thus, combining (A.21) and (A.22), Slutsky's lemma ensures the convergence in distribution of $\sqrt{n}U_n(\mathbb{X}_n)$ towards $\mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right)$.
Now, in order to obtain the convergence in the Wasserstein metric, one needs to check the convergence of the second order moments. Notice that

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

Let us consider all the cases where $i\neq i'$ and $j\neq j'$.
If $\#\{i,i',j,j'\}=4$, $\mathbb{E}\left[h(X_i,X_{i'})h(X_j,X_{j'})\right]=0$, by independence and $(\mathcal{A}_{Cent})$.
If $\#\{i,i',j,j'\}=3$, $\mathbb{E}\left[h(X_i,X_{i'})h(X_j,X_{j'})\right]=\sigma^2_{P^1\otimes P^2}/4$, by symmetry of h.
If $\#\{i,i',j,j'\}=2$, $\mathbb{E}\left[h(X_i,X_{i'})h(X_j,X_{j'})\right]=\mathbb{E}\left[\left(h(X_1,X_2)\right)^2\right]$. Therefore,

$$\mathbb{E}\left[\left(\sqrt{n}U_n(\mathbb{X}_n)\right)^2\right]=\frac{n-2}{n-1}\sigma^2_{P^1\otimes P^2}+\frac{2}{n-1}\mathbb{E}\left[\left(h(X_1,X_2)\right)^2\right]\underset{n\to+\infty}{\longrightarrow}\sigma^2_{P^1\otimes P^2},$$

which ends the proof of Proposition 3.5.

A.8 Proof of Corollary 3.1

By Proposition 3.5, we have that

$$\mathcal{L}\left(\sqrt{n}U_n,P^1\otimes P^2\right)\underset{n\to+\infty}{\Longrightarrow}\mathcal{N}(0,\sigma^2_{P^1\otimes P^2}), \tag{A.23}$$

where $\mathcal{N}(0,\sigma^2_{P^1\otimes P^2})$ has a continuous c.d.f. Therefore, by [6, Lemma 2.11],

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}\left(\sqrt{n}U_n(\mathbb{X}_n^{\perp})\leq z\right)-\Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right|\underset{n\to+\infty}{\longrightarrow}0. \tag{A.24}$$

Furthermore, since convergence w.r.t the d_2 distance implies weak convergence, Theorem 3.1 combined with (A.23) leads to

$$\mathcal{L}\left(\sqrt{n}U_n, P_n^1 \otimes P_n^2 \middle| \mathbb{X}_n\right) \underset{n\to+\infty}{\Longrightarrow} \mathcal{N}(0, \sigma^2_{P^1\otimes P^2}) \quad P\text{-a.s. in } (X_i)_i. \tag{A.25}$$

Hence,

$$\sup_{z\in\mathbb{R}} \left| \mathbb{P}\left(\sqrt{n}U_n(\mathbb{X}_n^*) \leq z | \mathbb{X}_n\right) - \Phi_{0,\sigma^2_{P^1\otimes P^2}}(z) \right| \underset{n\to+\infty}{\longrightarrow} 0 \ P\text{-a.s. in } (X_i)_i, \tag{A.26}$$

and the first part of the corollary is obtained.

Moreover, [6, Lemma 21.2] can then be applied to both (A.23) and (A.25), to obtain that on the event where (A.25) holds

$$q^*_{\eta,n}(\mathbb{X}_n) \underset{n\to+\infty}{\longrightarrow} \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(\eta) \quad P\text{-a.s. in } (X_i)_i, \tag{A.27}$$

and that $q^{\perp}_{\eta,n}$ also converges to $\Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(\eta)$.

A.9 Proof of Theorem 3.2

Let us focus on the sequence of upper-tailed tests in $\Gamma(q^*)$, the proof for the other tests being similar.

Under (H_0), from Proposition 3.5 and (A.27), by Slutsky's lemma, $(\sqrt{n}U_n(\mathbb{X}_n), q^*_{1-\alpha,n}(\mathbb{X}_n))$ converges in distribution to $(Z, \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(1-\alpha))$, where $Z \sim \mathcal{N}(0, \sigma^2_{P^1\otimes P^2})$. Therefore, under (H_0),

$$\mathbb{P}(\sqrt{n}U_n(\mathbb{X}_n) > q^*_{1-\alpha,n}(\mathbb{X}_n)) \to_{n\to+\infty} \alpha,$$

which proves $(\mathcal{P}_{size})$.

Under any alternative such that $\int h(x,x')dP(x)dP(x') > 0$, by Proposition 3.4,

$$U_n(\mathbb{X}_n) \underset{n\to+\infty}{\longrightarrow} \int h(x,x')dP(x)dP(x') > 0, \ P\text{-a.s. in } (X_i)_i.$$

Furthermore, due to (A.27), $q^*_{1-\alpha,n}(\mathbb{X}_n)/\sqrt{n} \to_{n\to+\infty} 0$ P-a.s. in $(X_i)_i$. Hence, $\mathbb{P}(\sqrt{n}U_n(\mathbb{X}_n) \leq q^*_{1-\alpha,n}(\mathbb{X}_n)) \to_{n\to+\infty} 0$, and thus $(\mathcal{P}_{consist.})$ is proved.

A.10 Proof of Proposition 3.6

As above, we focus on the sequence of upper-tailed tests in $\Gamma(q^*_{MC})$. Let $Z \sim \mathcal{N}(0,1)$ and define for z in $\mathbb{R}$,

$$F^*_{n,\mathbb{X}_n}(z) = \mathbb{P}\left(\sqrt{n}U_n(\mathbb{X}_n^*) \leq z | \mathbb{X}_n\right), \ F^{*B_n}_{n,\mathbb{X}_n}(z) = \frac{1}{B_n}\sum_{b=1}^{B_n} \mathbb{1}_{\sqrt{n}U_n(\mathbb{X}_n^{*b})\leq z}.$$

By the Dvoretzky-Kiefer-Wolfowitz inequality (see [6]), for $n \geq 2$ and $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{z\in\mathbb{R}}\left|F^{*B_n}_{n,\mathbb{X}_n}(z)-F^{*}_{n,\mathbb{X}_n}(z)\right|>\varepsilon\right)=\mathbb{E}\left[\mathbb{P}\left(\sup_{z\in\mathbb{R}}\left|F^{*,B_n}_{n,\mathbb{X}_n}(z)-F^{*}_{n,\mathbb{X}_n}(z)\right|>\varepsilon\middle|\mathbb{X}_n\right)\right]$$
$$\leq 2e^{-2B_n\varepsilon^2}\underset{n\to+\infty}{\longrightarrow}0,$$

that is $\sup_{z\in\mathbb{R}}|F^{*B_n}_{n,\mathbb{X}_n}(z)-F^{*}_{n,\mathbb{X}_n}(z)|\underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}}0$. With (A.26), this leads to

$$\sup_{z\in\mathbb{R}}\left|F^{*B_n}_{n,\mathbb{X}_n}(z)-\Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right|\underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}}0. \tag{A.28}$$

We finish the proof using similar arguments as in [6, Lemma 21.2], combined with a subsequence argument [3, Theorem 9.2.1]. Let ϕ_0 be an extraction. Then, by (A.28), 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,

$$\sup_{z\in\mathbb{R}}\left|F^{*B_{\phi_1\circ\phi_0(n)}}_{\phi_1\circ\phi_0(n),\mathbb{X}_{\phi_1\circ\phi_0(n)}}(\omega)(z)-\Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right|\underset{n\to+\infty}{\longrightarrow}0.$$

From now on, fix ω in Ω_0. In particular, this fixes a realisation of $\mathbb{X}_n$, and a realisation of $\left(\mathbb{X}^{*1}_n,\dots,\mathbb{X}^{*B_n}_n\right)$ and thus, $F^{*B_n}_{n,\mathbb{X}_n}(\omega)$ is deterministic. Hence, $F^{*B_{\phi_1\circ\phi_0(n)}}_{\phi_1\circ\phi_0(n),\mathbb{X}_{\phi_1\circ\phi_0(n)}}(\omega)(Z)\underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}}\Phi_{0,\sigma^2_{P^1\otimes P^2}}(Z)$, and for η in $(0,1)$,

$$\Phi_{0,1}\left(\left(F^{*B_{\phi_1\circ\phi_0(n)}}_{\phi_1\circ\phi_0(n),\mathbb{X}_{\phi_1\circ\phi_0(n)}}(\omega)\right)^{-1}(\eta)\right)=\mathbb{P}\left(F^{*B_{\phi_1\circ\phi_0(n)}}_{\phi_1\circ\phi_0(n),\mathbb{X}_{\phi_1\circ\phi_0(n)}}(\omega)(Z)<\eta\right)$$
$$\underset{n\to+\infty}{\longrightarrow}\mathbb{P}\left(\Phi_{0,\sigma^2_{P^1\otimes P^2}}(Z)<\eta\right)=\Phi_{0,1}\left(\left(\Phi_{0,\sigma^2_{P^1\otimes P^2}}\right)^{-1}(\eta)\right).$$

