



HAL
open science

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

Baptiste Broto, François Bachoc, Marine Depecker

► **To cite this version:**

Baptiste Broto, François Bachoc, Marine Depecker. Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution. 2018. hal-01962010v1

HAL Id: hal-01962010

<https://hal.science/hal-01962010v1>

Preprint submitted on 20 Dec 2018 (v1), last revised 5 Feb 2020 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Baptiste Broto¹, François Bachoc², and Marine Depecker¹

¹CEA, LIST, Université Paris-Saclay, F-91120, Palaiseau, France

²Institut 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 [18] and suggest a new one with a lower variance. Then, when the distribution of the inputs is unknown, we extend these estimators. Finally, we provide asymptotic properties of the estimators studied in this article.

1 Introduction

Sensitivity indices are important tools in sensitivity analysis. They aim at 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 an input variable can lead to an important variation of the output or not.

In global sensitivity analysis, the input variables X_1, \dots, X_p are assumed to be random variables. Sobol defined the first sensitivity indices for a general framework, called the Sobol indices, in [17]. Many other sensitivity indices have been defined and studied (see [1] 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 Sobol indices have been suggested for dependent input variables (see for example [11], [12] and [3]).

Recently, Owen defined new sensitivity indices in [14] called "Shapley effects" that have good 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 indice 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 [16]. The Shapley value has been widely studied ([4], [5]) and applied in different fields (see for example [13] or [9]). However, only few articles focus on the Shapley effects in sensitivity analysis (see [14], [18], [15], [10], [2]).

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 the quantities that we call the "conditional elements" which the Shapley effects depend on. The second part is the way to estimate the Shapley effects when we have estimates of the conditional elements. We call this part the W -aggregation procedure. We refer to Sections 3 and 4 for more details on these two parts. Song et al. suggested an algorithm to estimate the Shapley effects in [18] that is implemented in the R package "sensitivity". The algorithm of [18] uses a W -aggregation procedure based on permutations of $\{1, \dots, p\}$. We study this W -aggregation procedure and explain why it minimizes the variance of the estimate of the Shapley effect corresponding to a single input variable. We then suggest a new W -aggregation procedure, based on the subsets of $\{1, \dots, p\}$, 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. This provides an improved accuracy, compared to the first W -aggregation procedure in [18]. The comparison between the two W -aggregation procedures is illustrated with numerical experiments.

Then, we focus on the estimation of the conditional elements with two different estimators: the double Monte-Carlo estimator (used in the algorithm of [18]) and the Pick-and-Freeze estimator (see [8] for the independent case) that we extend to the case where the inputs are dependent. We give their definition when it is possible to sample from the conditional distributions of the input vector. Then we extend these estimators to the case where we just observe an i.i.d. sample from the input variables. The extension relies on nearest-neighbour techniques. To the best of our knowledge, the estimators we suggest are the first that do not require exact samples from the conditional distribution of the input variables. One of our main result is the consistency of these estimators under some mild assumptions, and their rate of convergence under additional regularity assumptions. We conclude giving 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. To highlight the efficiency of these estimators, we provide numerical experiments in the two following cases: where it is possible to sample from the conditional distributions of the input vector and where we just observe an i.i.d. sample from the input variables. We observe that in the second case, the estimators of the Shapley effects have a similar accuracy as in the first case.

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 study the W -aggregation procedure used by

the algorithm of [18] and we suggest a new one. In Section 4, we assume that the input distribution is known. We give two methods to estimate the conditional elements and we illustrate the various estimators of the Shapley effects with numerical applications. In Section 5, 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. We conclude in Section 6. All the proofs are provided in the appendix.

2 The Shapley effects

We let $X = (X_1, \dots, X_p)$ be the input random vector on $E = E_1 \times \dots \times E_p$ with distribution \mathbb{P}_X . We assume that there is an output variable Y defined by

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

with $f \in L^2(\mathbb{P}_X)$. We write $[1 : p]$ for the set $\{1, 2, \dots, p\}$. 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:

$$V_u := \text{Var}(\text{E}(Y|X_u)) \tag{2}$$

and

$$E_u := \text{E}(\text{Var}(Y|X_{-u})), \tag{3}$$

where $-u := [1 : p] \setminus u$. 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 [18] by:

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

where $-i$ is the subset $[1 : p] \setminus \{i\}$.

Remark 1. *As explained in [18], 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. *We have that W_\emptyset and $W_{[1:p]}$ are equal to 0 and $\text{Var}(Y)$ respectively. The variance of Y is easy to estimate, so we assume without loss of generality that we know the theoretical value $\text{Var}(Y)$.*

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.

Proposition 1. *We have*

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

where \mathcal{S}_p is the set of permutations of $[1 : p]$ and $P_i(\sigma) := \{\sigma(j), j \in [1 : i]\}$.

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 represent two different W -aggregation procedures of the Shapley effects.

To simplify notations, if $u \subset -i$, we write $u + i$ for $u \cup \{i\}$.

3 W -aggregation procedures

As we can see in Equation (4) or in Equation (5), the Shapley effects are functions of the conditional elements $(W_u)_u$. For now, we do not focus on the estimation of W_u . We assume that we have a random function $\Psi : u \mapsto \widehat{W}_u$ which, for all $u \subset [1 : p]$, gives a random estimate \widehat{W}_u of W_u . From Remark 2, we let $\Psi(\emptyset) = 0$ and $\Psi([1 : p]) = \text{Var}(Y)$. We call "W-aggregation procedure" an algorithm that estimates the Shapley effects from a such function. We begin to explain the W -aggregation procedure of [18]. Then we suggest an improved W -aggregation procedure.

3.1 The random-permutation procedure

In this section, we present the "random-permutation W -aggregation procedure" suggested in [18].

3.1.1 Estimator

The W -aggregation procedure of the algorithm of [18] is based on Equation (5). Thus we could estimate η_i by

$$\widehat{\eta}_i = \frac{1}{p! \text{Var}(Y)} \sum_{\sigma \in \mathcal{S}_p} \left(\widehat{W}_{P_i(\sigma)+i} - \widehat{W}_{P_i(\sigma)} \right). \quad (6)$$

However, as the number of permutations is $p!$, there are too many summands and [18] suggests to replace the sum over all the $p!$ permutations by the sum over M ($M < p!$) random uniformly distributed permutations. Thus, we can estimate η_i by

$$\widehat{\eta}_i = \frac{1}{M \text{Var}(Y)} \sum_{m=1}^M \left(\widehat{W}_{P_i(\sigma_m)+i} - \widehat{W}_{P_i(\sigma_m)} \right). \quad (7)$$

where the σ_m are i.i.d. and uniformly distributed on \mathcal{S}_p .

