Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution

Baptiste Broto, François Bachoc, Marine Depecker

Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution

Baptiste Broto[1], François Bachoc [2], and Marine Depecker[1]

[1]CEA, LIST, Université Paris-Saclay, F-91120, Palaiseau, France
[2]Institut de Mathématiques de Toulouse, Université Paul Sabatier, F-31062 Toulouse, France

Abstract

The Shapley effects are global sensitivity indices: they quantify the impact of each input variable on the output variable in a model. In this work, we suggest new estimators of these sensitivity indices. When the input distribution is known, we investigate the already existing estimator defined in [SNS16] and suggest a new one with a lower variance. Then, when the distribution of the inputs is unknown, we extend these estimators. We provide asymptotic properties of the estimators studied in this article. We also apply one of these estimators to a real data set.

1 Introduction

Sensitivity indices are important tools in sensitivity analysis. They aim to quantify the impact of the input variables on the output of a model. In this way, they give a better understanding of numerical models and improve their interpretability. For example, the sensitivity indices enable to know if the variation of a specific input variable can lead to an important variation of the output or not.

In global sensitivity analysis, the input variables $X_1, ..., X_p$ are assumed to be random variables. Sobol defined the first sensitivity indices for a general framework, called the Sobol indices, in [Sob93]. Many other sensitivity indices have been defined and studied (see [BHP16] for a general review of these indices). Nevertheless, many of these indices suffer from a lack of interpretation when the input variables are dependent. To overcome this lack of interpretation, many variants of the Sobol indices have been suggested for dependent input variables (see for example [JLD06], [MT12] and [Cha13]).

Recently, Owen defined new sensitivity indices in [Owe14] called "Shapley effects" that have beneficial properties and that are easy to interpret, even in

the dependent case. The main advantages of these sensitivity indices compared to the Sobol indices (and their variants) are: they remain positive, their sum is equal to one and there is exactly one index for each input (and there are no indices for groups of variables). The Shapley effects are based on the notion of "Shapley value", that originates from game theory in [Sha53]. The Shapley value has been widely studied ([CBSV16], [FWJ08]) and applied in different fields (see for example [MvLG$^+$08] or [HI03]). However, only few articles focus on the Shapley effects in sensitivity analysis (see [Owe14, SNS16, OP17, IP17, BBDM19, BEDC19]). Song et al. suggested an algorithm to estimate the Shapley effects in [SNS16] that is implemented in the R package "`sensitivity`".

In this paper, we work on the Shapley effects and their estimation. We divide this estimation into two parts. The first part is the estimation of quantities that we call the "conditional elements", on which the Shapley effects depend. The second part consists in aggregating the estimates of the conditional elements in order to obtain estimates of the Shapley effects. We call this part the W-aggregation procedure. We refer to Sections 3 and 4 for more details on these two parts.

First, we focus on the estimation of the conditional elements with two different estimators: the double Monte-Carlo estimator (used in the algorithm of [SNS16]) and the Pick-and-Freeze estimator (see [HS96] for the independent case) that we extend to the case where the inputs are dependent. We present the two estimators when it is possible to sample from the conditional distributions of the input vector. Then we suggest a new W-aggregation procedure, based on the subsets of $\{1, ..., p\}$, to estimate all the Shapley effects (for all the input variables) at the same time. We choose the best parameters to minimize the sum of the variances of all the Shapley effects estimators. The algorithm of [SNS16] uses a W-aggregation procedure based on permutations of $\{1, ..., p\}$. We study this W-aggregation procedure and explain how it minimizes the variance of the estimates of the Shapley effects. Our suggested W-aggregation procedure provides an improved accuracy, compared to the W-aggregation procedure in [SNS16], using all the estimates of the conditional elements for all the estimates of the Shapley effects. The comparison between the two W-aggregation procedures is illustrated with numerical experiments. These experiments also show that the double Monte-Carlo estimator provides better results than the Pick-and-Freeze estimator.

Then, we extend the estimators of the conditional elements (the double Monte-Carlo estimator and the Pick-and-Freeze estimator) to the case where we only observe an i.i.d. sample from the input variables. The extension relies on nearest-neighbour techniques, which are widely used for many non-parametric estimation problems [BS19, BSY19]. To the best of our knowledge, the estimators we suggest are the first that do not require exact samples from the conditional distributions of the input variables. One of our main results is the consistency of these estimators under some mild assumptions, and their rate of convergence under additional regularity assumptions. We then give the consistency of the estimators of the Shapley effects with the two W-aggregation procedures and using the double Monte-Carlo estimator or the Pick-and-Freeze

estimator. We observe, in numerical experiments, that the estimators of the Shapley effects have a similar accuracy as when it is possible to sample from the conditional distributions. We also apply one of these estimators on meteorological data, more specifically on the output of three different metamodels predicting the ozone concentration in function of nine input variables (with some categorical variables and some continuous variables). This application enables to study the influence of the inputs variables on black-box machine learning procedures.

The paper is organized as follows. In Section 2, we define the framework of global sensitivity analysis and we recall the definition and some properties of the Shapley effects. In Section 3, we assume that the input distribution is known and we present the two methods to estimate the conditional elements. In Section 4, we suggest a new W-aggregation procedure and we study the W-aggregation procedure used by the algorithm of [SNS16]. In Section 5, we summarize the four estimators of the Shapley effects, give their consistency and we illustrate them with numerical applications. In Section 6, we assume that the input distribution is unknown and that we just observe a sample of the input vector. We give consistent estimators of the conditional elements and thus consistent estimators of the Shapley effects in this case, and we illustrate this with numerical experiments. In Section 7, we apply one of our estimators to a real data set. We conclude in Section 8. All the proofs are provided in the appendix.

2 The Shapley effects

We let $\mathbf{X} = (X_1, ..., X_p)$ be the input random vector on the input domain $\mathcal{X} = \mathcal{X}_1 \times ... \times \mathcal{X}_p$ with distribution $\mathbb{P}_{\mathbf{X}}$. We assume that there is an output variable Y in $\mathbb{R}$ defined by

$$Y = f(\mathbf{X}), \tag{1}$$

with $f \in L^2(\mathbb{P}_{\mathbf{X}})$. We write $[1 : p]$ for the set $\{1, 2, ..., p\}$. For any non-empty $u \subset [1 : p]$, letting $u = \{i_1, ..., i_r\}$ with $i_1 < i_2 < ... < i_r$, we define $\mathbf{X}_u = (X_{i_1}, ..., X_{i_r})$. We can now define the conditional elements $(W_u)_{u\subset[1:p]}$ and the Shapley effects $(\eta_i)_{i\in[1:p]}$.

For all $u \subset [1 : p]$, we define:

$$V_u := \mathrm{Var}(\mathrm{E}(Y|\mathbf{X}_u)) \tag{2}$$

and

$$E_u := \mathrm{E}(\mathrm{Var}(Y|\mathbf{X}_{-u})), \tag{3}$$

where $-u := [1 : p]\backslash u$. We let by convention $\mathrm{E}(Y|\mathbf{X}_\emptyset) = \mathrm{E}(Y)$ and $\mathrm{Var}(Y|\mathbf{X}_\emptyset) = \mathrm{Var}(Y)$. We define the conditional elements $(W_u)_{u\subset[1:p]}$ as being either $(V_u)_{u\subset[1:p]}$ or $(E_u)_{u\subset[1:p]}$. For all $i \in [1 : p]$, we define the Shapley effect η_i as in [SNS16] by:

$$\eta_i := \frac{1}{p\mathrm{Var}(Y)} \sum_{u\subset -i} \binom{p-1}{|u|}^{-1} (W_{u\cup\{i\}} - W_u), \tag{4}$$

where we define $-i$ as the subset $[1:p] \setminus \{i\}$ and $|u|$ as the cardinality of u.

Remark 1. *As explained in [SNS16], the Shapley effects do not depend on whether* $(W_u)_{u\subset[1:p]}$ *denotes* $(V_u)_{u\subset[1:p]}$ *or* $(E_u)_{u\subset[1:p]}$.

Remark 2. *The quantities* $W_\emptyset$ *and* $W_{[1:p]}$ *are equal to* 0 *and* $\mathrm{Var}(Y)$ *respectively. The variance of* Y *is easy to estimate, so we assume without loss of generality that we know the theoretical value* $\mathrm{Var}(Y)$.

We can notice that the Shapley effects are a sum over the subsets $u \subset -i$. Another classical way to compute the Shapley effects is to sum over the permutations of $[1:p]$, see Proposition 1. We let $\mathcal{S}_p$ be the set of permutations of $[1:p]$. An element $\sigma \in \mathcal{S}_p$ is a bijective function from $[1:p]$ to $[1:p]$. We let σ^{-1} be its inverse function. As in [SNS16], for $i \in [1:p]$ and $\sigma \in \mathcal{S}_p$, we let $P_i(\sigma) := \{\sigma(j) |\ j \in [1:\sigma^{-1}(i)-1]\}$.

Proposition 1. *[Equation (11) in [SNS16], Section 4.1 (see also [CGT09])] We have*

$$\eta_i = \frac{1}{p!\mathrm{Var}(Y)} \sum_{\sigma\in\mathcal{S}_p} (W_{P_i(\sigma)\cup\{i\}} - W_{P_i(\sigma)}). \tag{5}$$

Our aim is to estimate the Shapley effects. We have seen two different ways to compute the Shapley effects, given by Equation (4) (with a sum over the subsets) and Equation (5) (with a sum over the permutations). These two equations will correspond to two different W-aggregation procedures of the Shapley effects.

3 Estimation of the conditional elements

We explain now how to estimate these $(W_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$ in a restricted setting (recall that $W_\emptyset = 0$ and $W_{[1:p]} = \mathrm{Var}(Y)$ are known). The restricted setting is the following: as in [SNS16], we will assume that for any $\emptyset \subsetneq u \subsetneq [1:p]$ and $\mathbf{x}_u \in \mathcal{X}_u := \prod_{i\in u} \mathcal{X}_i$, it is feasible to generate an i.i.d. sample from the distribution of $\mathbf{X}_{-u}$ conditionally to $\mathbf{X}_u = \mathbf{x}_u$. Moreover, we assume that we have access to the computer code of f.

To estimate W_u, we suggest two different estimators. The first one consists in a double Monte-Carlo procedure to estimate E_u, and it is the estimator used in the algorithm of [SNS16]. The other one is the well-known Pick-and-Freeze estimator (see [HS96] for the first definition, [GJKL14, GJK$^+$16] for theoretical studies) for V_u, that we extend to the case where the input variables $(X_i)_{i\in[1:p]}$ are not independent.

Finally, we assume that each evaluation of f is costly, so we define the cost of each estimator $\widehat{W_u}$ as the number of evaluations of f.

3.1 Double Monte-Carlo

A first way to estimate $E_u = \mathrm{E}(\mathrm{Var}(Y|\mathbf{X}_{-u}))$ is using double Monte-Carlo: a first Monte-Carlo step of size N_I for the conditional variance, another one of

size N_u for the expectation. Thus, the estimator of E_u suggested in [SNS16] is

$$\widehat{E}_{u,MC} := \frac{1}{N_u}\sum_{n=1}^{N_u}\frac{1}{N_I-1}\sum_{k=1}^{N_I}\left(f(\mathbf{X}_{-u}^{(n)},\mathbf{X}_u^{(n,k)}) - \overline{f(\mathbf{X}_{-u}^{(n)})}\right)^2, \tag{6}$$

where for $n = 1, ..., N_u$, $\overline{f(\mathbf{X}_{-u}^{(n)})} := N_I^{-1}\sum_{k=1}^{N_I} f(\mathbf{X}_{-u}^{(n)},\mathbf{X}_u^{(n,k)})$, $(\mathbf{X}_{-u}^{(n)})_{n\in[1:N_u]}$ is an i.i.d. sample with the distribution of $\mathbf{X}_{-u}$ and $(\mathbf{X}_u^{(n,k)})_{k\in[1:N_I]}$ conditionally to $\mathbf{X}_{-u}^{(n)}$ is i.i.d. with the distribution of $\mathbf{X}_u$ conditionally to $\mathbf{X}_{-u} = \mathbf{X}_{-u}^{(n)}$. For all $n \in [1 : N_u]$, the computation of

$$\frac{1}{N_I-1}\sum_{k=1}^{N_I}\left(f(\mathbf{X}_{-u}^{(n)},\mathbf{X}_u^{(n,k)}) - \overline{f(\mathbf{X}_{-u}^{(n)})}\right)^2$$

requires the values of $\left(f(\mathbf{X}_{-u}^{(n)},\mathbf{X}_u^{(n,k)})\right)_{k\in[1:N_I]}$. We will take $N_I = 3$, as suggested in [SNS16]. Thus, the double Monte-Carlo estimator given in Equation (6) has a cost (number of evaluations of f) of $3N_u$.

We remark that for $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in \mathcal{X}$ and for $\emptyset \subsetneq u \subsetneq [1 : p]$, we let $(\mathbf{x}_u^{(1)}, \mathbf{x}_{-u}^{(2)})$ be the element $\mathbf{v} \in \mathcal{X}$ such that $\mathbf{v}_u = \mathbf{x}_u^{(1)}$ and $\mathbf{v}_{-u} = \mathbf{x}_{-u}^{(2)}$, and we let $f(\mathbf{x}_u^{(1)}, \mathbf{x}_{-u}^{(2)}) := f(\mathbf{v})$. We use this notation throughout the paper.

Remark 3. *The estimator of Equation* (6) *is an unbiased estimator of* $E_u = \mathrm{E}(\mathrm{Var}(Y|\mathbf{X}_{-u}))$.

3.2 Pick-and-freeze

We now provide a second estimator of W_u: the Pick-and-Freeze estimator for V_u. We have

$$V_u = \mathrm{Var}(\mathrm{E}(Y|\mathbf{X}_u)) = \mathrm{E}(\mathrm{E}(Y|\mathbf{X}_u)^2) - \mathrm{E}(Y)^2.$$

Remark that $\mathrm{E}(Y)$ is easy to estimate so we assume without loss of generality that we know the value of $\mathrm{E}(Y)$ (for the numerical applications, we will take the empirical mean). It remains to estimate $\mathrm{E}(\mathrm{E}(Y|\mathbf{X}_u)^2)$, which seems to be complicated. We prove the following proposition that enables to simplify the formulation of this quantity.

Proposition 2. *Let* $\mathbf{X} = (\mathbf{X}_u, \mathbf{X}_{-u})$ *and* $\mathbf{X}^u = (\mathbf{X}_u, \mathbf{X}'_{-u})$ *of distribution* $\mathbb{P}_{\mathbf{X}}$ *such that, a.s.* $\mathbb{P}_{(\mathbf{X}_{-u},\mathbf{X}'_{-u})|\mathbf{X}_u=\mathbf{x}_u} = \mathbb{P}_{\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{x}_u} \otimes \mathbb{P}_{\mathbf{X}_{-u}|\mathbf{X}_u=\mathbf{x}_u}$. *We have*

$$\mathrm{E}(\mathrm{E}(Y|\mathbf{X}_u)^2) = \mathrm{E}(f(\mathbf{X})f(\mathbf{X}^u)). \tag{7}$$

Remark that Proposition 2 enables to write a double expectation as one single expectation, that we estimate by a simple Monte-Carlo. Thus, we suggest the Pick-and-Freeze estimator, for $\emptyset \subsetneq u \subsetneq [1 : p]$,

$$\widehat{V}_{u,PF} := \frac{1}{N_u}\sum_{n=1}^{N_u} f\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,1)}\right) f\left(\mathbf{X}_u^{(n)},\mathbf{X}_{-u}^{(n,2)}\right) - \mathrm{E}(Y)^2, \tag{8}$$

where $(\mathbf{X}_u^{(n)})_{n\in[1:N_u]}$ is an i.i.d. sample with the distribution of $\mathbf{X}_u$ and where $\mathbf{X}_{-u}^{(n,1)}$ and $\mathbf{X}_{-u}^{(n,2)}$ conditionally to $\mathbf{X}_u^{(n)}$ are independent with the distribution of $\mathbf{X}_{-u}$ conditionally to $\mathbf{X}_u = \mathbf{X}_u^{(n)}$. This estimator has a cost of $2N_u$.

4 W-aggregation procedures

As we can see in Equation (4) or in Equation (5), the Shapley effects are functions of the conditional elements $(W_u)_{u\subset[1:p]}$. In Section 3, we have seen how to estimate these conditional elements when it is possible to sample from the conditional distributions of the input vector. In this section, we assume that we have estimators $(\widehat{W}_u)_{u\subset[1:p]}$. From Remark 2, we let $\widehat{W}_\emptyset = W_\emptyset = 0$ and $\widehat{W}_{[1:p]} = W_{[1:p]} = \mathrm{Var}(Y)$. We also add the following assumption that will be needed for the theoretical results that we will prove.

Assumption 1. *For all $\emptyset \subsetneq u \subsetneq [1:p]$, $\widehat{W}_u$ is computed with a cost κN_u by $\widehat{W}_u = \frac{1}{N_u}\sum_{n=1}^{N_u} \widehat{W}_u^{(n)}$ where the $(\widehat{W}_u^{(n)})_{n\in[1:N_u]}$ are independent and identically distributed. The $(\widehat{W}_u)_{u\subset[1:p]}$ are independent. The integer $\kappa \in \mathbb{N}^*$ is the number of evaluations of the computer code f (i.e. the cost) for each $\widehat{W}_u^{(n)}$.*

Assumption 1 means that we estimate the $(W_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$ by Monte-Carlo, independently and with different costs $(\kappa N_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$. The accuracy N_u corresponds to computing N_u independent and identically distributed estimators $\widehat{W}_u^{(1)}, ..., \widehat{W}_u^{(N_u)}$ that are averaged. We have seen in Section 3 two estimators that satisfy Assumption 1: the double Monte-Carlo estimator (with $\kappa = 3$) and the Pick-and-Freeze estimator (with $\kappa = 2$).

We call "W-aggregation procedure" an algorithm that estimates the Shapley effects from the estimates $(\widehat{W}_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$ and that selects the values of the accuracies $(N_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$. We first suggest a new W-aggregation procedure. Then we obtain a theoretical insight on the W-aggregation procedure of [SNS16].

4.1 The subset procedure

In this section, we suggest a new W-aggregation procedure for the Shapley effects. This procedure consists in computing once for all the estimates $\widehat{W}_u$ for all $u \subset [1:p]$, and to store them. Then, we use these estimates to estimate all the Shapley effects.

4.1.1 The W-aggregation procedure

We suggest to estimate the Shapley effects $(\eta_i)_{i\in[1:p]}$ by using the following W-aggregation procedure:

1. For all $u \subset [1:p]$, compute $\widehat{W}_u$.

2. For all $i \in [1:p]$, estimate η_i by

$$\widehat{\eta}_i := \frac{1}{p\mathrm{Var}(Y)} \sum_{u \subset -i} \binom{p-1}{|u|}^{-1} (\widehat{W}_{u\cup\{i\}} - \widehat{W}_u). \tag{9}$$

We call this W-aggregation procedure "subset W-aggregation procedure". We can note that each estimate $\widehat{W}_u$ is used for all the estimates $(\widehat{\eta}_i)_{i\in[1:p]}$. It remains to choose the values of the accuracies $(N_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$.

4.1.2 Choice of the accuracy of each $\widehat{W}_u$

In this section, we explain how to choose the values of the accuracies $(N_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$. In the following proposition, we give the best choice of the accuracies $(N_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$ to minimize $\sum_{i=1}^p \mathrm{Var}(\widehat{\eta}_i)$ for a fixed total cost $\kappa\sum_{\emptyset\subsetneq u\subsetneq[1:p]} N_u$.

Proposition 3. *Let a total cost $N_{tot} \in \mathbb{N}$ be fixed. Under Assumption 1, if the Shapley effects are estimated with the subset W-aggregation procedure, the solution of the relaxed program (i.e. the problem without the constraint of letting the $(N_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$ be integers)*

$$\min_{(N_u)_{\emptyset\subsetneq u\subsetneq[1:p]}\in(0,+\infty)^{2^p-2}} \sum_{i=1}^p \mathrm{Var}(\widehat{\eta}_i) \qquad \textit{subject to} \quad \kappa \sum_{\emptyset\subsetneq u\subsetneq[1:p]} N_u = N_{tot} \tag{10}$$

is $(N_u^*)_{\emptyset\subsetneq u\subsetneq[1:p]}$ with for all $\emptyset \subsetneq u \subsetneq [1:p]$

$$N_u^* = \frac{N_{tot}}{\kappa} \frac{\sqrt{(p-|u|)!|u|!(p-|u|-1)!(|u|-1)!\mathrm{Var}(\widehat{W}_u^{(1)})}}{\sum_{\emptyset\subsetneq v\subsetneq[1:p]} \sqrt{(p-|v|)!|v|!(p-|v|-1)!(|v|-1)!\mathrm{Var}(\widehat{W}_v^{(1)})}}.$$

Usually, we do not know the values of $\mathrm{Var}(\widehat{W}_u^{(1)})$ for $\emptyset \subsetneq u \subsetneq [1:p]$, but we need them to compute the value of N_u^*. In practice, we will assume that these values are equal in order to compute N_u^*. Furthermore, the sum over the subsets v such that $\emptyset \subsetneq v \subsetneq [1:p]$ can be too costly to compute. Hence, we make the following approximations in practice:

$$N_u^* \approx \frac{\frac{N_{tot}}{\kappa}\binom{p}{|u|}^{-\frac{1}{2}}\binom{p}{|u|-1}^{-\frac{1}{2}}}{\sum_{\emptyset\subsetneq v\subsetneq[1:p]}\binom{p}{|v|}^{-\frac{1}{2}}\binom{p}{|v|-1}^{-\frac{1}{2}}} \approx \frac{\frac{N_{tot}}{\kappa}\binom{p}{|u|}^{-1}}{\sum_{\emptyset\subsetneq v\subsetneq[1:p]}\binom{p}{|v|}^{-1}} = \frac{N_{tot}}{\kappa}\frac{\binom{p}{|u|}^{-1}}{p-1}. \tag{11}$$

Hence, when implementing the subset W-aggregation procedure, we will choose N_u^* as

$$N_u^* := \mathrm{Round}\left(N_{tot}\kappa^{-1}\binom{p}{|u|}^{-1}(p-1)^{-1}\right) \tag{12}$$

for $\emptyset \subsetneq u \subsetneq [1:p]$, where Round is the nearest integer function. In this way, for a fixed total cost, we take the accuracies $(N_u)_{\emptyset \subsetneq u \subsetneq [1:p]}$ near the optimal choice that minimizes $\sum_{i=1}^{p} \mathrm{Var}(\widehat{\eta}_i)$. Hence, the parameter N_{tot} is now the only parameter left to choose. In practice, this parameter is often imposed as a global budget constraint.

Remark 4. *With the approximation discussed above, the real total cost* $\kappa \sum_{\emptyset \subsetneq u \subsetneq [1:p]} N_u$ *can be different from the* N_{tot} *chosen (because of the approximations and the choice of the closest integer). In this case, we suggest to adapt the value of* N_{tot} *in order to make the total cost* $\kappa \sum_{\emptyset \subsetneq u \subsetneq [1:p]} N_u^*$ *take the desired value.*

Remark 5. *In order to compute the* $(N_u^*)_{\emptyset \subsetneq u \subsetneq [1:p]}$ *in practice, we assume that the values of* $\mathrm{Var}(\widehat{W}_u^{(1)})$*, for* $\emptyset \subsetneq u \subsetneq [1:p]$*, are equal. We can see on unreported numerical experiments that this choice of* N_u *gives much better results than if we choose the same value of* N_u *for all* $\emptyset \subsetneq u \subsetneq [1:p]$*. However, it seems difficult to obtain theoretical results on the values of* $\mathrm{Var}(\widehat{W}_u^{(1)})$*, as they depend on the conditional distributions of* $\mathbf{X}$ *in a complicated way.*

Hence, this assumption is more a convenient heuristic to compute the best accuracies $(N_u^*)_{\emptyset \subsetneq u \subsetneq [1:p]}$ *than a real property satisfied in many cases. Proposition 3 and the heuristic in Equation* (11) *justify the choice of* $(N_u^*)_{\emptyset \subsetneq u \subsetneq [1:p]}$ *given in Equation* (12)*, and we make this choice even if the assumption of equal values of the* $(\mathrm{Var}(\widehat{W}_u^{(1)}))_{\emptyset \subsetneq u \subsetneq [1:p]}$ *is not satisfied.*

4.1.3 Consistency

A straightforward consequence of the subset W-aggregation procedure and Equation (9) is that the consistency of $(\widehat{W}_u)_{u \subset [1:p]}$ implies the consistency of $(\widehat{\eta}_i)_{i \in [1:p]}$ (Assumption 1 is not necessary).

Proposition 4. *Assume that for all* $\emptyset \subsetneq u \subsetneq [1:p]$*, we have estimators* $\widehat{W}_u$ *that converge to* W_u *in probability (resp. almost surely) when* N_u *goes to* $+\infty$*, where* κN_u *is the cost of* $\widehat{W}_u$*. If we use the subset* W*-aggregation procedure with the choice of* $(N_u^*)_{\emptyset \subsetneq u \subsetneq [1:p]}$ *given by Equation* (12)*, the estimators of the Shapley effects converge to the Shapley effects in probability (resp. almost surely) when* N_{tot} *goes to* $+\infty$ *(where* N_{tot} *is the total cost of the subset* W*-aggregation procedure).*

4.2 The random-permutation procedure

In this section, we present and study the "random-permutation W-aggregation procedure" suggested in [SNS16].

4.2.1 The W-aggregation procedure

The W-aggregation procedure of the algorithm of [SNS16] is based on Equation (5). Because of the equation, one could estimate η_i by

$$\widehat{\eta}_i = \frac{1}{p!\mathrm{Var}(Y)} \sum_{\sigma \in \mathcal{S}_p} \left(\widehat{W}_{P_i(\sigma)\cup\{i\}} - \widehat{W}_{P_i(\sigma)} \right), \tag{13}$$

for $i \in [1:p]$. In Equation (13), informally, $(\widehat{W}_u)_{\emptyset \subsetneq u \subsetneq [1:p]}$ are estimators. However, as the number of permutations is $p!$, there are too many summands and [SNS16] suggests to replace the sum over all the $p!$ permutations by the sum over M ($M < p!$) random uniformly distributed permutations. Thus, for a fixed $i \in [1:p]$, the estimator of η_i suggested in [SNS16] is

$$\widehat{\eta}_i = \frac{1}{M\mathrm{Var}(Y)} \sum_{m=1}^{M} \left(\widehat{W}_{P_i(\sigma_m)\cup\{i\}}(m) - \widehat{W}_{P_i(\sigma_m)}(m) \right), \tag{14}$$

where $(\sigma_m)_{m\in[1:M]}$ are independent and uniformly distributed on $\mathcal{S}_p$. If $m, m' \in [1:M]$ with $m \neq m'$ and $P_i(\sigma_m) = P_i(\sigma_{m'}) =: u$, [SNS16] estimates twice the same W_u. To formalize these different estimations, we write $\widehat{W}_u(m)$ the estimation of W_u at step m in Equation (14).

Finally, [SNS16] reduces the computation cost using the following idea. The authors of [SNS16] notice that for $1 \leq i < p$, for any permutation $\sigma \in \mathcal{S}_p$ and for $i \in [1:p]$, we have $P_{\sigma(i+1)}(\sigma) = P_{\sigma(i)}(\sigma) \cup \{\sigma(i)\}$. Thus, the algorithm of [SNS16] uses every estimate $\widehat{W}_{P_{\sigma_m(i)}(\sigma_m)\cup\{\sigma_m(i)\}}(m)$ for $\widehat{\eta}_{\sigma_m(i)}$ (as an estimator of $W_{P_{\sigma_m(i)}(\sigma_m)\cup\{\sigma_m(i)\}}$) and for $\widehat{\eta}_{\sigma_m(i+1)}$ (as an estimator of $W_{P_{\sigma_m(i+1)}(\sigma_m)}$). With this improvement, the number of estimations of W_u (for $\emptyset \subsetneq u \subsetneq [1:p]$) is divided by two when estimating all the Shapley effects $\eta_1, ..., \eta_p$. The W-aggregation procedure is then

1. Let $\widehat{\eta}_1 = ... = \widehat{\eta}_p = 0$.
2. For all $m = 1, 2, ..., M$
   (a) Generate σ_m uniformly distributed on $\mathcal{S}_p$.
   (b) Let $prevC = 0$.
   (c) For all $i = 1, 2, ..., p$
      i. Let $u = P_{\sigma_m(i)}(\sigma_m)$.
      ii. Compute $\widehat{W}_{u\cup\{\sigma_m(i)\}}(m)$.
      iii. Compute $\widehat{\Delta} = \widehat{W}_{u\cup\{\sigma_m(i)\}}(m) - prevC$.
      iv. Update $\widehat{\eta}_{\sigma_m(i)} = \widehat{\eta}_{\sigma_m(i)} + \widehat{\Delta}$.
      v. Set $prevC = \widehat{W}_{P_{\sigma_m(i+1)}(\sigma_m)}$.
3. Let $\widehat{\eta}_i = \widehat{\eta}_i/(\mathrm{Var}(Y)M)$ for all $i = 1, ..., p$.