Finally, as $\Phi_{0,1}$ is a one-to-one function and $\Phi^{-1}_{0,1}$ is continuous,

$$\sqrt{\phi_1\circ\phi_0(n)}U^{*(\lceil\eta(B_{\phi_1\circ\phi_0(n)})\rceil)}(\omega)$$
$$=\left(F^{*B_{\phi_1\circ\phi_0(n)}}_{\phi_1\circ\phi_0(n),\mathbb{X}_{\phi_1\circ\phi_0(n)}}(\omega)\right)^{-1}(\eta)\underset{n\to+\infty}{\longrightarrow}\Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(\eta), \tag{A.29}$$

and this for all ω in Ω_0, and any initial extraction ϕ_0. Therefore, we obtain that $\sqrt{n}U^{*(\lceil\eta B_n\rceil)}\underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}}\Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(\eta)$. We conclude as for Theorem 3.2.

A.11 Proof of Theorem 4.1

For the sake of clarity and a better readability, we first present a sketch of the proof of this Theorem in Subsection A.11.1. A complete version is detailed in Subsection A.11.2.

A.11.1 Sketch of proof of Theorem 4.1

Let d_{BL} denote the bounded Lipschitz metric, which metrizes the weak convergence [3, Prop. 11.3.2 and Th. 11.3.3]. For any variable Z_n depending on $\mathbb{X}_n$ and Π_n, $\mathcal{L}(Z_n|\mathbb{X}_n)$ denotes the conditional distribution of Z_n given $\mathbb{X}_n$ and for any integrable function f, $\mathbb{E}_{P^1\otimes P^2}[f] = \mathbb{E}[f(X_1^1, X_2^2)]$.

• The first step of the proof consists in decomposing $\sqrt{n}U_n\left(\mathbb{X}_n^{\Pi_n}\right)$ in

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

where

- $M_n^{\Pi_n}(\mathbb{X}_n) = \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i\neq j}\mathbb{1}_{\Pi_n(i)=j}C_{i,j}$,
- $R_n^{\Pi_n}(\mathbb{X}_n) = \displaystyle\sum_{i=1}^{n}\left(\mathbb{1}_{\Pi_n(i)=i} - \frac{1}{n}\right)C_{i,i}$,
- $T_n(\mathbb{X}_n) = \dfrac{1}{n}\displaystyle\sum_{i\neq j}C_{i,j}$,

with

$$C_{i,j} = \varphi\left(X_i^1, X_j^2\right) - \mathbb{E}\left[\varphi\left(X_i^1, X^2\right)\middle|X_i^1\right] - \mathbb{E}\left[\varphi\left(X^1, X_j^2\right)\middle|X_j^2\right] + \mathbb{E}_{P^1\otimes P^2}[\varphi],$$

$X = (X^1, X^2)$ being P-distributed and independent of $(X_i)_i$.
We then prove from Cauchy-Schwarz inequality that

$$\mathbb{E}\left[\left(\mathbb{E}\left[\frac{\left|R_n^{\Pi_n}(\mathbb{X}_n)\right|}{\sqrt{n}}\middle|\mathbb{X}_n\right]\right)^2\right] \underset{n\to+\infty}{\longrightarrow} 0 \text{ and } \mathbb{E}\left[\left(\frac{T_n(\mathbb{X}_n)}{\sqrt{n}}\right)^2\right] \underset{n\to+\infty}{\longrightarrow} 0,$$

therefore from Markov's inequality,

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

From the definition of d_{BL}, this allows us to derive that

$$d_{BL}\left(\mathcal{L}\left(\sqrt{n}U_n\left(\mathbb{X}_n^{\Pi_n}\right)\middle|\mathbb{X}_n\right), \mathcal{L}\left(\frac{n}{n-1}M_n^{\Pi_n}(\mathbb{X}_n)\middle|\mathbb{X}_n\right)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0. \tag{A.30}$$

• The second, and most difficult, step of the proof consists in proving that

$$d_{BL}\left(\mathcal{L}\left(M_n^{\Pi_n}(\mathbb{X}_n)\middle|\mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0. \tag{A.31}$$

Consider

$$Y_{n,i} = \frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}\left(\mathbb{1}_{\Pi_n(i)=j}C_{i,j} + \mathbb{1}_{\Pi_n(j)=i}C_{j,i}\right), \tag{A.32}$$

and for Π'_n another uniformly distributed random permutation with values in $\mathfrak{S}_n$, independent of Π_n and $\mathbb{X}_n$, define accordingly $Y'_{n,i}$ by replacing Π_n by Π'_n in (A.32), so that $M_n^{\Pi_n}(\mathbb{X}_n) = \sum_{i=1}^n Y_{n,i}$ and similarly for $M_n^{\Pi'_n}(\mathbb{X}_n)$.
Denoting $\mathcal{F}_{n,i} = \sigma(\Pi_n, \Pi'_n, X_1, X_2, \ldots, X_i)$ for $n \geq i \geq 2$, we prove through technical computations that for a, b in $\mathbb{R}$, $\left(aY_{n,i} + bY'_{n,i}, \mathcal{F}_{n,i}\right)_{2 \leq i \leq n}$ is a martingale difference array which satisfies the assumptions of the following result, commonly attributed to Brown [1].

Theorem A.1. *Let $(X_{n,k})_{k \in \{1,\ldots,p_n\}, n \in \mathbb{N}^*}$ be a martingale difference array, i.e. such that there exists an array of σ-algebra $(\mathcal{F}_{n,k})_{k \in \{1,\ldots,p_n\}, n \in \mathbb{N}^*}$ that is increasing w.r.t. k such that for all $k = 1, \ldots, p_n$, $\mathbb{E}[X_{n,k} | \mathcal{F}_{n,k-1}] = 0$.*
Let $A_n = \sum_{k=1}^{p_n} \mathbb{E}\left[X_{n,k}^2 | \mathcal{F}_{n,k-1}\right]$, and assume that

- $A_n \xrightarrow[n \to +\infty]{\mathbb{P}} \sigma^2 > 0$,
- $\forall \varepsilon > 0$, $\displaystyle\sum_{k=1}^{p_n} \mathbb{E}\left[X_{n,k}^2 \mathbb{1}_{|X_{n,k}| > \varepsilon}\right] \to_{n \to +\infty} 0$.

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

Thus, given a, b in $\mathbb{R}$, we obtain that

$$\mathcal{L}\left(aM_n^{\Pi_n}(\mathbb{X}_n) + bM_n^{\Pi'_n}(\mathbb{X}_n)\right) \underset{n \to +\infty}{\Longrightarrow} \mathcal{N}\left(0, \left(a^2 + b^2\right)\sigma^2_{P^1 \otimes P^2}\right),$$

which, according to the Cramér-Wold device, leads to Lemma A.1 below.

Lemma A.1. *Considering the above notation,*

$$\mathcal{L}\left(\left(M_n^{\Pi_n}(\mathbb{X}_n), M_n^{\Pi'_n}(\mathbb{X}_n)\right)'\right) \underset{n \to +\infty}{\Longrightarrow} \mathcal{N}_2\left(0, \begin{pmatrix} \sigma^2_{P^1 \otimes P^2} & 0 \\ 0 & \sigma^2_{P^1 \otimes P^2} \end{pmatrix}\right),$$

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

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

$$\begin{cases} \mathbb{P}\left(M_n^{\Pi_n}(\mathbb{X}_n) \leq t\right) \underset{n \to +\infty}{\longrightarrow} \Phi_{0, \sigma^2_{P^1 \otimes P^2}}(t), \\ \mathbb{P}\left(M_n^{\Pi_n}(\mathbb{X}_n) \leq t, M_n^{\Pi'_n}(\mathbb{X}_n) \leq t\right) \underset{n \to +\infty}{\longrightarrow} \Phi^2_{0, \sigma^2_{P^1 \otimes P^2}}(t). \end{cases}$$

Using Chebychev's inequality, with the fact (see [3, 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 subsequence of any initial subsequence, we prove that this leads to (A.31), and therefore,

$$d_{BL}\left(\mathcal{L}\left(\sqrt{n} U_n\left(\mathbb{X}_n^{\Pi_n}\right) \middle| \mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1 \otimes P^2}\right)\right) \xrightarrow[n \to +\infty]{\mathbb{P}} 0.$$

• The third, and final, step of the proof consists in deriving, by direct computations and the strong law of large numbers of Hœffding [4], the convergence of the conditional second order moments

$$\mathbb{E}\left[\left(\sqrt{n}U_n\left(\mathbb{X}_n^{\Pi_n}\right)\right)^2\Big|\mathbb{X}_n\right] \underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}} \sigma^2_{P^1\otimes P^2},$$

which ends the proof.