Finally, [18] reduces the computation cost using the following idea. They notice that for $i < p$ and for any permutation σ , we have $P_{\sigma(i+1)}(\sigma) = P_{\sigma(i)}(\sigma) + i$. Thus, the algorithm of [18] uses every estimate $\widehat{W}_{P_{\sigma(i+1)}(\sigma_m)}$ for $\widehat{\eta}_{\sigma(i)}$ and for $\widehat{\eta}_{\sigma(i+1)}$. So, the number of estimations of \widehat{W}_u (for $u \neq \emptyset$) is divided by two when estimating the Shapley effects. The W -aggregation procedure is

1. For all $m = 1, 2, \dots, M$
 - (a) Generate σ_m
 - (b) Let $prevC = 0$
 - (c) For all $i = 1, 2, \dots, p$
 - i. Estimate $W_{P_{\sigma_m(i)}(\sigma_m)+i}$ by $\widehat{W}_{P_{\sigma_m(i)}(\sigma_m)+i} = \Psi(P_{\sigma_m(i)}(\sigma_m) + i)$.
 - ii. Compute $\widehat{\Delta} = \widehat{W}_{P_{\sigma_m(i)}(\sigma_m)+i} - prevC$
 - iii. Update $\widehat{\eta}_{\sigma_m(j)} = \widehat{\eta}_{\sigma_m(j)} + \widehat{\Delta}_{\sigma_m(j)}c(\sigma_m)$.
 - iv. Set $prevC = \widehat{W}_{P_{\sigma_m(i)}(\sigma_m)+i}$.
2. Let $\widehat{\eta}_i = \widehat{\eta}_i / (\text{Var}(Y)M)$ for all $i = 1, \dots, p$.

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

3.1.2 The precision of the estimate of W_u is adapted with u

In this section, we explain how the Shapley effect estimation given by Equation (7) is equivalent to an estimation given by

$$\widehat{\eta}_i := \frac{1}{p\text{Var}(Y)} \sum_{u \subset -i} \binom{p-1}{|u|}^{-1} (\widetilde{W}_{u+i} - \widetilde{W}_u), \quad (8)$$

where \widetilde{W}_u is some estimation of W_u with a precision adapted with u . The W_u which have a large coefficient in (4) will be better estimated than the W_u that have a small coefficient.

First of all, let us show how to link Equation (7) with Equation (8). We have

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

with

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

Yet, the map $P_i : S_p \rightarrow \mathcal{P}([1 : p] \setminus \{i\})$ is not injective. So, the number of m such that $P_i(\sigma_m) = u$ can be larger than 1, even if the random permutations

$(\sigma_k)_{k \leq M}$ are two by two distinct. Actually, if $u \subset -i$, the number $\tilde{N}_u = \tilde{N}_{u,M}$ (resp. $\tilde{N}_{u+i} = \tilde{N}_{u+i,M}$) of estimates \widehat{W}_u (resp. \widehat{W}_{u+i}) used for \tilde{W}_u (resp. \tilde{W}_{u+i}) is the random variable $|\{m \leq M, P_i(\sigma_m) = u\}| \sim \mathcal{B}(M, \frac{|u|(p-1-|u)|}{p!})$, where \mathcal{B} is the binomial distribution. To simplify notations, if $u \subset -i$, let us write

$$p_u := p_{u+i} := \frac{|u|(p-1-|u)|}{p!}. \quad (9)$$

We want to study the cost-accuracy tradeoff of the different estimators of the $(W_u)_u$ to minimize the variance of the Shapley effect estimator $\hat{\eta}_i$. To that end, we introduce the following assumption.

Assumption 1. Each \widehat{W}_u is computed with the same cost N_O by $\widehat{W}_u = \frac{1}{N_O} \sum_{k=1}^{N_O} \widehat{W}_u^{(k)}$ where the $(\widehat{W}_u^{(k)})_k$ are independent and identically distributed. The $(\widehat{W}_u)_u$ are independent.

Thus, under Assumption 1, each \tilde{W}_u has a random cost \tilde{N}_u with

$$\mathbb{E}(\tilde{N}_u) = N_O M p_u$$

and

$$\text{Var}(\tilde{N}_u) = N_O^2 M p_u (1 - p_u).$$

To conclude, the estimation of $\hat{\eta}_i$ given by Equation (7) is equivalent to an estimation given by (8) where each \tilde{W}_u is computed with a random cost \tilde{N}_u that depends on u . We show in (2) that this random cost is actually a good choice. Before giving the proposition, let us introduce another assumption.

Assumption 2. Each \tilde{W}_u is computed with a cost N_u by $\tilde{W}_u = \frac{1}{N_u} \sum_{k=1}^{N_u} \widehat{W}_u^{(k)}$ where the $(\widehat{W}_u^{(k)})_k$ are independent and identically distributed. The $(\tilde{W}_u)_u$ are independent.

Proposition 2. Let $i \in [1 : p]$ be fixed. Assume that we estimate η_i by Equation (8) under Assumption 2 and assume that $\text{Var}(\widehat{W}_u^{(1)})$ does not depend on u . Then, the solution of the relaxed problem (i.e. the problem without the constraint of letting the $(N_u)_u$ be integers) $\min_{(N_u)_u} \text{Var}(\hat{\eta}_i)$ subject to $\sum_u N_u = 2MN_O$ is $N_u^* = N_O M p_u = \mathbb{E}(\tilde{N}_u)$.

Note that when we want to estimate only η_i using (7) and under Assumption 1, the total cost for $\hat{\eta}_i$ is $2MN_O$. So, according to Proposition 2, the average of the costs of $(\tilde{W}_u)_u$ is optimal to minimize $\text{Var}(\hat{\eta}_i)$.

Remark 3. When the number of inputs p is small, [18] suggest to take all the permutations of $[1 : p]$ instead of choosing random permutations. In this case, the number \tilde{N}_u is no longer random and is exactly $N_O p! p_u$, which is the exact solution of the problem $\min_{(N_u)_u} \text{Var}(\hat{\eta}_i)$ subject to $\sum_u N_u = 2p! N_O$. However, this algorithm requires small values of p and the total cost is a multiple of $2p!$ (so there are very restricted possible values). Hence, the algorithm with all the permutations is not explicitly detailed in [18].

3.1.3 Choice of N_O

In this section, we explain why we should choose $N_O = 1$ under Assumption 1.

Proposition 3 is actually given in [18] in the particular case of the algorithm of [18], with more complicated computations than in the proof of Proposition 3 (they explain how to choose other parameters). We just give the result in a more general case, which totally arises from [18].

Proposition 3. *Under Assumption 1 and assuming that for all $u \subset [1 : p]$, $E(\widehat{W}_u^{(1)}) = W_u$, in order to minimize, over N_O and M , the variance of $\widehat{\eta}_i$ with a fixed cost $2MN_O =: 2C$, we have to choose $N_O = 1$ and $M = C$.*

Now, we give another reason for choosing $N_O = 1$. According to Proposition 2, for all u , $E(\tilde{N}_u) = N_u^*$ is the best cost of \tilde{W}_u for minimizing $\text{Var}(\widehat{\eta}_i)$. Thus, we would want N_u to be close to its mean. So, we want to minimize its variance $\text{Var}(\tilde{N}_u) = N_O^2 M p_u (1 - p_u)$. When the total cost $2MN_O = 2C$ is fixed, in order to minimize $\text{Var}(\tilde{N}_u)$, we have to choose $M = C$ and $N_O = 1$.