We write this W-aggregation procedure "random-permutation W-aggregation procedure".

Remark 6. *Recall that in the subset W-aggregation procedure, each estimation of W_u was used for the estimation of all the $(\eta_i)_{i\in[1:p]}$ (and not only for two of them). Thus the subset W-aggregation procedure seems to be more efficient.*

Remark 7. *When the number of inputs p is small, [SNS16] suggests to take all the permutations of $[1:p]$ instead of choosing random permutations in Step 2a of the random-permutation W-aggregation procedure. However, this algorithm requires small values of p and the total cost is a multiple of $p!$ (so there are very restricted possible values). Furthermore, this method still remains very costly due to the computation of $(p-1)!$ conditional variances. For example, in the linear Gaussian framework with $p = 10$ (where the computation of the conditional elements is immediate, see [BBDM19]) it spends more than ten minutes computing the Shapley effects. Hence, the algorithm with all the permutations is not explicitly detailed in [SNS16].*

4.2.2 Choice of the accuracy of each $\widehat{W}_u$

As in Section 4.1.2, we suggest a choice of the accuracies $(N_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$.

In order to avoid a random total cost, we require for all $\emptyset \subsetneq u \subsetneq [1:p]$ that the accuracy N_u of the $\left(\widehat{W}_u(m)\right)_m$ depends only on $|u|$, and we write $N_{|u|} := N_u$. In this case, the total cost of the random-permutation W-aggregation procedure is equal to $N_{tot} = \kappa M \sum_{k=1}^{p-1} N_k$. Moreover, we assume that the total cost $N_{tot} = \kappa M \sum_{k=1}^{p-1} N_k$ is proportional to $(p-1)$, and thus can be written $N_{tot} = \kappa M N_O(p-1)$ for some fixed $N_O \in \mathbb{N}^*$. As the permutations $(\sigma_m)_{m\in[1:M]}$ are random, we choose to minimize $\mathrm{E}\left[\sum_{i=1}^p \mathrm{Var}\left(\widehat{\eta}_i|\,(\sigma_m)_{m\in[1:M]}\right)\right]$.

To compute the optimal values of $(N_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$, we introduce the following assumption.

Assumption 2. *For all $\emptyset \subsetneq u \subsetneq [1:p]$ and all $m \in [1:M]$, $\widehat{W}_u(m)$ is computed with a cost $\kappa N_{|u|}$ by $\widehat{W}_u(m) = \frac{1}{N_{|u|}} \sum_{n=1}^{N_{|u|}} \widehat{W}_u^{(n)}(m)$ where the $(\widehat{W}_u^{(n)}(m))_{n\in[1:N_u]}$ are independent and identically distributed. The $(\widehat{W}_u(m))_{\emptyset\subsetneq u\subsetneq[1:p],\, m\in[1:M]}$ are independent.*

When it is possible to sample from the conditional distributions of the input vector, we can generate i.i.d. double Monte-Carlo estimators $(\widehat{E}_{u,MC}(m))_{m\in[1:M]}$ or Pick-and-Freeze estimators $(\widehat{V}_{u,PF}(m))_{m\in[1:M]}$. Hence, they satisfy Assumption 2 by taking $N_u = N_{|u|}$ for all $\emptyset \subsetneq u \subsetneq [1:p]$.

Proposition 5. *Assume that we estimate the Shapley effects with the random-permutation W-aggregation procedure under Assumption 1 and that the variances $(\mathrm{Var}(\widehat{W}_u^{(1)}(1)))_{\emptyset\subsetneq u\subsetneq[1:p]}$ are equal. Then, the solution of the problem*

$$\min_{(N_k)_{k\in[1:p-1]}\in(0,+\infty)^{p-1}} \mathrm{E}\left[\sum_{i=1}^p \mathrm{Var}\left(\widehat{\eta}_i|\,(\sigma_m)_{m\in[1:M]}\right)\right] \quad \text{subject to} \quad \kappa M \sum_{k=1}^{p-1} N_k = \kappa M N_O(p-1)$$

is $(N_k^{**})_{k\in[1:p-1]}$ *with for all* $k \in [1 : p-1]$,

$$N_k^{**} = N_O.$$

Hence, from now on, with the random permutation W-aggregation procedure, we will choose the accuracy $N_u = N_O$ for all subset u.

Remark 8. *As in Remark 5, we assume in Proposition 5 that the variances* $(\mathrm{Var}(\widehat{W}_u^{(1)}(1)))_{\emptyset\subsetneq u\subsetneq[1:p]}$ *are equal. This assumption is not easy to check, but is required technically to prove Proposition 5.*

Remark 9. *With the exact-permutation* W*-aggregation procedure (see Remark 7),* $N_k^* = N_O p!$ *is the solution of the problem* $\sum_{i=1}^p \mathrm{Var}\,(\widehat{\eta}_i)$ *subject to* $\sum_{k=1}^{p-1} N_k = p! N_O(p-1)$.

There are now two parameters to choose: the number of permutations M and the accuracy N_O of the estimations of the $(W_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$. Typically, their product MN_O is imposed by budget constraints.

4.2.3 Choice of N_O

We have seen that for all $\emptyset \subsetneq u \subsetneq [1 : p]$, we choose $N_u = N_{|u|}^{**} = N_O$. In this section, we explain why we should choose $N_O = 1$ under Assumption 1 and M as large as possible.

Proposition 6 generalizes the result given in [SNS16], Appendix B. Its proof is given in the supplementary material, which is much simpler than the arguments in [SNS16].

Assumption 3. *Assumption 2 holds and for all* $\emptyset \subsetneq u \subsetneq [1 : p]$, *we have* $\mathrm{E}(\widehat{W}_u^{(1)}(1)) = W_u$.

Assumption 3 ensures that the estimators have a zero bias. Recall that the double Monte-Carlo estimator and the Pick-and-Freeze estimator have a zero bias. Hence, they satisfy Assumption 3 by generating i.i.d. $(\widehat{W}_u(m))_{m\in[1:M]}$ and by taking $N_u = N_{|u|}$ for all $\emptyset \subsetneq u \subsetneq [1 : p]$.

Proposition 6. *Let* $i \in [1 : p]$ *be fixed. Under Assumption 3, in order to minimize, over* N_O *and* M, *the variance of* $\widehat{\eta}_i$ *with a fixed cost* $\kappa M N_O \times (p-1) = \kappa C \times (p-1)$ *(for some* $C \in \mathbb{N}^*$*), we have to choose* $N_O = 1$ *and* $M = C$.

From now on, we assume that $N_u = N_O = 1$ when we use the random-permutation W-aggregation procedure and we will let M, the number of random permutations, go to infinity. Then, the total cost N_{tot} of the random-permutation W-aggregation procedure is equal to $N_{tot} = \kappa M(p-1)$, for estimating the p Shapley effects $\eta_1, \ldots, \eta_p$. Hence, under Assumption 2 or Assumption 3, $\widehat{W}_u(m) = \widehat{W}_u^{(1)}(m)$ and has now a cost κ.

4.2.4 Consistency

We give here two sufficient conditions for the consistency of the estimators of the Shapley effects given by the random-permutation W-aggregation procedure. We introduce a general assumption.

Assumption 4. *For all u such that $\emptyset \subsetneq u \subsetneq [1:p]$, $\left(\widehat{W}_u(m)\right)_{m\in[1:M]}$ have a cost κ (since we chose $N_u = 1$) and are identically distributed with a distribution that depends on an integer N such that $\mathrm{E}\left(\widehat{W}_u(1)\right) \underset{N\to+\infty}{\longrightarrow} W_u$. Moreover, for all u such that $\emptyset \subsetneq u \subsetneq [1:p]$, we have*

$$\frac{1}{M^2}\sum_{m,m'=1}^{M} cov\left(\widehat{W}_u(m), \widehat{W}_u(m')\right) \underset{N,M\to+\infty}{\longrightarrow} 0.$$

Assumption 4 is more general than Assumption 3. Indeed, it enables the estimators to have a bias and a covariance which go to zero. This assumption will be useful to prove the consistency results in Section 6.2. Remark that in Assumption 4, for all $\emptyset \subsetneq u \subsetneq [1:p]$, each estimate $\left(\widehat{W}_u(m)\right)_{m\in[1:M]}$ has a cost κ, as in Assumption 3 since we fixed $N_u = N_O = 1$.

Proposition 7. *Assume that we estimate the Shapley effects using the random-permutation W-aggregation procedure. Let $N_{tot} = \kappa M(p-1)$ be the total cost of the random-permutation W-aggregation procedure.*

1. *Under Assumption 3, the estimates of the Shapley effects converge to the Shapley effects in probability when N_{tot} goes to $+\infty$.*
2. *Under Assumption 4, the estimates of the Shapley effects converge to the Shapley effects in probability when N_{tot} and N go to $+\infty$.*

5 Estimators of the Shapley effects

5.1 Four consistent estimators of the Shapley effects

Recall that in Section 3, we have seen two estimators of the $(W_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$: double Monte-Carlo (used in the algorithm of [SNS16]) and Pick-and-Freeze. In Section 4, we have studied two W-aggregation procedures for the Shapley effects using estimators of the $(W_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$: the subset W-aggregation procedure and the random-permutation W-aggregation procedure (used in the algorithm of [SNS16]). To sum up, four estimators of the Shapley effects are available:

- subset W-aggregation procedure with double Monte-Carlo;
- subset W-aggregation procedure with Pick-and-Freeze;
- random-permutation W-aggregation procedure with double Monte-Carlo, which is the already existing algorithm of [SNS16];

- random-permutation W-aggregation procedure with Pick-and-Freeze.

With the random-permutation W-aggregation procedure, we have seen that we need different estimators $(\widehat{W}_u(m))_{m\in[1:M]}$ of the same W_u. In this case, we choose i.i.d. realizations of the estimator of W_u. Moreover, we have seen in Section 4.2.3 that when we use the random-permutation W-aggregation procedure, we choose $N_u = N_O = 1$.

By Propositions 4 and 7, all these four estimators are consistent when the global budget N_{tot} goes to $+\infty$. Indeed, by Proposition 4, the consistency of the $(\widehat{W}_u)_{\emptyset\subsetneq u\subsetneq[1:p]}$ is sufficient for the consistency with the subset procedure and by Proposition 7, unbiased and i.i.d. estimators $(\widehat{W}_u(m))_{m\in[1:M]}$ for all $\emptyset \subsetneq u \subsetneq [1:p]$ provide the consistency with the random-permutation procedure.

5.2 Numerical comparison of the different algorithms

In this section, we carry out numerical experiments on the different algorithms in the restricted framework (where the exact conditional samples are available).

To compare these estimators, we use the linear Gaussian framework: $\mathcal{X} = \mathbb{R}^p$, $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ and $Y = \sum_{i=1}^p \beta_i X_i$. In this case, the theoretical values are easily computable (see [OP17, IP17, BBDM19]). We choose $p = 10$, $\beta_i = 1$ for all $i \in [1:p]$ and $\boldsymbol{\Gamma} = \mathbf{A}^T\mathbf{A}$ where $\mathbf{A}$ is a $p \times p$ matrix which components are realisations of p^2 i.i.d. Gaussian variables with zero mean and unit variance. To compare these different estimators, we fix a total cost (number of evaluations of f) of $N_{tot} = 54000$. We compute 1000 realizations of each estimator.

In Figure 1, we plot the theoretical values of the Shapley effects together with the boxplots of the 1000 realizations of each estimator.

In Figure 2, we plot the sum over $i \in [1:p]$ of the quadratic risks: $\sum_{i=1}^p \mathrm{E}\left((\widehat{\eta}_i - \eta_i)^2\right)$ (estimated with 1000 realizations) of each estimator.

We can see that the subset W-aggregation procedure gives better results than the random-permutation W-aggregation procedure, and that the double Monte-Carlo estimator is better than the Pick-and-Freeze estimator.

Remark 10. *It appears that double Monte-Carlo is numerically more efficient than Pick-and-Freeze for estimating the Shapley effects. Indeed, if we focus only on the estimation of one W_u for a fixed $\emptyset \subsetneq u \subsetneq [1:p]$, we can see numerically that the Pick-and-Freeze estimator has a larger variance than the double Monte-Carlo estimator. This finding appears to be difficult to confirm theoretically in the general case. Nevertheless, we can obtain such a theoretical confirmation in a simple, specific example. Let $\mathbf{X} \sim \mathcal{N}(0, \mathbf{I}_2)$, $Y = X_1 + X_2$. Remark that, in this example, the variances of $\widehat{W}_u^{(1)}$, $\emptyset \subsetneq u \subsetneq [1:p]$, are equal. In this case, and for $u = \{1\}$, we can easily get $\mathrm{Var}(\widehat{V}_{u,PF}) = \frac{40}{9}\mathrm{Var}(\widehat{E}_{u,MC})$ for the same cost (number of evaluations of f), and choosing $N_I = 3$ for the double Monte-Carlo estimator. This could be surprising since [JKLR$^+$14] proved that some Pick-and-Freeze estimator is asymptotically efficient in the independent case. However, this result and our finding are not contradictory for two reasons: the authors of [JKLR$^+$14] estimate the variance of Y so their result does not apply*

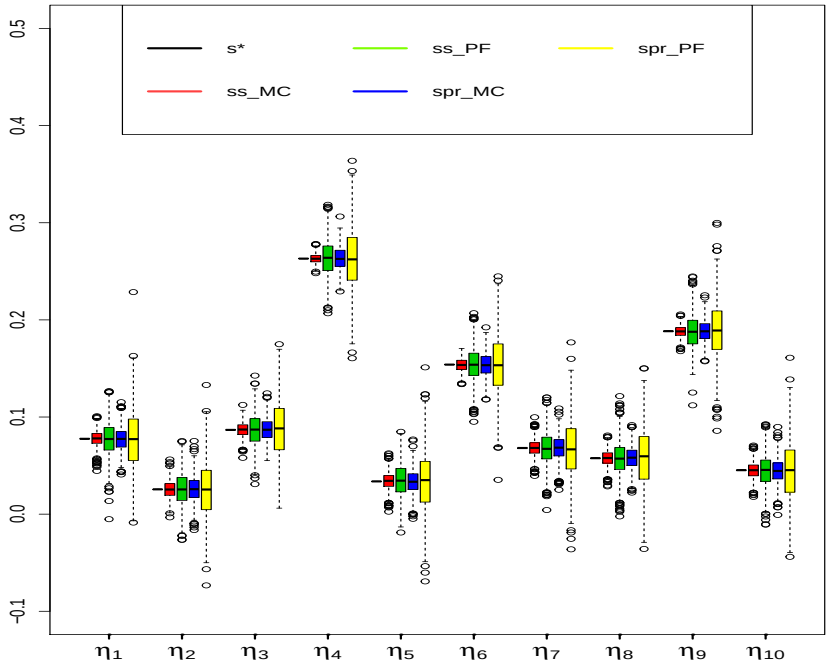

Figure 1: Estimation of the Shapley effects in the linear Gaussian framework. In black (s*) we show the theoretical values, in red (ss_MC) the estimates from the subset W-aggregation procedure with the double Monte-Carlo estimator, in green (ss_PF) the estimates from the subset W-aggregation procedure with the Pick-and-Freeze estimator, in blue (spr_MC) the estimates from the random-permutation W-aggregation procedure with the double Monte-Carlo estimator and in yellow (spr_PF) the estimates from the random-permutation W-aggregation procedure with the Pick-and-Freeze estimator.

here and the double Monte-Carlo estimator is based on different observations from the Pick-and-Freeze estimator.

To conclude, we improved the already existing algorithm of [SNS16] (random-permutation W-aggregation procedure with double Monte-Carlo) by the estimator given by the subset W-aggregation procedure with double Monte-Carlo.

6 Extension when we observe an i.i.d. sample

In Section 5, we have considered a restricted framework: we assumed that for all $\emptyset \subsetneq u \subsetneq [1:p]$ and all $\mathbf{x}_u \in \mathcal{X}_u$, we could generate an i.i.d. sample from

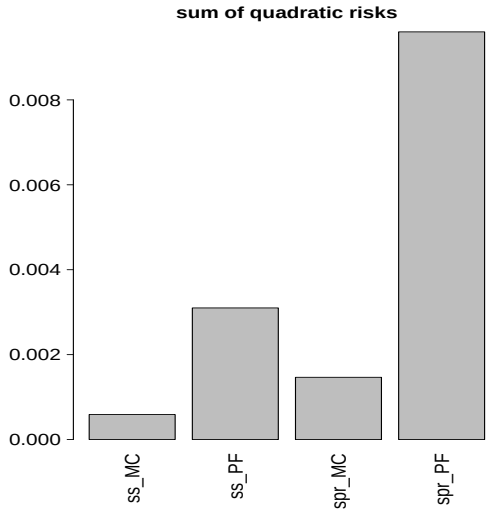

Figure 2: Sum over $i \in [1:p]$ of the estimated quadratic risks of the four estimators of the Shapley effects in the linear Gaussian framework.

the distribution of $\mathbf{X}_{-u}$ conditionally to $\mathbf{X}_u = \mathbf{x}_u$. However, in many cases, we can not generate this sample, as we only observe an i.i.d. sample of $\mathbf{X}$. In this section, we assume that we only observe an i.i.d. sample $(\mathbf{X}^{(n)})_{n\in[1:N]}$ of $\mathbf{X}$ and that we have access to the computer code f. We extend the double Monte-Carlo and Pick-and-Freeze estimators in this general case and show their consistency and rates of convergence. We then give the consistency of the implied estimators of the Shapley effects (obtained from the W-aggregation procedures studied previously). To the best of our knowledge, these suggested estimators are the first estimators of Shapley effects in this general framework. We conclude giving numerical experiments.

We choose a very general framework to prove the consistency of the estimators. This framework is given in the following assumption.

Assumption 5. *For all $i \in [1:p]$, $(\mathcal{X}_i, d_i)$ is a Polish space with metric d_i and $\mathbf{X} = (X_1, ..., X_p)$ has a density $f_{\mathbf{X}}$ with respect to a finite measure $\mu = \bigotimes_{i=1}^p \mu_i$ which is bounded and $\mathbb{P}_X$-almost everywhere continuous.*

This assumption is really general. Actually, it enables to have some continuous variables (with the Euclidean distance), some categorical variables in countable ordered or unordered sets and some variables in separable Hilbert spaces (for example $L^2(\mathbb{R}^d)$, for some $d \in \mathbb{N}^*$). The fact that $\mathbf{X}$ has a continuous density $f_{\mathbf{X}}$ with respect to a finite measure $\mu = \bigotimes \mu_i$ means that the distribution of $\mathbf{X}$ is smooth. Assumption 5 is satisfied in many realistic cases. The assumption of a bounded density which is $\mathbb{P}_{\mathbf{X}}$-almost everywhere continuous may be less realistic in some cases but is needed in the proofs. It would be interesting to alleviate it in future work.

To prove rates of convergence, we will need the following stronger assumption.

Assumption 6. *The function f is $\mathcal{C}^1$, $\mathcal{X}$ is compact in $\mathbb{R}^p$, $\mathbf{X}$ has a density $f_{\mathbf{X}}$ with respect to the Lebesgue measure λ_p on $\mathcal{X}$ such that λ_p-a.s. on $\mathcal{X}$, we have $0 < C_{\inf} \leq f_{\mathbf{X}} \leq C_{\sup} < +\infty$. Furthermore, $f_{\mathbf{X}}$ is Lipschitz continuous on $\mathcal{X}$.*

Assumption 6 is more restrictive than Assumption 5. It requires all the input variables to be continuous and real-valued. Moreover, their values are restricted to a compact set where the density is lower-bounded. Assumption 6 will be satisfied in some realistic cases (for instance with uniform or truncated Gaussian input random variables). Nevertheless, there also exist realistic cases where the input density is not lower-bounded (for instance with triangular input random variables). We remark that the assumption of a lower-bounded density is common in the field of non-parametric statistics [Gho01]. Here, it enables us to control the order of magnitude of conditional densities.

6.1 Estimators of the conditional elements

As far as we know, only [VG13] suggests a consistent estimator of W_u when we only observe an i.i.d. sample and when the input variables can be dependent, but only for V_u with $|u| = 1$. The estimator suggested in [VG13] is asymptotically efficient but the fact that u has to be a singleton prevents us to use this estimator for the Shapley effects (because we have to estimate W_u for all $\emptyset \subsetneq u \subsetneq [1:p]$). We can find another estimator of the $(V_u)_{u \subset [1:p]}$ in [Pli10] (but no theoretical results on the convergence are given). Finally, note that [PBS13] provides an estimator of different sensitivity indices, with convergence proofs.

In this section we introduce two consistent estimators of $(W_u)_{\emptyset \subsetneq u \subsetneq [1:p]}$ when we observe only an i.i.d. sample of $\mathbf{X}$, and which are easy to implement. These two estimators follow the principle of the double Monte-Carlo and Pick-and-Freeze estimators, but replacing exact samples from the conditional distributions by approximate ones based on nearest-neighbours methods.

To that end, we have to introduce notation. Let $N \in \mathbb{N}$ and $(\mathbf{X}^{(n)})_{n \in [1:N]}$ be an i.i.d. sample of $\mathbf{X}$. If $\emptyset \subsetneq v \subsetneq [1:p]$, let us write $k_N^v(l,n)$ for the index such that $\mathbf{X}_v^{(k_N^v(l,n))}$ is the (or one of the) n-th closest element to $\mathbf{X}_v^{(l)}$ in $(\mathbf{X}_v^{(i)})_{i \in [1:N]}$, and such that $(k_N^v(l,n))_{n \in [1:N_I]}$ are two by two distinct.

The index $k_N^v(l,n)$ could be not uniquely defined if there exist different observations $\mathbf{X}_v^{(i)}$ at equal distance from $\mathbf{X}_v^{(l)}$. In this case, we will choose $k_N^v(l,n)$ uniformly over the indices of these observations, with the following independence assumption.

Assumption 7. *Conditionally to $(\mathbf{X}_v^{(n)})_{n \in [1:N]}$, $k_N^v(l,i)$ is randomly and uniformly chosen over the indices of all the i-th nearest neighbours of $\mathbf{X}_v^{(l)}$ in $(\mathbf{X}_v^{(n)})_{n \in [1:N]}$ and the $(k_N^v(l,i))_{i[1:N_I]}$ are two by two distinct. Furthermore, conditionally to $(\mathbf{X}_v(n))_{n \in [1:N]}$, for all $l \in [1:N]$, the random vector $(k_N(l,i))_{i \in [1:N_I]}$ is independent on all the other random variables.*

To summarize the idea of Assumption 7, we can say that the nearest neighbours of $\mathbf{X}_v^{(l)}$ are chosen uniformly among the possible choices and independently on the other variables. Assumption 7 actually only formalizes the random choice of the nearest neighbours where there can be equalities of the distances and this choice is easy to implement in practice.

When $\mathbf{X}_v$ is absolutely continuous with respect to the Lebesgue measure, distance equalities can not happen and $k_N^v(l,n)$ is uniquely defined. Thus, Assumption 7 trivially holds in this case. Assumption 7 is thus specific to the case where some input variables are not continuous.

6.1.1 Double Monte-Carlo

We write $(s(l))_{l\in[1:N_u]}$ a sample of uniformly distributed integers in $[1:N]$ (with or without replacement) independent of the other random variables. Then, we define two slightly different versions of the double Monte-Carlo estimator by

$$\widehat{E}_{u,MC}^{mix} = \frac{1}{N_u}\sum_{l=1}^{N_u} \widehat{E}_{u,s(l),MC}^{mix}, \tag{15}$$

and

$$\widehat{E}_{u,MC}^{knn} = \frac{1}{N_u}\sum_{l=1}^{N_u} \widehat{E}_{u,s(l),MC}^{knn}, \tag{16}$$

with

$$\widehat{E}_{u,s(l),MC}^{mix} = \frac{1}{N_I-1}\sum_{i=1}^{N_I}\left[f\left(\mathbf{X}_{-u}^{(s(l))}, \mathbf{X}_u^{(k_N^{-u}(s(l),i))}\right) - \frac{1}{N_I}\sum_{h=1}^{N_I} f\left(\mathbf{X}_{-u}^{(s(l))}, \mathbf{X}_u^{(k_N^{-u}(s(l),h))}\right)\right]^2 \tag{17}$$

and

$$\widehat{E}_{u,s(l),MC}^{knn} = \frac{1}{N_I-1}\sum_{i=1}^{N_I}\left[f\left(\mathbf{X}^{(k_N^{-u}(s(l),i))}\right) - \frac{1}{N_I}\sum_{h=1}^{N_I} f\left(\mathbf{X}^{(k_N^{-u}(s(l),h))}\right)\right]^2. \tag{18}$$

The double Monte-Carlo estimator has two sums: one of size N_I for the conditional variance, one other of size N_u for the expectation. The integer N_I is also the number of nearest neighbours and it is a fixed parameter to choose. For example, we can choose $N_I = 3$ (as in the case where the conditional samples are available).

Remark 11. *If we observe the sample $(\mathbf{X}^{(n)})_{n\in[1:N]}$ and if the values of $(f(\mathbf{X}^{(n)}))_{n\in[1:N]}$ have to be assessed, the cost of the estimators $\widehat{E}_{u,MC}^{mix}$ and $\widehat{E}_{u,MC}^{knn}$ remains the number of evaluations of f (which is $N_I N_u$). If we observe the sample $(\mathbf{X}^{(n)}, f(\mathbf{X}^{(n)}))_{n\in[1:N]}$, the estimator $\widehat{E}_{u,MC}^{knn}$ does not require evaluations of f but the cost remains proportional to N_u (for the search of the nearest neighbours and for the elementary operations).*

Remark 12. *The integer N is the size of the sample of $\mathbf{X}$ (that enables us to estimate implicitly its conditional distributions through the nearest neighbours) and the integer N_u is the accuracy of the estimator $\widehat{E}_{u,MC}$ from the estimated distribution of $\mathbf{X}$. Of course, it would be intuitive to take $N_u = N$ and $(s(l))_{l\in[1:N]} = (l)_{l\in[1:N]}$, but this framework would not be general enough for the subset W-aggregation procedure (in which the accuracy N_u of $\widehat{E}_{u,MC}$ depends on u) and for the proof of the consistency when using the random-permutation W-aggregation procedure in Section 6.2. Furthermore, we may typically have to take N_u smaller than N.*

Remark that we give two versions of the double Monte-Carlo estimator. The "mix" version seems more accurate but requires to call the computer code of f at new inputs. For the "knn" version, it is sufficient to have an i.i.d. sample $(\mathbf{X}^{(n)}, f(\mathbf{X}^{(n)}))_{n\in[1:N]}$.

Now that we defined these two versions of the double Monte-Carlo estimator for an unknown input distribution, we give the consistency of these estimators in Theorem 1. We let $\widehat{E}_{u,MC}$ be given by Equation (15) or Equation (16). In the asymptotic results below, N_I is fixed and N and N_u go to infinity.

Theorem 1. *Assume that Assumption 5 holds and Assumption 7 holds for $v = -u$. If f is bounded, then $\widehat{E}_{u,MC}$ converges to E_u in probability when N and N_u go to $+\infty$.*

Furthermore, with additional regularity assumptions, we can give the rate of convergence of these estimators in Theorem 2 and Corollary1.