A.11.2 Complete proof of Theorem 4.1

Recall that d_{BL} denotes the bounded Lipschitz metric which metrizes the weak convergence, defined by

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

where, as defined in [3], BL is the set of bounded Lipschitz functions on $\mathbb{R}$, and $\|f\|_{BL} = \|f\|_\infty + \sup_{x\neq y}\dfrac{|f(x)-f(y)|}{|x-y|}$.
Recall that the proof consists of three steps presented in Section A.11.1. We give below a complete proof for each of these steps.

First step: decomposition of $\sqrt{n}U_n\big(\mathbb{X}_n^{\Pi_n}\big)$ in the *Linear case* It is obvious that by the definition (2.6) of h_φ,

$$U_n\big(\mathbb{X}_n^{\Pi_n}\big) = \frac{1}{n-1}U_n^{\Pi_n}, \tag{A.33}$$

where $U_n^{\Pi_n} = \sum_{i=1}^n \varphi\Big(X_i^1, X^2_{\Pi_n(i)}\Big) - \dfrac{1}{n}\sum_{i,j=1}^n \varphi\big(X_i^1, X_j^2\big)$, so,

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

On the one hand, if $\mathbb{E}_P[f]$ and $\mathbb{E}_{P^1\otimes P^2}[f]$ respectively denote $\mathbb{E}\left[f\left(X_1^1, X_1^2\right)\right]$, and $\mathbb{E}\left[f\left(X_1^1, X_2^2\right)\right]$, for any integrable function f, then

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

On the other hand,

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

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

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

Therefore, $U_n^{\Pi_n}$ is equal to

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

As a consequence, setting

$$C_{i,j} = \varphi\big(X_i^1, X_j^2\big) - \mathbb{E}\left[\varphi\big(X_i^1, X^2\big)\big|X_i^1\right] - \mathbb{E}\left[\varphi\big(X^1, X_j^2\big)\big|X_j^2\right] + \mathbb{E}_{P^1\otimes P^2}[\varphi],$$

$$\sqrt{n}U_n\big(\mathbb{X}_n^{\Pi_n}\big) = \frac{n}{n-1}\left(M_n^{\Pi_n}(\mathbb{X}_n) + \frac{R_n^{\Pi_n}(\mathbb{X}_n)}{\sqrt{n}} - \frac{T_n(\mathbb{X}_n)}{\sqrt{n}}\right), \tag{A.34}$$

with

$$\begin{aligned} M_n^{\Pi_n}(\mathbb{X}_n) &= \frac{1}{\sqrt{n}}\sum_{i\neq j}\mathbb{1}_{\Pi_n(i)=j}C_{i,j}, \\ R_n^{\Pi_n}(\mathbb{X}_n) &= \sum_{i=1}^{n}\left(\mathbb{1}_{\Pi_n(i)=i} - \frac{1}{n}\right)C_{i,i}, \\ T_n(\mathbb{X}_n) &= \frac{1}{n}\sum_{i\neq j}C_{i,j}. \end{aligned}$$

Let us now prove that

$$d_{BL}\left(\mathcal{L}\left(\sqrt{n}U_n\big(\mathbb{X}_n^{\Pi_n}\big)\big|\mathbb{X}_n\right), \mathcal{L}\left(\frac{n}{n-1}M_n^{\Pi_n}(\mathbb{X}_n)\middle|\mathbb{X}_n\right)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0. \tag{A.35}$$

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

$$\begin{aligned} &\left|\mathbb{E}\left[f\left(\sqrt{n}U_n\big(\mathbb{X}_n^{\Pi_n}\big)\right)\big|\mathbb{X}_n\right] - \mathbb{E}\left[f\left(\frac{n}{n-1}M_n^{\Pi_n}(\mathbb{X}_n)\right)\middle|\mathbb{X}_n\right]\right| \\ &\qquad\leq \mathbb{E}\left[\left|\sqrt{n}U_n\big(\mathbb{X}_n^{\Pi_n}\big) - \frac{n}{n-1}M_n^{\Pi_n}(\mathbb{X}_n)\right|\middle|\mathbb{X}_n\right] \\ &\qquad\leq \frac{n}{n-1}\left(\mathbb{E}\left[\frac{\left|R_n^{\Pi_n}(\mathbb{X}_n)\right|}{\sqrt{n}}\middle|\mathbb{X}_n\right] + \frac{|T_n(\mathbb{X}_n)|}{\sqrt{n}}\right). \end{aligned}$$

Hence, taking the supremum over $\{f\in BL; \|f\|_{BL}\leq 1\}$,

$$\begin{aligned} &d_{BL}\left(\mathcal{L}\left(\sqrt{n}U_n\big(\mathbb{X}_n^{\Pi_n}\big)\big|\mathbb{X}_n\right), \mathcal{L}\left(\frac{n}{n-1}M_n^{\Pi_n}(\mathbb{X}_n)\middle|\mathbb{X}_n\right)\right) \\ &\qquad\leq \frac{n}{n-1}\left(\mathbb{E}\left[\frac{\left|R_n^{\Pi_n}(\mathbb{X}_n)\right|}{\sqrt{n}}\middle|\mathbb{X}_n\right] + \frac{|T_n(\mathbb{X}_n)|}{\sqrt{n}}\right). \end{aligned} \tag{A.36}$$

Moreover, on the one hand, since Π_n is independent of $(X_i)_i$, by Cauchy-Schwarz inequality,

$$\mathbb{E}\left[\left(\mathbb{E}\left[\frac{\left|R_n^{\Pi_n}(\mathbb{X}_n)\right|}{\sqrt{n}}\middle|\mathbb{X}_n\right]\right)^2\right] \leq \frac{1}{n}\mathbb{E}\left[\left(R_n^{\Pi_n}(\mathbb{X}_n)\right)^2\right],$$

and

$$\begin{aligned}
&\mathbb{E}\left[\left(R_n^{\Pi_n}(\mathbb{X}_n)\right)^2\right] \\
&\leq \sum_{i,j=1}^{n} \mathbb{E}\left[\left(\mathbb{1}_{\Pi_n(i)=i} - \frac{1}{n}\right)\left(\mathbb{1}_{\Pi_n(j)=j} - \frac{1}{n}\right)\right] \mathbb{E}\left[C_{i,i}C_{j,j}\right] \\
&\leq C\left(\mathbb{E}_P\left[\varphi^2\right] + \mathbb{E}_{P^1\otimes P^2}\left[\varphi^2\right]\right) \sum_{i,j=1}^{n}\left(\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=i}\mathbb{1}_{\Pi_n(j)=j}\right] - \frac{1}{n^2}\right) \\
&\leq C\left(\mathbb{E}_P\left[\varphi^2\right] + \mathbb{E}_{P^1\otimes P^2}\left[\varphi^2\right]\right)\left(\sum_{i=1}^{n}\left(\frac{1}{n} - \frac{1}{n^2}\right) + \sum_{i\neq j}\left(\frac{1}{n(n-1)} - \frac{1}{n^2}\right)\right) \\
&\leq C\left(\mathbb{E}_P\left[\varphi^2\right] + \mathbb{E}_{P^1\otimes P^2}\left[\varphi^2\right]\right) < +\infty.
\end{aligned}$$

Therefore, from Markov's inequality, we deduce that

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

On the other hand,

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

Notice that for $i \neq j$, $\mathbb{E}\left[C_{i,j}|X_i\right] = \mathbb{E}\left[C_{i,j}|X_j\right] = 0$.
If $\#\{i,j,k,l\} = 4$, then $\mathbb{E}\left[C_{i,j}C_{k,l}\right] = \left(\mathbb{E}\left[C_{i,j}\right]\right)^2 = 0$.
If i, j, l are all different, then

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

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

$$\mathbb{E}\left[C_{i,j}^2\right] = \sigma_{P^1\otimes P^2}^2, \text{ and } \mathbb{E}\left[C_{i,j}C_{j,i}\right] \leq \sigma_{P^1\otimes P^2}^2, \tag{A.37}$$

by the Cauchy-Schwarz inequality. Combining the above computations, we obtain that

$$\mathbb{E}\left[\left(\frac{T_n(\mathbb{X}_n)}{\sqrt{n}}\right)^2\right] \leq 2\frac{n(n-1)}{n^3}\sigma_{P^1\otimes P^2}^2 \underset{n\to+\infty}{\longrightarrow} 0,$$

and therefore,

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

Finally, from (A.36), we derive (A.35).