From now on, we assume that $N_O = 1$ when we use the random-permutation W -aggregation procedure and we will let M , the number of random permutations, go to infinity. So, for the same subset u , we will call several times the random function value $\Psi(u)$. We will write $(\widehat{W}_u^{(i)})_{i \geq 1}$ to differentiate the various calls to $\Psi(u)$ when using the random-permutation procedure. Notice that the cost of each $\widehat{W}_u^{(i)}$ is equal to one because $N_O = 1$.

Finally, since we let $\Psi(\emptyset) = 0$ and $\Psi([1 : p]) = \text{Var}(Y)$, their cost is equal to 0. Then, the total cost N_{tot} of the random-permutation W -aggregation procedure is equal to $N_{tot} = M(p - 1)$, for estimating the p Shapley effects η_1, \dots, η_p .

3.1.4 Consistency

We give sufficient conditions to have the consistency of the estimators of the Shapley effects given by the random-permutation W -aggregation procedure.

Proposition 4. *Assume that for all u such that $\emptyset \subsetneq u \subsetneq [1 : p]$, $(\widehat{W}_u^{(i)})_{i \geq 1}$ (the different calls to the function $\Psi(u)$) are identically distributed and that we estimate the Shapley effects using the random-permutation W -aggregation procedure. Let $N_{tot} = M(p - 1)$ be the total cost of the random-permutation W -aggregation procedure.*

1. *If for all u such that $\emptyset \subsetneq u \subsetneq [1 : p]$, the $(\widehat{W}_u^{(i)})_{i \geq 1}$ are unbiased and independent, then the estimates of the Shapley effects are consistent when N_{tot} goes to $+\infty$.*
2. *Assume that for all u such that $\emptyset \subsetneq u \subsetneq [1 : p]$, the distribution of the $\widehat{W}_u^{(i)}$ depends on an integer N such that*

$$E\left(\widehat{W}_u^{(1)}\right) \xrightarrow{N \rightarrow +\infty} W_u.$$

Assume that for all u such that $\emptyset \subsetneq u \subsetneq [1 : p]$,

$$\frac{1}{n^2} \sum_{i,j=1}^n \text{cov} \left(\widehat{W}_u^{(i)}, \widehat{W}_u^{(j)} \right) \xrightarrow{N, n \rightarrow +\infty} 0,$$

where n is the number of the various calls to the random function value $\Psi(u)$. Then, the estimates of Shapley effects are consistent when N_{tot} and N go to $+\infty$.

3.2 The subset procedure

In this section, we suggest a new W -aggregation procedure for the Shapley effects. We want to compute 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.

3.2.1 Estimator

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

- For all $u \subset [1 : p]$, estimate W_u by $\widehat{W}_u = \Psi(u)$ and store it.
- For all $i \in [1 : p]$, estimate η_i by

$$\widehat{\eta}_i := \frac{1}{p \text{Var}(Y)} \sum_{u \subset -i} \binom{p-1}{|u|}^{-1} (\widehat{W}_{u \cup \{i\}} - \widehat{W}_u). \quad (10)$$

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 $\widehat{\eta}_{i \in [1:p]}$. Recall that with the random-permutation W -aggregation procedure, each estimate \widehat{W}_u was used for the estimate of only two Shapley effects: more particularly $\widehat{W}_{P_{\sigma(i+1)}(\sigma_m)}$ was used for $\widehat{\eta}_{\sigma(i)}$ and for $\widehat{\eta}_{\sigma(i+1)}$. Thus the subset W -aggregation procedure seems to be more efficient.

Remark 4. When the number of inputs p is large (for example $p = 100$), it is too costly to estimate W_u for all $u \subset [1 : p]$. So, the subset W -aggregation procedure is, at first sight, limited to small to moderate values of p . Fortunately, we will suggest in Section 3.2.2 not to estimate all the $(W_u)_u$ with the same accuracy. Thus, when the number of variables p is large, most of the $(W_u)_u$ will be approximated by 0 and the subset W -aggregation procedure will work for any value of p .

3.2.2 Choice of the precision of each \widehat{W}_u

We have seen in Section 3.1.2 that the random-permutation W -aggregation procedure adapts the cost of each \widehat{W}_u with u . This cost is random but the

average cost is the optimal cost for minimizing $\text{Var}(\widehat{\eta}_i)$. We want to use this idea for the subset W -aggregation procedure: we will adapt the cost of each \widehat{W}_u in order to minimize the variance. However, for the subset W -aggregation procedure, we estimate all the $(\widehat{\eta}_i)_{i \in [1:p]}$ at the same time. Thus, we choose to minimize $\sum_i \text{Var}(\widehat{\eta}_i)$. In the following proposition, we give the best choice of the costs $(N_u)_u$ to minimize this sum of variances.

Proposition 5. *Under Assumption 2 (where \widetilde{W}_u in this assumption corresponds to \widehat{W}_u in Equation (10)), the solution of the relaxed program (i.e. the problem without the constraint of letting the $(N_u)_u$ be integers)*

$$\min_{(N_u)_{\emptyset \subsetneq u \subsetneq [1:p]}} \sum_{i=1}^p \text{Var}(\widehat{\eta}_i) \quad \text{subject to} \quad \sum_{\emptyset \subsetneq u \subsetneq [1:p]} N_u = N_{tot} \quad (11)$$

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

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

Usually, we do not know the values of $\text{Var}(\widehat{W}_u^{(1)})$. Practically, 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 long to compute. So, we do the following approximations in practice:

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

So, in the following, with the subset W -aggregation procedure, we will choose N_u as the closest integer to $N_{tot} \binom{p}{|u|}^{-1} (p-1)^{-1}$. In this way, for a fixed total cost, we take the costs $(N_u)_u$ near the optimal choice to minimize $\sum_i \text{Var}(\widehat{\eta}_i)$. Hence, the parameter N_{tot} is now the only parameter left to chose.

Remark 5. *With this method, the real total cost $\sum_{\emptyset \subsetneq u \subsetneq [1:p]} N_u$ 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 have the desired total cost $\sum_{\emptyset \subsetneq u \subsetneq [1:p]} N_u$.*

Remark 6. *For example, if the number of inputs is $p = 100$ and we want a total cost of 10^6 . The previous results suggest to choose $N_u = 1700$ if $|u| = 1$ or $|u| = p-1$, $N_u = 34$ if $|u| = 2$ or $|u| = p-2$, $N_u = 1$ if $|u| = 3$ or $|u| = p-3$ and $N_u = 0$ if $3 < |u| < p-3$. If $N_u = 0$, we take $\widehat{W}_u = 0$. So, there are 333500 quantities W_u to estimate, instead of $2^{100} - 2 \approx 10^{30}$ if we wanted to estimate all the W_u .*

3.2.3 Consistency

A straightforward consequence of the subset W -aggregation procedure and Equation (10) is that the consistency of $(\widehat{W}_u)_{u \subset [1:p]}$ implies the consistency of $(\widehat{\eta}_i)_{i \in [1:p]}$.