Theorem 2. *Under Assumption 6, for all $\varepsilon > 0$, $\varepsilon' > 0$, there exist fixed constants $C^{(1)}_{\sup}(\varepsilon')$ and $C^{(2)}_{\sup}$ such that*

$$\mathbb{P}\left(\left|\widehat{E}_{u,MC} - E_u\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\left(\frac{C^{(1)}_{\sup}(\varepsilon')}{N^{\frac{1}{p-|u|}-\varepsilon'}} + \frac{C^{(2)}_{\sup}}{N_u}\right). \tag{19}$$

Corollary 1. *Under Assumption 6, choosing $N_u \geq CN^{1/(p-|u|)}$ for some fixed $0 < C < +\infty$, we have for all $\delta > 0$,*

$$\left|\widehat{E}_{u,MC} - E_u\right| = o_p\left(\frac{1}{N^{\frac{1}{2(p-|u|)}-\delta}}\right).$$

We remark that for $|u| = p - 1$, we nearly obtain a parametric rate of convergence $N^{\frac{1}{2}}$. The rate of convergence decreases when $|u|$ decreases which can be interpreted by the fact that we estimate non-parametrically the function $\mathbf{x}_{-u} \mapsto \mathrm{Var}(f(\mathbf{X})|\mathbf{X}_{-u} = \mathbf{x}_{-u})$. The estimation problem is higher-dimensional when $|u|$ decreases.

6.1.2 Pick-and-Freeze

We now give similar results for the Pick-and-Freeze estimators. The number N_I of nearest neighbours that we need for the Pick-and-Freeze estimators is equal

to 2. Assume that $\mathrm{E}(Y)$ is known and let $(s(l))_{l\in[1:N_u]}$ be as in Section 6.1.1. Then, we define two slightly different versions of the Pick-and-Freeze estimator by

$$\widehat{V}_{u,PF}^{mix} = \frac{1}{N_u}\sum_{l=1}^{N_u} \widehat{V}_{u,s(l),PF}^{mix}, \tag{20}$$

and

$$\widehat{V}_{u,PF}^{knn} = \frac{1}{N_u}\sum_{l=1}^{N_u} \widehat{V}_{u,s(l),PF}^{knn}, \tag{21}$$

with

$$\widehat{V}_{u,s(l),PF}^{mix} = f\left((\mathbf{X}^{(k_N^u(s(l),1))}\right) f\left(\mathbf{X}_u^{(k_N^u(s(l),1))}, \mathbf{X}_{-u}^{(k_N^u(s(l),2))}\right) - \mathrm{E}(Y)^2 \tag{22}$$

and

$$\widehat{V}_{u,s(l),PF}^{knn} = f(\mathbf{X}^{(k_N^u(s(l),1))}) f(\mathbf{X}^{(k_N^u(s(l),2))}) - \mathrm{E}(Y)^2. \tag{23}$$

As for the double Monte-Carlo estimators, we give the consistency of the Pick-and-Freeze estimators in Theorem 3 and the rate of convergence in Theorem 4 and in Corollary2. We let $\widehat{V}_{u,PF}$ be given by Equation (20) or Equation (21).

Theorem 3. *Assume that Assumption 5 holds and Assumption 7 holds for $v = u$ and $N_I = 2$. If f is bounded, then $\widehat{V}_{u,PF}$ converges to V_u in probability when N and N_u go to $+\infty$.*

Theorem 4. *Under Assumption 6, if $|u| = 1$, for all $\varepsilon > 0$, $\varepsilon' > 0$,*

$$\mathbb{P}\left(\left|\widehat{V}_{u,PF} - V_u\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\left(\frac{C_{\sup}^{(1)}(\varepsilon')}{N^{1-\varepsilon'}} + \frac{C_{\sup}^{(2)}}{N_u}\right), \tag{24}$$

and if $|u| > 1$, for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\widehat{V}_{u,PF} - V_u\right| > \varepsilon\right) \leq \frac{C_{\sup}^{(3)}}{\varepsilon^2}\left(\frac{1}{N^{\frac{1}{|u|}}} + \frac{1}{N_u}\right), \tag{25}$$

with fixed constants $C_{\sup}^{(1)}(\varepsilon') < +\infty$, $C_{\sup}^{(2)} < +\infty$ and $C_{\sup}^{(3)} < +\infty$.

Corollary 2. *Under Assumption 6, choosing $N_u \geq CN^{1/|u|}$ for some fixed $0 < C < +\infty$, we have*

1. *for all u such that $|u| = 1$, for all $\delta > 0$,*

$$\left|\widehat{V}_{u,PF} - V_u\right| = o_p\left(\frac{1}{N^{\frac{1}{2}-\delta}}\right).$$

2. *for all u such that $|u| > 1$,*

$$\left|\widehat{V}_{u,PF} - V_u\right| = O_p\left(\frac{1}{N^{\frac{1}{2|u|}}}\right).$$

The interpretation of the rates of convergence is the same as for the double Monte-Carlo estimators.

6.2 Consistency of the Shapley effect estimators

Now that we have constructed estimators of W_u with an unknown input distribution, we can obtain estimators of the Shapley effects using the subset and random-permutation W-aggregation procedures. Note that for each W-aggregation procedure, we need to choose the accuracy N_u of the $(\widehat{W}_u)_{\emptyset \subsetneq u \subsetneq [1:p]}$. Although Assumption 1 does not hold with the estimators $\widehat{E}_{u,MC}$ and $\widehat{V}_{u,PF}$ (the summands of these estimators are not independent), we keep choosing $N_u = N_O = 1$ for the random-permutation W-aggregation procedure and N_u as the closest integer to $N_{tot}\kappa^{-1}\binom{p}{|u|}^{-1}(p-1)^{-1}$ with the subset W-aggregation procedure. To unify notation, let $N_I = 2$ when the estimators of the conditional elements $(W_u)_{\emptyset \subsetneq u \subsetneq [1:p]}$ are the Pick-and-freeze estimators (in this way, N_I is the number of nearest neighbours). With the double Monte-Carlo estimators, let N_I be a fixed integer (for example $N_I = 3$).

Finally, recall that for the random-permutation W-aggregation procedure, we need different estimators $(\widehat{W}_u(m))_{m\in[1:M]} = (\widehat{W}_u(m)^{(1)})_{m\in[1:M]}$ of W_u, with the notation of Assumption 2. In this case, we choose i.i.d. realizations of $\widehat{W}_u$ conditionally to $(\mathbf{X}^{(n)})_{n\in[1:N]}$. That is $(\widehat{W}_u(m))_{m\in[1:M]} = \left(\widehat{W}_{u,s(m)}\right)_{m\in[1:M]}$, where $\widehat{W}_{u,s(m)}$ is defined by either Equation (17), Equation (18), Equation (22) or Equation (23), and $(s(m))_{m\in[1:M]}$ are independent and uniformly distributed on $[1:N]$. This enables to have different estimators with a small covariance using the same sample $(\mathbf{X}^{(n)})_{n\in[1:N]}$. Indeed, to prove the consistency in Proposition 8 of the Shapley effects estimator with the random-permutation procedure, we show that Assumption 4 is satisfied.

Proposition 8. *Assume that Assumption*5 *holds and Assumption 7 holds for all subset u, $\emptyset \subsetneq u \subsetneq [1:p]$. If f is bounded, then the estimators of the Shapley effects defined by the random-permutation W-aggregation procedure or the subset W-aggregation procedure combined with $\widehat{W}_u = \widehat{E}_{u,MC}$ (resp. $\widehat{W}_u = \widehat{V}_{u,PF}$) converge to the Shapley effects in probability when N and N_{tot} go to $+\infty$.*

Remark 13. *The Sobol indices are functions of the $(W_u)_{u\subset[1:p]}$. Indeed, we can define the Sobol index of a group of variables $\mathbf{X}_u$ by either S_u as in [Cha13, BBDM19] or S_u^{cl} as in [IP17], where*

$$S_u := \frac{1}{\mathrm{Var}(Y)}\sum_{v\subset u}(-1)^{|u|-|v|}V_v, \quad S_u^{cl} := \frac{V_u}{\mathrm{Var}(Y)},$$

and where we note that $V_u = \mathrm{Var}(Y) - E_{-u}$ by the law of total variance. Thus, we get consistent estimators of the Sobol indices in the general setting of Assumption 5. Note that the sum over $u \subset [1:p]$ of the Sobol indices S_u^{cl} is not equal to 1, and when the inputs are dependent, the Sobol index S_u can take negatives values.

6.3 Numerical experiments

In this section, we compute numerically the estimators of the Shapley effects with an unknown input distribution. As in Section 5.2, we choose the linear Gaussian framework to compute the theoretical values of the Shapley effects.

We take the same parameters as in Section 5.2. The size N of the observed sample $(\mathbf{X}^{(n)})_{n\in[1:N]}$ is 10000. Each estimator is computed 200 times. We now have 8 consistent estimators given by:

- 2 different W-aggregation procedures: subset or random-permutation;
- 2 different estimators of W_u: double Monte-Carlo or Pick-and-Freeze;
- 2 slightly different versions of the estimators of W_u: "mix" or "knn".

In Figure 3, we plot the theoretical values of the Shapley effects, together with the boxplots of the 200 realizations of each estimator, and with a total cost $N_{tot} = 54000$ (we assume here that f is a costly computer code and that for all estimators, the cost is the number of evaluations of f).

Remark 14. *In the linear Gaussian framework, the function f is not bounded and the assumptions of Proposition 8 do not hold. We can thus not guarantee the consistency of the Shapley effects estimators. However, this framework enables to compute the theoretical Shapley effects and we can see numerically that the estimators seem to be consistent.*

We show the sums over $i \in [1 : p]$ of their quadratic risks (estimated with 200 realizations) in Figure 4. As in Section 5.2, the subset W-aggregation procedure is better than the random-permutation W-aggregation procedure and double Monte-Carlo is better than Pick-and-Freeze. Furthermore, there are no significant differences between the version "mix" and the version "knn". Recall that, in order to compute the estimators with the "mix" version, we need to call the computer code of f at new inputs whereas "knn" only needs an i.i.d. sample $(\mathbf{X}^{(n)}, f(\mathbf{X}^{(n)}))_{n\in[1:N]}$.

We now compare the sums over $i \in [1 : p]$ of the estimated quadratic risks of the estimators from the subset W-aggregation procedure with double Monte-Carlo when we know the distribution of $\mathbf{X}$ (results of Section 5.2) and when we just observe a sample of size 10000 (previous results of this section). These values are equal to $5.9\ 10^{-3}$ when we know the distribution of $\mathbf{X}$, to $6.6\ 10^{-3}$ when we only observe the sample with $\widehat{E}^{mix}_{u,MC}$ and to $7.4\ 10^{-3}$ when we only observe the sample with $\widehat{E}^{knn}_{u,MC}$. Thus, in dimension 10, replacing the knowledge of $\mathbf{X}$ by a sample of size 10000 does not seem to deteriorate significantly our estimates of the Shapley effects.

7 Application to real data

In this section, we apply the estimator of the Shapley effects given by the subset W-aggregation procedure and the double Monte-Carlo estimator $\widehat{E}^{knn}_{u,MC}$ in

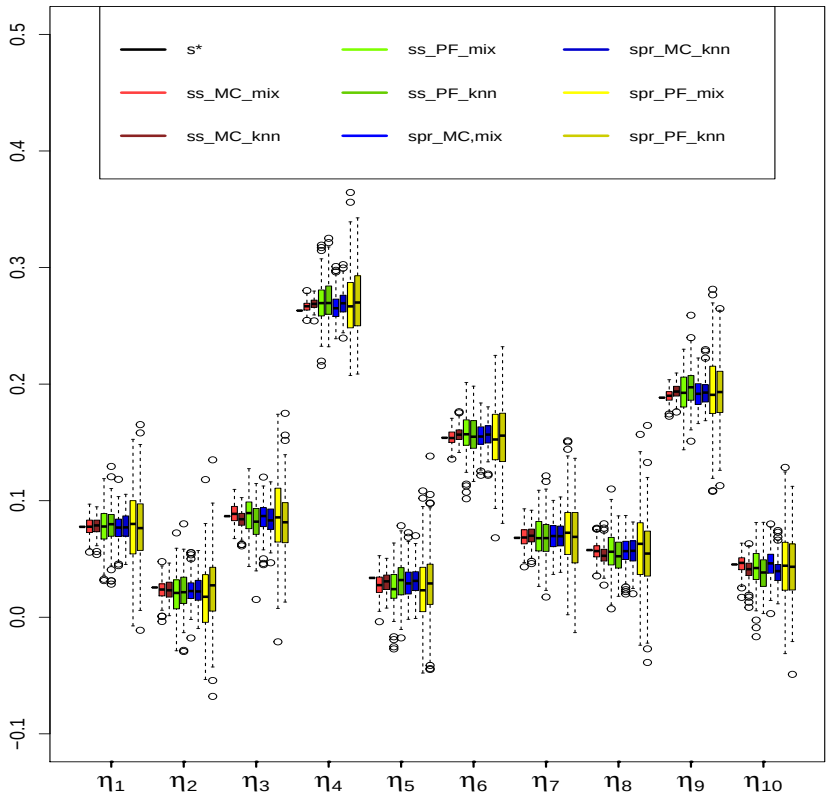

Figure 3: Estimation of the Shapley effects in the linear Gaussian framework when we only observe a sample of $\mathbf{X}$. In black (s*) we show the theoretical results, in red the estimates from the subset W-aggregation procedure with the double Monte-Carlo estimator (ss_MC_mix and ss_MC_knn), in green the estimates from the subset W-aggregation procedure with the Pick-and-Freeze estimator (ss_PF_mix and ss_PF_knn), in blue the estimates from the random-permutation W-aggregation procedure with the double Monte-Carlo estimator (spr_MC_mix and spr_MC_knn) and in yellow the estimates from the random-permutation W-aggregation procedure with the Pick-and-Freeze estimator (spr_PF_mix and spr_PF_knn).

Equation (16) to a real data set. We use the "depSeuil.dat" data, available at `http://www.math.univ-toulouse.fr/~besse/Wikistat/data` from [BMM+07]. This data set contains 10 variables with 1041 sample observations. The variables are:

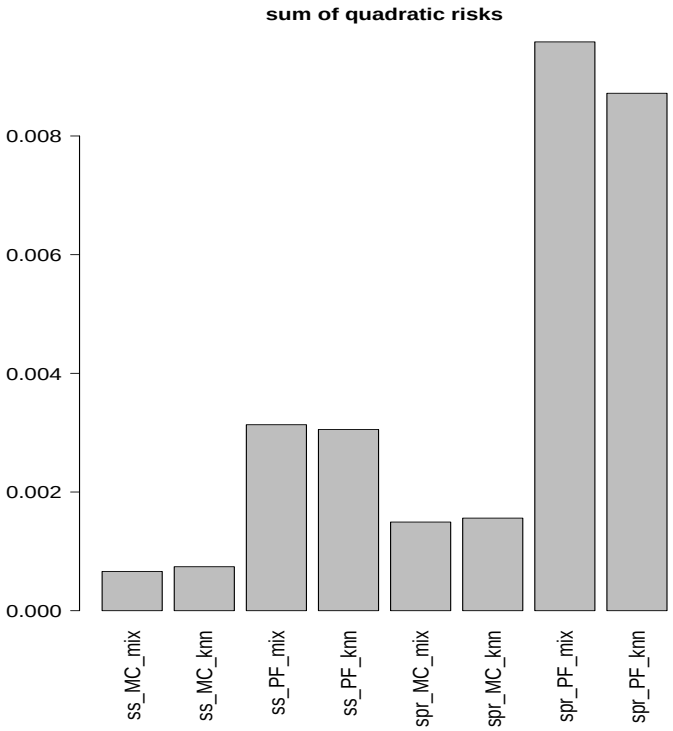

Figure 4: Sum over i of the estimated quadratic risks of the eight estimators of the Shapley effects in the linear Gaussian framework when we only observe a sample of $\mathbf{X}$.

- JOUR: type of day (holiday: 1, no holiday: 0);
- O3obs: observed ozone concentration;
- MOCAGE: ozone concentration predicted by a fluid mechanics model;
- TEMPE: temperature predicted by the official meteorology service of France;
- RMH2O: humidity ratio;
- NO2: nitrogen dioxide concentration;
- NO: nitrogen oxide concentration;
- STATION: site of observation (5 different sites);
- VentMOD: wind force;
- VentANG: wind direction.

Here, we focus on the ozone concentration O3obs in function of the nine other variables. Hence, let $\tilde{Y}$ be the random variable of the ozone concentration and let $\mathbf{X}$ be the random vector containing the nine other random variables. Using the estimator $\widehat{E}^{knn}_{\emptyset,MC}$ of $\mathrm{E}_\emptyset = \mathrm{E}(\mathrm{Var}(\tilde{Y}|\mathbf{X}))$ given by Equation (16), with $N_I = 3$ and $N_\emptyset = 1000$, we estimate the value of $\mathrm{Var}(\mathrm{E}(\tilde{Y}|\mathbf{X}))/\mathrm{Var}(\tilde{Y})$ to 0.57, whereas it would be equal to 1 if $\tilde{Y}$ was a function of $\mathbf{X}$. Thus, we can not assume that the ozone concentration is a function of the nine other random variables.

The theory and methodology of this article holds when $\tilde{Y}$ is a deterministic function of $\mathbf{X}$. Hence, we create metamodels of the ozone concentration in function of $\mathbf{X}$, and we write Y the output of the metamodel. In this case, Y is indeed a deterministic function of $\mathbf{X}$ and we can compute the Shapley effects, which now quantify the impact of the inputs on the metamodel prediction. In practice, we replace the output column by the fitted values given by the metamodel.

To study the impact of the metamodel on the Shapley effects, we estimate the Shapley effects corresponding to three metamodels:

- XGBoost, from the R package `xgboost`, with optimized parameter by cross-validation;
- generalized linear model (GLM);
- Random Forest, from the R package `randomForest`, which optimizes automatically the parameters by out-of-bag.

Remark 15. *Using the estimator* $\widehat{E}^{knn}_{\emptyset,MC}$*, we estimate the value of* $\mathrm{Var}(\mathrm{E}(Y|\mathbf{X}))/\mathrm{Var}(Y)$ *to 0.91, 0.93 and 0.90 where* Y *denotes the output of each of the three metamodels XGBoost, GLM and Random Forest respectively. In contrast, the value of* $\mathrm{Var}(\mathrm{E}(\tilde{Y}|\mathbf{X}))/\mathrm{Var}(\tilde{Y})$ *is 0.57 when* $\tilde{Y}$ *denotes the original observed ozone concentrations. This shows that the predicted values are different from the initial values of the ozone concentration. Moreover, this shows that the metamodels do not overfit the data, since the estimated values of* $\mathrm{Var}(\mathrm{E}(Y|\mathbf{X}))/\mathrm{Var}(Y)$ *are close to 1. Indeed, that means that the fitted values of the ozone concentration are much more explained by* $\mathbf{X}$ *and have been smoothed by the metamodels. Furthermore, if the metamodels were overfitting the noise contained in the observed ozone concentration values, their outputs could not be predicted well given* $\mathbf{X}$*, and the estimate of* $\mathrm{Var}(\mathrm{E}(Y|\mathbf{X}))/\mathrm{Var}(Y)$ *would then be small when* Y *is one of the metamodel outputs.*

For each metamodel, we estimate the Shapley effects with the subset W-aggregation procedure and the double Monte-Carlo estimator $\widehat{E}^{knn}_{u,MC}$, with $N_I = 3$ and $N_{tot} = 50000$ (but the real cost is actually 40176, see Remark 4). For each metamodel, the computation time of all the Shapley effects on a personal computer is around 5 minutes. The results are presented in Figure 5.

We remark that the three metamodels yield similar Shapley effects. This is reassuring, since observing different behaviours of the metamodels would be a

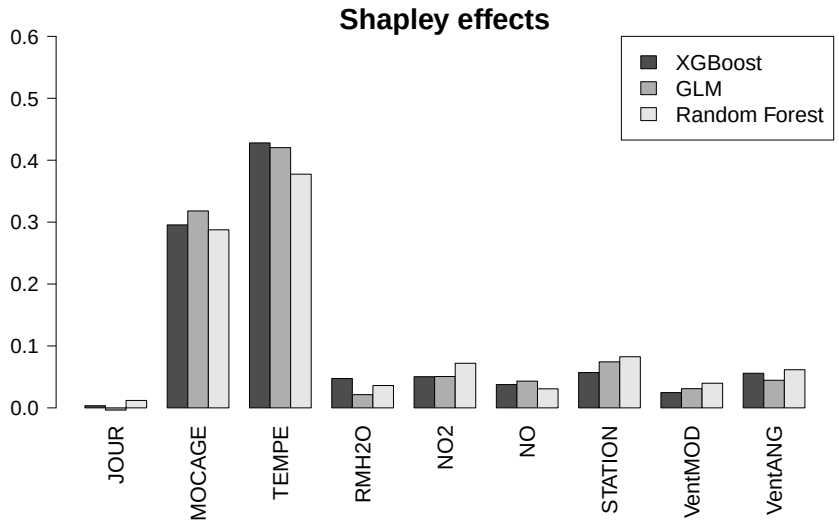

Figure 5: Estimation of the Shapley effects for three metamodels: XGBoost, GLM and Random Forest.

sign of inaccuracy for some of them. Only two variables have a significant impact on the ozone concentration: the predicted ozone concentration (MOCAGE) and the predicted temperature (TEMPE). This comforts the results of [BMM$^+$07] as they use regression trees whose two most important variables are the predicted ozone concentration and the predicted temperature. All the other variables have a much smaller impact. The Shapley effect of the predicted temperature is larger than the one of the predicted ozone concentration. It could be from the better accuracy of the predicted temperature (given by the official meteorology service of France) than the predicted ozone concentration (given by a fluid mechanics model). Finally, we remark that the type of the day (holiday or not) has no impact on the ozone concentration. The corresponding Shapley effect is even estimated by a slightly negative value for the GLM, which stems from the small error estimation.

To conclude, the Shapley effect estimator given by the subset W-aggregation procedure and the double Monte-Carlo estimator $\widehat{E}^{knn}_{u,MC}$ enables us to estimate the Shapley effects on real data. The estimator only requires a data frame of the inputs-output and handles heterogeneous data, with some categorical inputs and some continuous inputs. Here, the estimator was applied to a metamodel output. This illustrates the interest of the Shapley effects (and of sensitivity analysis) to understand and interpret the predictions of complex black-box machine learning procedures [RSG16, BGLR18].

This estimator has been implemented in the R package `sensitivity` as the function "shapleySubsetMc".

8 Conclusion

In this article, we focused on the estimation of the Shapley effects. We explained that this estimation is divided into two parts: the estimation of the conditional elements $(W_u)_{\emptyset \subsetneq u \subsetneq [1:p]}$ and the W-aggregation procedure. We suggested the new subset W-aggregation procedure and we explained how the already existing random-permutation W-aggregation procedure of [SNS16] minimizes the variance. However, the subset W-aggregation procedure is more efficient by using all the estimates of the conditional elements for each Shapley effect estimation. We highlighted this efficiency by numerical experiments. In a second part, we suggested various estimators of $(W_u)_{\emptyset \subsetneq u \subsetneq [1:p]}$ when the input distribution is unknown and when we only observe an i.i.d. sample of the input variables. We proved their consistency and gave the rates of convergence. Then, we used these new estimators to estimate the Shapley effects with consistency. We illustrated the efficiency of these estimators with numerical experiments and we tested one estimator on real heterogeneous data.

It is known that the Monte-Carlo algorithms for the estimation of the Sobol indices require many evaluations of f to be accurate (typically several thousands). If the evaluation cost is too high, it could be necessary to replace the function f with a metamodel $\widehat{f}$, such as Kriging [SWNW03]. An important field of research in Kriging is adaptive design of experiments [JSW98]. It would be interesting to study adaptive design of experiments in order to estimate the Shapley effects [FRK15], and to develop adaptive algorithms adapted to the estimators that we suggest in this article.

Acknowledgments

We are grateful to Vincent Prost for his helpful advises. We acknowledge the financial support of the Cross-Disciplinary Program on Numerical Simulation of CEA, the French Alternative Energies and Atomic Energy Commission. We would like to thank BPI France for co-financing this work, as part of the PIA (Programme d'Investissements d'Avenir) - Grand Défi du Numérique 2, supporting the PROBANT project. We are grateful to Philippe Besse and Olivier Mestre for enabling us to use the Ozone data set. We are very grateful to the associate editor and two reviewers, for their comments that lead to an improvement of this article.

A Proofs for the double Monte-Carlo and Pick-and-Freeze estimators: Theorems 1, 2, 3 and 4

To unify notation, let us write

$$\Phi_{MC}^{mix}: \quad (\mathbf{x}^{(1)},...,\mathbf{x}^{(N_I)}) \longmapsto \frac{1}{N_I-1}\sum_{k=1}^{N_I}\left(f(\mathbf{x}_{-u}^{(1)},\mathbf{x}_u^{(k)})-\frac{1}{N_I}\sum_{l=1}^{N_I}f(\mathbf{x}_{-u}^{(1)},\mathbf{x}_u^{(l)})\right)^2,$$

$$\Phi_{MC}^{knn}: \quad (\mathbf{x}^{(1)},...,\mathbf{x}^{(N_I)}) \longmapsto \frac{1}{N_I-1}\sum_{k=1}^{N_I}\left(f(\mathbf{x}^{(k)})-\frac{1}{N_I}\sum_{l=1}^{N_I}f(\mathbf{x}^{(l)})\right)^2,$$

$$\Phi_{PF}^{mix}: \quad (\mathbf{x}^{(1)},\mathbf{x}^{(2)}) \longmapsto f(\mathbf{x}^{(1)})f(\mathbf{x}_u^{(1)},\mathbf{x}_{-u}^{(2)})-\mathrm{E}(Y)^2,$$

$$\Phi_{PF}^{knn}: \quad (\mathbf{x}^{(1)},\mathbf{x}^{(2)}) \longmapsto f(\mathbf{x}^{(1)})f(\mathbf{x}^{(2)})-\mathrm{E}(Y)^2.$$

Remark that all these four functions are bounded as f is bounded. When we do not write the exponent mix or knn of Φ or of the estimators, it means that we refer to both of them (mix and knn). We write the proofs only for $\widehat{E}_{u,MC}$. For the estimators $\widehat{V}_{u,PF}$, it suffices to replace Φ_{MC} by Φ_{PF}, $-u$ by u (and vice-versa), E_u by V_u, $\mathrm{Var}(Y|X_{-u})$ by $\mathrm{E}(Y|X_u)^2-\mathrm{E}(Y)^2$ and N_I by 2. Hence, we shall only write the complete proofs for Theorems 1 and 2. To simplify notation, we will write $\widehat{E}_u$ for $\widehat{E}_{u,MC}$, $\widehat{E}_{u,l}$ for $\widehat{E}_{u,l,MC}$ and Φ for Φ_{MC}. N_I is a fixed integer. We also write $k_N(l,i) := k_N^{-u}(l,i)$, and the dependence on $-u$ is implicit.

A.1 Proof of consistency: Theorems 1 and 3

Recall that for all $i \in [1:p]$, $(\mathcal{X}_i, d_i)$ is a Polish space. Then, for all $v \subset [1:p]$, $\mathcal{X}_v := \prod_{i\in v}\mathcal{X}_i$ is a Polish space for the distance $d_v := \max_{i\in v} d_i$. We will write $B_v(\mathbf{x}_v, r)$ the open ball in $\mathcal{X}_v$ of radius r and center $\mathbf{x}_v$. We also let $\mu_v := \bigotimes_{i\in v}\mu_i$. Recall that the choice of the N_I-nearest neighbours could be not unique. In this case, conditionally to $(\mathbf{X}_{-u}^{(n)})_{n\le N}$, the $(k_N(l,i))_{l\in[1:N],i\in[1:N_I]}$ are random variables that we choose in the following way. Conditionally to $(\mathbf{X}_{-u}^{(n)})_{n\le N}$, we choose $k_N(l,i)$ uniformly over all the indices of the i-th nearest neighbours of $\mathbf{X}_{-u}^{(l)}$, such that the $(k_N(l,i))_{i\le N_I}$ are two by two distinct and independent of all the other random variables conditionally to $(\mathbf{X}_{-u}^{(n)})_{n\le N}$.