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

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

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

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

Let us now recall the result of Lemma A.1:

$$\mathcal{L}\left(\left(M_n^{\Pi_n}(\mathbb{X}_n), M_n^{\Pi_n'}(\mathbb{X}_n) \right)' \right) \underset{n \to +\infty}{\Longrightarrow} \mathcal{N}_2\left(0, \begin{pmatrix} \sigma^2_{P^1 \otimes P^2} & 0 \\ 0 & \sigma^2_{P^1 \otimes P^2} \end{pmatrix} \right).$$

Proof of Lemma A.1 According to the Cramér-Wold device, given a, b in $\mathbb{R}$, we aim at proving that

$$\mathcal{L}\left(a M_n^{\Pi_n}(\mathbb{X}_n) + b M_n^{\Pi_n'}(\mathbb{X}_n) \right) \underset{n \to +\infty}{\Longrightarrow} \mathcal{N}\left(0, \left(a^2 + b^2 \right) \sigma^2_{P^1 \otimes P^2} \right).$$

In order to deal with simpler mathematical expressions, we introduce below some additional notation.

- For $n \geq i \geq 2$, $\mathcal{F}_{n,i} = \sigma\left(\Pi_n, \Pi_n', X_1, X_2, \ldots, X_i \right)$.
- Let

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

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

so that $M_n^{\Pi_n}(\mathbb{X}_n) = \sum_{i=1}^n Y_{n,i}$ and $M_n^{\Pi_n'}(\mathbb{X}_n) = \sum_{i=1}^n Y_{n,i}'$.

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

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

In the same way, we have that $\mathbb{E}\left[Y'_{n,i}\middle|\mathcal{F}_{n,i-1}\right] = 0$, so $\mathbb{E}\left[aY_{n,i} + bY'_{n,i}\middle|\mathcal{F}_{n,i-1}\right] = 0$. From Theorem A.1, we thus deduce that if

$$(i) \qquad \sum_{i=2}^{n}\mathbb{E}\left[\left(aY_{n,i} + bY'_{n,i}\right)^2\middle|\mathcal{F}_{n,i-1}\right] \xrightarrow[n\to+\infty]{\mathbb{P}} (a^2+b^2)\sigma^2_{P^1\otimes P^2},$$

$$(ii) \qquad \sum_{i=2}^{n}\mathbb{E}\left[\left(aY_{n,i} + bY'_{n,i}\right)^2\mathbb{1}_{\left|aY_{n,i}+bY'_{n,i}\right|>\varepsilon}\right] \xrightarrow[n\to+\infty]{} 0 \text{ for any } \varepsilon > 0,$$

then

$$\mathcal{L}\left(aM_n^{\Pi_n}(\mathbb{X}_n) + bM_n^{\Pi'_n}(\mathbb{X}_n)\right) \underset{n\to+\infty}{\Longrightarrow} \mathcal{N}\left(0, \left(a^2+b^2\right)\sigma^2_{P^1\otimes P^2}\right).$$

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

Assumption (i). In all the following, only consider $n \geq 4$. Noticing that

$$\begin{aligned}
&\sum_{i=2}^{n}\mathbb{E}\left[\left(aY_{n,i} + bY'_{n,i}\right)^2\middle|\mathcal{F}_{n,i-1}\right] \\
&\qquad = (a^2+b^2)\sum_{i=2}^{n}\mathbb{E}\left[Y^2_{n,i}\middle|\mathcal{F}_{n,i-1}\right] + 2ab\sum_{i=2}^{n}\mathbb{E}\left[Y_{n,i}Y'_{n,i}\middle|\mathcal{F}_{n,i-1}\right], \quad \text{(A.38)}
\end{aligned}$$

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

The first point consists in proving that

$$\sum_{i=2}^{n}\mathbb{E}\left[Y^2_{n,i}\right] \xrightarrow[n\to+\infty]{} \sigma^2_{P^1\otimes P^2} \quad \text{and} \quad \mathbb{V}\text{ar}\left(\sum_{i=2}^{n}\mathbb{E}\left[Y^2_{n,i}\middle|\mathcal{F}_{n,i-1}\right]\right) \xrightarrow[n\to+\infty]{} 0,$$

which leads, thanks to Chebychev's inequality, to

$$\sum_{i=2}^{n}\mathbb{E}\left[Y^2_{n,i}\middle|\mathcal{F}_{n,i-1}\right] \xrightarrow[n\to+\infty]{\mathbb{P}} \sigma^2_{P^1\otimes P^2}.$$

The second point consists in proving that

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

so

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

• First point. On the one hand,

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

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

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

Thus,

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

so $\sum_{i=2}^{n} \mathbb{E}\left[Y_{n,i}^2\right] = \frac{n-1}{n}\mathbb{E}\left[C_{1,2}^2\right] + \frac{1}{n}\mathbb{E}\left[C_{1,2} C_{2,1}\right]$. From (A.37), we derive that

$$\sum_{i=2}^{n} \mathbb{E}\left[Y_{n,i}^2\right] \xrightarrow[n\to+\infty]{} \sigma^2_{P^1 \otimes P^2}. \tag{A.39}$$

On the other hand, we have that

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

Then,

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

with

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

Thus,

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}\left(\sum_{i=2}^{n}\left(\mathbb{E}\left[Y_{n,i}^2\middle|\mathcal{F}_{n,i-1}\right]\right)\right)\\
&\qquad\le 4\left(\mathbb{E}\left[A_{n,1}^2\right]+\mathbb{E}\left[A_{n,2}^2\right]+4\mathbb{E}\left[A_{n,3}^2\right]+4\mathbb{E}\left[A_{n,4}^2\right]\right). \quad \text{(A.40)}
\end{aligned}$$

Let us now control each term of the above right-hand side.
Convergence of $\mathbb{E}\left[A_{n,1}^2\right]$ *and* $\mathbb{E}\left[A_{n,2}^2\right]$.

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

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

$$\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=l}\right]\mathbb{E}\left[\mathbb{E}\left[C_{i,j}^2\middle|X_j\right]\mathbb{E}\left[C_{k,l}^2\middle|X_l\right]\right]=\frac{1}{n}\mathbb{E}\left[\left(\mathbb{E}\left[C_{2,1}^2\middle|X_1\right]\right)^2\right].$$

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

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

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

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

By combining these results, from (A.37) and under the assumption $(\mathcal{A}_{\varphi,Mmt})$, we obtain that

$$\begin{aligned}\mathbb{E}\left[A_{n,1}^2\right] &\leq \frac{n-1}{2n^2}\left(\mathbb{E}\left[\left(\mathbb{E}\left[C_{2,1}^2|X_1\right]\right)^2\right] - \frac{\sigma_{P^1\otimes P^2}^4}{n}\right) \\ &\quad + C\, n^2\left(\frac{1}{n(n-1)} - \frac{1}{n^2}\right)\sigma_{P^1\otimes P^2}^4 \underset{n\to+\infty}{\longrightarrow} 0.\end{aligned}$$

One can prove in the same way that $\mathbb{E}\left[A_{n,2}^2\right] \underset{n\to+\infty}{\longrightarrow} 0$.