Proposition 6. *Assume that for all u , we have consistent (resp. strongly consistent) estimators \widehat{W}_u when N_u goes to $+\infty$, where N_u is the cost of \widehat{W}_u . If we use the subset W -aggregation procedure with the choice of $(N_u)_u$ given in Section 3.2.2, the estimators of Shapley effects are consistent (resp. strongly consistent) when N_{tot} goes to $+\infty$ (where N_{tot} is the total cost of the subset W -aggregation procedure).*

4 Complete algorithms

We have seen in Section 3 two different procedures to estimate the Shapley effects considering that estimates of the conditional elements $(W_u)_u$ can be obtained. We explain now how to estimate these $(W_u)_u$ in a restricted setting and we give numerical experiments to compare the various estimators in this framework. The restricted setting is the following: as in [18], we will assume that for any u and x_u , it is feasible to compute an i.i.d. sample of law X_{-u} conditionally to $X_u = x_u$. Moreover, we assume that we have access to the computer code of f . 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 it requires.

Since W_\emptyset and $W_{[1:p]}$ are equal to 0 and $\text{Var}(Y)$ respectively, we assume that $|u| \in [1 : p - 1]$.

4.1 Estimators of W_u

To estimate W_u , we suggest two different estimators. The first consists in a double Monte-Carlo procedure to estimate E_u , and it is the estimator used in the algorithm of [18]. The other one is the well-known Pick-and-Freeze estimator (see [8] for the first definition, [6, 7] for theoretical studies) for V_u , that we extend to the case where the inputs variables $(X_i)_i$ are not independent.

4.1.1 Double Monte-Carlo

A first way to estimate $E_u = \text{E}(\text{Var}(Y|X_u))$ is using double Monte-Carlo: one of size N_I for the conditional variance, one other of size N_O for the expectation. Thus, the estimator of E_u suggested in [18] is

$$\widehat{E}_{u,MC} := \frac{1}{N_O} \sum_{n=1}^{N_O} \frac{1}{N_I - 1} \sum_{k=1}^{N_I} \left(f(X_{-u}^{(n)}, X_u^{(n,k)}) - \overline{f(X_{-u}^{(n)})} \right)^2, \quad (12)$$

where $\overline{f(X_{-u}^{(n)})} := N_I^{-1} \sum_{k=1}^{N_I} f(X_{-u}^{(n)}, X_u^{(n,k)})$, $(X_{-u}^{(n)})_n$ is an i.i.d. sample of law X_{-u} and $(X_u^{(n,k)})_k$ conditionally to $X_{-u}^{(n)}$ is i.i.d. of law X_u conditionally to

$X_{-u} = X_{-u}^{(n)}$. We will take $N_I = 3$, as suggested in [18]. Thus, this estimator has a cost (number of evaluations of f) of $3N_O$.

Remark 7. *It is an unbiased estimator of $E_u = E(\text{Var}(Y|X_u))$.*

Remark 8. *The algorithm of [18] is the combination of the random-permutation W -aggregation procedure with the double Monte-Carlo estimator $\widehat{E}_{u,MC}$, and it is suggested to choose $N_O = 1$ and $N_I = 3$.*

According to Propositions 4 and 6, the random-permutation W -aggregation procedure and the subset W -aggregation procedure are consistent when taking the double Monte-Carlo estimator $\widehat{W}_u = \widehat{E}_{u,MC}$.

4.1.2 Pick-and-freeze

We now give a second estimator of W_u : the pick-and-freeze estimator for V_u . Remark that

$$V_u = \text{Var}(E(Y|X_u)) = E(E(Y|X_u)^2) - E(Y)^2.$$

Remark that $E(Y)$ is easy to estimate so we do not focus on the estimation of $E(Y)$ (for the numerical applications, we will take the empirical mean). It remains to estimate $E(E(Y|X_u)^2)$, which seems to be complicated. We prove the following proposition that enables to simplify the formulation of this quantity.

Proposition 7. *Let $X = (X_u, X_{-u})$ and $X^u = (X_u, X'_{-u})$ of law $\mathcal{L}(X)$ such that, a.s. $\mathbb{P}_{(X_{-u}, X'_{-u})|X_u=x_u} = \mathbb{P}_{X_{-u}|X_u=x_u} \otimes \mathbb{P}_{X'_{-u}|X_u=x_u}$. We have*

$$E(E(Y|X_u)^2) = E(f(X)f(X^u)). \quad (13)$$

Remark that Proposition 7 enables to write a double expectation as one simple expectation, that we estimate by a simple Monte-Carlo. Thus, we suggest the pick-and-freeze estimator:

$$\widehat{V}_{u,PF} := \frac{1}{N_O} \sum_{n=1}^{N_O} f(X_u^{(n)}, X_{-u}^{(n,1)}) f(X_u^{(n)}, X_{-u}^{(n,2)}) - E(Y)^2, \quad (14)$$

where $(X_u^{(n)})_n$ is an i.i.d. sample of law X_u and where $(X_{-u}^{(n,1)}, X_{-u}^{(n,2)})$ conditionally to $X_u^{(n)}$ is i.i.d. of law X_{-u} conditionally to $X_u = X_u^{(n)}$. This estimator has a cost of $2N_O$.

As we have seen in Section 3.1.3, when we use the random-permutation W -aggregation procedure, we choose $N_O = 1$. According to Propositions 4 and 6, the random-permutation W -aggregation procedure and the subset W -aggregation procedure are consistent with the Pick-and-Freeze estimator $\widehat{W}_u = \widehat{V}_{u,PF}$.

4.2 Numerical comparison of the different algorithms

In this section, we carry out numerical experiments on the different algorithms in the restricted framework (where the conditional samples are available). Recall that in Section 3, we have seen two W -aggregation procedures for the Shapley effects using estimators of the $(W_u)_u$: the random-permutation W -aggregation procedure (used in the algorithm of [18]) and the subset W -aggregation procedure. In Section 4.1, we have seen two estimators of the $(W_u)_u$: double Monte-Carlo (used in the algorithm of [18]) and Pick-and-Freeze. To sum up, we have four consistent estimators of the Shapley effects:

- subset W -aggregation procedure with double Monte-Carlo, that we write "subset double Monte-Carlo";
- subset W -aggregation procedure with Pick-and-Freeze, that we write "subset Pick-and-Freeze";
- random-permutation W -aggregation procedure with double Monte-Carlo, that we write "random-permutation double Monte-Carlo", and which is the already existing algorithm of [18];
- random-permutation W -aggregation procedure with Pick-and-Freeze, that we write "random-permutation Pick-and-Freeze".

To compare these estimators, we use the linear Gaussian framework: $X \sim \mathcal{N}(\mu, \Gamma)$ and $Y = \sum_{i=1}^p \beta_i X_i$. In this case, the theoretical values are easily computable (see [15, 10, 2]). We choose $p = 10$, $\beta_i = 1$ for all i and $\Gamma = A^T A$ where $A \in \mathcal{M}_p$ is a realisation of p^2 i.i.d. Gaussian variables with zero mean and unit variance. To compare these different estimators, we fix a total cost 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 1000 realizations of each estimator.

In Figure 2, we plot the sum over i of the quadratic risk: $\sum_{i=1}^p \mathbb{E}((\hat{\eta}_i - \eta_i)^2)$ (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 the estimator double Monte-Carlo is better than the estimator Pick-and-freeze. To conclude, we improved the already existing algorithm of [18] (random-permutation W -aggregation procedure with double Monte-Carlo) by the estimator given by the subset W -aggregation procedure with double Monte-Carlo.