In particular, as we want to prove asymptotic results, we assume (without loss of generality) that we have an infinite i.i.d. sample $(\mathbf{X}^{(n)})_{n\in\mathbb{N}^*}$, and we assume that for all $N \in \mathbb{N}^*$, conditionally to $(\mathbf{X}_{-u}^{(n)})_{n\le N}$, $(k_N(l,i))_{i\le N_I} \perp\!\!\!\perp \sigma\left((\mathbf{X}_u^{(n)})_{n\le N}, (\mathbf{X}^{(n)})_{n>N}, (k_{N'}(l',i'))_{(N$

Hence, for all $N \in \mathbb{N}^*$ and $l \in [1:N]$, conditionally to $(\mathbf{X}_{-u}^{(n)})_{n\in\mathbb{N}}$, we have

$$(k_N(l,i))_{i\le N_I} \perp\!\!\!\perp \sigma\left((\mathbf{X}_u^{(n)})_{n\in\mathbb{N}}, (k_{N'}(l',i'))_{(N',l')\neq(N,l),\ i'\in[1:N_I]}\right).$$

To simplify notation, let us write $k_N(i) := k_N(1,i)$ (the index of one i-th neighbour of $\mathbf{X}_{-u}^{(1)}$) and $k'_N(i) := k_N(2,i)$ (the index of one i-th neighbour of $\mathbf{X}_{-u}^{(2)}$). Remark that $\mathbf{X}_{-u}^{(k_N(i))}$ does not depend on $k_N(i)$. Let $\mathbf{k} := (k_N(i))_{i\leq N_I, N\in\mathbb{N}^*}$ and $\mathbf{k}_N := (k_N(i))_{i\leq N_I}$. We will use the letter $\mathbf{h}$ for the realizations of the variable $\mathbf{k}$.

To begin with, let us recall two well-known results that we will use in the following.

Lemma 1. *Let A be a real random variable. If $\mathcal{H}$ is independent of $\sigma(\sigma(A),\mathcal{G})$, then*

$$\mathrm{E}(A|\sigma(\mathcal{G},\mathcal{H})) = \mathrm{E}(A|\mathcal{G}).$$

Lemma 2. *Let A, B be random variables. For all measurable ϕ,*

$$\mathcal{L}(\phi(A,B)|A=a) = \mathcal{L}(\phi(a,B)|A=a)$$

and if B is independent of A, then

$$\mathcal{L}(\phi(A,B)|A=a) = \mathcal{L}(\phi(a,B)).$$

Now, to prove Theorem 1, we need to prove several intermediate results.

Lemma 3. *For all $l\in\mathbb{N}^*$,*

$$\mathbf{X}_{-u}^{(k_N(l))} \xrightarrow[N\to+\infty]{a.s.} \mathbf{X}_{-u}^{(1)}. \tag{26}$$

Proof. First, let us show that for all $\varepsilon>0$, $\mathbb{P}(d_{-u}(\mathbf{X}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)})<\varepsilon)>0$. Indeed, as $\mathcal{X}_{-u}$ is a Polish space, its support has measure 1. Thus

$$\begin{aligned}
\mathbb{P}(d_{-u}(\mathbf{X}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)})<\varepsilon) &= \int_{\mathcal{X}_{-u}^2} \mathbb{1}_{d_{-u}(\mathbf{x}_{-u},\mathbf{x}'_{-u})<\varepsilon} d\mathbb{P}_{\mathbf{X}_{-u}}\otimes\mathbb{P}_{\mathbf{X}_{-u}}(\mathbf{x}_{-u},\mathbf{x}'_{-u}) \\
&= \int_{\mathcal{X}_{-u}} \mathbb{P}_{\mathbf{X}_{-u}}(B_{-u}(\mathbf{x}_{-u},\varepsilon)) d\mathbb{P}_{\mathbf{X}_{-u}}(\mathbf{x}_{-u}) \\
&= \int_{supp(\mathcal{X}_{-u})} \mathbb{P}_{\mathbf{X}_{-u}}(B_{-u}(\mathbf{x}_{-u},\varepsilon)) d\mathbb{P}_{\mathbf{X}_{-u}}(\mathbf{x}_{-u}) \\
&> 0,
\end{aligned}$$

because if $\mathbf{x}_{-u}\in supp(\mathcal{X}_{-u})$, then $B_{-u}(\mathbf{x}_{-u},\varepsilon)\not\subset supp(\mathcal{X}_{-u})^c$ and $\mathbb{P}_{\mathbf{X}_{-u}}(B_{-u}(\mathbf{x}_{-u},\varepsilon))>0$.

Next, remark that

$$\mathbf{X}_{-u}^{(k_N(l))} \xrightarrow[N\to+\infty]{a.s.} \mathbf{X}_{-u}^{(1)} \iff \mathbf{X}_{-u}^{(k_N(2))} \xrightarrow[N\to+\infty]{a.s.} \mathbf{X}_{-u}^{(1)},$$

and,

$$\begin{aligned}
\mathbb{P}\left(\left\{\mathbf{X}_{-u}^{(k_N(2))} \underset{N\to+\infty}{\longrightarrow} \mathbf{X}_{-u}^{(1)}\right\}^c\right) &= \mathbb{P}\left(\bigcup_{k\geq 1}\bigcap_{n\geq 2} d_{-u}(\mathbf{X}_{-u}^{(n)}, \mathbf{X}_{-u}^{(1)}) \geq \frac{1}{k}\right) \\
&\leq \sum_{k\geq 1} \mathbb{P}\left(\bigcap_{n\geq 2} d_{-u}(\mathbf{X}_{-u}^{(n)}, \mathbf{X}_{-u}^{(1)}) \geq \frac{1}{k}\right) \\
&= \sum_{k\geq 1} \lim_{N\to+\infty} \mathbb{P}\left(d_{-u}(\mathbf{X}_{-u}^{(2)}, \mathbf{X}_{-u}^{(1)}) \geq \frac{1}{k}\right)^N \\
&= \sum_{k\geq 1} \lim_{N\to+\infty} \left[1 - \mathbb{P}\left(d_{-u}(\mathbf{X}_{-u}^{(2)}, \mathbf{X}_{-u}^{(1)}) < \frac{1}{k}\right)\right]^N \\
&= \sum_{k\geq 1} 0 \\
&= 0.
\end{aligned}$$

□

Lemma 4. *There exists a version of*

$$\mathcal{L}(\mathbf{X}_u|\mathbf{X}_{-u} = \cdot) : (\mathcal{X}_{-u}, d_{-u}) \longrightarrow (\mathcal{M}_1(\mathcal{X}_u), \mathcal{T}(weak))$$

which is continuous $\mathbb{P}_{\mathbf{X}_{-u}}$-a.e., where $\mathcal{M}_1(\mathcal{X}_u)$ is the set of probability measures on $\mathcal{X}_u$ and $\mathcal{T}(weak)$ is the topology of weak convergence.

Proof. We assumed that there exists a version of $f_{\mathbf{X}}$ which is bounded and $\mathbb{P}_{\mathbf{X}}$-a.e. continuous. Let

$$f_{\mathbf{X}_{-u}}(\mathbf{x}_{-u}) := \int_{\mathcal{X}_u} f_{\mathbf{X}}(\mathbf{x}_u, \mathbf{x}_{-u}) d\mu_u(\mathbf{x}_u),$$

which is bounded by $\mu_u(\mathcal{X}_u)\|f_{\mathbf{X}}\|_\infty$ and is a $\mathbb{P}_{\mathbf{X}_{-u}}$-a.e. continuous (thanks to the dominated converging Theorem) version of the density of $\mathbf{X}_{-u}$ with respect to μ_{-u}. Let $\mathbf{x}_{-u} \in \mathcal{X}_{-u}$ such that $f_{\mathbf{X}_{-u}}(\mathbf{x}_{-u}) \leq \|f_{\mathbf{X}_{-u}}\|_\infty$, $f_{\mathbf{X}_{-u}}(\mathbf{x}_{-u}) > 0$ and such that $f_{\mathbf{X}_{-u}}$ is continuous at $\mathbf{x}_{-u}$. We have that

$$f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}(\mathbf{x}_u) := \frac{f_{\mathbf{X}}(\mathbf{x}_u, \mathbf{x}_{-u})}{f_{\mathbf{X}_{-u}}(\mathbf{x}_{-u})}$$

is a version of the density of $\mathbf{X}_u$ conditionally to $\mathbf{X}_{-u} = \mathbf{x}_{-u}$ (defined for almost all $\mathbf{x}_{-u}$). Let $(\mathbf{x}_{-u}^{(n)})$ be a sequence converging to $\mathbf{x}_{-u}$. There exists n_0 such that for all $n \geq n_0$, $f_{\mathbf{X}_{-u}}(\mathbf{x}_{-u}^{(n)}) > 0$. Thus, by continuity of f which respect to $\mathbf{x}_{-u}$ and of $f_{\mathbf{X}_{-u}}$, we have $f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}(\mathbf{x}_u) = \lim_{n\to+\infty} f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(n)}}(\mathbf{x}_u)$ for almost all $\mathbf{x}_u$. Then, using the dominated converging Theorem,

$$\mathcal{L}(\mathbf{X}_u|\mathbf{X}_{-u} = \mathbf{x}_{-u}^{(n)}) \underset{N\to+\infty}{\overset{weakly}{\longrightarrow}} \mathcal{L}(\mathbf{X}_u|\mathbf{X}_{-u} = \mathbf{x}_{-u}).$$

□

Remark 16. *The assumption "*$\mathbf{X} = (\mathbf{X}_u, \mathbf{X}_{-u})$ *has a bounded density* $f_{\mathbf{X}}$ *with respect to a finite measure* $\mu = \bigotimes_{i=1}^{p} \mu_i$*, which is continuous* $\mathbb{P}_{\mathbf{X}}$*-a.e." is only used in the proof of Lemma 4. It would be interesting in future work to prove 4 with a weaker assumption.*

Remark 17. *There exists a different proof of Lemma 4 if we assume that* μ *is regular. Theorem 8.1 of [Tju74] ensures that the conditional distribution in the sense of Tjur is defined for all* $\mathbf{x}_{-u}$ *such that* $f_{\mathbf{X}_{-u}} > 0$ *(and not only for almost all* $\mathbf{x}_{-u}$*) and the continuity of* $f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}(\mathbf{x}_u)$ *with respect to* $\mathbf{x}_{-u}$ *comes from Theorem 22.1 of [Tju74].*

Remark 18. *To avoid confusion, we can now define* $\mathcal{L}(\mathbf{X}_u|\mathbf{X}_{-u} = \mathbf{x}_{-u})$ *as the probability measure of density* $\frac{f(\cdot,\mathbf{x}_{-u})}{f_{\mathbf{X}_{-u}}(\mathbf{x}_{-u})}$*, which is defined for all (and not "almost all")* $\mathbf{x}_{-u}$ *in* $\{f_{\mathbf{X}_{-u}} > 0\}$*.*

Proposition 9. *If*

$$\mathcal{L}(\mathbf{X}_u|\mathbf{X}_{-u} = .) : (\mathcal{X}_{-u}, d_{-u}) \longrightarrow (\mathcal{M}_1(\mathcal{X}_u), \mathcal{T}(weak))$$

is continuous (where $\mathcal{T}(weak)$ *is the topology of weak convergence) almost everywhere, then, for almost all* $\left((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}\right)$*, we have*

$$\mathrm{E}\left(\widehat{E}_{u,1} \,\middle|\, (\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}\right) \underset{N\to+\infty}{\longrightarrow} \mathrm{Var}(Y|\mathbf{X}_{-u} = \mathbf{x}_{-u}^{(1)}) \tag{27}$$

and,

$$\mathrm{E}(\widehat{E}_{u,1}) \underset{N\to+\infty}{\longrightarrow} E_u. \tag{28}$$

Proof. Let $\mathbf{Z} = (\mathbf{Z}_1, ..., \mathbf{Z}_{N_I}) : (\Omega, \mathcal{A}) \to (\mathcal{X}^{N_I}, \mathcal{E}^{\otimes N_I})$ measurable, where $\mathcal{E}$ is the σ-algebra on $\mathcal{X}$, such that for almost all $\left((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}\right)$, we have

$$\mathcal{L}\left(\mathbf{Z}|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}\right) = \bigotimes_{i=1}^{N_I} \mathcal{L}(\mathbf{X}^{(1)}|\mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)}).$$

It suffices to show that, for almost all $\left((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}\right)$,

$$(\mathbf{X}^{(k_N(i))})_{i\le N_I} \underset{N\to+\infty}{\overset{\mathcal{L}_{|(\mathbf{X}_{-u}^{(n)})_n=(\mathbf{x}_{-u}^{(n)})_n,\mathbf{k}=\mathbf{h}}}{\longrightarrow}} \mathbf{Z}. \tag{29}$$

Indeed, if Equation (29) is true, then, using that Φ is bounded,

$$\begin{aligned}
& \mathrm{E}\left(\widehat{E}_{u,1} \,\middle|\, (\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}\right) \\
= \quad & \mathrm{E}\left[\Phi\left((\mathbf{X}^{(k_N(i))})_{i\le N_I}\right) \,\middle|\, (\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}\right] \\
\underset{N\to+\infty}{\longrightarrow} \quad & \mathrm{E}(\Phi(\mathbf{Z}) \,|\, (\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}) \\
= \quad & \mathrm{Var}(Y|\mathbf{X}_{-u} = \mathbf{x}_{-u}^{(1)}),
\end{aligned}$$

by definition of $\mathbf{Z}$ and of Φ. Thus, we have Equation 27. Furthermore, using dominated convergence theorem, integrating on $\left((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}\right)$, we obtain Equation 28.

Thus, it remains to show that conditionally to $(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}$, the random vector $(\mathbf{X}^{(k_N(i))})_{i \leq N_I}$ converges in distribution to $\mathbf{Z}$. We prove this convergence step by step.

Lemma 5. *For almost all* $(\mathbf{x}_{-u}^{(n)})_n$,

$$\mathcal{L}((\mathbf{X}_u^{(n)})_n | (\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n) = \bigotimes_{n \geq 1} \mathcal{L}(\mathbf{X}_u | \mathbf{X}_{-u} = \mathbf{x}_{-u}^{(n)}).$$

Proof. Let $(\tilde{\mathbf{X}}_{-u}^{(n)})_n : \Omega \to \mathcal{X}_{-u}^{\mathbb{N}}$ be an i.i.d. sequence of distribution $\mathcal{L}(\mathbf{X}_{-u})$. Then, we let $(\tilde{\mathbf{X}}_u^{(n)})_n : \Omega \to \mathcal{X}_u^{\mathbb{N}}$ be a sequence with conditional distribution

$$\mathcal{L}((\tilde{\mathbf{X}}_u^{(n)})_n | (\tilde{\mathbf{X}}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n) = \bigotimes_{n \geq 1} \mathcal{L}(\mathbf{X}_u | \mathbf{X}_{-u} = \mathbf{x}_{-u}^{(n)}).$$

We just have to prove that $(\tilde{\mathbf{X}}^{(n)})_n$ is an i.i.d. sample of distribution $\mathcal{L}(\mathbf{X})$.

Each $\tilde{\mathbf{X}}^{(n)}$ has a distribution $\mathcal{L}(\mathbf{X})$ because for all bounded measurable ϕ,

$$\begin{aligned}
\mathrm{E}(\phi(\tilde{\mathbf{X}}^{(n)})) &= \int_\Omega \phi(\tilde{\mathbf{X}}^{(n)}(\omega)) d\mathbb{P}(\omega) \\
&= \int_{\mathcal{X}_u \times \mathcal{X}_{-u}} \phi(\mathbf{x}_u, \mathbf{x}_{-u}) d\mathbb{P}_{(\tilde{\mathbf{X}}_u, \tilde{\mathbf{X}}_{-u})}(\mathbf{x}_u, \mathbf{x}_{-u}) \\
&= \int_{\mathcal{X}_{-u}} \left(\int_{\mathcal{X}_u} \phi(\mathbf{x}_u, \mathbf{x}_{-u}) d\mathbb{P}_{\mathbf{X}_u | \mathbf{X}_{-u} = \mathbf{x}_{-u}}(\mathbf{x}_u) \right) d\mathbb{P}_{\mathbf{X}_{-u}}(\mathbf{x}_{-u}) \\
&= \int_{\mathcal{X}} \phi(\mathbf{x}) d\mathbb{P}_{\mathbf{X}}(\mathbf{x}).
\end{aligned}$$

Moreover, $(\tilde{\mathbf{X}}^{(n)})_n$ are independent because if $n \neq m$, then, for all bounded

Borel functions ϕ_1 and ϕ_2, we have:

$$
\begin{aligned}
& \mathrm{E}(\phi_1(\tilde{\mathbf{X}}^{(n)})\phi_2(\tilde{\mathbf{X}}^{(m)})) \\
= & \int_{\mathcal{X}_u^2\times\mathcal{X}_{-u}^2} \phi_1(\mathbf{x}_u^{(n)},\mathbf{x}_{-u}^{(n)})\phi_2(\mathbf{x}_u^{(m)},\mathbf{x}_{-u}^{(m)}) d\mathbb{P}_{(\tilde{\mathbf{X}}_u^{(n)},\tilde{\mathbf{X}}_u^{(m)},\tilde{\mathbf{X}}_{-u}^{(n)},\tilde{\mathbf{X}}_{-u}^{(m)})}(\mathbf{x}_u^{(n)},\mathbf{x}_u^{(m)},\mathbf{x}_{-u}^{(n)},\mathbf{x}_{-u}^{(m)}) \\
= & \int_{\mathcal{X}_{-u}^2}\left(\int_{\mathcal{X}_u^2} \phi_1(\mathbf{x}_u^{(n)},\mathbf{x}_{-u}^{(n)})\phi_2(\mathbf{x}_u^{(m)},\mathbf{x}_{-u}^{(m)}) d\mathbb{P}_{(\tilde{\mathbf{X}}_u^{(n)},\tilde{\mathbf{X}}_u^{(m)})|(\tilde{\mathbf{X}}_{-u}^{(n)},\tilde{\mathbf{X}}_{-u}^{(m)})=(\mathbf{x}_{-u}^{(n)},\mathbf{x}_{-u}^{(m)})}(\mathbf{x}_u^{(n)},\mathbf{x}_u^{(m)})\right) d\mathbb{P}_{(\tilde{\mathbf{X}}_{-u}^{(n)},\tilde{\mathbf{X}}_{-u}^{(m)})}(\mathbf{x}_{-} \\
= & \int_{\mathcal{X}_{-u}^2}\left(\int_{\mathcal{X}_u^2} \phi_1(\mathbf{x}_u^{(n)},\mathbf{x}_{-u}^{(n)})\phi_2(\mathbf{x}_u^{(m)},\mathbf{x}_{-u}^{(m)}) d\mathbb{P}_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(n)}}\otimes\mathbb{P}_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(m)}}(\mathbf{x}_u^{(n)},\mathbf{x}_u^{(m)})\right) d\mathbb{P}_{\mathbf{X}_{-u}}^{\otimes 2}(\mathbf{x}_{-u}^{(n)},\mathbf{x}_{-u}^{(m)}) \\
= & \int_{\mathcal{X}_{-u}^2}\left(\int_{\mathcal{X}_u} \phi_1(\mathbf{x}_u^{(n)},\mathbf{x}_{-u}^{(n)}) d\mathbb{P}_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(n)}}(\mathbf{x}_u^{(n)})\right) \\
& \left(\int_{\mathcal{X}_u} \phi_2(\mathbf{x}_u^{(m)},\mathbf{x}_{-u}^{(m)}) d\mathbb{P}_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(m)}}(\mathbf{x}_u^{(m)})\right) d\mathbb{P}_{\mathbf{X}_{-u}}^{\otimes 2}(\mathbf{x}_{-u}^{(n)},\mathbf{x}_{-u}^{(m)}) \\
= & \int_{\mathcal{X}_{-u}}\left(\int_{\mathcal{X}_u} \phi_1(\mathbf{x}_u^{(n)},\mathbf{x}_{-u}^{(n)}) d\mathbb{P}_{\tilde{\mathbf{X}}_u|\tilde{\mathbf{X}}_{-u}=\mathbf{x}_{-u}^{(n)}}(\mathbf{x}_u^{(n)})\right) d\mathbb{P}_{\tilde{\mathbf{X}}_{-u}}(\mathbf{x}_{-u}^{(n)}) \\
& \left(\int_{\mathcal{X}_u} \phi_2(\mathbf{x}_u^{(m)},\mathbf{x}_{-u}^{(m)}) d\mathbb{P}_{\tilde{\mathbf{X}}_u|\tilde{\mathbf{X}}_{-u}=\mathbf{x}_{-u}^{(m)}}(\mathbf{x}_u^{(m)})\right) d\mathbb{P}_{\tilde{\mathbf{X}}_{-u}}(\mathbf{x}_{-u}^{(m)}) \\
= & \mathrm{E}(\phi_1(\tilde{\mathbf{X}}^{(n)}))\mathrm{E}(\phi_2(\tilde{\mathbf{X}}^{(m)})).
\end{aligned}
$$

The above calculation can be extended to finite products of more than two terms. That concludes the proof of Lemma 5. □

Lemma 6. *For almost all* $\left((\mathbf{x}_{-u}^{(n)})_n,\mathbf{h}\right)$*, we have:*

$$
\mathcal{L}\left((\mathbf{X}_u^{(k_N(i))})_{i\le N_I}|(\mathbf{X}_{-u}^{(n)})_n=(\mathbf{x}_{-u}^{(n)})_n,\mathbf{k}=\mathbf{h}\right)=\bigotimes_{i=1}^{N_I}\mathcal{L}\left(\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(h_N(i))}\right).
$$

Proof. For all bounded Borel function ϕ,

$$
\begin{aligned}
& \mathrm{E}\left(\phi((\mathbf{X}_u^{(k_N(i))})_{i\le N_I})|(\mathbf{X}_{-u}^{(n)})_n=(\mathbf{x}_{-u}^{(n)})_n,\mathbf{k}=\mathbf{h}\right) \\
= & \mathrm{E}\left(\phi\left((\mathbf{X}_u^{(k_N(i))})_{i\le N_I}\right)\middle|(\mathbf{X}_{-u}^{(n)})_n=(\mathbf{x}_{-u}^{(n)})_n,(k_{N'}(i))_{i\le N_I,N'\in\mathbb{N}^*}=(h_{N'}(i))_{i\le N_I,N'\in\mathbb{N}^*}\right) \\
= & \mathrm{E}\left(\phi\left((\mathbf{X}_u^{(k_N(i))})_{i\le N_I}\right)\middle|(\mathbf{X}_{-u}^{(n)})_n=(\mathbf{x}_{-u}^{(n)})_n,(k_N(i))_{i\le N_I}=(h_N(i))_{i\le N_I}\right) \\
= & \mathrm{E}\left(\phi\left((\mathbf{X}_u^{(h_N(i))})_{i\le N_I}\right)|(\mathbf{X}_{-u}^{(n)})_n=(\mathbf{x}_{-u}^{(n)})_n\right),
\end{aligned}
$$

using Lemmas 1 and 2 conditionally to $(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n$. Then,

$$\begin{aligned}
& \mathrm{E}\left(\phi\left((\mathbf{X}_u^{(h_N(i))})_{i\leq N_I}\right)|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n\right) \\
= & \int_{\mathcal{X}_u^{N_I}} \phi(\mathbf{x}_u^{(1)},...,\mathbf{x}_u^{(N_I)}) d\mathbb{P}_{(\mathbf{X}_u^{(h_N(i))})_{i\leq N_I}|(\mathbf{X}_{-u}^{(n)})_n=(\mathbf{x}_{-u}^{(n)})_n}(\mathbf{x}_u^{(1)},...,\mathbf{x}_u^{(N_I)}) \\
= & \int_{\mathcal{X}_u^{N_I}} \phi(\mathbf{x}_u^{(1)},...,\mathbf{x}_u^{(N_I)}) d\bigotimes_{i=1}^{N_I} \mathbb{P}_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(h_N(i))}}(\mathbf{x}_u^{(1)},...,\mathbf{x}_u^{(N_I)}).
\end{aligned}$$

That concludes the proof of Lemma 6. □

Recall that $\mathbf{X}_{-u}^{(k_N(i)))} \underset{N\to+\infty}{\longrightarrow} \mathbf{X}_{-u}^{(1)}$ $\mathbb{P}$-a.e., thus, for almost all $\left((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}\right)$,

$$\mathbf{x}_{-u}^{(h_N(i))} \underset{N\to+\infty}{\longrightarrow} \mathbf{x}_{-u}^{(1)}.$$

Thus, using the continuity of the conditional distribution given by Lemma 4, for almost all $\left((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}\right)$, we have,

$$\mathcal{L}(\mathbf{X}_u|\mathbf{X}_{-u} = \mathbf{x}_{-u}^{(h_N(i))}) \overset{weakly}{\underset{N\to+\infty}{\longrightarrow}} \mathcal{L}(\mathbf{X}_u|\mathbf{X}_{-u} = \mathbf{x}_{-u}^{(1)}).$$

Thus, for almost all $\left((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}\right)$,

$$\bigotimes_{i=1}^{N_I} \mathcal{L}(\mathbf{X}_u|\mathbf{X}_{-u} = \mathbf{x}_{-u}^{(h_N(i))}) \overset{weakly}{\underset{N\to+\infty}{\longrightarrow}} \bigotimes_{i=1}^{N_I} \mathcal{L}(\mathbf{X}_u|\mathbf{X}_{-u} = \mathbf{x}_{-u}^{(1)}) = \mathcal{L}(\mathbf{Z}_u|\mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)}).$$

So, using Lemma 6, for almost all $\left((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}\right)$,

$$\mathcal{L}\left((\mathbf{X}_u^{(k_N(i))})_{i\leq N_I}|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}\right) \overset{weakly}{\underset{N\to+\infty}{\longrightarrow}} \mathcal{L}(\mathbf{Z}_u|\mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)}).$$

So, for almost all $\left((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}\right)$,

$$\mathcal{L}\left((\mathbf{X}_u^{(k_N(i))})_{i\leq N_I}|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}\right) \overset{weakly}{\underset{N\to+\infty}{\longrightarrow}} \mathcal{L}\left(\mathbf{Z}_u|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}\right).$$

Using Slutsky lemma, for almost all $\left((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}\right)$,

$$\mathcal{L}\left((\mathbf{X}^{(k_N(i))})_{i\leq N_I}|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}\right) \overset{weakly}{\underset{N\to+\infty}{\longrightarrow}} \mathcal{L}\left(\mathbf{Z}|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}\right),$$

that concludes the proof of Proposition 9. □

Lemma 7. *The value of* $\mathrm{Var}(\widehat{E}_{u,1,MC})$ *is bounded by* $128\|f\|_\infty^4$.

Proof. As f is bounded, Φ is bounded by $\frac{1}{N_I-1}\sum_{k=1}^{N_I}(2\|f\|_\infty)^2 = \frac{N_I}{N_I-1}4\|f\|_\infty^2 \leq 8\|f\|_\infty^2$ so $\mathrm{Var}(\widehat{E}_{u,1})$ is bounded by $2\|\Phi\|_\infty^2 \leq 128\|f\|_\infty^4$. □

Proposition 10. *We have*

$$cov(\widehat{E}_{u,1}, \widehat{E}_{u,2}) \underset{N\to+\infty}{\longrightarrow} 0.$$

Proof. We use the law of total covariance

$$cov(\widehat{E}_{u,1}, \widehat{E}_{u,2}) = \mathrm{E}\left(cov\left(\widehat{E}_{u,1}, \widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)}\right)\right) + cov\left(\mathrm{E}(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)}), \mathrm{E}(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)})\right). \tag{30}$$

We will show that both terms go to 0 as N goes to $+\infty$. Let us compute the second term. Using Proposition 9,

$$\begin{aligned} & cov\left(\mathrm{E}(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)}), \mathrm{E}(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)})\right) \\ = \quad & \mathrm{E}\left(\mathrm{E}(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)})\mathrm{E}(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)})\right) - \mathrm{E}(\widehat{E}_{u,1})\mathrm{E}(\widehat{E}_{u,2}) \\ \underset{N\to+\infty}{\longrightarrow} \quad & \mathrm{E}\left(\mathrm{Var}(Y|\mathbf{X}_{-u} = \mathbf{X}_{-u}^{(1)})\mathrm{Var}(Y|\mathbf{X}_{-u} = \mathbf{X}_{-u}^{(2)})\right) - E_u^2 \\ = \quad & 0. \end{aligned}$$

It remains to prove that $\mathrm{E}\left(cov\left(\widehat{E}_{u,1}, \widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)}\right)\right)$ goes to 0. By dominated convergence theorem, it suffices to show that for almost all $(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)})$,

$$cov\left(\widehat{E}_{u,1}, \widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}^{(2)}\right) \underset{N\to+\infty}{\longrightarrow} 0. \tag{31}$$

From now on, we aim to proving Equation (31).