Convergence of $\mathbb{E}\left[A_{n,3}^2\right]$. We easily prove that

$$\mathbb{E}\left[A_{n,3}^2\right] = \frac{1}{n^2}\sum_{1\leq j<i\leq n}\sum_{1\leq l<k\leq n}\kappa_{i,j,k,l} - \frac{1}{4n^2}\left(\mathbb{E}\left[C_{1,2}C_{2,1}\right]\right)^2,$$

where

$$\begin{aligned}\kappa_{i,j,k,l} = \mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(j)=i}\mathbb{1}_{\Pi_n(k)=l}\mathbb{1}_{\Pi_n(l)=k}\right]\times \\ \mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{j,i}|X_j\right]\mathbb{E}\left[C_{k,l}C_{l,k}|X_l\right]\right].\end{aligned}$$

Let us again consider $\kappa_{i,j,k,l}$ in all the cases where $1\leq j<i\leq n$, and $1\leq l<k\leq n$.
If $i=k$ and $j=l$, then

$$\kappa_{i,j,k,l} = \frac{1}{n(n-1)}\mathbb{E}\left[\left(\mathbb{E}\left[C_{2,1}C_{1,2}|X_1\right]\right)^2\right].$$

If $i=k$ and $j\neq l$, or if $i\neq k$ and $j=l$, then $\kappa_{i,j,k,l}=0$.
If $i\neq k$ and $j\neq l$, then

$$\kappa_{i,j,k,l} = \frac{\left(\mathbb{E}\left[C_{1,2}C_{2,1}\right]\right)^2}{n(n-1)(n-2)(n-3)}.$$

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

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

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

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

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 $\#\{j,k,p,q\}\geq 3$, 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,

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

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

$$\mathbb{E}\left[\mathbb{E}\left[C_{i,j}C_{k,i}|X_j,X_k\right]\mathbb{E}\left[C_{l,p}C_{q,l}|X_p,X_q\right]\right]=0. \tag{A.41}$$

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

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

and

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

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

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

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

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

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

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

and

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

By combining these results, we obtain that

$$\begin{aligned}
\mathbb{E}\left[A_{n,4}^2\right] \leq C\frac{n^3}{n^2}\frac{\mathbb{E}\left[\left(\mathbb{E}\left[C_{3,1}C_{2,3}|X_1,X_2\right]\right)^2\right]}{n(n-1)} \\
+ C'\,\frac{n^4}{n^2}\frac{(n-4)!}{n!}\mathbb{E}\left[C_{3,1}C_{2,3}C_{4,2}C_{1,4}\right] \underset{n\to+\infty}{\longrightarrow} 0.
\end{aligned}$$

From (A.40), and the above results of convergence towards 0 for $\mathbb{E}\left[A_{n,1}^2\right]$, $\mathbb{E}\left[A_{n,2}^2\right]$, $\mathbb{E}\left[A_{n,3}^2\right]$, and $\mathbb{E}\left[A_{n,4}^2\right]$, we firstly derive that

$$\mathbb{V}\mathrm{ar}\left(\sum_{i=2}^{n}\left(\mathbb{E}\left[Y_{n,i}^2\middle|\mathcal{F}_{n,i-1}\right]\right)\right)\underset{n\to+\infty}{\longrightarrow}0.$$

• Second point. Notice that

$$\mathbb{E}\left[Y_{n,i}Y'_{n,i}\middle|\mathcal{F}_{n,i-1}\right]=B_{n,1}+B_{n,2}+B_{n,3}+B_{n,4},$$

with

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

and

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

Thus,

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{i=2}^{n}\mathbb{E}\left[Y_{n,i}Y'_{n,i}\middle|\mathcal{F}_{n,i-1}\right]\right)^2\right]&\\
\le 4\left(\mathbb{E}\left[B_{n,1}^2\right]+\mathbb{E}\left[B_{n,2}^2\right]+\mathbb{E}\left[B_{n,3}^2\right]+\mathbb{E}\left[B_{n,4}^2\right]\right).&\quad\text{(A.42)}
\end{aligned}$$

Convergence of $\mathbb{E}\left[B_{n,1}^2\right]$ *and* $\mathbb{E}\left[B_{n,2}^2\right]$. It can be proved that

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

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

$$\mathbb{E}\left[B_{n,1}^2\right]\le\frac{n-1}{2n^3}\mathbb{E}\left[\left(\mathbb{E}\left[C_{1,2}^2\middle|X_2\right]\right)^2\right]+C\frac{\sigma_{P^1\otimes P^2}^4}{n-1}\underset{n\to+\infty}{\longrightarrow}0.$$

In the same way, we also prove that $\mathbb{E}\left[B_{n,2}^2\right] \underset{n\to+\infty}{\longrightarrow} 0$.

Convergence of $\mathbb{E}\left[B_{n,3}^2\right]$. We also have that

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

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

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

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

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

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

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

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

then $B_{n,4} = B_{n,4,1} + B_{n,4,2} + B_{n,4,3} + B_{n,4,4}$ and in particular,

$$\mathbb{E}\left[{B_{n,4}}^2\right] \leq 4\left(\mathbb{E}\left[{B_{n,4}}^2\right] + \mathbb{E}\left[{B_{n,4}}^2\right] + \mathbb{E}\left[{B_{n,4}}^2\right] + \mathbb{E}\left[{B_{n,4}}^2\right]\right).$$

Yet,

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

Now, consider all the cases where $1 \leq j \neq k < i \leq n$, $1 \leq p \neq q < l \leq n$.
If $\#\{j,k,p,q\} \geq 3$, using a similar argument as in (A.41), we obtain that

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

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

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

and

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

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

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

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

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

and

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

By combining these results, we obtain that

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

Following the same lines of proof, we furthermore obtain that $\mathbb{E}\left[B_{n,4,2}^2\right]$, $\mathbb{E}\left[B_{n,4,3}^2\right]$, and $\mathbb{E}\left[B_{n,4,4}^2\right]$ also converge towards 0. Hence, $\mathbb{E}\left[B_{n,4}^2\right]\underset{n\to+\infty}{\longrightarrow}0$. From (A.42), and the convergence towards 0 of $\mathbb{E}\left[B_{n,1}^2\right]$, $\mathbb{E}\left[B_{n,2}^2\right]$, $\mathbb{E}\left[B_{n,3}^2\right]$, and $\mathbb{E}\left[B_{n,4}^2\right]$, we derive that

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

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

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

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

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

Since $Y_{n,i} = n^{-1/2}\left(\mathbb{1}_{\Pi_n(i)<i} C_{i,\Pi_n(i)} + \mathbb{1}_{\Pi_n^{-1}(i)<i} C_{\Pi_n^{-1}(i),i}\right)$,

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

We thus obtain that

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

where the right-hand side tends to 0 as soon as $\mathbb{E}\left[C_{1,2}^4\right] < +\infty$.
This last condition is ensured by $(\mathcal{A}_{\varphi,Mmt})$, which allows to confirm that assumption (ii) is also checked, and that

$$\mathcal{L}\left(aM_n^{\Pi_n}(\mathbb{X}_n) + bM_n^{\Pi'_n}(\mathbb{X}_n)\right) \underset{n\to+\infty}{\Longrightarrow} \mathcal{N}\left(0, \left(a^2+b^2\right)\sigma^2_{P^1\otimes P^2}\right).$$

This ends the proof of Lemma A.1.

Recall that we aim at proving that

$$d_{BL}\left(\mathcal{L}\left(M_n^{\Pi_n}(\mathbb{X}_n)\middle|\mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1\otimes P^2}\right)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0.$$

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

$$\begin{cases}
\mathbb{P}\left(M_n^{\Pi_n}(\mathbb{X}_n) \leq t\right) \underset{n\to+\infty}{\longrightarrow} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(t), \\
\mathbb{P}\left(M_n^{\Pi_n}(\mathbb{X}_n) \leq t, M_n^{\Pi'_n}(\mathbb{X}_n) \leq t\right) \underset{n\to+\infty}{\longrightarrow} \Phi^2_{0,\sigma^2_{P^1\otimes P^2}}(t).
\end{cases}$$

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

$$\begin{cases}
\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{M_n\leq t}|\mathbb{X}_n\right]\right] \underset{n\to+\infty}{\longrightarrow} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(t), \\
\mathbb{E}\left[\left(\mathbb{E}\left[\mathbb{1}_{M_n\leq t}|\mathbb{X}_n\right]\right)^2\right] \underset{n\to+\infty}{\longrightarrow} \Phi^2_{0,\sigma^2_{P^1\otimes P^2}}(t).
\end{cases} \tag{A.43}$$