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

In Section 4, we have considered a restricted framework: we assumed that for all u and all x_u , we could generate an i.i.d. sample of law X_{-u} conditionally to $X_u = x_u$. However, in many cases, we can not generate such samples, as we only

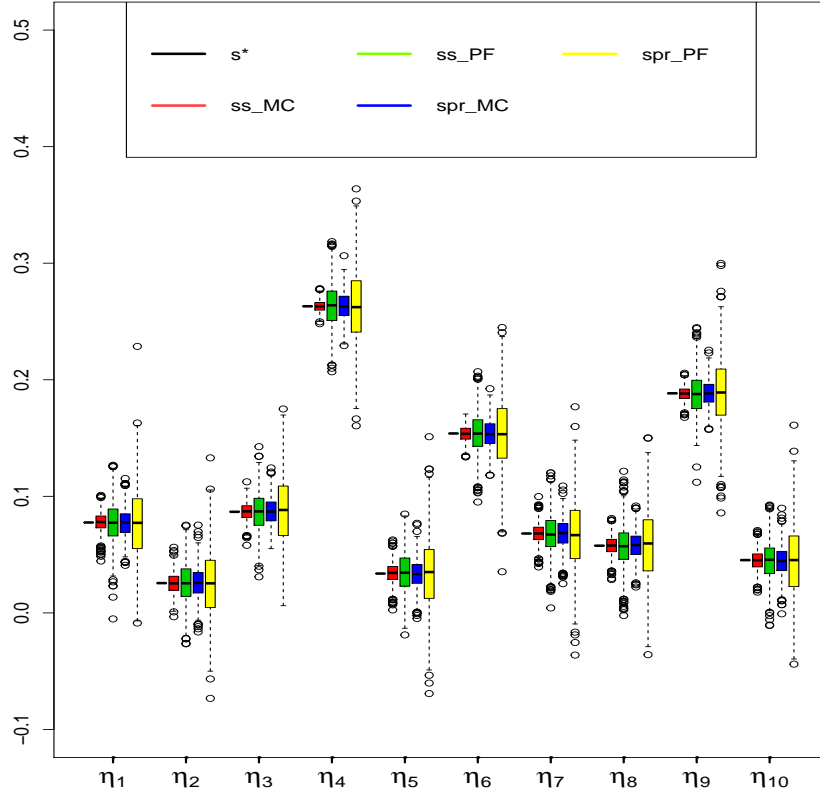


Figure 1: Estimation of the Shapley effects in the linear Gaussian framework. In black (s^*) we show the theoretical values, in red (ss_MC) the estimates from the subset W -aggregation procedure with the double Monte-Carlo estimator, in green (ss_PF) the estimates from the subset W -aggregation procedure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

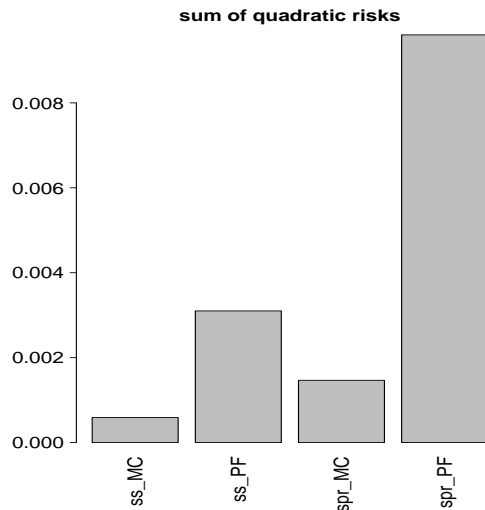


Figure 2: Sum over i of the estimated quadratic risks of the four estimators of the Shapley effects in the linear Gaussian framework.

observe an i.i.d. sample of X . In this section, we assume that we just observe a sample of X and that we have access to the computer code f . We extend the double Monte-Carlo and Pick-and-Freeze estimator 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.

5.1 Estimators of W_u

As far as we know, only [20] suggest an estimator of W_u when we only observe an i.i.d. sample and when the input variables can be dependent, and only for V_u with $|u| = 1$. The estimator suggested in [20] 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 $u \subset [1 : p]$). In this section we introduce two consistent estimators of W_u when we observe only an i.i.d. sample of 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 the i.i.d. sample of law X_{-u} conditionally to $X_u = x_u$ by the observations $(X_{-u}^{(n_1)}, \dots, X_{-u}^{(n_k)})$ such that $(X_u^{(n_1)}, \dots, X_u^{(n_k)})$ are the k nearest neighbours of x_u . For each estimator, we give the consistency and the rate of convergence.

5.1.1 Double Monte-Carlo

In order to define the double Monte-Carlo estimator of W_u with only an i.i.d. sample of X , we introduce the following notations.

Let $N \in \mathbb{N}$ and $(X_n)_{n \leq N}$ be an i.i.d. sample of X . We write $(s(l))_{l \leq N_O}$ a sample of uniformly distributed integers in $[1 : N]$ (with or without replacement) independent of the other random variables. Let us write $k(l, n)$ for the index of the observation such that $X_{-u}^{(k(l, n))}$ is the (or one of the) n -th closest element to $X_{-u}^{(l)}$ in $(X_{-u}^{(i)})_{i \leq N}$, and such that $(k_N(l, n))_{n \leq N_I}$ are two by two distinct. Then, we define two slightly different versions of the double Monte-Carlo estimator by

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

and

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

with

$$\widehat{E}_{u, s(l), MC}^{mix} = \frac{1}{N_I - 1} \sum_{i=1}^{N_I} \left[f \left(X_{-u}^{(s(l))}, X_u^{(k(s(l), i))} \right) - \frac{1}{N_I} \sum_{h=1}^{N_I} f \left(X_{-u}^{(s(l))}, X_u^{(k(s(l), h))} \right) \right]^2$$

and

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

Remark 9. *The index $k(l, n)$ could be not well-defined if there exist different observations $X_{-u}^{(i)}$ at equal distance from $X_{-u}^{(l)}$. In this case, we will choose $k(l, n)$ uniformly over the indices of these observations, as it is explained in Theorem 1. However, when X_{-u} is absolutely continuous with respect to the Lebesgue measure, this situation can not happen and $k(l, n)$ is uniquely defined.*

The index $k(l, n)$ depends on N but this dependency is implicit. However, to avoid confusion, we could write $k_N(l, n)$.

The double Monte-Carlo estimator has two sums: one of size N_I for the conditional variance, one other of size N_O 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 10. *The integer N is the size of the sample of X (that enables us to estimate implicitly its law through the nearest neighbours) and the integer N_O is the accuracy of the estimator $\widehat{W}_{u, MC}$ from the estimated law of X . Of course, it would be intuitive to take $N_O = N$ and $(s(l))_l$ without replacement,*

but this framework would not be general enough for the random-permutation W -aggregation procedure (in which there are replacements and the accuracy \tilde{N}_u is a random variable, as it is explained in Section 3.1.2) or for the subset W -aggregation procedure (in which the accuracy N_u of $\widehat{W}_{u,PF}$ depends on u).