First, we want to prove Equation (31) for $\mathbf{x}_{-u}^{(1)} \neq \mathbf{x}_{-u}^{(2)}$. Using dominated convergence theorem and Proposition 9, it will suffice to show that (conditionally to $\mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)},\ \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}^{(2)}$), for almost all $((\mathbf{x}_{-u}^{(n)})_{n\geq 3}, \mathbf{h}, \mathbf{h}')$,

$$\mathrm{E}\left(\widehat{E}_{u,1}, \widehat{E}_{u,2}|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}, \mathbf{k}' = \mathbf{h}'\right) \underset{N\to+\infty}{\longrightarrow} \mathrm{Var}\left(Y|\mathbf{X}_{-u} = \mathbf{x}_{-u}^{(1)}\right)\mathrm{Var}\left(Y|\mathbf{X}_{-u} = \mathbf{x}_{-u}^{(2)}\right).$$

Let

$$A := \left\{\left((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}, \mathbf{h}'\right) \mid \mathbf{x}_{-u}^{h_N(N_1)} \underset{N\to+\infty}{\longrightarrow} \mathbf{x}_{-u}^{(1)},\ \mathbf{x}_{-u}^{h'_N(N_1)} \underset{N\to+\infty}{\longrightarrow} \mathbf{x}_{-u}^{(2)}\right\}.$$

The set A has probability 1 thanks to Lemma 3. Let $\left((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}, \mathbf{h}'\right) \in A$ be such that $\mathbf{x}_{-u}^{(1)} \neq \mathbf{x}_{-u}^{(2)}$ and let $\delta := d_{-u}(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)})/2$. There exists N_1 such that for all $N \geq N_1$,

$$d_{-u}\left(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(h_N(N_I))}\right) < \frac{\delta}{2},\ \ d_{-u}\left(\mathbf{x}_{-u}^{(2)}, \mathbf{x}_{-u}^{(h'_N(N_I))}\right) < \frac{\delta}{2}.$$

Thus, for all $N \geq N_1$,

$$\begin{aligned}
& E(\widehat{E}_{u,1}\widehat{E}_{u,2}|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}, \mathbf{k}' = \mathbf{h}') \\
= \quad & \mathrm{E}\left[\Phi\left((\mathbf{X}^{k_N(i)})_{i\leq N_I}\right)\Phi\left((\mathbf{X}^{k'_N(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}, \mathbf{k}' = \mathbf{h}'\right] \\
= \quad & \mathrm{E}\left[\Phi\left((\mathbf{X}^{k_N(i)})_{i\leq N_I}\right)\Phi\left((\mathbf{X}^{k'_N(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k}_N = \mathbf{h}_N, \mathbf{k}'_N = \mathbf{h}'_N\right] \\
= \quad & \mathrm{E}\left[\Phi\left((\mathbf{X}^{h_N(i)})_{i\leq N_I}\right)\Phi\left((\mathbf{X}^{k'_N(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k}'_N = \mathbf{h}'_N\right] \\
= \quad & \mathrm{E}\left[\Phi\left((\mathbf{X}^{h_N(i)})_{i\leq N_I}\right)\Phi\left((\mathbf{X}^{h'_N(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n\right] \\
= \quad & \mathrm{E}\left[\Phi\left((\mathbf{x}_{-u}^{h_N(i)})_{i\leq N_I}, (\mathbf{X}_u^{h_N(i)})_{i\leq N_I}\right)\Phi\left((\mathbf{x}_{-u}^{h'_N(i)})_{i\leq N_I}, (\mathbf{X}_u^{h'_N(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n\right] \\
= \quad & \mathrm{E}\left[\Phi\left((\mathbf{x}_{-u}^{h_N(i)})_{i\leq N_I}, (\mathbf{X}_u^{h_N(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n\right] \\
& \mathrm{E}\left[\Phi\left((\mathbf{x}_{-u}^{h'_N(i)})_{i\leq N_I}, (\mathbf{X}_u^{h'_N(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n\right] \\
= \quad & \mathrm{E}\left[\widehat{E}_{u,1}\middle|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k} = \mathbf{h}\right]\mathrm{E}\left[\widehat{E}_{u,2}\middle|(\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k}' = \mathbf{h}'\right] \\
\underset{N\to+\infty}{\longrightarrow} \quad & \mathrm{Var}\left[Y|\,\mathbf{X}_{-u} = \mathbf{x}_{-u}^{(1)}\right]\mathrm{Var}\left[Y|\,\mathbf{X}_{-u} = \mathbf{x}_{-u}^{(2)}\right],
\end{aligned}$$

thanks to Proposition 9.

Assume now that $\mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}$. We can assume without lost of generality that $\mathbb{P}(\mathbf{X}_{-u} = \mathbf{x}_{-u}) > 0$ because if we write $H := \{\mathbf{x}_{-u}|\ \mathbb{P}(\mathbf{X}_{-u} = \mathbf{x}_{-u}) = 0\}$, we have $\mathbb{P}(\mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} \in H) = 0$. We have to show that

$$\mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}\right) - \mathrm{Var}(Y|\mathbf{X}_{-u} = \mathbf{x}_{-u})^2 \underset{N\to+\infty}{\longrightarrow} 0.$$

Let $\varepsilon > 0$.

Let M_N the number of observations which are equal to $\mathbf{x}_{-u}$,

$$M_N := \#\{n \leq N\ :\ \mathbf{X}_{-u}^{(n)} = \mathbf{x}_{-u}\},$$

and let H_N be the number of nearest neighbours (up to N_I-nearest) shared by $\mathbf{X}_{-u}^{(1)}$ and $\mathbf{X}_{-u}^{(2)}$,

$$H_N := \#\left[\{k_N(i)\ :\ i \leq N_I\} \cap \{k'_N(i)\ :\ i \leq N_I\}\right].$$

If $M_n = m \geq 2N_I$, $\mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u} = \mathbf{X}_{-u}^{(2)}$, then the N_I-nearest neighbours $\mathbf{k}_N$ of $\mathbf{X}_{-u}^{(1)}$ and $\mathbf{k}'_N$ of $\mathbf{X}_{-u}^{(2)}$ are independent and are samples of uniformly distributed

variables on the same set of cardinal m, without replacement. Thus,

$$\begin{aligned}
& \mathbb{P}(H_N = 0 | M_N = m, \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}) \\
= & \frac{\binom{m - N_I}{N_I}}{\binom{m}{N_I}} \\
= & \frac{(m - 2N_I + 1)(m - 2N_I + 2)...(m - N_I)}{(m - N_I + 1)(m - N_I + 2)...m} \\
\underset{m \to +\infty}{\longrightarrow} & 1.
\end{aligned}$$

Thus, there exists m_1 such that

$$\alpha_{m_1} := \mathbb{P}(H_N = 0 | M_N \geq m_1, \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}) > 1 - \frac{\varepsilon}{5\|\Phi\|_\infty^2}. \tag{32}$$

So,

$$\begin{aligned}
& \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}\right) \\
= & \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N < m_1\right) \mathbb{P}(M_N < m_1 | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}) \\
& + \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1\right) \mathbb{P}(M_N \geq m_1 | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}).
\end{aligned}$$

Let

$$\beta_N := \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N < m_1\right) \mathbb{P}(M_N < m_1 | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}).$$

Conditionally to $\mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}$, we know that $M_N - 2 \sim \mathcal{B}\left(N - 2, \mathbb{P}(\mathbf{X}_{-u} = \mathbf{x}_{-u})\right)$, the binomial distribution. Thus, there exists N_1 such that for all $N \geq N_1$,

$$\mathbb{P}\left(M_N < m_1 | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}\right) < \frac{\varepsilon}{5\max(1, \|\Phi\|_\infty^2)}, \tag{33}$$

and so, for all $N \geq N_1$, $\beta_N < \varepsilon/5$. Furthermore

$$\begin{aligned}
& \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1\right) \\
= & \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1, H_N = 0\right) \mathbb{P}(H_N = 0 | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1) \\
& + \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1, H_N \geq 1\right) \mathbb{P}(H_N \geq 1 | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1).
\end{aligned}$$

Let

$$\gamma_N := \mathbb{P}\left(M_N \geq m_1 | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}\right).$$

Moreover, conditionally to $\mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1$, $H_N = 0$ implies that $\widehat{E}_{u,1} \perp\!\!\!\perp \widehat{E}_{u,2}$ thanks to Lemma 8.

Lemma 8. *Conditionally to* $\mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1,\ H_N = 0$, *the vector* $\left((\mathbf{X}^{(k_N(i))})_{i\leq N_I}, (\mathbf{X}^{(k'_N(i))})_{i\leq N_I}\right)$ *is composed of* $2N_I$ *i.i.d. random variables of distribution* $\mathbf{X}$ *conditionally to* $\mathbf{X}_{-u} = \mathbf{x}_{-u}$.

Proof. We know that, conditionally to $\mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1,\ H_N = 0$, the vector $\left((\mathbf{X}_{-u}^{(k_N(i))})_{i\leq N_I}, (\mathbf{X}_{-u}^{(k'_N(i))})_{i\leq N_I}\right)$ is constant equal to $(\mathbf{x}_{-u})_{i\leq 2N_I}$. It suffices to show that, conditionally to $\mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1,\ H_N = 0$, the vector $\left((\mathbf{X}_{u}^{(k_N(i))})_{i\leq N_I}, (\mathbf{X}_{u}^{(k'_N(i))})_{i\leq N_I}\right)$ is composed of $2N_I$ i.i.d. random variables of distribution $\mathbf{X}$ conditionally to $\mathbf{X}_{-u} = \mathbf{x}_{-u}$. Let $((\mathbf{x}_{-u}^{(n)})_n, \mathbf{h}_N, \mathbf{h}'_N)$ such that $\mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1$ and $H_N = 0$. As $M_N \geq m_1 \geq N_I$, for all $i \leq N_I$, we have $\mathbf{x}_{-u}^{(k_N(i))} = \mathbf{x}_{-u} = \mathbf{x}_{-u}^{(k'_N(i))}$. As $H_N = 0$, then, for all i and j smaller than N_I, $h_N(i) \neq h'_N(j)$. Thus, we have for any bounded Borel function ϕ,

$$
\begin{aligned}
& \mathrm{E}\left(\phi\left[(\mathbf{X}_{u}^{(k_N(i))})_{i\leq N_I}, (\mathbf{X}_{u}^{(k'_N(i))})_{i\leq N_I}\right] \middle| (\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k}_N = \mathbf{h}_N, \mathbf{k}'_N = \mathbf{h}'_N\right) \\
= \ & \mathrm{E}\left(\phi\left[(\mathbf{X}_{u}^{(h_N(i))})_{i\leq N_I}, (\mathbf{X}_{u}^{(k'_N(i))})_{i\leq N_I}\right] \middle| (\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n, \mathbf{k}'_N = \mathbf{h}'_N\right) \\
= \ & \mathrm{E}\left(\phi\left[(\mathbf{X}_{u}^{(h_N(i))})_{i\leq N_I}, (\mathbf{X}_{u}^{(h'_N(i))})_{i\leq N_I}\right] \middle| (\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n,\right) \\
= \ & \mathrm{E}\left(\phi\left[(\mathbf{X}_{u}^{(h_N(i))})_{i\leq N_I}, (\mathbf{X}_{u}^{(h'_N(i))})_{i\leq N_I}\right] \middle| (\mathbf{X}_{-u}^{(h_N(i))})_{i\leq N_I} = (\mathbf{x}_{-u})_{i\leq N_I}, (\mathbf{X}_{-u}^{(h'_N(i))})_{i\leq N_I} = (\mathbf{x}_{-u})_{i\leq N_I}\right) \\
= \ & \mathrm{E}\left(\phi\left[(\mathbf{X}_{u}^{(i)})_{i\leq N_I}, (\mathbf{X}_{u}^{(i+N_I)})_{i\leq N_I}\right] \middle| (\mathbf{X}_{-u}^{(i)})_{i\leq 2N_I} = (\mathbf{x}_{-u})_{i\leq 2N_I}\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& \mathrm{E}\left(\phi\left[(\mathbf{X}_{u}^{(k_N(i))})_{i\leq N_I}, (\mathbf{X}_{u}^{(k'_N(i))})_{i\leq N_I}\right] \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1, H_N = 0\right) \\
= \ & \mathrm{E}\left\{\mathrm{E}\left(\phi\left[(\mathbf{X}_{u}^{(k_N(i))})_{i\leq N_I}, (\mathbf{X}_{u}^{(k'_N(i))})_{i\leq N_I}\right] \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1, H_N = 0, (\mathbf{X}_{-u}^{(n)})_n, \mathbf{k}, \mathbf{k}'\right)\right\} \\
= \ & \mathrm{E}\left\{\mathrm{E}\left(\phi\left[(\mathbf{X}_{u}^{(i)})_{i\leq N_I}, (\mathbf{X}_{u}^{(i+N_I)})_{i\leq N_I}\right] \middle| (\mathbf{X}_{-u}^{(i)})_{i\leq 2N_I} = (\mathbf{x}_{-u})_{i\leq 2N_I}\right)\right\} \\
= \ & \mathrm{E}\left(\phi\left[(\mathbf{X}_{u}^{(i)})_{i\leq N_I}, (\mathbf{X}_{u}^{(i+N_I)})_{i\leq N_I}\right] \middle| (\mathbf{X}_{-u}^{(i)})_{i\leq 2N_I} = (\mathbf{x}_{-u})_{i\leq 2N_I}\right),
\end{aligned}
$$

that concludes the proof of Lemma 8. □

Thus

$$
\begin{aligned}
& \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1,\ H_N = 0\right) \\
= \ & \mathrm{E}\left(\widehat{E}_{u,1} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1,\ H_N = 0\right)^2
\end{aligned}
$$

and so, using Proposition 9, there exists N_2 such that for all $N \geq N_2$,

$$
\left|\mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1,\ H_N = 0\right) - \mathrm{Var}(Y | \mathbf{X}_{-u} = \mathbf{x}_{-u})^2\right| < \frac{\varepsilon}{5}. \tag{34}
$$

Thus, for all $N \geq \max(N_1, N_2)$,

$$\begin{aligned}
& \left| \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}\right) - \mathrm{Var}(Y|\mathbf{X}_{-u} = \mathbf{x}_{-u})^2 \right| \\
\leq \quad & |\beta_N| + \left| \gamma_N \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1,\ H_N \geq 1\right)(1 - \alpha_{m_1}) \right| \\
& + \left| \gamma_N \alpha_{m_1} \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1,\ H_N = 0\right) - \mathrm{Var}(Y|\mathbf{X}_{-u} = \mathbf{x}_{-u})^2 \right|.
\end{aligned}$$

The upper-bound is a sum of three terms. The first one is bounded by $\varepsilon/5$ using Equation 33 and the second one is bounded by $\varepsilon/5$ using Equation 32. For the last one, we use that, for all $C \in \mathbb{R}$,

$$\gamma_N \alpha_{m_1} C = (\gamma_N \alpha_{m_1} - 1)C + C.$$

Thus,

$$\begin{aligned}
& \left| \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}\right) - \mathrm{Var}(Y|\mathbf{X}_{-u} = \mathbf{x}_{-u})^2 \right| \\
\leq \quad & \frac{\varepsilon}{5} + \frac{\varepsilon}{5} + |\gamma_N \alpha_{m_1} - 1| \, \|\Phi\|_\infty^2 + \left| \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}, M_N \geq m_1,\ H_N = 0\right) - \mathrm{Var}(Y|\mathbf{X}_{-u} = \mathbf{x} \right. \\
\leq \quad & \frac{3\varepsilon}{5} + (|\gamma_N - 1|\alpha_N + |\alpha_N - 1|) \, \|\Phi\|_\infty^2 \quad \text{using Equation 34} \\
\leq \quad & \varepsilon,
\end{aligned}$$

using Equation 33 and Equation 32. Finally, we proved that

$$\mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)} = \mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}\right) - \mathrm{Var}(Y|\mathbf{X}_{-u} = \mathbf{x}_{-u})^2 \underset{N \to +\infty}{\longrightarrow} 0.$$

Hence, Equation (31) is proved and the proof of Proposition 10 is concluded. □

Proposition 11. *We have*

$$\widehat{E}_u - \mathrm{E}\left(\widehat{E}_{u,1}\right) \underset{\substack{N \to +\infty, \\ N_u \to +\infty}}{\overset{\mathbb{P}}{\longrightarrow}} 0. \tag{35}$$

Proof. Let $\varepsilon > 0$. By Chebyshev's inequality,

$$\mathbb{P}\left(\left|\widehat{E}_u - \mathrm{E}\left(\widehat{E}_u\right)\right| > \varepsilon\right) \leq \frac{\mathrm{Var}(\widehat{E}_u)}{\varepsilon^2}. \tag{36}$$

If $(s(l))_{l \leq N_u}$ is a sample of uniformly distributed variables on $[1 : N]$ with

replacement, we remark that for all $i \neq j$,

$$\begin{aligned}
& cov\left(\widehat{E}_{u,s(i)}, \widehat{E}_{u,s(j)}\right) \\
= \; & \mathrm{E}(\widehat{E}_{u,s(i)}\widehat{E}_{u,s(j)}) - \mathrm{E}(\widehat{E}_{u,s(i)})\mathrm{E}(\widehat{E}_{u,s(j)}) \\
= \; & \mathrm{E}(\widehat{E}_{u,s(i)}\widehat{E}_{u,s(j)}|s(i) \neq s(j))\mathbb{P}(s(i) \neq s(j)) \\
& + \mathrm{E}(\widehat{E}_{u,s(i)}\widehat{E}_{u,s(j)}|s(i) = s(j))\mathbb{P}(s(i) = s(j)) - \mathrm{E}(\widehat{E}_{u,s(i)})\mathrm{E}(\widehat{E}_{u,s(j)}) \\
= \; & \left[\mathrm{E}(\widehat{E}_{u,s(i)}\widehat{E}_{u,s(j)}|s(i) \neq s(j)) - \mathrm{E}(\widehat{E}_{u,1})\mathrm{E}(\widehat{E}_{u,2})\right]\mathbb{P}(s(i) \neq s(j)) \\
& + \left[\mathrm{E}(\widehat{E}_{u,s(i)}\widehat{E}_{u,s(i)}|s(i) = s(j)) - \mathrm{E}(\widehat{E}_{u,1})^2\right]\mathbb{P}(s(i) = s(j)) \\
= \; & \left[\mathrm{E}(\widehat{E}_{u,1}\widehat{E}_{u,2}|s(i) = 1, s(j) = 2) - \mathrm{E}(\widehat{E}_{u,1})\mathrm{E}(\widehat{E}_{u,2})\right]\mathbb{P}(s(i) \neq s(j)) \\
& + \left[\mathrm{E}(\widehat{E}_{u,1}\widehat{E}_{u,1}|s(i) = s(j) = 1) - \mathrm{E}(\widehat{E}_{u,1})^2\right]\mathbb{P}(s(i) = s(j)) \\
= \; & cov\left(\widehat{E}_{u,1}, \widehat{E}_{u,2}\right)\mathbb{P}(s(i) \neq s(j)) + \mathrm{Var}\left(\widehat{E}_{u,1}\right)\mathbb{P}(s(i) = s(j)),
\end{aligned}$$

thus

$$\begin{aligned}
\mathrm{Var}(\widehat{E}_u) &= \frac{1}{N_u^2}\sum_{i,j=1}^{N_u} cov\left(\widehat{E}_{u,s(i)}, \widehat{E}_{u,s(j)}\right) \\
&= \frac{1}{N_u^2}\sum_{i\neq j=1}^{N_u} cov\left(\widehat{E}_{u,1}, \widehat{E}_{u,2}\right)\mathbb{P}(s(i) \neq s(j)) \\
&\quad + \frac{1}{N_u^2}\sum_{i\neq j=1}^{N_u} \mathrm{Var}\left(\widehat{E}_{u,1}\right)\mathbb{P}(s(i) = s(j)) + \frac{1}{N_u^2}\sum_{i=1}^{N_u} \mathrm{Var}\left(\widehat{E}_{u,s(i)}\right) \\
&\leq \frac{1}{N_u^2}\sum_{i\neq j=1}^{N_u} \left|cov\left(\widehat{E}_{u,1}, \widehat{E}_{u,2}\right)\right| \\
&\quad + \frac{1}{N_u^2}\sum_{i\neq j=1}^{N_u} \mathrm{Var}\left(\widehat{E}_{u,1}\right)\frac{1}{N} + \frac{1}{N_u^2}\sum_{i=1}^{N_u} \mathrm{Var}\left(\widehat{E}_{u,1}\right) \\
&\leq \left|cov\left(\widehat{E}_{u,1}, \widehat{E}_{u,2}\right)\right| + \mathrm{Var}\left(\widehat{E}_{u,1}\right)\left(\frac{1}{N} + \frac{1}{N_u}\right).
\end{aligned}$$

If $(s(l))_{l\leq N_u}$ is a sample of uniformly distributed variables on $[1:N]$ without replacement, we have

$$\begin{aligned}
\mathrm{Var}(\widehat{E}_u) &= \frac{1}{N_u^2}\sum_{i,j=1}^{N_u} cov\left(\widehat{E}_{u,s(i)}, \widehat{E}_{u,s(j)}\right) \\
&= \frac{1}{N_u^2}\sum_{i\neq j=1}^{N_u} cov\left(\widehat{E}_{u,s(i)}, \widehat{E}_{u,s(j)}\right) + \frac{1}{N_u^2}\sum_{i=1}^{N_u} \mathrm{Var}\left(\widehat{E}_{u,s(i)}\right) \\
&= \frac{N_u - 1}{N_u} cov\left(\widehat{E}_{u,1}, \widehat{E}_{u,2}\right) + \frac{1}{N_u}\mathrm{Var}\left(\widehat{E}_{u,1}\right).
\end{aligned}$$

In both cases (with or without replacement), thanks to Proposition 10, we have

$$\mathbb{P}\left(\left|\widehat{E}_u - \mathrm{E}\left(\widehat{E}_u\right)\right| > \varepsilon\right) \underset{\substack{N\to+\infty,\\ N_u\to+\infty}}{\longrightarrow} 0.$$

□

Now, to prove Theorem 1, we only have to use Proposition 9 (which can be applied thanks to Lemma 4) and Proposition 11.

A.2 Proof for rate of convergence: Theorems 2 and 4

We want to prove Theorems 2 and 4 about the rate of convergence of the double Monte-Carlo and Pick-and-Freeze estimators. We have to add some notation. We will write $C_{\sup}$ for a generic non-negative finite constant (depending only on u, f and the distribution of $\mathbf{X}$). The actual value of $C_{\sup}$ is of no interest and can change in the same sequence of equations. Similarly, we will write $C_{\inf}$ a generic strictly positive constant. We will write $C_{\sup}(\varepsilon)$ for a generic non-negative finite constant depending only on ε, u, f and the distribution of $\mathbf{X}$.

Recall that for all i, $\mathcal{X}_i$ is a compact subset of $\mathbb{R}$ and that f is $\mathcal{C}^1$. Moreover recall that $\mathbf{X}$ has a probability density $f_{\mathbf{X}}$ with respect to λ_p (the Lebesgue measure on $\mathbb{R}^p$) such that λ_p-a.e., we have $0 < C_{\inf} \leq f_{\mathbf{X}} \leq C_{\sup}$, and such that $f_{\mathbf{X}}$ is Lipschitz continuous.

Note that with these assumptions, Φ is $\mathcal{C}^1$ on the compact set $\mathcal{X}$ and so Lipschitz continuous. For all n, we will write d for the euclidean distance on $\mathbb{R}^n$ (for any value of n) and $B(\mathbf{x}, r)$ for the open ball of radius r and center $\mathbf{x}$ in $\mathbb{R}^n$. We also let $\mathcal{S}(\mathbf{x}, r)$ be the sphere of center $\mathbf{x}$ and radius r.

Remark that

$$\begin{aligned}
& \mathbb{P}\left(d(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)}) = d(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(3)})\right) \\
= & \int_{\mathcal{X}_{-u}^2} \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) = d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(3)})\right) d\mathbb{P}_{\mathbf{X}_{-u}}^{\otimes 2}(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \\
\leq & C_{\sup} \int_{\mathcal{X}_{-u}^2} \lambda_{|-u|}\left(\mathcal{S}(\mathbf{x}_{-u}^{(1)}, d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}))\right) d\mathbb{P}_{\mathbf{X}_{-u}}^{\otimes 2}(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \\
= & 0,
\end{aligned}$$

because the Lebesgue measure of the sphere $\mathcal{S}(\mathbf{x}_{-u}^{(1)}, d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}))$ is zero. Thus, almost everywhere, for all l and all $i \neq j$,

$$d\left(\mathbf{X}_{-u}^{(l)}, \mathbf{X}_{-u}^{(i)}\right) \neq d\left(\mathbf{X}_{-u}^{(l)}, \mathbf{X}_{-u}^{(j)}\right).$$

Thus, the indices of the nearest neighbours $(k_N(l,i))_{l,i}$ are constant random variables conditionally to $(\mathbf{X}_{-u}^{(n)})_n$ or to $(\mathbf{X}_{-u}^{(n)})_{n\leq N}$. In particular, for all N and

l, $k_N(l,1) = l$. Thanks to Doob-Dynkin lemma, we can write, abusing notation, $k_N(l,i)(\omega) = k_N(l,i)[(\mathbf{X}_{-u}^{(n)}(\omega))_n] = k_N(l,i)[(\mathbf{X}_{-u}^{(n)}(\omega))_{n\leq N}]$. To simplify notation, let us write $k_N(i) := k_N(1,i)$ (the index of one i-th neighbour of $\mathbf{X}_{-u}^{(1)}$) and $k'_N(i) := k_N(2,i)$ (the index of one i-th neighbour of $\mathbf{X}_{-u}^{(2)}$).

Remark 19. *We can prove the rate of convergence in a more general framework than the Euclidean space with the Lebesgue measure. It suffices to have a compact set $\mathcal{X}$ with a dominating finite measure $\mu = \bigotimes \mu_i$ such that for μ_i-almost all $\mathbf{x}_i \in \mathcal{X}_i$ and for all $\delta > 0$,*

$$C_{\inf}\delta \leq \mu_i(B(\mathbf{x}_i,\delta)) = \mu_i(\overline{B}(\mathbf{x}_i,\delta)) \leq C_{\sup}\delta.$$

We prove Theorems 2 and 4 step by step.