In a separable metric space, convergence in probability is metrizable (see [3, Th. 9.2.1] for instance), therefore it is equivalent to almost sure convergence of a subsequence of any initial subsequence. So, let us fix an initial extraction $\phi_0 : \mathbb{N} \to \mathbb{N}$, which defines a subsequence $\left(M_{\phi_0(n)}\right)_{n\in\mathbb{N}}$ 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 (A.43), we derive that

$$\begin{cases}
\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{M_{\phi_0(n)}\leq q_m}\middle|\mathbb{X}_{\phi_0(n)}\right]\right] \underset{n\to+\infty}{\longrightarrow} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(q_m), \\
\mathbb{E}\left[\left(\mathbb{E}\left[\mathbb{1}_{M_{\phi_0(n)}\leq q_m}\middle|\mathbb{X}_{\phi_0(n)}\right]^2\right)\right] \underset{n\to+\infty}{\longrightarrow} \Phi^2_{0,\sigma^2_{P^1\otimes P^2}}(q_m),
\end{cases}$$

which leads (by Chebychev's inequality) to

$$\mathbb{E}\left[\mathbb{1}_{M_{\phi_0(n)}\leq q_m}\middle|\mathbb{X}_{\phi_0(n)}\right] \xrightarrow[n\to+\infty]{\mathbb{P}} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(q_m). \tag{A.44}$$

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

$$\mathbb{E}\left[\mathbb{1}_{M_{\phi_1\circ\phi(n)}\leq q_1}\middle|\mathbb{X}_{\phi_1\circ\phi(n)}\right](\omega) \xrightarrow[n\to+\infty]{} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(q_1).$$

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

$$\mathbb{E}\left[\mathbb{1}_{M_{\phi_m\circ\phi_{m-1}\circ\ldots\circ\phi_0(n)}\leq q_m}\middle|\mathbb{X}_{\phi_m\circ\phi_{m-1}\circ\ldots\circ\phi_0(n)}\right](\omega) \xrightarrow[n\to+\infty]{} \Phi_{0,\sigma^2_{P^1\otimes P^2}}(q_m).$$

Then, from (A.44), 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 ω in Ω_{m+1},

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

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

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

Since $\Phi_{0,\sigma^2_{P^1\otimes P^2}}$ is a continuous distribution function, it can be proved that this follows

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

To conclude, we actually proved that

$$d_{BL}\left(\mathcal{L}\left(M_n^{\Pi_n}(\mathbb{X}_n)\middle|\mathbb{X}_n\right),\mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0,$$

which, combined with (A.35), leads to

$$d_{BL}\left(\mathcal{L}\left(\sqrt{n}U_n\left(\mathbb{X}_n^{\Pi_n}\right)\middle|\mathbb{X}_n\right),\mathcal{N}\left(0,\sigma^2_{P^1\otimes P^2}\right)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Third step: convergence of conditional second order moments Recall that from (A.33), $U_n\left(\mathbb{X}_n^{\Pi_n}\right)=\frac{1}{n-1}U_n^{\Pi_n}$, where

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

Therefore,

$$\mathbb{E}\left[\left(\sqrt{n}U_n\left(\mathbb{X}_n^{\Pi_n}\right)\right)^2\middle|\mathbb{X}_n\right] = \frac{n^2}{(n-1)^2}\left(\frac{1}{n}\mathbb{E}\left[\left(U_n^{\Pi_n}\right)^2\middle|\mathbb{X}_n\right]\right), \tag{A.45}$$

and if $C_{i,j,k,l} = \left(\mathbb{E}\left[\mathbb{1}_{\Pi_n(i)=j}\mathbb{1}_{\Pi_n(k)=l}\right] - \frac{1}{n^2}\right)\varphi\left(X_i^1, X_j^2\right)\varphi\left(X_k^1, X_l^2\right)$,

$$\frac{1}{n}\mathbb{E}\left[\left(U_n^{\Pi_n}\right)^2\middle|\mathbb{X}_n\right] = \frac{1}{n}\sum_{i,j=1}^{n}\sum_{k,l=1}^{n} C_{i,j,k,l}.$$

Firstly,

$$\frac{1}{n}\sum_{\substack{i,j,k,l\in\{1,\dots,n\}\\ \#\{i,j,k,l\}=4}} C_{i,j,k,l} = \frac{(n-2)(n-3)}{n^2}U_{n,1},$$

where

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

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

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

Secondly,

$$\frac{1}{n}\sum_{\substack{i,j,k,l\in\{1,\dots,n\}\\ \#\{i,j,k,l\}=3\\ i=j, i=l, j=k, \text{ or } k=l}} C_{i,j,k,l} = \frac{2(n-2)}{n^2}U_{n,2},$$

where

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

is a U-statistic of order 3 which converges almost surely, so

$$\frac{2(n-2)}{n^2}U_{n,2} \xrightarrow[n\to+\infty]{a.s.} 0.$$

Thirdly,

$$\frac{1}{n}\sum_{\substack{i,j,k,l\in\{1,\dots,n\}\\ \#\{i,j,k,l\}=3\\ i=k, \text{ or } j=l}} C_{i,j,k,l} = -\frac{n(n-1)(n-2)}{n^3}U_{n,3},$$

where

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

is a U-statistic of order 3. So,

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

Fourthly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=2 \\ i=j=k, i=j=l, \\ i=k=l, \text{ or } j=k=l}} C_{i,j,k,l} = -\frac{2(n-1)}{n^2} U_{n,4},$$

where

$$U_{n,4} = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} \left(\varphi\left(X_i^1, X_i^2\right) \varphi\left(X_i^1, X_j^2\right) + \varphi\left(X_i^1, X_i^2\right) \varphi\left(X_j^1, X_i^2\right)\right)$$

is a U-statistic of order 2, so

$$-\frac{2(n-1)}{n^2} U_{n,4} \xrightarrow[n\to+\infty]{a.s.} 0.$$

Fifthly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=2 \\ i=j\ne k=l, \text{ or } i=l\ne j=k}} C_{i,j,k,l} = \frac{1}{n^2} U_{n,5},$$

where

$$U_{n,5} = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} \left(\varphi\left(X_i^1, X_i^2\right) \varphi\left(X_j^1, X_j^2\right) + \varphi\left(X_i^1, X_j^2\right) \varphi\left(X_j^1, X_i^2\right)\right)$$

is a U-statistic of order 2, so

$$\frac{1}{n^2} U_{n,5} \xrightarrow[n\to+\infty]{a.s.} 0.$$

Sixthly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=2 \\ i=k\ne j=l}} C_{i,j,k,l} = \frac{(n-1)^2}{n^2} U_{n,6},$$

where

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

is a U-statistic of order 2, so

$$\frac{(n-1)^2}{n^2} U_{n,6} \xrightarrow[n \to +\infty]{a.s.} \mathbb{E}\left[\varphi^2\big(X_1^1, X_2^2\big)\right].$$

Seventhly,

$$\frac{1}{n} \sum_{\substack{i,j,k,l \in \{1,\dots,n\} \\ \#\{i,j,k,l\}=1}} C_{i,j,k,l} = \frac{n-1}{n^3} \sum_{i=1}^{n} \varphi\big(X_i^1, X_i^2\big),$$

which almost surely tends to 0 thanks to the strong law of large numbers.
By combining all these results, and the fact that

$$\begin{aligned}\sigma^2_{P^1 \otimes P^2} = \mathbb{E}\left[\varphi^2\big(X_1^1, X_2^2\big)\right] + \left(\mathbb{E}\left[\varphi\big(X_1^1, X_2^2\big)\right]\right)^2 \\ - \mathbb{E}\left[\left(\mathbb{E}\left[\varphi(X_1^1, X_2^2)\middle|X_1\right]\right)^2\right] - \mathbb{E}\left[\left(\mathbb{E}\left[\varphi(X_1^1, X_2^2)\middle|X_2\right]\right)^2\right],\end{aligned}$$

we finally obtain that

$$\frac{1}{n}\mathbb{E}\left[\left(U_n^{\Pi_n}\right)^2 \middle| \mathbb{X}_n\right] \xrightarrow[n \to +\infty]{a.s.} \sigma^2_{P^1 \otimes P^2},$$

and from (A.45), we deduce that

$$\mathbb{E}\left[\left(\sqrt{n} U_n\big(\mathbb{X}_n^{\Pi_n}\big)\right)^2 \middle| \mathbb{X}_n\right] \xrightarrow[n \to +\infty]{a.s.} \sigma^2_{P^1 \otimes P^2}.$$