Now that we defined the 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 (5.1.1) or (5.1.1). In the asymptotic results below, N_I is fixed and N and N_O go to infinity.

Theorem 1. *Assume that for all i , (E_i, d_i) is a Polish space, that conditionally to $(X_{-u}^{(n)})_n$, the $k_N(l, i)$ are uniformly chosen over the indices of all the i -th nearest neighbours of $X_{-u}^{(l)}$ in $(X_{-u}^{(n)})_{n \in [1:N]}$, and such that for all N , $(k_{N'}(l, i))_{i \leq N_I, N' \neq N} \perp\!\!\!\perp \sigma((X_u^{(n)})_n, (k_N(l, i))_{i \leq N_I})$ and $(k_N(l, i))_{i \leq N_I} \perp\!\!\!\perp \sigma((X_u^{(n)})_n, (k_N(l', i))_{i \leq N_I})$ for all $l \neq l'$. Assume that $X = (X_u, X_{-u})$ has a continuous density f_X with respect to a finite measure $\mu = \mu_u \otimes \mu_{-u}$. If f is bounded, then $\widehat{E}_{u,MC}$ is consistent when N and N_O go to $+\infty$.*

Furthermore, with additional regularity assumptions, we can give the rate of convergence of these estimators in Theorems 2 and 1.

Theorem 2. *Assume that f is \mathcal{C}^1 , E is compact, X has a density f_X with respect to the Lebesgue measure λ_p on \mathbb{R}^p such that λ_p a.s. we have $0 < C_{\inf} \leq f_X \leq C_{\sup} < +\infty$ and such that f_X is Lipschitz continuous. Then, for all $\varepsilon > 0$, $\varepsilon' > 0$, there exist fixed constants $C_{\sup}^{(1)}(\varepsilon')$ and $C_{\sup}^{(2)}$ such that*

$$\mathbb{P}\left(\left|\widehat{E}_{u,MC} - E_u\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \left(\frac{C_{\sup}^{(1)}(\varepsilon')}{N^{\frac{1}{p-|u|}-\varepsilon'}} + \frac{C_{\sup}^{(2)}}{N_O} \right). \quad (17)$$

Corollary 1. *Under the assumptions of Theorem 2, choosing $N_O \geq CN^{1/(p-|u|)}$, 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 $x_{-u} \mapsto \text{Var}(f(X)|X_{-u} = x_{-u})$. The estimation problem is high-dimensional when $|u|$ decreases.

5.1.2 Pick-and-Freeze

We now give similar results for the Pick-and-Freeze estimators. Let $N \in \mathbb{N}$ and $(X_n)_{n \leq N}$ be an i.i.d. sample of X . We write $(s(l))_{l \leq N_O}$ a sample of random integers uniformly distributed on $[1 : N]$ (with or without replacement) independent of the other random variables. Let $k(m, 2)$ be the (or one of the)

index in $(X_u^{(n)})_{n \leq N}$ of the second nearest neighbour of $X_u^{(m)}$, thus, the index of the nearest neighbour of $X_u^{(m)}$ different from m . Notice that in Section 5.1.2, $k(m, 2)$ refers to distances relative to X_u , while $k(l, n)$ refers to distances relative to X_{-u} in Section 5.1.1. Assume that $E(Y)$ is known. Then, we define two slightly different versions of the Pick-and-Freeze estimator by

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

and

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

with

$$\widehat{V}_{u,s(l),PF}^{mix} = f\left(X^{(s(l))}\right) f\left(X_u^{(s(l))}, X_{-u}^{(k(s(l),2))}\right) - E(Y)^2 \quad (18)$$

and

$$\widehat{V}_{u,s(l),PF}^{knn} = f\left(X^{(s(l))}\right) f\left(X^{(k(s(l),2))}\right) - E(Y)^2. \quad (19)$$

Remark 11. *The index $k(m, 2)$ depends on N but this dependency is implicit. However, to avoid confusion, we could write $k_N(m, 2)$.*

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 Corollary 2. We let $\widehat{V}_{u,PF}$ be given by (5.1.2) or (5.1.2).

Theorem 3. *Assume that for all i , (E_i, d_i) is a Polish space, that conditionally to $(X_u^{(n)})_n$, the $k(l, 2)$ are chosen uniformly over all the indices of the nearest neighbours of $X_u^{(l)}$ in $(X_u^{(n)})_{n \in [1:N] \setminus \{l\}}$, such that for all $l \neq l'$, $k_N(l, 2) \perp\!\!\!\perp \left((X_{-u}^{(n)})_n, k_N(l', 2)\right)$ and such that*

$(k_{N'}(l, 2))_{N' \neq N} \perp\!\!\!\perp \left((X_{-u}^{(n)})_n, k_N(l, 2)\right)$. Assume that $X = (X_u, X_{-u})$ has a continuous density f_X with respect to a finite measure $\mu = \mu_u \otimes \mu_{-u}$. If f is bounded, then $\widehat{V}_{u,PF}$ is consistent when N and N_O go to $+\infty$.

Theorem 4. *Assume that f is C^1 , E is compact, X has a density f_X with respect to the Lebesgue measure λ_p on \mathbb{R}^p such that λ_p a.s. we have $0 < C_{\inf} \leq f_X \leq C_{\sup} < +\infty$ and such that f_X is Lipschitz continuous. Then, if $|u| = 1$, for all $\varepsilon > 0$, $\varepsilon' > 0$,*

$$\mathbb{P}\left(\left|\widehat{V}_{u,PF} - V_u\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \left(\frac{C_{\sup}^{(1)}(\varepsilon')}{N^{1-\varepsilon'}} + \frac{C_{\sup}^{(2)}}{N_O}\right), \quad (20)$$

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

$$\mathbb{P}\left(\left|\widehat{V}_{u,PF} - V_u\right| > \varepsilon\right) \leq \frac{C_{\sup}^{(3)}}{\varepsilon^2} \left(\frac{1}{N^{\frac{1}{|u|}}} + \frac{1}{N_O}\right), \quad (21)$$

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

Corollary 2. *Under the assumptions of Theorem 4, choosing $N_O \geq CN^{1/|u|}$, 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.

5.2 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_O of the $(\widehat{W}_u)_u$. Although Assumptions 1 and 2 do not hold with the estimators $\widehat{E}_{u,MC}$ and $\widehat{V}_{u,PF}$, we keep choosing $N_O = 1$ with the random-permutation W -aggregation procedure and N_O as the closest integer to $N_{tot} \binom{p}{|u|}^{-1} (p+1)^{-1}$ with the subset W -aggregation procedure. To unify notations, let $N_I = 2$ when estimators of the conditional elements $(W_u)_u$ are the Pick-and-freeze estimators (in this way, N_I is the number of the nearest neighbours). With the double Monte-Carlo estimators, let N_I be a fixed integer (for example $N_I = 3$).