Lemma 9. *Assume that $(a_i)_i$ and $(b_i)_i$ are sequences such that for all i, $|a_i| \leq M$, $|b_i| \leq M$ and $|a_i - b_i| \leq \varepsilon$. Then, for all $N \in \mathbb{N}^*$*

$$\left|\prod_{i=1}^N a_i - \prod_{i=1}^N b_i\right| \leq NM^{N-1}\varepsilon.$$

Proof. By induction. □

Lemma 10. *If for all $i \leq N$, $d(\mathbf{x}_{-u}^{(i)}, \mathbf{y}_{-u}^{(i)}) < \varepsilon$, then, for all $(\mathbf{a}_{-u}^{(i)})_{i\leq N_I} \in \mathcal{X}_{-u}^{N_I}$,*

$$\begin{aligned}
&\left|\mathrm{E}\left[\Phi\left((\mathbf{a}_{-u}^{(i)})_{i\leq N_I}, (\mathbf{X}_u^{(i)})_{i\leq N_I}\right) \middle| (\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{x}_{-u}^{(i)})_{i\leq N_I}\right]\right. \\
-\quad &\left.\mathrm{E}\left[\Phi\left((\mathbf{a}_{-u}^{(i)})_{i\leq N_I}, (\mathbf{X}_u^{(i)})_{i\leq N_I}\right) \middle| (\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{y}_{-u}^{(i)})_{i\leq N_I}\right]\right| \quad \leq C_{\sup}\varepsilon.
\end{aligned}$$

Proof.

$$\begin{aligned}
&\left|\mathrm{E}\left[\Phi\left((\mathbf{a}_{-u}^{(i)})_{i\leq N_I}, (\mathbf{X}_u^{(i)})_{i\leq N_I}\right) \middle| (\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{x}_{-u}^{(i)})_{i\leq N_I}\right]\right. \\
&\left.- \mathrm{E}\left[\Phi\left((\mathbf{a}_{-u}^{(i)})_{i\leq N_I}, (\mathbf{X}_u^{(i)})_{i\leq N_I}\right) \middle| (\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{y}_{-u}^{(i)})_{i\leq N_I}\right]\right| \\
= &\left|\int_{\mathcal{X}_u^{N_I}} \Phi((\mathbf{a}_{-u}^{(i)})_{i\leq N_I}, (\mathbf{x}_u^{(i)})_{i\leq N_I}) \left(f_{(\mathbf{X}_u^{(i)})_{i\leq N_I}|(\mathbf{X}_{-u}^{(i)})_{i\leq N_I}=(\mathbf{x}_{-u}^{(i)})_{i\leq N_I}}((\mathbf{x}_u^{(i)})_{i\leq N_I})\right.\right. \\
&\left.\left. - f_{(\mathbf{X}_u^{(i)})_{i\leq N_I}|(\mathbf{X}_{-u}^{(i)})_{i\leq N_I}=(\mathbf{y}_{-u}^{(i)})_{i\leq N_I}}((\mathbf{x}_u^{(i)})_{i\leq k})\right) d((\mathbf{x}_u^{(i)})_{i\leq N_I})\right| \\
\leq &\ C_{\sup}\int_{\mathcal{X}_u^{N_I}} \left|\prod_{i=1}^{N_I} f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(i)}}(\mathbf{x}_u^{(i)}) - \prod_{i=1}^{N_I} f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{y}_{-u}^{(i)}}(\mathbf{x}_u^{(i)})\right| d((\mathbf{x}_u^{(i)})_{i\leq N_I}).
\end{aligned}$$

We know that,

$$\begin{aligned}
& \left|f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}}(\mathbf{x}_u) - f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{y}_{-u}}(\mathbf{x}_u)\right| \\
\leq & \left|\frac{f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{x}_{-u})}{\int_{\mathcal{X}_u} f_{\mathbf{X}}(\mathbf{x}'_u,\mathbf{x}_{-u})d(\mathbf{x}'_u)} - \frac{f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{y}_{-u})}{\int_{\mathcal{X}_u} f_{\mathbf{X}}(\mathbf{x}'_u,\mathbf{y}_{-u})d(\mathbf{x}'_u)}\right| \\
\leq & \frac{1}{\int_{\mathcal{X}_u} f_{\mathbf{X}}(\mathbf{x}'_u,\mathbf{x}_{-u})d(\mathbf{x}'_u)} \left|f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{x}_{-u}) - f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{y}_{-u})\right| \\
& + f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{y}_{-u}) \left|\frac{1}{\int_{\mathcal{X}_u} f_{\mathbf{X}}(\mathbf{x}'_u,\mathbf{x}_{-u})d(\mathbf{x}'_u)} - \frac{1}{\int_{\mathcal{X}_u} f_{\mathbf{X}}(\mathbf{x}'_u,\mathbf{y}_{-u})d(\mathbf{x}'_u)}\right| \\
\leq & C_{\sup}\left|f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{x}_{-u}) - f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{y}_{-u})\right| + C_{\sup}\left|f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{x}_{-u}) - f_{\mathbf{X}}(\mathbf{x}_u,\mathbf{y}_{-u})\right| \\
\leq & C_{\sup} d(\mathbf{x}_{-u},\mathbf{y}_{-u}).
\end{aligned}$$

Thus, for all $i \in [1 : N_i]$ and for all $\mathbf{x}_u^{(i)}$,

$$\left|f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(i)}}(\mathbf{x}_u^{(i)}) - f_{\mathbf{X}_u|\mathbf{X}_{-u}=\mathbf{y}_{-u}^{(i)}}(\mathbf{x}_u^{(i)})\right| \leq C_{\sup}\varepsilon.$$

Thus, using Lemma 9,

$$\begin{aligned}
& \left|\mathrm{E}\left[\Phi\left((\mathbf{a}_{-u}^{(i)})_{i\leq N_I},(\mathbf{X}_u^{(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{x}_{-u}^{(i)})_{i\leq N_I}\right]\right. \\
& \left. - \mathrm{E}\left[\Phi\left((\mathbf{a}_{-u}^{(i)})_{i\leq N_I},(\mathbf{X}_u^{(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{y}_{-u}^{(i)})_{i\leq N_I}\right]\right| \leq C_{\sup}\varepsilon.
\end{aligned}$$

□

Lemma 11. *If for all i, $d(\mathbf{x}_{-u}^{(i)},\mathbf{y}_{-u}^{(i)}) < \varepsilon$, then*

$$\begin{aligned}
& \left|\mathrm{E}\left[\Phi\left((\mathbf{x}_{-u}^{(i)})_{i\leq N_I},(\mathbf{X}_u^{(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{x}_{-u}^{(i)})_{i\leq N_I}\right]\right. \\
- & \left. \mathrm{E}\left[\Phi\left((\mathbf{y}_{-u}^{(i)})_{i\leq N_I},(\mathbf{X}_u^{(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{y}_{-u}^{(i)})_{i\leq N_I}\right]\right| \leq C_{\sup}\varepsilon.
\end{aligned}$$

Proof.

$$\begin{aligned}
& \left|\mathrm{E}\left[\Phi\left((\mathbf{x}_{-u}^{(i)})_{i\leq N_I},(\mathbf{X}_u^{(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{x}_{-u}^{(i)})_{i\leq N_I}\right]\right. \\
& \left. - \mathrm{E}\left[\Phi\left((\mathbf{y}_{-u}^{(i)})_{i\leq N_I},(\mathbf{X}_u^{(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{y}_{-u}^{(i)})_{i\leq N_I}\right]\right| \\
\leq & \left|\mathrm{E}\left[\Phi\left((\mathbf{x}_{-u}^{(i)})_{i\leq N_I},(\mathbf{X}_u^{(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{x}_{-u}^{(i)})_{i\leq N_I}\right]\right. \\
& \left. - \mathrm{E}\left[\Phi\left((\mathbf{x}_{-u}^{(i)})_{i\leq N_I},(\mathbf{X}_u^{(i)})_{i\leq N_I}\right)\middle|(\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{y}_{-u}^{(i)})_{i\leq N_I}\right]\right| \\
& + \left|\mathrm{E}\left[\Phi\left((\mathbf{x}_{-u}^{(i)})_{i\leq N_I},(\mathbf{X}_u^{(i)})_{i\leq N_I}\right) - \Phi\left((\mathbf{y}_{-u}^{(i)})_{i\leq N_I},(\mathbf{X}_u^{(i)})_{i\leq N_I}\right)|(\mathbf{X}_{-u}^{(i)})_{i\leq N_I} = (\mathbf{y}_{-u}^{(i)})_{i\leq N_I}\right]\right| \\
\leq & C_{\sup}\varepsilon + C_{\sup}\varepsilon,
\end{aligned}$$

using Lemma 10 and using that Φ is Lipschitz continuous on $\mathcal{X}$. □

Lemma 12. *There exists $C_{\sup} < +\infty$ such that for all $a > 0$,*

$$\mathbb{P}\left(d\left(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}\right) \geq a \middle| \mathbf{X}_{-u}^{(1)}\right) \leq C_{\sup} N^{N_I} (1 - C_{\inf} a^{|-u|})^{N-N_I}. \quad (37)$$

Proof. Let $K(a) := \#\{n \in [2:N],\ d(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(n)}) < a\}$. Conditionally to $\mathbf{X}_{-u}^{(1)}$, $K(a) \sim \mathcal{B}(N-1, p(a, \mathbf{X}_{-u}^{(1)}))$, writing $p(a, \mathbf{X}_{-u}^{(1)}) := \mathbb{P}(d(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)}) < a | \mathbf{X}_{-u}^{(1)})$. Thus,

$$\begin{aligned}
& \mathbb{P}\left(d\left(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}\right) \geq a \middle| \mathbf{X}_{-u}^{(1)}\right) \\
= \ & \mathbb{P}\left(K(a) \leq N_I - 1 | \mathbf{X}_{-u}^{(1)}\right) \\
= \ & \sum_{k=0}^{N_I-1} \binom{N-1}{k} p(a, \mathbf{X}_{-u}^{(1)})^k (1 - p(a, \mathbf{X}_{-u}^{(1)}))^{N-1-k} \\
\leq \ & N_I \binom{N-1}{N_I-1} (1 - p(a, \mathbf{X}_{-u}^{(1)}))^{N-N_I} \\
\leq \ & C_{\sup} N^{N_I} (1 - p(a, \mathbf{X}_{-u}^{(1)}))^{N-N_I}.
\end{aligned}$$

We know that

$$\begin{aligned}
p(a, \mathbf{X}_{-u}^{(1)}) &= \int_{B(\mathbf{X}_{-u}^{(1)}, a)} f_{\mathbf{X}_{-u}}(\mathbf{x}_{-u}) d\mathbf{x}_{-u} \\
&\geq C_{\inf} \lambda_{|-u|}\left(B(\mathbf{X}_{-u}^{(1)}, a)\right) \\
&\geq C_{\inf} a^{|-u|}.
\end{aligned}$$

Thus

$$\mathbb{P}\left(d\left(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}\right) \geq a \middle| \mathbf{X}_{-u}^{(1)}\right) \leq C_{\sup} N^{N_I} (1 - C_{\inf} a^{|-u|})^{N-N_I}. \quad (38)$$

□

Remark 20. *For the estimators $\widehat{V}_{u,PF}$, we choose only one nearest neighbour different from $\mathbf{X}_u^{(1)}$ in $\widehat{V}_{u,1,PF}$, which is $\mathbf{X}_u^{(k_N(2))}$. Thus, in the previous computation, we do not have the N^{N_I}. Remark that this is also true for $\widehat{E}_{u,MC}$ taking $N_I = 2$.*

Lemma 13. *For all $\varepsilon > 0$, there exists $C_{\sup}(\varepsilon)$ such that*

$$\mathrm{E}\left(d\left(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}\right)\right) \leq \frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}, \quad (39)$$

and for all $\mathbf{x}_{-u}^{(1)}$,

$$\mathrm{E}\left(d\left(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}\right) \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)}\right) \leq \frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}. \quad (40)$$

Proof. Using Lemma 12, we have

$$\begin{aligned}
&\mathrm{E}\left((N-N_I)^{\frac{1}{|-u|}-\varepsilon} d\left(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}\right) \middle| \mathbf{X}_{-u}^{(1)}\right) \\
= &\int_0^{+\infty} \mathbb{P}\left((N-N_I)^{\frac{1}{|-u|}-\varepsilon} d\left(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}\right) > t \middle| \mathbf{X}_{-u}^{(1)}\right) dt \\
\leq &1 + \int_1^{+\infty} \mathbb{P}\left(d\left(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}\right) > t(N-N_I)^{-\frac{1}{|-u|}+\varepsilon} \middle| \mathbf{X}_{-u}^{(1)}\right) dt \\
= &1 + \frac{1}{|-u|}\int_1^{+\infty} s^{\frac{1}{|-u|}-1} \mathbb{P}\left(d\left(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}\right) > s^{\frac{1}{|-u|}}(N-N_I)^{-\frac{1}{|-u|}+\varepsilon} \middle| \mathbf{X}_{-u}^{(1)}\right) ds \\
\leq &1 + \frac{1}{|-u|}\int_1^{+\infty} C_{\sup} N^{N_I} (1 - C_{\inf} s (N-N_I)^{|-u|\varepsilon-1})^{N-N_I} ds,
\end{aligned}$$

and

$$\begin{aligned}
(1 - C_{\inf} s (N-N_I)^{|-u|\varepsilon-1})^{N-N_I} &= \exp\left[(N-N_I)\ln\left(1 - C_{\inf} s (N-N_I)^{|-u|\varepsilon-1}\right)\right] \\
&\leq \exp\left[(N-N_I)\left(-C_{\inf} s (N-N_I)^{|-u|\varepsilon-1}\right)\right] \\
&= \exp(-C_{\inf} s (N-N_I)^{|-u|\varepsilon}).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathrm{E}\left((N-N_I)^{\frac{1}{|-u|}-\varepsilon} d\left(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}\right) \middle| \mathbf{X}_{-u}^{(1)}\right) \\
\leq &1 + C_{\sup}\int_1^{+\infty} N^{N_I} \exp(-C_{\inf} s (N-N_I)^{|-u|\varepsilon}) ds \\
\leq &1 + C_{\sup}\left[N^{N_I}\exp(-C_{\inf}\frac{1}{2}(N-N_I)^{|-u|\varepsilon})\right]\int_1^{+\infty} \exp(-C_{\inf}\frac{s}{2}(N-N_I)^{|-u|\varepsilon}) ds \\
\leq &1 + C_{\sup}(\varepsilon).
\end{aligned}$$

Indeed, the values $N^{N_I}\exp(-C_{\inf}\frac{1}{2}(N-N_I)^{|-u|\varepsilon}$ and $\int_1^{+\infty}\exp(-C_{\inf}\frac{s}{2}(N-N_I)^{|-u|\varepsilon}) ds$ go to 0 when N do $+\infty$. Thus

$$\mathrm{E}\left(d\left(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}\right) \middle| \mathbf{X}_{-u}^{(1)}\right) \leq \frac{1 + C_{\sup}(\varepsilon)}{(N-N_I)^{\frac{1}{p-|u|}-\varepsilon}} \leq \frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}.$$

That concludes the proof of Lemma 13. □

Remark 21. *For the estimators $\widehat{V}_{u,PF}$, we do not have the N^{N_I}. Thus, we can choose $\varepsilon = 0$ up to Proposition 12.*

Proposition 12. *For all $\varepsilon > 0$, there exists $C_{\sup}(\varepsilon)$ such that*

$$\left|\mathrm{E}\left(\widehat{E}_u\right) - E_u\right| \leq \frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}} \tag{41}$$

and for almost all $\mathbf{x}_{-u}^{(1)}$,

$$\left|\mathrm{E}\left(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)}\right) - \mathrm{Var}(Y|\mathbf{X}_{-u} = \mathbf{x}_{-u}^{(1)})\right| \leq \frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}. \tag{42}$$

Proof. For almost all $(\mathbf{x}_{-u}^{(n)})_n$, using the definition of the random variable $\mathbf{Z}$ (in the proof of Proposition 9) and using Lemma 6,

$$\begin{aligned}
& \left| \mathrm{E}\left(\Phi\left((\mathbf{X}_{-u}^{(k_N(i)[(\mathbf{X}_{-u}^{(n)})_n])})_{i\le N_I}, (\mathbf{X}_u^{(k_N(i)[(\mathbf{X}_{-u}^{(n)})_n])})_{i\le N_I} \right) \middle| (\mathbf{X}_{-u}^{(n)})_n = (\mathbf{x}_{-u}^{(n)})_n \right) \right. \\
& \left. -\mathrm{E}\left(\Phi(\mathbf{Z}) | \, \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)} \right) \right| \\
= & \left| \mathrm{E}\left(\Phi\left((\mathbf{x}_{-u}^{(k_N(i)[(\mathbf{x}_{-u}^{(n)})_n])})_{i\le N_I}, (\mathbf{X}_u^{(i)})_{i\le N_I} \right) \middle| (\mathbf{X}_{-u}^{(i)})_{i\le N_I} = (\mathbf{x}_{-u}^{(k_N(i)[(\mathbf{x}_{-u}^{(n)})_n])})_{i\le N_I} \right) \right. \\
& \left. -\mathrm{E}\left(\Phi\left((\mathbf{x}_{-u}^{(1)})_{i\le N_I}, (\mathbf{X}_u^{(i)})_{i\le N_I} \right) \middle| (\mathbf{X}_{-u}^{(i)})_{i\le N_I} = (\mathbf{x}_{-u}^{(1)})_{i\le N_I} \right) \right| \\
\le & C_{\sup} d\left(\mathbf{x}_{-u}^{(k_N(N_I)[(\mathbf{x}_{-u}^{(n)})_n])}, \mathbf{x}_{-u}^{(1)} \right),
\end{aligned}$$

thanks to Lemma 11. Thus, using Lemma 13, for all $\varepsilon > 0$,

$$\begin{aligned}
\left| \mathrm{E}\left(\widehat{E}_{u,1} | \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)} \right) - \mathrm{Var}(Y | \mathbf{X}_{-u} = \mathbf{x}_{-u}^{(1)}) \right| &\le C_{\sup} \mathrm{E}\left(d\left(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))} \right) \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)} \right) \\
&\le C_{\sup} \frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}.
\end{aligned}$$

□

In the following, to simplify notation, we may write "$\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}$" for "$\mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)}$ and $\mathbf{X}_{-u}^{(2)} = \mathbf{x}_{-u}^{(2)}$".

Lemma 14. *For almost all $(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)})$ and for all $a \ge 0$, we have*

$$\mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}) \ge a \middle| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)} \right) \le \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_{N-1}(N_I))}) \ge a \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)} \right),$$

and thus, integrating a on $\mathbb{R}_+$,

$$\mathrm{E}\left(d\left(\mathbf{X}_{-u}^{(k_N(N_I))}, \mathbf{X}_{-u}^{(1)} \right) \middle| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)} \right) \le \mathrm{E}\left(d\left(\mathbf{X}_{-u}^{(k_{N-1}(N_I))}, \mathbf{X}_{-u}^{(1)} \right) \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)} \right).$$

Proof. Let $g_N(i)$ be the index of the i-th nearest neighbour of $\mathbf{X}_{-u}^{(1)}$ in $(\mathbf{X}_{-u}^{(n)})_{n\in[1:N]\setminus\{2\}}$. For almost all $(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)})$, we have

$$\begin{aligned}
& \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}) \ge a \middle| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)} \right) \\
= & \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}) \ge a \middle| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) > d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \right) \\
& \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) > d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \middle| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)} \right) \\
& + \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}) \ge a \middle| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \le d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \right) \\
& \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \le d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \middle| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)} \right).
\end{aligned}$$

Moreover, conditionally to $\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}$, if $d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) > d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))})$, then the N_I-nearest neighbours of $\mathbf{X}_{-u}^{(1)}$ do not change if we do not take into account $\mathbf{X}_{-u}^{(2)}$. Thus

$$\begin{aligned}
& \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}) \geq a \middle| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) > d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \right) \\
= \quad & \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \geq a \middle| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) > d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \right) \\
= \quad & \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \geq a \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)}, d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) > d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \right).
\end{aligned}$$

Similarly, conditionally to $\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}$, if $d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \leq d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}$, then $\mathbf{x}_{-u}^{(2)}$ is one of the N_I-nearest neighbours of $\mathbf{X}_{-u}^{(1)}$. Thus

$$\begin{aligned}
& \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}) \geq a \middle| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \leq d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \right) \\
\leq \quad & \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \geq a \middle| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \leq d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \right) \\
= \quad & \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \geq a \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)}, d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \leq d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \right).
\end{aligned}$$

Finally,

$$\begin{aligned}
& \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_N(N_I))}) \geq a \middle| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)} \right) \\
\leq \quad & \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \geq a \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)}, d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) > d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \right) \\
& \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) > d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)} \right) \\
& + \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \geq a \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)}, d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \leq d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \right) \\
& \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \leq d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)} \right) \\
= \quad & \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(g_N(N_I))}) \geq a \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)} \right) \\
= \quad & \mathbb{P}\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_{N-1}(N_I))}) \geq a \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)} \right),
\end{aligned}$$

and we proved Lemma 14. □

Proposition 13. *For all $\varepsilon > 0$, there exists $C_{\sup}(\varepsilon)$ such that*

$$\left| cov(\widehat{E}_{u,1}, \widehat{E}_{u,2}) \right| \leq \frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}. \tag{43}$$

Proof. We use the law of total covariance,

$$cov(\widehat{E}_1, \widehat{E}_2) = \mathrm{E}\left[cov\left(\widehat{E}_1, \widehat{E}_2 \middle| \mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)} \right) \right] + cov\left[\mathrm{E}\left(\widehat{E}_{u,1} | \mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)} \right), \mathrm{E}\left(\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)} \right) \right]. \tag{44}$$

Part 1: First, we will bound the second term of Equation 44. Thanks to Lemma 11, we have

$$\begin{aligned}
&\left|\mathrm{E}\left(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)}=\mathbf{x}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}=\mathbf{x}_{-u}^{(2)}\right)-\mathrm{Var}\left(Y|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(1)})\right)\right| \\
\leq\ & \mathrm{E}\left\{\left|\mathrm{E}\left[\Phi\left((\mathbf{X}^{(k_N(i))})_{i\leq N_I}\right)\middle|\mathbf{X}_{-u}^{(1)}=\mathbf{x}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}=\mathbf{x}_{-u}^{(2)},(\mathbf{X}_{-u}^{(n)})_{n\geq 3}\right]-\mathrm{Var}\left(Y|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(1)})\right)\right|\right\} \\
\leq\ & C_{\sup}\mathrm{E}\left(d\left(\mathbf{X}_{-u}^{(1)},\mathbf{X}_{-u}^{(k_N(N_I))}\right)\middle|\mathbf{X}_{-u}^{(1)}=\mathbf{x}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}=\mathbf{x}_{-u}^{(2)}\right) \quad \text{using Lemma 11,} \\
\leq\ & C_{\sup}\mathrm{E}\left(d\left(\mathbf{X}_{-u}^{(1)},\mathbf{X}_{-u}^{(k_{N-1}(N_I))}\right)\middle|\mathbf{X}_{-u}^{(1)}=\mathbf{x}_{-u}^{(1)}\right) \quad \text{using Lemma 14,} \\
\leq\ & \frac{C_{\sup}(\varepsilon)}{(N-1)^{\frac{1}{p-|u|}-\varepsilon}} \quad \text{using Lemma 13,} \\
\leq\ & \frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}.
\end{aligned}$$

Similarly,

$$\left|\mathrm{E}\left(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)}=\mathbf{x}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}=\mathbf{x}_{-u}^{(2)}\right)-\mathrm{Var}\left(Y|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(2)})\right)\right|\leq\frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}.$$

Thus, using that Φ is bounded,

$$\begin{aligned}
&\left|\mathrm{E}\left(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)}=\mathbf{x}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}=\mathbf{x}_{-u}^{(2)}\right)\mathrm{E}\left(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)}=\mathbf{x}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}=\mathbf{x}_{-u}^{(2)}\right)\right. \\
&\qquad\left.-\mathrm{Var}\left(Y|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(1)}\right)\mathrm{Var}\left(Y|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(2)})\right)\right|\leq\frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}.
\end{aligned}$$

Moreover, using Proposition 12, we have

$$\begin{aligned}
&\left|\mathrm{E}\left(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)}=\mathbf{x}_{-u}^{(1)}\right)\mathrm{E}\left(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(2)}=\mathbf{x}_{-u}^{(2)}\right)\right. \\
&\left.-\mathrm{Var}\left(Y|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(1)})\right)\mathrm{Var}\left(Y|\mathbf{X}_{-u}=\mathbf{x}_{-u}^{(2)})\right)\right|\leq\frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\left|\mathrm{E}\left(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)}=\mathbf{x}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}=\mathbf{x}_{-u}^{(2)}\right)\mathrm{E}\left(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)}=\mathbf{x}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}=\mathbf{x}_{-u}^{(2)}\right)\right. \\
&\qquad\left.-\mathrm{E}\left(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)}=\mathbf{x}_{-u}^{(1)}\right)\mathrm{E}\left(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(2)}=\mathbf{x}_{-u}^{(2)}\right)\right| \leq\frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}.
\end{aligned}$$

Finally,

$$\begin{aligned}
&\left|cov\left[\mathrm{E}\left(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}\right),\mathrm{E}\left(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}\right)\right]\right| \\
=\ & \left|\mathrm{E}\left[\mathrm{E}\left(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}\right)\mathrm{E}\left(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}\right)\right]-\mathrm{E}\left[\mathrm{E}\left(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)}\right)\mathrm{E}\left(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(2)}\right)\right]\right| \\
\leq\ & \mathrm{E}\left[\left|\mathrm{E}\left(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}\right)\mathrm{E}\left(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1)},\mathbf{X}_{-u}^{(2)}\right)-\mathrm{E}\left(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1)}\right)\mathrm{E}\left(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(2)}\right)\right|\right] \\
\leq\ & \frac{C_{\sup}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}.
\end{aligned}$$

Remark 22. *In this Part 1, we can choose $\varepsilon = 0$ for the estimators $\widehat{V}_{u,PF}$ or for $\widehat{E}_{u,MC}$ if we take $N_I = 2$.*

Part 2: Let $\varepsilon > 0$. We will bound the first term of Equation 44: $\mathrm{E}\left[cov\left(\widehat{E}_1, \widehat{E}_2 \middle| \mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(2)}\right)\right]$. We want to prove that

$$\left| \int_{\mathcal{X}_{-u}^2} \mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right) - \mathrm{E}(\widehat{E}_{u,1} | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}) \mathrm{E}(\widehat{E}_{u,2} | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}) d\mathbb{P}_{\mathbf{X}_{-u}}^{\otimes 2}(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \right| \leq \frac{C_{\sup}(\varepsilon}{N^{1-\varepsilon}}$$

Let us write

$$l(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) := \min\left(d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)})/2,\ \frac{1}{N^{\frac{1}{|-u|} - \delta}} \right)$$

where $\delta = \varepsilon/(4|-u|)$, and

$$G(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) := \Big\{ (\mathbf{x}_{-u}^{(n)})_{n \in [3:N]} \big| \quad d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(k_N(N_I)[(\mathbf{x}_{-u}^{(n)})_{n \leq N}])}) < l(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}),$$
$$d(\mathbf{x}_{-u}^{(2)}, \mathbf{x}_{-u}^{(k'_N(N_I)[(\mathbf{x}_{-u}^{(n)})_{n \leq N}])}) < l(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \Big\}.$$

Part 2.A: We prove the following lemmas.