Since $d_{BL}\left(\mathcal{L}\left(\sqrt{n} U_n\big(\mathbb{X}_n^{\Pi_n}\big)\middle|\mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1 \otimes P^2}\right)\right) \xrightarrow[n \to +\infty]{\mathbb{P}} 0$, this allows to conclude that

$$d_2\left(\mathcal{L}\left(\sqrt{n} U_n\big(\mathbb{X}_n^{\Pi_n}\big)\middle|\mathbb{X}_n\right), \mathcal{N}\left(0, \sigma^2_{P^1 \otimes P^2}\right)\right) \xrightarrow[n \to +\infty]{\mathbb{P}} 0.$$

A.12 Proof of Corollary 4.1

Here, unlike the bootstrap approach, we only have in Theorem 4.1 a consistency result in probability. Thus, as for Proposition 3.6, we use an argument of subsequences. So let $\phi_0 : \mathbb{N} \to \mathbb{N}$ be an extraction defining a subsequence. By Theorem 4.1, there exists an extraction ϕ_1 such that P-a.s. in $(X_i)_i$,

$$\mathcal{L}\left(\sqrt{\phi_1 \circ \phi_0(n)} U_{\phi_1 \circ \phi_0(n)}\left(\mathbb{X}^{\Pi_{\phi_1 \circ \phi_0(n)}}_{\phi_1 \circ \phi_0(n)}\right)\middle|\mathbb{X}_{\phi_1 \circ \phi_0(n)}\right) \underset{n \to +\infty}{\Longrightarrow} \mathcal{N}\left(0, \sigma^2_{P^1 \otimes P^2}\right). \tag{A.46}$$

In particular, applying [6, Lemma 21.2] on the event where the convergence is true, we obtain that for η in $(0,1)$, $q^{\star}_{\eta, \phi_1 \circ \phi_0(n)}\left(\mathbb{X}_{\phi_1 \circ \phi_0(n)}\right) \xrightarrow[n \to +\infty]{a.s.} \Phi^{-1}_{0, \sigma^2_{P^1 \otimes P^2}}(\eta)$, which ends the proof by [3, Theorem 9.2.1].

A.13 Proof of Theorem 4.2

As in the proof of Theorem 3.2, we focus on the upper-tailed tests of $\Gamma(q^\star)$, the other cases being similar. The proof of Theorem 4.2 is very similar to the one of Theorem 3.2, just replacing the argument of (A.27) by $q^\star_{1-\alpha,n}(\mathbb{X}_n) \xrightarrow[n\to+\infty]{\mathbb{P}} \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(1-\alpha)$, which is derived from Corollary 4.1.

Now for the test with the Monte Carlo approximation of the quantiles, as above, we focus on the upper-tailed test of $\Gamma(q^\star_{MC})$. The proof is very similar to the one of Proposition 3.6. We first show, following the proof of [6, Lemma 21.2], that

$$\sqrt{n}U^{\star(\lceil(1-\alpha)(B_n+1)\rceil)} \xrightarrow[n\to+\infty]{\mathbb{P}} \Phi^{-1}_{0,\sigma^2_{P^1\otimes P^2}}(1-\alpha), \tag{A.47}$$

then, we conclude using the same arguments as in the proof of Theorem 4.2.

Let $F^\star_{n,\mathbb{X}_n}$ be the c.d.f of $\mathcal{L}\left(\sqrt{n}U_n, P^\star_n|\mathbb{X}_n\right)$, and let us first show that

$$\sup_{z\in\mathbb{R}}\left|F^\star_{n,\mathbb{X}_n}(z) - \Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0. \tag{A.48}$$

As Theorem 4.1 provides only a convergence in probability, similar arguments of subsequences as in the proof of Corollary 4.1, have to be used. So, let ϕ_0 be an initial extraction and ϕ_1 be the extraction such that (A.46) is satisfied. As convergence in the d_{BL} metric is equivalent to a weak convergence (see [3, Proposition 11.3.3] for instance), and as the limit is continuous, by [6, Lemma 2.11] we obtain that

$$\sup_{z\in\mathbb{R}}\left|F^\star_{\phi_1\circ\phi_0(n),\mathbb{X}_{\phi_1\circ\phi_0(n)}}(z) - \Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right| \xrightarrow[n\to+\infty]{a.s.} 0.$$

This being true for any initial extraction ϕ_0, we obtain (A.48).
Let $F^{\star B_n}_{n,\mathbb{X}_n}$ denote the empirical c.d.f of $\mathcal{L}\left(\sqrt{n}U_n, P^\star_n|\mathbb{X}_n\right)$ associated with the sample $\left(\Pi^1_n,\dots,\Pi^{B_n}_n\right)$, that is

$$\forall z\in\mathbb{R},\quad F^{\star B_n}_{n,\mathbb{X}_n}(z) = \frac{1}{B_n}\sum_{b=1}^{B_n}\mathbb{1}_{\sqrt{n}U_n\left(\mathbb{X}^{\Pi^b_n}_n\right)\le z}.$$

Then, by the DKW inequality, we obtain as in the proof of Proposition 3.6,

$$\sup_{z\in\mathbb{R}}\left|F^{\star B_n}_{n,\mathbb{X}_n}(z) - F^\star_{n,\mathbb{X}_n}(z)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0. \tag{A.49}$$

Finally, let

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

Since $G^{\star B_n}_{n,\mathbb{X}_n}(z) = \frac{1}{B_n+1}\left(\mathbb{1}_{\sqrt{n}U_n(\mathbb{X}_n)\le z} + B_nF^{\star B_n}_{n,\mathbb{X}_n}(z)\right)$,

$$\sup_{z\in\mathbb{R}}\left|G^{\star B_n}_{n,\mathbb{X}_n}(z) - F^{\star B_n}_{n,\mathbb{X}_n}(z)\right| \le \frac{2}{B_n+1} \xrightarrow[n\to+\infty]{} 0. \tag{A.50}$$

Combining (A.48), (A.49) and (A.50), we obtain that

$$\sup_{z\in\mathbb{R}}\left|G^{\star B_n}_{n,\mathbb{X}_n}(z)-\Phi_{0,\sigma^2_{P^1\otimes P^2}}(z)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Since

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

we obtain (A.47) and we conclude as in the proof of Proposition 3.6.

B Additional Results

B.1 About the non-degeneracy of the U-statistic

We focus on the *Linear case* with $\varphi=\varphi^w$ given by (2.2). Define

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

Recall that in this case, degeneracy is equivalent to stating that for $X=(X^1,X^2)$ with distribution $P^1\otimes P^2$, $Z(X)$ is a random variable which is almost surely null under (H_0). Since $\mathbb{E}[Z(X)]=0$, $Z(X)=0$ a.s. is equivalent to $\mathbb{V}\mathrm{ar}(Z(X))=0$. Here we provide a computation of $\mathbb{V}\mathrm{ar}(Z(X))$.

Let us introduce $dM_1^{[1]}(u)$ and $dM_2^{[1]}(v)$ the mean measures of respectively X^1 with distribution P^1 and X^2 of distribution P^2 [2, Chapter 5], then one can rewrite

$$\begin{aligned}Z(X)=&\int w(u,v)dN_{X^1}(u)dN_{X^2}(v)+\int w(u,v)dM_1^{[1]}(u)dM_2^{[1]}(v)\\&-\int w(u,v)dN_{X^1}(u)dM_2^{[1]}(v)-\int w(u,v)dM_1^{[1]}(u)dN_{X^2}(v).\end{aligned}$$

Therefore, $\mathbb{E}[Z(X)]=0$, and

$$\begin{aligned}\mathbb{V}\mathrm{ar}(Z(X)) &= \mathbb{E}\left[Z(X)^2\right]\\ &= \int_{[0,1]^4} w(u,v)w(s,t)\mathbb{E}\left[dN_{X^1}(u)dN_{X^1}(s)\right]\mathbb{E}\left[dN_{X^2}(v)dN_{X^2}(t)\right]\\ &\quad-\int_{[0,1]^4} w(u,v)w(s,t)\mathbb{E}\left[dN_{X^1}(u)dN_{X^1}(s)\right]dM_2^{[1]}(v)dM_2^{[1]}(t)\\ &\quad-\int_{[0,1]^4} w(u,v)w(s,t)dM_1^{[1]}(u)dM_1^{[1]}(s)\mathbb{E}\left[dN_{X^2}(v)dN_{X^2}(t)\right]\\ &\quad+\int_{[0,1]^4} w(u,v)w(s,t)dM_1^{[1]}(u)dM_1^{[1]}(s)dM_2^{[1]}(v)dM_2^{[1]}(t).\end{aligned}$$