Proposition 8. *Assume that for all i , (E_i, d_i) is a polish space, that conditionally to $(X_{-u}^{(n)})_n$ (resp. $(X_u^{(n)})_n$), the $k_N(l, i)$ are uniformly chosen over the indices of all the i -th nearest neighbours of $X_{-u}^{(l)}$ (resp. $X_u^{(l)}$) in $(X_{-u}^{(n)})_{n \in [1:N]}$ (resp. $(X_u^{(n)})_{n \in [1:N]}$), and such that for all N , $(k_{N'}(l, i))_{i \leq N_I, N' \neq N} \perp\!\!\!\perp \sigma((X_u^{(n)})_n, (k_N(l, i))_{i \leq N_I})$ and $(k_N(l, i))_{i \leq N_I} \perp\!\!\!\perp \sigma((X_u^{(n)})_n, (k_N(l', i))_{i \leq N_I})$ for all $l \neq l'$ (resp. such that for all $l \neq l'$, $k_N(l, 2) \perp\!\!\!\perp ((X_{-u}^{(n)})_n, k_N(l', 2))$ and such that $(k_{N'}(l, 2))_{N' \neq N} \perp\!\!\!\perp ((X_{-u}^{(n)})_n, k_N(l, 2))$). Assume that $X = (X_1, \dots, X_p)$ has a continuous density f_X with respect to a finite measure $\mu = \otimes \mu_i$. If f is bounded, then the estimator 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. $\widehat{W}_u = \widehat{V}_{u,PF}$) are consistent when N and N_{tot} go to $+\infty$.*

5.3 Numerical experiments

In this section, we compute numerically the estimators of the Shapley effects with an unknown input distribution. As in Section 4.2, we choose the linear Gaussian framework to compute the theoretical values of the Shapley effects.

We take the same parameters as in Section 4.2. The size N of the observed sample $(X^{(n)})_{n \leq N}$ 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 200 realizations of each estimator.

Remark 12. *In the linear Gaussian framework, the function f is not bounded and the assumptions of Proposition 8 do not hold. We can 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 boxplots of the 8 estimators in Figure 3 and the sums over $i \in [1 : p]$ of their quadratic risks (estimated with 200 realizations) in Figure 4. As in Section 4.2, 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". We can remark 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 $(X_n, f(X)_n)_n$.

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 law of X (results of Section 4.2) and when we just observe a sample of size 10000 (previous results of this section). These values are equal to $5.9 \cdot 10^{-3}$ when we know the law of X , to $6.6 \cdot 10^{-3}$ when we only observe the sample with $\hat{E}_{u,MC}^{mix}$ and to $7.4 \cdot 10^{-3}$ when we only observe the sample with $\hat{E}_{u,MC}^{knn}$. Thus, in dimension 10, replacing the knowledge of X by a sample of size 10000 does not seem to deteriorate significantly our estimates of the Shapley effects.

6 Conclusion

In this article, we focused on the estimation of the Shapley effects. We explained that this estimation is divided into two parts: the W -aggregation procedure and the estimation of the conditional elements $(W_u)_{u \subset [1,p]}$. Based on an

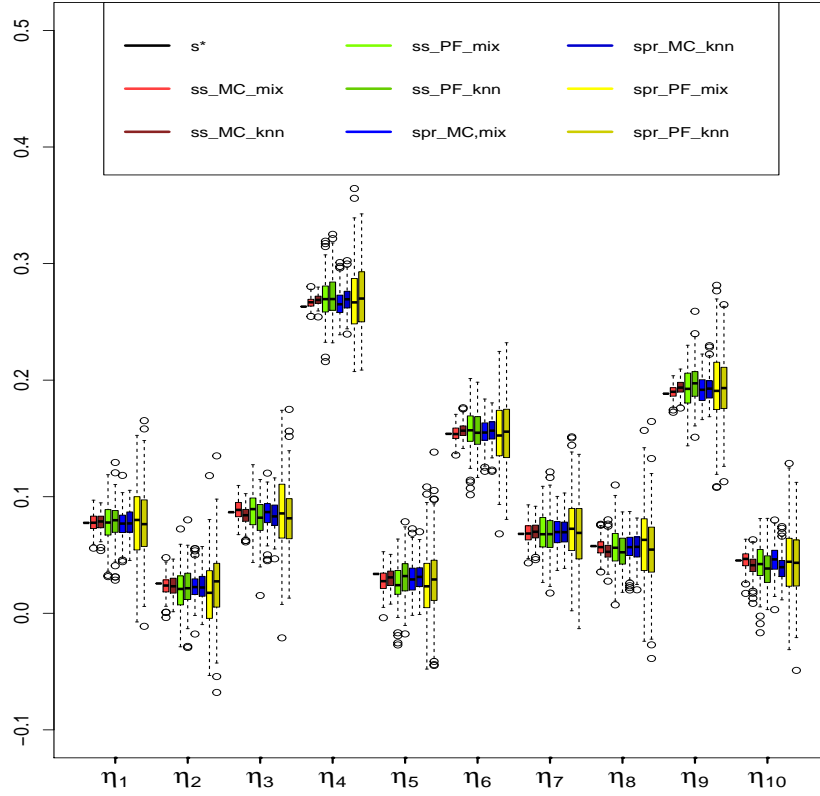


Figure 3: Estimation of the Shapley effects in the linear Gaussian framework when we only observe a sample of X . In black (s^*) we show the theoretical results, in red the estimates from the subset W -aggregation procedure with the double Monte-Carlo estimator (ss_MC_mix and ss_MC_knn), in green the 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).

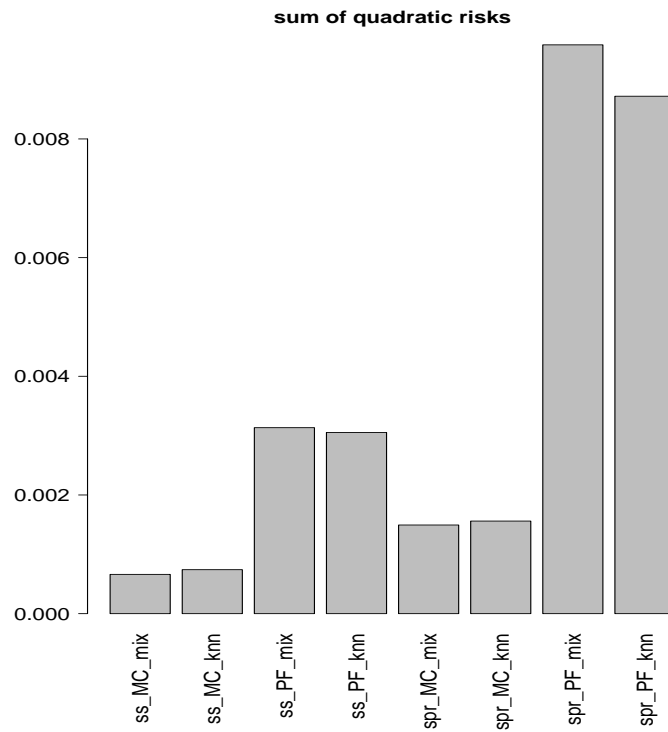


Figure 4: Sum over i of the estimated quadratic risks of the eight estimators of the Shapley effects in the linear Gaussian framework when we only observe a sample of X .

analysis of the already existing algorithm of [18], we suggested the new subset W -aggregation procedure that is theoretically more efficient. We highlighted this efficiency by numerical experiments. In a second part, we suggested various estimators of $(W_u)_u$ when the input distribution is unknown and when we just 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.