Lemma 15. *For all $\varepsilon > 0$, there exists $C_{\sup}(\varepsilon)$ such that,*

$$\int_{\mathcal{X}_{-u}^2} \mathbb{P}\left(d(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_{N-1}(N_I))}) \geq d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)})/2 \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)} \right) d\mathbb{P}_{\mathbf{X}_{-u}}^{\otimes 2}(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \leq \frac{C_{\sup}(\varepsilon)}{N^{1-\varepsilon}}. \tag{45}$$

Proof. We divide $\mathcal{X}_{-u}^2$ in $F_1 := \{(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \in \mathcal{X}_{-u}^2,\ d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) < (N - N_I - 1)^{\frac{-1+\varepsilon}{|-u|}}\}$ and $F_2 := \{(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \in \mathcal{X}_{-u}^2,\ d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \geq (N - N_I - 1)^{\frac{-1+\varepsilon}{|-u|}}\}$.

$$\begin{aligned}
& \int_{F_1} \mathbb{P}\left(d(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_{N-1}(N_I))}) \geq d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)})/2 \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)} \right) d\mathbb{P}_{\mathbf{X}_{-u}}^{\otimes 2}(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \\
\leq\ & C_{\sup} \lambda_{|-u|}^{\otimes 2}(F_1) \\
\leq\ & C_{\sup} \int_{\mathcal{X}_{-u}} \lambda_{|-u|}\left(B\left[\mathbf{x}_{-u}, (N - N_I - 1)^{\frac{-1+\varepsilon}{|-u|}} \right] \right) d\mathbf{x}_{-u} \\
\leq\ & C_{\sup} \int_{\mathcal{X}_{-u}} (N - N_I - 1)^{\frac{-1+\varepsilon}{|-u|}|-u|} dx_{-u} \\
\leq\ & C_{\sup} (N - N_I - 1)^{-1+\varepsilon} \\
\leq\ & \frac{C_{\sup}}{N^{1-\varepsilon}}.
\end{aligned}$$

Furthermore, using Lemma 12, we have

$$\begin{aligned}
&\int_{F_2} \mathbb{P}\left(d(\mathbf{X}_{-u}^{(1)}, \mathbf{X}_{-u}^{(k_{N-1}(N_I))}) \geq d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)})/2 \middle| \mathbf{X}_{-u}^{(1)} = \mathbf{x}_{-u}^{(1)} \right) d\mathbb{P}_{\mathbf{X}_{-u}}^{\otimes 2}(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \\
\leq &\int_{F_2} C_{\sup}(N-1)^{N_I}(1 - C_{\inf} d(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)})^{|-u|})^{N-1-N_I} d\mathbb{P}_{\mathbf{X}_{-u}}^{\otimes 2}(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \\
\leq &\lambda_{|-u|}(\mathcal{X}_{-u})^2 C_{\sup}(N-1)^{N_I}(1 - C_{\inf}(N - N_I - 1)^{\frac{-1+\varepsilon}{|-u|}|-u|})^{N-1-N_I} \\
\leq &C_{\sup}(N-1)^{N_I}(1 - C_{\inf}(N - N_I - 1)^{-1+\varepsilon})^{N-1-N_I} \\
\leq &C_{\sup}(N-1)^{N_I} \exp\left[(N-1-N_I)\ln\left(1 - C_{\inf}(N - N_I - 1)^{-1+\varepsilon}\right)\right] \\
\leq &C_{\sup}(N-1)^{N_I} \exp\left[-C_{\inf}(N - N_I - 1)^{\varepsilon} + o((N - N_I - 1)^{\varepsilon})\right] \\
\leq &\frac{C_{\sup}(\varepsilon)}{N^{1-\varepsilon}}.
\end{aligned}$$

□

Remark 23. *In Lemma 15, we need $\varepsilon > 0$ even for the Pick-and-Freeze estimators. That explains the rate of convergence when $|u| = 1$ for the Pick-and-Freeze estimators.*

Lemma 16. *For all $\varepsilon > 0$, there exists $C_{\sup}(\varepsilon)$ such that,*

$$\int_{\mathcal{X}_{-u}^2} \mathbb{P}_{\mathbf{X}_{-u}}^{\otimes(N-2)}(G(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)})^c) d\mathbb{P}_{\mathbf{X}_{-u}}^{\otimes 2}(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \leq \frac{C_{\sup}(\varepsilon)}{N^{1-\varepsilon}}. \tag{46}$$

Proof. Using Lemma 12, we have

$$\mathbb{P}\left(d(\mathbf{X}_{-u}^{(k_{N-1}(N_I))}, \mathbf{x}_{-u}^{(1)}) \geq N^{-\frac{1}{|-u|}+\delta} | \mathbf{X}_{-u}^{(1)} \right) \leq C_{\sup}(N-1)^{N_I}(1 - C_{\inf} N^{-1+\delta|-u|})^{N-1-N_I},$$

so

$$\mathbb{P}\left(d(\mathbf{X}_{-u}^{(k_{N-1}(N_I))}, \mathbf{x}_{-u}^{(1)}) \geq N^{-\frac{1}{|-u|}+\delta} | \mathbf{X}_{-u}^{(1)} \right) \leq \frac{C_{\sup}(\varepsilon)}{N}. \tag{47}$$

Thus, we have

$$
\begin{aligned}
&\int_{\mathcal{X}^2_{-u}} \mathbb{P}^{\otimes(N-2)}_{\mathbf{X}_{-u}}(G(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u})^c) d\mathbb{P}^{\otimes 2}_{\mathbf{X}_{-u}}(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u}) \\
\leq\ &\int_{\mathcal{X}^2_{-u}} \mathbb{P}\left(d(\mathbf{X}^{(1)}_{-u},\mathbf{X}^{(k_N(N_I))}_{-u}) \geq d(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u})/2 \middle| \mathbf{X}^{(1,2)}_{-u} = \mathbf{x}^{(1,2)}_{-u} \right) d\mathbb{P}^{\otimes 2}_{\mathbf{X}_{-u}}(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u}) \\
&+\int_{\mathcal{X}^2_{-u}} \mathbb{P}\left(d(\mathbf{X}^{(2)}_{-u},\mathbf{X}^{(k'_N(N_I))}_{-u}) \geq d(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u})/2 \middle| \mathbf{X}^{(1,2)}_{-u} = \mathbf{x}^{(1,2)}_{-u} \right) d\mathbb{P}^{\otimes 2}_{\mathbf{X}_{-u}}(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u}) \\
&+\int_{\mathcal{X}^2_{-u}} \mathbb{P}\left(d(\mathbf{X}^{(1)}_{-u},\mathbf{X}^{(k_N(N_I))}_{-u}) \geq N^{-\frac{1}{|-u|}+\delta} \middle| \mathbf{X}^{(1,2)}_{-u} = \mathbf{x}^{(1,2)}_{-u} \right) d\mathbb{P}^{\otimes 2}_{\mathbf{X}_{-u}}(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u}) \\
&+\int_{\mathcal{X}^2_{-u}} \mathbb{P}\left(d(\mathbf{X}^{(2)}_{-u},\mathbf{X}^{(k'_N(N_I))}_{-u}) \geq N^{-\frac{1}{|-u|}+\delta} \middle| \mathbf{X}^{(1,2)}_{-u} = \mathbf{x}^{(1,2)}_{-u} \right) d\mathbb{P}^{\otimes 2}_{\mathbf{X}_{-u}}(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u}) \\
\leq\ &\int_{\mathcal{X}^2_{-u}} \mathbb{P}\left(d(\mathbf{X}^{(1)}_{-u},\mathbf{X}^{(k_{N-1}(N_I))}_{-u}) \geq d(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u})/2 \middle| \mathbf{X}^{(1)}_{-u} = \mathbf{x}^{(1)}_{-u} \right) d\mathbb{P}^{\otimes 2}_{\mathbf{X}_{-u}}(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u}) \\
&+\int_{\mathcal{X}^2_{-u}} \mathbb{P}\left(d(\mathbf{X}^{(2)}_{-u},\mathbf{X}^{(k'_{N-1}(N_I))}_{-u}) \geq d(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u})/2 \middle| \mathbf{X}^{(2)}_{-u} = \mathbf{x}^{(2)}_{-u} \right) d\mathbb{P}^{\otimes 2}_{\mathbf{X}_{-u}}(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u}) \\
&+\int_{\mathcal{X}^2_{-u}} \mathbb{P}\left(d(\mathbf{X}^{(1)}_{-u},\mathbf{X}^{(k_{N-1}(N_I))}_{-u}) \geq N^{-\frac{1}{|-u|}+\delta} \middle| \mathbf{X}^{(1)}_{-u} = \mathbf{x}^{(1)}_{-u} \right) d\mathbb{P}^{\otimes 2}_{\mathbf{X}_{-u}}(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u}) \\
&+\int_{\mathcal{X}^2_{-u}} \mathbb{P}\left(d(\mathbf{X}^{(2)}_{-u},\mathbf{X}^{(k'_{N-1}(N_I))}_{-u}) \geq N^{-\frac{1}{|-u|}+\delta} \middle| \mathbf{X}^{(2)}_{-u} = \mathbf{x}^{(2)}_{-u} \right) d\mathbb{P}^{\otimes 2}_{\mathbf{X}_{-u}}(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u}),
\end{aligned}
$$

and we conclude the proof of Lemma 16 using Lemma 15 and Equation 47. □

For $i = 1, 2$, let B_i be the ball of center $\mathbf{x}^{(i)}_{-u}$ and of radius $l(\mathbf{x}^{(1)}_{-u},\mathbf{x}^{(2)}_{-u})$, let p_i be the probability of B_i and N_i be the number of observations $(\mathbf{X}^{(n)}_{-u})_{n\in[3:N]}$ in the ball B_i. Remark that

$$p_i \leq \frac{C_{\sup}}{N^{1-\delta|-u|}}.$$

We have the two following lemmas.

Lemma 17. *Conditionally to $\mathbf{X}^{(1,2)}_{-u} = \mathbf{x}^{(1,2)}_{-u}$, the random variable N_i is binomial $\mathcal{B}(N-2, p_i)$.*
Conditionally to $\mathbf{X}^{(1,2)}_{-u} = \mathbf{x}^{(1,2)}_{-u}, N_j = n_j$, the random variable N_i is binomial $\mathcal{B}(N-2-n_j, p_i(1-p_j)^{-1})$.

Proof. For the first assertion, we use that the $(\mathbf{X}^{(n)}_{-u})_n$ are i.i.d. For the second assertion, we compute $\mathbb{P}(N_i = n_i | \mathbf{X}^{(1,2)}_{-u} = \mathbf{x}^{(1,2)}_{-u}, N_j = n_j)$ with Bayes' theorem. □

Lemma 18. *If $N_i = n_i$, let $\mathbf{X}^{(\mathbf{M}_i)}_{-u}$ be the random vector composed of the n_i observations in B_i of $(\mathbf{X}^{(n)}_{-u})_{n\in[3:N]}$ and $\mathbf{M}_i \in [3:N]^{n_i}$ the vector containing*

the corresponding indices. We have:

$$\begin{aligned} & \mathcal{L}\left(\mathbf{X}^{(\mathbf{M}_1)}, \mathbf{X}^{(\mathbf{M}_2)} | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, N_1 = n_1, N_2 = n_2\right) \\ = \quad & \mathcal{L}\left(\mathbf{X}^{(\mathbf{M}_1)} | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, N_1 = n_1\right) \otimes \mathcal{L}\left(\mathbf{X}^{(\mathbf{M}_2)} | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, N_2 = n_2\right). \end{aligned}$$

Proof. For any bounded Borel functions $\phi_1,\ \phi_2$, we have

$$\begin{aligned} & \mathrm{E}\left(\phi_1(\mathbf{X}^{(\mathbf{M}_1)})\phi_2(\mathbf{X}^{(\mathbf{M}_2)}) | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, N_1 = n_1, N_2 = n_2\right) \\ = \quad & \frac{\mathrm{E}\left(\phi_1(\mathbf{X}^{(\mathbf{M}_1)})\phi_2(\mathbf{X}^{(\mathbf{M}_2)}) \mathbb{1}_{N_1=n_1} \mathbb{1}_{N_2=n_2} | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right)}{\mathbb{P}(N_1 = n_1, N_2 = n_2 | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)})}. \end{aligned}$$

Let

$$\mathcal{P}([3:N], n_1) := \{(k_1, \cdots, k_{n_1}) \in [3:N]^{n_1} \ : \ k_i < k_j \text{ for } i, j \in [1:n_1],\ i < j\}$$

be the set of all possible two-by-two distinct elements in $[3:N]$. To simplify notation, we also consider an element of $\mathcal{P}([3:N], n_1)$ with the subset of $[3:N]$ that contains its indices. We have

$$\begin{aligned} & \mathrm{E}\left(\phi_1(\mathbf{X}^{(\mathbf{M}_1)})\phi_2(\mathbf{X}^{(\mathbf{M}_2)}) \mathbb{1}_{N_1=n_1} \mathbb{1}_{N_2=n_2} \Big| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right) \\ = \quad & \sum_{\mathbf{m}_1 \in \mathcal{P}([3:N], n_1)} \sum_{\mathbf{m}_2 \in \mathcal{P}([3:N] \setminus \mathbf{m}_1, n_2)} \mathrm{E}\Big(\phi_1(\mathbf{X}^{(\mathbf{m}_1)})\phi_2(\mathbf{X}^{(\mathbf{m}_2)}) \mathbb{1}_{\mathbf{X}_{-u}^{(\mathbf{m}_1)} \in B_1^{n_1}} \mathbb{1}_{\mathbf{X}_{-u}^{(\mathbf{m}_2)} \in B_2^{n_2}} \\ & \times \prod_{i \in [3:N] \setminus (\mathbf{m}_1 \cup \mathbf{m}_2)} \mathbb{1}_{\mathbf{X}_{-u}^{(i)} \notin B_1 \cup B_2} \Big| \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\Big) \end{aligned}$$

Now, using the independence of $(\mathbf{X}^{(n)})_n$ and summing over $\mathbf{m}_1$ and $\mathbf{m}_2$, we

have, for any value of $\mathbf{m}_1 \in \mathcal{P}([3:N], n_1)$ and $\mathbf{m}_2 \in \mathcal{P}([3:N], n_2)$,

$$
\begin{aligned}
& \mathrm{E}\left(\phi_1(\mathbf{X}^{(\mathbf{M}_1)})\phi_2(\mathbf{X}^{(\mathbf{M}_2)})|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, N_1 = n_1, N_2 = n_2\right) \\
= \quad & \binom{N-2}{n_1}\binom{N-2-n_1}{n_2}(1-p_1-p_2)^{N-2-n_1-n_2} \\
& \frac{\mathrm{E}\left(\phi_1(\mathbf{X}^{(\mathbf{m}_1)})\mathbb{1}_{\mathbf{X}_{-u}^{(\mathbf{m}_1)}\in B_1^{n_1}}\middle|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right)\mathrm{E}\left(\phi_2(\mathbf{X}^{(\mathbf{m}_2)})\mathbb{1}_{\mathbf{X}_{-u}^{(\mathbf{m}_2)}\in B_2^{n_2}}\middle|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right)}{\binom{N-2}{n_1}\binom{N-2-n_1}{n_2}p_1^{n_1}p_2^{n_2}(1-p_1-p_2)^{N-2-n_1-n_2}} \\
= \quad & \frac{\mathrm{E}\left(\phi_1(\mathbf{X}^{(\mathbf{m}_1)})\mathbb{1}_{\mathbf{X}_{-u}^{(\mathbf{m}_1)}\in B_1^{n_1}}\middle|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right)}{p_1^{n_1}}\frac{\mathrm{E}\left(\phi_2(\mathbf{X}^{(\mathbf{m}_2)})\mathbb{1}_{\mathbf{X}_{-u}^{(\mathbf{m}_2)}\in B_2^{n_2}}\middle|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right)}{p_2^{n_2}} \\
= \quad & \frac{\mathrm{E}\left(\phi_1(\mathbf{X}^{(\mathbf{m}_1)})\mathbb{1}_{\mathbf{X}_{-u}^{(\mathbf{m}_1)}\in B_1^{n_1}}\prod_{i\in[3:N]\setminus\mathbf{m}_1}\mathbb{1}_{\mathbf{X}_{-u}^{(i)}\notin B_1}\middle|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right)}{p_1^{n_1}(1-p_1)^{n_1}} \\
& \frac{\mathrm{E}\left(\phi_2(\mathbf{X}^{(\mathbf{m}_2)})\mathbb{1}_{\mathbf{X}_{-u}^{(\mathbf{m}_2)}\in B_2^{n_2}}\prod_{i\in[3:N]\setminus\mathbf{m}_2}\mathbb{1}_{\mathbf{X}_{-u}^{(i)}\notin B_2}\middle|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right)}{p_2^{n_2}(1-p_2)^{n_2}} \\
= \quad & \mathrm{E}\left(\phi_1(\mathbf{X}^{(\mathbf{M}_1)})\middle|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, N_1 = n_1\right)\mathrm{E}\left(\phi_2(\mathbf{X}^{(\mathbf{M}_2)})\middle|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}, N_2 = n_2\right).
\end{aligned}
$$

That concludes the proof of Lemma 18. □

Part 2.B: We aim to proving that

$$
\left|\int_{\mathcal{X}_{-u}^2}\mathrm{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right) - \mathrm{E}(\widehat{E}_{u,1}|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)})\mathrm{E}(\widehat{E}_{u,2}|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)})d\mathbb{P}_{\mathbf{X}_{-u}}^{\otimes 2}(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)})\right| \leq \frac{C_{\sup}(\varepsilon}{N^{1-\varepsilon}}
$$

To simplify notation, let $\mathbf{X}^{(\mathbf{k}_N)} := (\mathbf{X}^{(k_N(i))})_{i\leq N_I}$ and $\mathbf{X}^{(\mathbf{k}'_N)} := (\mathbf{X}^{(k'_N(i))})_{i\leq N_I}$.
We have

$$
\begin{aligned}
& \mathrm{E}\left(\Phi(\mathbf{X}^{(\mathbf{k}_N)})\Phi(\mathbf{X}^{(\mathbf{k}'_N)})|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right) \\
= \quad & \sum_{n_1,n_2=0}^{N-2}\mathrm{E}\left(\Phi(\mathbf{X}^{(\mathbf{k}_N)})|N_1 = n_1, \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right)\mathrm{E}\left(\Phi(\mathbf{X}^{(\mathbf{k}'_N)})|N_2 = n_2, \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right) \\
& \times\mathbb{P}(N_1 = n_1, N_2 = n_2|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}).
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
& \mathrm{E}\left(\Phi(\mathbf{X}^{(\mathbf{k}_N)})|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right)\mathrm{E}\left(\Phi(\mathbf{X}^{(\mathbf{k}'_N)})|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right) \\
= \quad & \sum_{n_1,n_2=0}^{N-2}\mathrm{E}\left(\Phi(\mathbf{X}^{(\mathbf{k}_N)})|N_1 = n_1, \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right)\mathrm{E}\left(\Phi(\mathbf{X}^{(\mathbf{k}'_N)})|N_2 = n_2, \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}\right) \\
& \times\mathbb{P}(N_1 = n_1|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)})\mathbb{P}(N_2 = n_2|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}).
\end{aligned}
$$

Thus, using that Φ is bounded and using Lemma 16, it suffices to show that

$$\sum_{n_1,n_2=N_I-1}^{N-2} \Big|\mathbb{P}(N_1=n_1,N_2=n_2|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)}) - \mathbb{P}(N_1=n_1|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})\mathbb{P}(N_2=n_2|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})\Big| \leq \frac{C_{\sup}(\varepsilon)}{N^{1-\varepsilon}}.$$

Let $K_N := \lfloor N^\alpha \rfloor$, where $\alpha = \varepsilon/3$. We divide the previous sum into two sums:

$$A(\mathbf{x}_{-u}^{(1)},\mathbf{x}_{-u}^{(2)}) := \sum_{n_1,n_2=N_I-2}^{K_N} \Big|\mathbb{P}(N_1=n_1,N_2=n_2|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)}) - \mathbb{P}(N_1=n_1|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})\mathbb{P}(N_2=n_2|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})\Big|,$$

$$B(\mathbf{x}_{-u}^{(1)},\mathbf{x}_{-u}^{(2)}) := \sum_{\substack{n_1,n_2=N_I-1,\\ n_1>K_N \text{ or } n_2>K_N}}^{N-2} \Big|\mathbb{P}(N_1=n_1,N_2=n_2|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)}) - \mathbb{P}(N_1=n_1|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})\mathbb{P}(N_2=n_2|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})\Big|.$$

Let us bound these two terms.

First, we have

$$A(\mathbf{x}_{-u}^{(1)},\mathbf{x}_{-u}^{(2)}) = \sum_{n_1,n_2=N_I-1}^{K_N} \mathbb{P}(N_1=n_1|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})\mathbb{P}(N_2=n_2|N_1=n_1,\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)}) \times \left|1-\frac{\mathbb{P}(N_2=n_2|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})}{\mathbb{P}(N_2=n_2|N_1=n_1,\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})}\right|.$$

Thus, it suffices to bound

$$\left|1-\frac{\mathbb{P}(N_2=n_2|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})}{\mathbb{P}(N_2=n_2|N_1=n_1,\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})}\right| \leq \frac{C_{\sup}(\varepsilon)}{N^{1-\varepsilon}}.$$

Thus, it suffices to show

$$\left|\log\left(\frac{\mathbb{P}(N_2=n_2|\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})}{\mathbb{P}(N_2=n_2|N_1=n_1,\mathbf{X}_{-u}^{(1,2)}=\mathbf{x}_{-u}^{(1,2)})}\right)\right| \leq \frac{C_{\sup}(\varepsilon)}{N^{1-\varepsilon}}.$$

To simplify notation, let $T = N - 2$. Thanks to Lemma 17, we have,

$$\begin{aligned}
& \log\left(\frac{\mathbb{P}(N_2 = n_2|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)})}{\mathbb{P}(N_2 = n_2|N_1 = n_1, \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)})}\right) \\
= & \log\left(\frac{T(T-1)...(T-n_1+1)}{(T-n_2)(T-n_2-1)...(T-n_2-n_1+1)}\frac{(1-p_1)^{T-n_1}(1-p_2)^{T-n_2}}{(1-p_1-p_2)^{T-n_1-n_2}}\right) \\
= & \log\left(1(1-\frac{1}{T})...(1-\frac{n_1-1}{T})\right) - \log\left((1-\frac{n_2}{T})(1-\frac{n_2+1}{T})...(1-\frac{n_2+n_1-1}{T})\right) \\
& (T-n_1)\log(1-p_1) + (T-n_2)\log(1-p_2) - (T-n_1-n_2)\log(1-p_1-p_2) \\
= & -\frac{n_1(n_1-1)}{2T} + n_1 O(\frac{n_1^2}{T^2}) + \frac{n_1(n_1+2n_2-1)}{2T} + n_1 O(\frac{(n_1+n_2)^2}{T^2}) \\
& -(T-n_2)p_2 + (T-n_2)O(p_2^2) - (T-n_1)p_1 + (T-n_1)O(p_1^2) \\
& +(T-n_1-n_2)(p_1+p_2) + (T-n_1-n_2)O((p_1+p_2)^2) \\
= & \frac{n_1 n_2}{T} + O(\frac{n_1^3}{T}) + O(\frac{n_1(n_1+n_2)^2}{T^2}) - n_2 p_1 - n_1 p_2 \\
& +(T-n_2)O(p_1^2) + (T-n_1)O(p_2^2) + (T-n_1-n_2)O((p_1+p_2)^2).
\end{aligned}$$

We know that

$$K_N p_i \leq \frac{C_{\sup}}{N^{1-\delta|-u|-\alpha}} \leq \frac{C_{\sup}}{N^{1-\varepsilon}}.$$

So, for all $n_1 \leq K_N$ and all $n_2 \leq K_N$,

$$\left|\log\left(\frac{\mathbb{P}(N_2 = n_2|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)})}{\mathbb{P}(N_2 = n_2|N_1 = n_1, \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)})}\right)\right| \leq \frac{C_{\sup}(\varepsilon)}{N^{1-\varepsilon}}.$$

Thus, we have shown that we have

$$A(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \leq \frac{C_{\sup}}{N^{1-\varepsilon}}.$$

Now, let us bound $B(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)})$. Remark that $\{(n_1, n_2) \in [N_I - 1 : N-2] |\ n_1 > K_N \text{ or } n_2 > K_N\}$ is a subset of

$$([K_N + 1 : N-2] \times [N_I - 1 : N-2]) \cup ([N_I - 1 : N-2] \times [K_N + 1 : N-2]).$$

Thus, it suffices to bound

$$\begin{aligned}
\sum_{n_1=K_N+1}^{N-2}\sum_{n_2=N_I-1}^{N-2} & \left|\mathbb{P}(N_1 = n_1, N_2 = n_2|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)})\right. \\
& \left.-\mathbb{P}(N_1 = n_1|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)})\mathbb{P}(N_2 = n_2|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)})\right| \\
= & \sum_{n_1=K_N+1}^{N-2} \mathbb{P}(N_1 = n_1|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}) \\
& \sum_{n_2=N_I-1}^{N-2} \left|\mathbb{P}(N_2 = n_2|N_1 = n_1, \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}) - \mathbb{P}(N_2 = n_2|\mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)})\right|.
\end{aligned}$$

Thus, it suffices to bound

$$\sum_{n_1=K_N+1}^{N-2} \mathbb{P}(N_1 = n_1 | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}).$$

Let $T := N-2$. We know that N_1 has a binomial distribution with parameters T and p_1. Thus,

$$\mathrm{E}(N_1) = p_1 T \leq C_{\sup} N^{\delta|-u|} \leq C_{\sup} N^{\frac{\varepsilon}{4}}.$$

Thus, there exists N_ε such that for $N \geq N_\varepsilon$, we have that, $\mathrm{E}(N_1) \leq K_T + 1$. Thus, for N large enough and for all $n_1 > K_T$ and, we have

$$\mathbb{P}(N_1 = n_1 | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}) \leq \mathbb{P}(N_1 = K_T + 1 | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}).$$

Thus, for $N \geq N_\varepsilon$,

$$\begin{aligned}
&\sum_{n_1=K_N+1}^{N-2} \mathbb{P}(N_1 = n_1 | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}) \\
\leq\ & (T-K_T)\mathbb{P}(N_1 = K_T+1 | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}) \\
=\ & (T-K_T)\frac{T!}{(T-K_T-1)!(K_T+1)!} p_1^{K_T+1}(1-p_1)^{T-K_T+1} \\
\leq\ & (T-K_T)\frac{T!}{(T-K_T-1)!(K_T+1)!} p_1^{K_T+1} \\
\leq\ & C_{\sup}\frac{(T-K_T)\sqrt{2\pi T}\left(\frac{T}{e}\right)^T \left(\frac{C_{\sup}}{T^{1-\delta|-u|}}\right)^{K_T+1}}{\sqrt{2\pi(K_T+1)}\left(\frac{K_T+1}{e}\right)^{(K_T+1)}\sqrt{2\pi(T-K_T-1)}\left(\frac{T-K_T-1}{e}\right)^{(T-K_T-1)}} \\
\leq\ & C_{\sup}\frac{(T-K_T)\sqrt{T}T^T C_{\sup}^{K_T+1}}{\sqrt{(K_T+1)(T-K_T-1)}(K_T+1)^{K_T+1}(T-K_T-1)^{T-K_T-1}T^{(1-\delta|-u|)(K_T+1)}} \\
\leq\ & C_{\sup}(T-K_T)^{K_T+\frac{3}{2}-T}(K_T+1)^{-K_T-\frac{3}{2}}T^{T-\frac{1}{2}+\delta|-u|(K_T+1)-K_T}C_{\sup}^{K_T+1}.
\end{aligned}$$

Using the Taylor expansion of $x \mapsto \log(1-x)$ at 0, we can see that

$$(T-K_T)^{-T}T^T \leq C_{\sup}\exp(K_T) \leq C_{\sup}^{K_T}.$$

Moreover, we have

$$(K_T+1)T^{1-\delta|-u|} \geq T^{\frac{\varepsilon}{3}}T^{1-\frac{\varepsilon}{4}} = T^{1+\frac{\varepsilon}{12}},$$

and so

$$\begin{aligned}
(T-K_T)^{K_T}(K_T+1)^{-K_T}T^{-K_T(1-\delta|-u|)}C_{\sup}^{K_T} &\leq \exp\left(K_T \log\left[C_{\sup}\frac{T-K_T}{T^{1+\frac{\varepsilon}{12}}}\right]\right) \\
&\leq C_{\sup}(\varepsilon)e^{-K_T}.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\sum_{n_1=K_N+1}^{N-2} \mathbb{P}(N_1 = n_1 | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}) \\
\leq\ & C_{\sup}(\varepsilon) e^{-K_T} (T - K_T)^{\frac{3}{2}} (K_T + 1)^{-\frac{3}{2}} T^{-\frac{1}{2}+\delta|-u|} \\
\leq\ & \frac{C_{\sup}(\varepsilon)}{T} \\
\leq\ & \frac{C_{\sup}(\varepsilon)}{N}.
\end{aligned}$$

Finally, we have

$$A(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \leq \frac{C_{\sup}}{N^{1-\varepsilon}}, \quad \text{and} \quad B(\mathbf{x}_{-u}^{(1)}, \mathbf{x}_{-u}^{(2)}) \leq \frac{C_{\sup}(\varepsilon)}{N}.$$

Thus

$$\begin{aligned}
\textstyle\sum_{n_1,n_2=N_I}^{N} \quad & \left| \mathbb{P}(N_1 = n_1, N_2 = n_2 | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}) \right. \\
& \left. - \mathbb{P}(N_1 = n_1 | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}) \mathbb{P}(N_2 = n_2 | \mathbf{X}_{-u}^{(1,2)} = \mathbf{x}_{-u}^{(1,2)}) \right| \leq \frac{C_{\sup}(\varepsilon)}{N^{1-\varepsilon}}.
\end{aligned}$$

So, we have proved Proposition 13. □

We conclude by the proof of Theorem 2.