By assuming that $\#X^1$ (resp. $\#X^2$) has second order moment, (see also Section 3.2 for comment on this assumption), one can introduce the second factorial moment measure associated with X^1 (resp. X^2), and denoted by $dM_1^{[2]}(u,s)$ (resp. $dM_2^{[2]}(v,t)$). Then straightforward computations show that

$$\begin{aligned}\mathbb{V}\mathrm{ar}(Z(X)) &= \int_{[0,1]^2} w(u,v)^2 dM_1^{[1]}(u) dM_2^{[1]}(v) \\ &+ \int_{[0,1]^3} w(u,v)w(u,t) dM_1^{[1]}(u) \left(dM_2^{[2]}(v,t) - dM_2^{[1]}(v) dM_2^{[1]}(t) \right) \\ &+ \int_{[0,1]^3} w(u,v)w(s,v) \left(dM_1^{[2]}(u,s) - dM_1^{[1]}(u) dM_1^{[1]}(s) \right) dM_2^{[1]}(v) \\ &+ \int_{[0,1]^4} w(u,v)w(s,t) \left(dM_1^{[2]}(u,s) - dM_1^{[1]}(u) dM_1^{[1]}(s) \right) \times \\ &\quad \left(dM_2^{[2]}(v,t) - dM_2^{[1]}(v) dM_2^{[1]}(t) \right).\end{aligned}$$

In particular, for Poisson processes, $dM^{[2]}(u,s) = dM^{[1]}(u) dM^{[1]}(s)$ and

$$\mathbb{V}\mathrm{ar}\left(Z(X)\right) = \int_{[0,1]^2} w(u,v)^2 dM_1^{[1]}(u) dM_2^{[1]}(v) > 0,$$

as soon as the Poisson processes have non zero intensities since for $j = 1, 2$, $dM_j^{[1]}(u) = \lambda_j(u)du$, with λ_j the intensity of X^j.

B.2 About the empirical centering assumption

Recall that

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

On the one hand, in the *Linear case*, that is if $h = h_\varphi$, then for $n \geq 1$ and for $x_1 = (x_1^1, x_1^2), \ldots, x_n = (x_n^1, x_n^2)$ in $\mathcal{X}^2$,

$$\begin{aligned}&\sum_{i,i',j,j'=1}^n h\left(\left(x_i^1, x_{i'}^2\right), \left(x_j^1, x_{j'}^2\right) \right) \\ &\quad = \frac{1}{2} \sum_{i,i',j,j'=1}^n \left(\varphi\left(x_i^1, x_{i'}^2\right) + \varphi\left(x_j^1, x_{j'}^2\right) - \varphi\left(x_i^1, x_{j'}^2\right) - \varphi\left(x_j^1, x_{i'}^2\right) \right) \\ &\quad = \frac{n^2}{2} \left(\sum_{i,i'=1}^n \varphi(x_i^1, x_{i'}^2) + \sum_{j,j'=1}^n \varphi(x_j^1, x_{j'}^2) - \sum_{i,j'=1}^n \varphi(x_i^1, x_{j'}^2) - \sum_{j,i'=1}^n \varphi(x_j^1, x_{i'}^2) \right) \\ &\quad = 0.\end{aligned}$$

So $(\mathcal{A}^*_{Cent})$ is immediately satisfied in the *Linear case*.

On the other hand, $(\mathcal{A}^*_{Cent})$ does not imply that h is of the form h_φ. Indeed, consider

$$h\left(\left(x^1,x^2\right),\left(y^1,y^2\right)\right)=\#x^1\cdot\#x^2\cdot\#y^1\cdot\#y^2\left[\left(\#x^1-\#y^1\right)\left(\#x^2-\#y^2\right)\right].$$

- The kernel h is obviously symmetric.
- The kernel h satisfies $(\mathcal{A}^*_{Cent})$. Indeed, let

  $$f\left(x^1,y^1\right)=\#x^1\cdot\#y^1\left(\#x^1-\#y^1\right).$$

  First, notice that $f\left(x^1,x^1\right)=0$ and $f\left(x^1,y^1\right)=-f\left(y^1,x^1\right)$.
  Moreover, $h\left(\left(x^1,x^2\right),\left(y^1,y^2\right)\right)=f\left(x^1,y^1\right)f\left(x^2,y^2\right)$. Thus

  $$\begin{aligned}
  &\sum_{i,i',j,j'=1}^{n}h\left(\left(x_i^1,x_{i'}^2\right),\left(x_j^1,x_{j'}^2\right)\right)\\
  &\quad=\sum_{i,i',j,j'=1}^{n}f\left(x_i^1,x_j^1\right)f\left(x_{i'}^2,x_{j'}^2\right)\\
  &\quad=\left(\sum_{i,j=1}^{n}f\left(x_i^1,x_j^1\right)\right)\left(\sum_{i',j'=1}^{n}f\left(x_{i'}^2,x_{j'}^2\right)\right)\\
  &\quad=\left(\sum_{i=1}^{n}\underbrace{f\left(x_i^1,x_i^1\right)}_{0}+\sum_{1\leq i<j\leq n}\underbrace{f\left(x_i^1,x_j^1\right)+f\left(x_j^1,x_i^1\right)}_{0}\right)\left(\sum_{i',j'=1}^{n}f\left(x_{i'}^2,x_{j'}^2\right)\right)\\
  &\quad=0,
  \end{aligned}$$

  and thus $(\mathcal{A}^*_{Cent})$ is satisfied by h.
- The kernel h cannot be written as an h_φ.
  On the one hand, first notice that for any $\varphi:\mathcal{X}^2\to\mathbb{R}$, the difference

  $$D_{h_\varphi}:=h_\varphi\left(\left(x^1,x^2\right),\left(y^1,y^2\right)\right)-h_\varphi\left(\left(\tilde{x}^1,x^2\right),\left(y^1,y^2\right)\right)$$

  does not depend on y^1. Indeed,

  $$\begin{aligned}
  D_{h_\varphi}&=\frac{1}{2}\Big(\varphi\left(x^1,x^2\right)+\varphi\left(y^1,y^2\right)-\varphi\left(x^1,y^2\right)-\varphi\left(y^1,x^2\right)\\
  &\quad-\varphi\left(\tilde{x}^1,x^2\right)-\varphi\left(y^1,y^2\right)+\varphi\left(\tilde{x}^1,y^2\right)+\varphi\left(y^1,x^2\right)\Big)\\
  &=\frac{1}{2}\left(\varphi\left(x^1,x^2\right)-\varphi\left(\tilde{x}^1,x^2\right)+\varphi\left(\tilde{x}^1,y^2\right)-\varphi\left(x^1,y^2\right)\right).
  \end{aligned}$$

  On the other hand, for the kernel h introduced above, the difference D_h does depend on y^1. Indeed

  $$\begin{aligned}
  D_h&=h\left(\left(x^1,x^2\right),\left(y^1,y^2\right)\right)-h\left(\left(\tilde{x}^1,x^2\right),\left(y^1,y^2\right)\right)\\
  &=\#x^2\cdot\#y^1\cdot\#y^2\left(\#x^2-\#y^2\right)\times\\
  &\quad\left[\#x^1\cdot\left(\#x^1-\#y^1\right)-\#\tilde{x}^1\cdot\left(\#\tilde{x}^1-\#y^1\right)\right],
  \end{aligned}$$

and if for instance, $\#x^1 = \#y^2 = 1$ and $\#\tilde{x}^1 = \#x^2 = 2$, then

$$D_h = 2\#y^1 \left[-3 + \#y^1\right],$$

which clearly depends on y^1.

So finally, there does not exist any φ such that $h = h_\varphi$.

References

[1] B. M. Brown. Martingale central limit theorems. *Ann. Math. Statist.*, 42(1):59–66, 1971.

[2] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I.* Probability and its Applications. Springer-Verlag, New York, 2003.

[3] R. M. Dudley. *Real Analysis and Probability.* Cambridge University Press, 2002.

[4] W. Hoeffding. The strong law of large numbers for U-statistics. Institute of Statistics, Mimeograph Series No. 302, 1961.

[5] R. L. Schilling. *Measures, integrals and martingales.* Cambridge Univ. Press, 2005.

[6] A. W. Van der Vaart. *Asymptotic statistics.* Cambridge series in Statistical and Probabilistic Mathematics, 1998.

[7] V. S. Varadarajan. On the convergence of sample probability distributions. *Sankhyà*, 19(1–2):23–26, 1958.