Proof.

$$\begin{aligned}
& \mathbb{P}\left(\left| \widehat{E}_u - E_u \right| > \varepsilon \right) \\
\leq\ & \mathbb{P}\left(\left| \widehat{E}_u - \mathrm{E}(\widehat{E}_u) \right| > \frac{\varepsilon}{2} \right) + \mathbb{P}\left(\left| \mathrm{E}(\widehat{E}_u) - E_u \right| > \frac{\varepsilon}{2} \right).
\end{aligned}$$

Then, we use the proof of Proposition 11. If $(s(l))_{l \leq N_u}$ is a sample of uniformly distributed variables on $[1:N]$ with replacement, then for all $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}\left(\left| \widehat{E}_u - \mathrm{E}(\widehat{E}_u) \right| > \frac{\varepsilon}{2} \right) &\leq \frac{4}{\varepsilon^2} \left(\left| cov\left(\widehat{E}_{u,1}, \widehat{E}_{u,2} \right) \right| + \mathrm{Var}\left(\widehat{E}_{u,1} \right) \left(\frac{1}{N} + \frac{1}{N_u} \right) \right) \\
&\leq \frac{1}{\varepsilon^2} \left(\frac{C_{\sup}(\varepsilon')}{N^{\frac{1}{p-|u|}-\varepsilon'}} + \frac{C_{\sup}}{N_u} \right),
\end{aligned}$$

for all $\varepsilon' > 0$, thanks to Proposition 13. If $(s(l))_{l \leq N_u}$ is a sample of uniformly distributed variables on $[1:N]$ without replacement, then for all $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}\left(\left| \widehat{E}_u - \mathrm{E}(\widehat{E}_u) \right| > \frac{\varepsilon}{2} \right) &\leq \frac{4}{\varepsilon^2} \left(\frac{N_u - 1}{N_u} cov\left(\widehat{E}_{u,1}, \widehat{E}_{u,2} \right) + \frac{1}{N_u} \mathrm{Var}\left(\widehat{E}_{u,1} \right) \right) \\
&\leq \frac{1}{\varepsilon^2} \left(\frac{C_{\sup}(\varepsilon')}{N^{\frac{1}{p-|u|}-\varepsilon'}} + \frac{C_{\sup}}{N_u} \right),
\end{aligned}$$

for all $\varepsilon' > 0$, thanks to Proposition 13. Moreover, for all $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(\left|\widehat{E}_u - E_u\right| > \frac{\varepsilon}{2}\right) &\leq \frac{2}{\varepsilon}\left|\mathrm{E}(\widehat{E}_u) - E_u\right| \\ &\leq \frac{C_{\sup}(\varepsilon')}{\varepsilon N^{\frac{1}{p-|u|}-\varepsilon'}},\end{aligned}$$

for all $\varepsilon' > 0$, thanks to Proposition 12. Finally, for all $\varepsilon > 0$, $\varepsilon' > 0$, we have

$$\mathbb{P}\left(\left|\widehat{E}_u - E_u\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\left(\frac{C_{\sup}(\varepsilon')}{N^{\frac{1}{p-|u|}-\varepsilon'}} + \frac{C_{\sup}}{N_u}\right).$$

That concludes the proof. □

B Other proofs

Proof of Proposition 2

Proof.

$$\begin{aligned}&\mathrm{E}(f(\mathbf{X})f(\mathbf{X}^u)) \\ =\ &\mathrm{E}(\mathrm{E}(f(\mathbf{X})f(\mathbf{X}^u)|\mathbf{X}_u)) \\ =\ &\mathrm{E}\left(\int_{\mathcal{X}_{-u}^2} f(\mathbf{X}_u, \mathbf{x}_{-u})f(\mathbf{X}_u, \mathbf{x}'_{-u})d\mathbb{P}_{\mathbf{X}_{-u}|\mathbf{X}_u} \otimes \mathbb{P}_{\mathbf{X}_{-u}|\mathbf{X}_u}(\mathbf{x}_{-u}, \mathbf{x}'_{-u})\right) \\ =\ &\mathrm{E}\left(\int_{\mathcal{X}_{-u}} f(\mathbf{X}_u, \mathbf{x}_{-u})d\mathbb{P}_{\mathbf{X}_{-u}|\mathbf{X}_u}(\mathbf{x}_{-u}) \int_{\mathcal{X}_{-u}} f(\mathbf{X}_u, \mathbf{x}'_{-u})d\mathbb{P}_{\mathbf{X}_{-u}|\mathbf{X}_u}(\mathbf{x}'_{-u})\right) \\ =\ &\mathrm{E}\left(\mathrm{E}(f(\mathbf{X})|\mathbf{X}_u)^2\right).\end{aligned}$$

That concludes the proof of Proposition 2. □

Proof of Proposition 3

Proof. Let

$$A_{i,u} := \begin{cases} -\frac{1}{p}\binom{p-1}{|u|}^{-1} & \text{if } i \notin u \\ \frac{1}{p}\binom{p-1}{|u|-1}^{-1} & \text{if } i \in u. \end{cases}$$

Under Assumption 1, we have

$$\begin{aligned}
\mathrm{Var}(Y)^2 \sum_{i=1}^{p} \mathrm{Var}(\widehat{\eta}_i) &= \sum_{i=1}^{p} \sum_{\emptyset \subsetneq u \subsetneq [1:p]} A_{i,u}^2 \mathrm{Var}(\widehat{W}_u) \\
&= \sum_{\emptyset \subsetneq u \subsetneq [1:p]} \mathrm{Var}(\widehat{W}_u) \sum_{i=1}^{p} A_{i,u}^2 \\
&= \sum_{\emptyset \subsetneq u \subsetneq [1:p]} \frac{\mathrm{Var}(\widehat{W}_u^{(1)})}{N_u} \sum_{i=1}^{p} A_{i,u}^2.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\sum_{i=1}^{p} A_{i,u}^2 &= \sum_{i \in -u} \frac{1}{p^2} \binom{p-1}{|u|}^{-2} + \sum_{i \in u} \frac{1}{p^2} \binom{p-1}{|u|-1}^{-2} \\
&= \frac{1}{p!^2} \left((p-|u|)|u|!^2 (p-|u|-1)!^2 + |u|(|u|-1)!^2 (p-|u|)!^2 \right) \\
&= \frac{(p-|u|)!|u|!}{p!^2} (p-|u|-1)!(|u|-1)!(|u|+p-|u|) \\
&= \frac{(p-|u|)!|u|!}{p!} \frac{(p-|u|-1)!(|u|-1)!}{(p-1)!} \\
&=: C(|u|,p).
\end{aligned}$$

Thus, we want to minimize

$$\sum_{\emptyset \subsetneq u \subsetneq [1:p]} \frac{\mathrm{Var}(\widehat{W}_u^{(1)})}{N_u} C(|u|,p)$$

subject to

$$\sum_{\emptyset \subsetneq u \subsetneq [1:p]} N_u = \frac{N_{tot}}{\kappa}.$$

Let $U = (\mathbb{R}_+^*)^{2^p-2}$. If $\mathbf{x} \in U$, we index the components of $\mathbf{x}$ by the subsets $\emptyset \subsetneq u \subsetneq [1:p]$ and we write $\mathbf{x} = (x_u)_{\emptyset \subsetneq u \subsetneq [1:p]}$. Let h be the C^1 function on U defined by $h(\mathbf{x}) = \sum_{\emptyset \subsetneq u \subsetneq [1:p]} \frac{C(|u|,p)\mathrm{Var}(\widehat{W}_u^{(1)})}{x_u}$, let g be the C^1 function on U defined by $g(\mathbf{x}) = (\sum_{\emptyset \subsetneq u \subsetneq [1:p]} x_u) - N_{tot}/\kappa$ and let $A = g^{-1}(\{0\})$. Using the method of Lagrange multipliers, if $h_{|A}$ has a local minimum in $\mathbf{a}$, there exists c such that $Dh(\mathbf{a}) = cDg(\mathbf{a})$, i.e. $\nabla h(\mathbf{a}) = \nabla g(\mathbf{a})$ i.e.

$$\mathbf{a} = \frac{N_{tot}}{\kappa \sum_{\emptyset \subsetneq v \subsetneq [1:p]} \sqrt{C(|v|,p)\mathrm{Var}(\widehat{W}_v^{(1)})}} \left(\sqrt{C(|u|,p)\mathrm{Var}(\widehat{W}_u^{(1)})} \right)_{\emptyset \subsetneq u \subsetneq [1:p]}.$$

Moreover, note that h is strictly convex and the set A is convex, thus $h_{|A}$ is strictly convex. Thus $\mathbf{a}$ is the strict global minimum point of $h_{|A}$. □

Proof of Proposition 5

Proof. Let us write $V := \mathrm{Var}(\widehat{W}_u^{(1)})$ that does not depend on u by assumption. To simplify notation, let $N_0 = N_p = +\infty$. In this way, we have, for all $u \subset [1:p]$, $\mathrm{Var}(\widehat{W}_u(m)) = V/N_{|u|}$.

We have

$$\begin{aligned}
\mathrm{Var}\left(\widehat{\eta}_i|\left(\sigma_m\right)_{m\leq M}\right) &= \frac{1}{p^2\mathrm{Var}(Y)^2}\sum_{u\subset -i}\frac{1}{M^2}\sum_{m=1}^{M}\left[\mathrm{Var}\left(\widehat{W}_{u\cup\{i\}}(m)\right)+\mathrm{Var}\left(\widehat{W}_u(m)\right)\right]\mathbb{1}_{P_i(\sigma_m)=u} \\
&= \frac{V}{p^2\mathrm{Var}(Y)^2}\sum_{u\subset -i}\frac{1}{M^2}\sum_{m=1}^{M}\left[\frac{1}{N_{|u\cup\{i\}|}}+\frac{1}{N_u}\right]\mathbb{1}_{P_i(\sigma_m)=u}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{E}\left[\mathrm{Var}\left(\widehat{\eta}_i|\left(\sigma_m\right)_{m\leq M}\right)\right] &= \frac{V}{p^2\mathrm{Var}(Y)^2}\sum_{u\subset -i}\frac{1}{M^2}\sum_{m=1}^{M}\left[\frac{1}{N_{|u\cup\{i\}|}}+\frac{1}{N_u}\right]\mathbb{P}(P_i(\sigma_m)=u) \\
&= \frac{V}{p^2\mathrm{Var}(Y)^2}\sum_{u\subset -i}\frac{1}{M^2}\sum_{m=1}^{M}\frac{1}{p}\binom{p-1}{|u|}^{-1}\left[\frac{1}{N_{|u\cup\{i\}|}}+\frac{1}{N_{|u|}}\right] \\
&= \frac{V}{p^2\mathrm{Var}(Y)^2}\sum_{u\subset[1:p]}a_{i,u}\frac{1}{N_{|u|}},
\end{aligned}$$

where

$$a_{i,u} := \begin{cases} \frac{1}{p}\binom{p-1}{|u|}^{-1} & \text{if } i\notin u \\ \binom{p-1}{|u|-1}^{-1} & \text{if } i\in u. \end{cases}$$

Remark that $\sum_{i=1}^{p} a_{i,u} = 2\binom{p}{|u|}^{-1}$. Then,

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{p}\mathrm{Var}\left(\widehat{\eta}_i|\left(\sigma_m\right)_{m\leq M}\right)\right] &= \sum_{i=1}^{p}\frac{V}{p^2\mathrm{Var}(Y)^2}\sum_{u\subset[1:p]}a_{i,u}\frac{1}{N_{|u|}} \\
&= \frac{V}{p^2\mathrm{Var}(Y)^2}\sum_{u\subset[1:p]}\frac{1}{N_{|u|}}\sum_{i=1}^{p}a_{i,u} \\
&= \frac{2V}{p^2\mathrm{Var}(Y)^2}\sum_{u\subset[1:p]}\frac{1}{N_{|u|}}\binom{p}{|u|}^{-1} \\
&= \frac{2V}{p^2\mathrm{Var}(Y)^2}\sum_{k=1}^{p-1}\frac{1}{N_k}
\end{aligned}$$

We get the relaxed problem

$$\min_{(N_k)_{k\in[1:p-1]}} \frac{2V}{p^2\mathrm{Var}(Y)^2}\sum_{k=1}^{p-1}\frac{1}{N_k}$$

subject to $M\sum_{k=1}^{p-1}N_k = MN_O(p-1)$. Let $U=(\mathbb{R}_+^*)^{p-1}$. Let h be the C^1 function on U defined by $h(\mathbf{x}) = \frac{2V}{p^2\mathrm{Var}(Y)^2}\sum_{k=1}^{p-1}\frac{1}{x_k}$, g be the C^1 function on U defined by $g(\mathbf{x}) = M\sum_{k=1}^{p-1}x_k - MN_O(p-1)$. Finally, let $A = g^{-1}(\{0\})$. Using the method of Lagrange multipliers, if $h_{|A}$ has a local minimum in $\mathbf{a}$, there exists c such that $Dh(\mathbf{a}) = cDg(\mathbf{a})$, i.e. $\nabla h(\mathbf{a}) = \nabla g(\mathbf{a})$ i.e. $\forall u$, $-\frac{1}{a_u^2} = c'$ i.e. $a_u = c''$. To sum up, if $h_{|A}$ has a local minimum, it is in $\mathbf{a}$ defined by

$$a_u = N_O M p_u.$$

Moreover, note that h is strictly convex and the set A is convex, thus $h_{|A}$ is strictly convex. Thus $\mathbf{a}$ is the strict global minimum point of $h_{|A}$. Thus, $\mathbf{a}$ is the global minimum on the constraint problem (where the inputs are integers). □

Proof of Proposition 6 This proof totally arises from the appendix of [SNS16]. The computations are the same.

Proof. Under Assumption 3, we have

$$\begin{aligned}
\mathrm{Var}(\widehat{\eta}_i) &= \frac{1}{M\mathrm{Var}(Y)^2}\left(\mathrm{Var}\left(\widehat{W}_{P_i(\sigma_1)\cup\{i\}}\right)+\mathrm{Var}\left(\widehat{W}_{P_i(\sigma_1)}\right)\right)\\
&= \frac{1}{M\mathrm{Var}(Y)^2}\Bigg(\mathrm{Var}(\mathrm{E}(\widehat{W}_{P_i(\sigma_1)\cup\{i\}}|\sigma_1)) + \mathrm{E}(\mathrm{Var}(\widehat{W}_{P_i(\sigma_1)\cup\{i\}}|\sigma_1))\\
&\quad +\mathrm{Var}(\mathrm{E}(\widehat{W}_{P_i(\sigma_1)}|\sigma_1)) + \mathrm{E}(\mathrm{Var}(\widehat{W}_{P_i(\sigma_1)}|\sigma_1))\Bigg)\\
&= \frac{1}{C\mathrm{Var}(Y)^2}\Bigg(N_O\mathrm{Var}(W_{P_i(\sigma_1)\cup\{i\}}) + N_O\mathrm{Var}(W_{P_i(\sigma_1)})\\
&\quad +\mathrm{E}(\mathrm{Var}(\widehat{W}^{(1)}_{P_i(\sigma_1)\cup\{i\}}|\sigma_1)) + \mathrm{E}(\mathrm{Var}(\widehat{W}^{(1)}_{P_i(\sigma_1)}|\sigma_1))\Bigg).
\end{aligned}$$

Thus, the minimum is with $N_O = 1$. □

Proof of Proposition 7

Proof. We only prove the second item. The first one is easier and uses the same idea. Let $i\in[1:p]$. Remark that

$$\begin{aligned}
\widehat{\eta}_i &= \frac{1}{M\mathrm{Var}(Y)}\sum_{m=1}^{M}\left(\widehat{W}_{P_i(\sigma_m)\cup\{i\}}(m) - \widehat{W}_{P_i(\sigma_m)}(m)\right)\\
&= \frac{1}{p\mathrm{Var}(Y)}\sum_{u\subset -i}\binom{p-1}{|u|}^{-1}\left(\tilde{W}_{u\cup\{i\},i} - \tilde{W}_{u,i}\right)
\end{aligned}$$

with

$$\tilde{W}_{u,i} := \binom{p-1}{|u|} \frac{p}{M} \sum_{m|\ P_i(\sigma_m)=u} \widehat{W}_u(m) \quad \text{and} \quad \tilde{W}_{u\cup\{i\},i} := \binom{p-1}{|u|} \frac{p}{M} \sum_{m|\ P_i(\sigma_m)=u} \widehat{W}_{u\cup\{i\}}(m),$$

where we sum over all the integers $m \in [1:M]$ such that $P_i(\sigma_m) = u$. Thus, for all u,

$$\tilde{W}_{u,i} \sim \binom{p-1}{|u \setminus \{i\}|} \frac{p}{M} \tilde{N}_{u,i,M} \widehat{W}_u^{\tilde{N}_{u,i,M}},$$

where

$$\widehat{W}_u^{\tilde{N}_{u,i,M}} := \frac{1}{\tilde{N}_{u,i,M}} \sum_{k=1}^{\tilde{N}_{u,i,M}} \widehat{W}_u(k),$$

and $\tilde{N}_{u,i,M} = \tilde{N}_{u\cup\{i\},i,M} \sim \mathcal{B}(M, \frac{|u|!(p-1-|u|)!}{p!})$ (the binomial distribution). Now, remark that M goes to $+\infty$ when N_{tot} goes to $+\infty$ (recall that $N_{tot} = \kappa M(p-1)$). Hence,

$$\binom{p-1}{|u \setminus \{i\}|} \frac{p}{M} \tilde{N}_{u,i,M} \underset{N_{tot}\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 1.$$

It suffices to show that for all $u \subset [1:p]$, the estimator $\omega \mapsto \widehat{W}_u^{\tilde{N}_{u,i,M}(\omega)}(\omega)$ converges to W_u in probability when N and N_{tot} go to $+\infty$ and we could conclude by

$$\begin{aligned} \widehat{\eta}_i \quad &= \quad \frac{1}{p\mathrm{Var}(Y)} \sum_{u\subset -i} \binom{p-1}{|u|}^{-1} \left(\tilde{W}_{u\cup\{i\},i} - \tilde{W}_{u,i}\right) \\ &\underset{N\to+\infty}{\overset{\mathbb{P}}{\underset{N_{tot}\to+\infty}{\longrightarrow}}} \quad \frac{1}{p\mathrm{Var}(Y)} \sum_{u\subset -i} \binom{p-1}{|u|}^{-1} \left(W_{u\cup\{i\}} - W_u\right) \\ &= \quad \eta_i. \end{aligned}$$

Let $\varepsilon > 0$ and $\delta > 0$. Using the assumptions and Chebyshev's inequality, we have that $(\widehat{W}_u^{N_O})_{N_O,N}$ is consistent, thus there exists N_{O1} and N_1 such that for all $N_O \geq N_{O1}$ and all $N \geq N_1$,

$$\mathbb{P}\left(\left|\widehat{W}_u^{N_O} - W_u\right| > \delta\right) < \frac{\varepsilon}{2}.$$

Moreover,

$$\mathbb{P}(\tilde{N}_{u,M} \leq N_{O1}) \underset{M\to+\infty}{\longrightarrow} 0.$$

Thus, there exists M_1 such that for all $M \geq M_1$,

$$\mathbb{P}(\tilde{N}_{u,M} \leq N_{O1}) < \frac{\varepsilon}{2}.$$

Thus, there exists N_{tot1} such that for all $N_{tot} \geq N_{tot1}$,

$$\mathbb{P}(\tilde{N}_{u,M} \leq N_{O1}) < \frac{\varepsilon}{2}.$$

Finally, for all $N_{tot} \geq N_{tot1}$ and $N \geq N_1$, we have

$$\begin{aligned}\mathbb{P}\left(\left|\widehat{W}_u^{\tilde{N}_{u,M}} - W_u\right| > \delta\right) &\leq \mathbb{P}\left(\left|\widehat{W}_u^{\tilde{N}_{u,M}} - W_u\right| > \delta,\ \ \tilde{N}_{u,M} \geq N_{O1}\right) + \mathbb{P}(\tilde{N}_{u,M} \leq N_{O1}) \\ &< \varepsilon.\end{aligned}$$

That proves that the estimator $\omega \mapsto \widehat{W}_u^{\tilde{N}_{u,i,M}(\omega)}(\omega)$ converges to W_u in probability when N and N_{tot} go to $+\infty$. □

Proof of Corollary 1 and Corollary 2

We do the proof for Corollary 1. The proof of Corollary 2 uses the same idea.

Proof. Let $\delta > 0$. Thanks to Theorem 2, with $\varepsilon' = \delta$, we have

$$\mathbb{P}\left(N^{\frac{1}{2(p-|u|)}-\delta}\left|\widehat{E}_{u,MC} - E_u\right| > \varepsilon\right) \leq \frac{C_{\sup}(\delta) N^{\frac{1}{p-|u|}-2\delta}}{\varepsilon^2 N^{\frac{1}{p-|u|}-\delta}} \underset{N\to+\infty}{\longrightarrow} 0.$$

That concludes the proof of Corollary 1. □

Proof of Proposition 8

Proof. If we use the subset W-aggregation procedure, we just have to use the consistency of $\widehat{W}_u$ from Theorems 1 and 3 and to use Proposition 4.

If we use the subset W-aggregation procedure, the consistency of the estimators of the Shapley effects comes from the second part of Proposition 7. We just have to verify Assumption 4. Let $\widehat{W}_u(m)$ of Proposition 7 be $\widehat{E}_{u,s(m),MC}$ or $\widehat{V}_{u,s(m),PF}$ defined in Section 6.1, where $(s(m))_m$ are independent and uniformly distributed on $[1:N]$. Then, following the end of the proof of Theorems 1 and 3, we obtain

$$\frac{1}{M^2}\sum_{m,m'=1}^{M} cov\left(\widehat{W}_u(m), \widehat{W}_u(m')\right) \underset{N,M\to+\infty}{\longrightarrow} 0,$$

and, by Proposition 9, we have

$$\mathrm{E}\left(\widehat{W}_u(1)\right) = \mathrm{E}\left(\widehat{W}_u^{(1)}\right) \underset{N\to+\infty}{\longrightarrow} W_u.$$

Thus, Assumption 4 holds. □

References

[BBDM19] Baptiste Broto, Franois Bachoc, Marine Depecker, and Jean-Marc Martinez. Sensitivity indices for independent groups of variables. *Mathematics and Computers in Simulation*, 163:19–31, September 2019.

[BEDC19] Nazih Benoumechiara and Kevin Elie-Dit-Cosaque. Shapley effects for sensitivity analysis with dependent inputs: bootstrap and kriging-based algorithms. *ESAIM: Proceedings and Surveys*, 65:266–293, 2019.

[BGLR18] Francois Bachoc, Fabrice Gamboa, Jean-Michel Loubes, and Laurent Risser. Entropic Variable Boosting for Explainability and Interpretability in Machine Learning. *arXiv preprint arXiv:1810.07924*, 2018.

[BHP16] Emanuele Borgonovo, Gordon B. Hazen, and Elmar Plischke. A Common Rationale for Global Sensitivity Measures and Their Estimation. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, 36(10):1871–1895, October 2016.

[BMM+07] Philippe Besse, Hlne Milhem, Olivier Mestre, Anne Dufour, and Vincent-Henri Peuch. Comparaison de techniques de Data Mining pour l'adaptation statistique des prvisions d'ozone du modle de chimie-transport MOCAGE. *Pollution Atmosphrique*, 49(195):285–292, 2007.

[BS19] Thomas B. Berrett and Richard J. Samworth. Efficient two-sample functional estimation and the super-oracle phenomenon. *arXiv:1904.09347 [math, stat]*, April 2019.

[BSY19] Thomas B. Berrett, Richard J. Samworth, and Ming Yuan. Efficient multivariate entropy estimation via k-nearest neighbour distances. *The Annals of Statistics*, 47(1):288–318, February 2019.

[CBSV16] Riccardo Colini-Baldeschi, Marco Scarsini, and Stefano Vaccari. Variance allocation and Shapley value. *Methodology and Computing in Applied Probability*, pages 1–15, 2016.

[CGT09] Javier Castro, Daniel Gmez, and Juan Tejada. Polynomial calculation of the Shapley value based on sampling. *Computers & Operations Research*, 36(5):1726–1730, May 2009.

[Cha13] Galle Chastaing. *Indices de Sobol gnraliss pour variables dpendantes*. phdthesis, Universit de Grenoble, September 2013.

[FRK15] Jana Fruth, Olivier Roustant, and Sonja Kuhnt. Sequential designs for sensitivity analysis of functional inputs in computer experiments. *Reliability Engineering & System Safety*, 134:260–267, 2015.

[FWJ08] Shaheen S. Fatima, Michael Wooldridge, and Nicholas R. Jennings. A linear approximation method for the Shapley value. *Artificial Intelligence*, 172(14):1673–1699, September 2008.

[Gho01] Subhashis Ghosal. Convergence rates for density estimation with Bernstein polynomials. *The Annals of Statistics*, 29(5):1264–1280, 2001.

[GJK+16] Fabrice Gamboa, Alexandre Janon, Thierry Klein, A. Lagnoux, and Clmentine Prieur. Statistical inference for Sobol pick-freeze Monte Carlo method. *Statistics*, 50(4):881–902, 2016.

[GJKL14] Fabrice Gamboa, Alexandre Janon, Thierry Klein, and Agns Lagnoux. Sensitivity analysis for multidimensional and functional outputs. *Electronic Journal of Statistics*, 8(1):575–603, 2014.

[HI03] Franz Hubert and Svetlana Ikonnikova. Strategic investment and bargaining power in supply chains: A Shapley value analysis of the Eurasian gas market. March 2003.

[HS96] Toshimitsu Homma and Andrea Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, 52(1):1–17, 1996.

[IP17] Bertrand Iooss and Clmentine Prieur. Shapley effects for sensitivity analysis with dependent inputs: comparisons with Sobol' indices, numerical estimation and applications. *arXiv:1707.01334 [math, stat]*, July 2017.

[JKLR+14] Alexandre Janon, Thierry Klein, Agnes Lagnoux-Renaudie, Malle Nodet, and Clmentine Prieur. Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364, October 2014.

[JLD06] Julien Jacques, Christian Lavergne, and Nicolas Devictor. Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering & System Safety*, 91(10-11):1126–1134, 2006.

[JSW98] Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

[MT12] Thierry A. Mara and Stefano Tarantola. Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety*, 107:115–121, November 2012.

[MvLG+08] Stefano Moretti, Danitsja van Leeuwen, Hans Gmuender, Stefano Bonassi, Joost van Delft, Jos Kleinjans, Fioravante Patrone, and Domenico Franco Merlo. Combining Shapley value and statistics

to the analysis of gene expression data in children exposed to air pollution. *BMC Bioinformatics*, 9:361, September 2008.

[OP17] Art B. Owen and Clmentine Prieur. On Shapley value for measuring importance of dependent inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):986–1002, 2017.

[Owe14] Art B. Owen. Sobol' Indices and Shapley Value. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):245–251, January 2014.

[PBS13] Elmar Plischke, Emanuele Borgonovo, and Curtis L. Smith. Global sensitivity measures from given data. *European Journal of Operational Research*, 226(3):536–550, May 2013.

[Pli10] Elmar Plischke. An effective algorithm for computing global sensitivity indices (EASI). *Reliability Engineering & System Safety*, 95(4):354–360, April 2010.

[RSG16] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Why should I trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144. ACM, 2016.

[Sha53] L. S. Shapley. A value for n-person games. Contribution to the Theory of Games. *Annals of Mathematics Studies*, 2:28, 1953.

[SNS16] Eunhye Song, Barry L. Nelson, and Jeremy Staum. Shapley Effects for Global Sensitivity Analysis: Theory and Computation. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1060–1083, January 2016.

[Sob93] Ilya M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1(4):407–414, 1993.

[SWNW03] Thomas J. Santner, Brian J. Williams, William Notz, and Brain J. Williams. *The design and analysis of computer experiments*, volume 1. Springer, 2003.

[Tju74] Tue Tjur. *Conditional Probability Distributions*. Institute of Mathematical Statistics, University of Copenhagen, 1974.

[VG13] Sbastien Da Veiga and Fabrice Gamboa. Efficient estimation of sensitivity indices. *Journal of Nonparametric Statistics*, 25(3):573–595, 2013.