References

- [1] E. BORGONOVO, G. B. HAZEN, AND E. PLISCHKE, *A Common Rationale for Global Sensitivity Measures and Their Estimation*, Risk Analysis: An Official Publication of the Society for Risk Analysis, 36 (2016), pp. 1871–1895, <https://doi.org/10.1111/risa.12555>.
- [2] B. BROTO, F. BACHOC, M. DEPECKER, AND J.-M. MARTINEZ, *Sensitivity indices for independent groups of variables*. Jan. 2018, <https://hal.archives-ouvertes.fr/hal-01680687> (accessed 2018-05-18).
- [3] G. CHASTAING, *Indices de Sobol généralisés pour variables dépendantes*, phdthesis, Université de Grenoble, Sept. 2013, <https://tel.archives-ouvertes.fr/tel-00930229/document> (accessed 2017-03-27).
- [4] R. COLINI-BALDESCHI, M. SCARSINI, AND S. VACCARI, *Variance allocation and Shapley value*, Methodology and Computing in Applied Probability, (2016), pp. 1–15, <http://link.springer.com/article/10.1007/s11009-016-9540-5>.
- [5] S. S. FATIMA, M. WOOLDRIDGE, AND N. R. JENNINGS, *A linear approximation method for the Shapley value*, Artificial Intelligence, 172 (2008), pp. 1673–1699, [https://doi.org/Fatima, ShaheenS., Wooldridge, MichaelandJennings, NicholasR. \(2008\) AlinearaapproximationmethodfortheShapleyvalueArtificialIntelligence, 172, \(14\), pp.1673-1699. \(doi:10.1016/j.artint.2008.05.003<http://dx.doi.org/10.1016/j.artint.2008.05.003>\).](https://doi.org/Fatima,ShaheenS.,Wooldridge,MichaelandJennings,NicholasR.(2008)AlinearaapproximationmethodfortheShapleyvalueArtificialIntelligence,172,(14),pp.1673-1699.(doi:10.1016/j.artint.2008.05.003<http://dx.doi.org/10.1016/j.artint.2008.05.003>).), <https://eprints.soton.ac.uk/265802/> (accessed 2017-07-17).
- [6] F. GAMBOA, A. JANON, T. KLEIN, AND A. LAGNOUX, *Sensitivity analysis for multidimensional and functional outputs*, Electronic Journal of Statistics, 8 (2014), pp. 575–603.
- [7] F. GAMBOA, A. JANON, T. KLEIN, A. LAGNOUX, AND C. PRIEUR, *Statistical inference for Sobol pick-freeze Monte Carlo method*, Statistics, 50 (2016), pp. 881–902.

- [8] T. HOMMA AND A. SALTELLI, *Importance measures in global sensitivity analysis of nonlinear models*, Reliability Engineering & System Safety, 52 (1996), pp. 1–17, <http://www.sciencedirect.com/science/article/pii/0951832096000026>.
- [9] F. HUBERT AND S. IKONNIKOVA, *Strategic investment and bargaining power in supply chains: A Shapley value analysis of the Eurasian gas market*, (2003).
- [10] B. IOOSS AND C. PRIEUR, *Shapley effects for sensitivity analysis with dependent inputs: comparisons with Sobol’ indices, numerical estimation and applications*, arXiv:1707.01334 [math, stat], (2017), <http://arxiv.org/abs/1707.01334> (accessed 2017-11-28). arXiv: 1707.01334.
- [11] J. JACQUES, C. LAVERGNE, AND N. DEVICTOR, *Sensitivity analysis in presence of model uncertainty and correlated inputs*, Reliability Engineering & System Safety, 91 (2006), pp. 1126–1134.
- [12] T. A. MARA AND S. TARANTOLA, *Variance-based sensitivity indices for models with dependent inputs*, Reliability Engineering & System Safety, 107 (2012), pp. 115–121, <https://doi.org/10.1016/j.res.2011.08.008>, <http://www.sciencedirect.com/science/article/pii/S0951832011001724> (accessed 2018-11-26).
- [13] S. MORETTI, D. VAN LEEUWEN, H. GMUENDER, S. BONASSI, J. VAN DELFT, J. KLEINJANS, F. PATRONE, AND D. F. MERLO, *Combining Shapley value and statistics to the analysis of gene expression data in children exposed to air pollution*, BMC Bioinformatics, 9 (2008), p. 361, <https://doi.org/10.1186/1471-2105-9-361>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2556684/> (accessed 2017-11-28).
- [14] A. OWEN, *Sobol’ Indices and Shapley Value*, SIAM/ASA Journal on Uncertainty Quantification, 2 (2014), pp. 245–251, <https://doi.org/10.1137/130936233>, <http://epubs.siam.org/doi/abs/10.1137/130936233> (accessed 2017-03-29).
- [15] A. B. OWEN AND C. PRIEUR, *On Shapley value for measuring importance of dependent inputs*, SIAM/ASA Journal on Uncertainty Quantification, 5 (2017), pp. 986–1002.
- [16] L. S. SHAPLEY, *A value for n -person games. Contribution to the Theory of Games*, Annals of Mathematics Studies, 2 (1953), p. 28.
- [17] I. M. SOBOL, *Sensitivity estimates for nonlinear mathematical models*, Mathematical Modelling and Computational Experiments, 1 (1993), pp. 407–414, <http://max2.ese.u-psud.fr/epc/conservation/MODE/Sobol%20original%20Paper.pdf>.

- [18] E. SONG, B. NELSON, AND J. STAUM, *Shapley Effects for Global Sensitivity Analysis: Theory and Computation*, SIAM/ASA Journal on Uncertainty Quantification, 4 (2016), pp. 1060–1083, <https://doi.org/10.1137/15M1048070>, <http://epubs.siam.org/doi/abs/10.1137/15M1048070> (accessed 2017-03-29).
- [19] T. TJUR, *Conditional Probability Distributions*, Institute of Mathematical Statistics, University of Copenhagen, 1974. Google-Books-ID: wuenAAAA-IAAJ.
- [20] S. D. VEIGA AND F. GAMBOA, *Efficient estimation of sensitivity indices*, Journal of Nonparametric Statistics, 25 (2013), pp. 573–595, <https://doi.org/10.1080/10485252.2013.784762>, <https://doi.org/10.1080/10485252.2013.784762>, <https://arxiv.org/abs/https://doi.org/10.1080/10485252.2013.784762>.

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

To unify notations, let us write

$$\begin{aligned} \Phi_{MC}^{mix} : (x^{(1)}, \dots, x^{(N_I)}) &\mapsto \frac{1}{N_I - 1} \sum_{k=1}^{N_I} \left(f(x_{-u}^{(1)}, x_u^{(k)}) - \frac{1}{N_I} \sum_{l=1}^{N_I} f(x_{-u}^{(1)}, x_u^{(l)}) \right)^2, \\ \Phi_{MC}^{knn} : (x^{(1)}, \dots, x^{(N_I)}) &\mapsto \frac{1}{N_I - 1} \sum_{k=1}^{N_I} \left(f(x^{(k)}) - \frac{1}{N_I} \sum_{l=1}^{N_I} f(x^{(l)}) \right)^2, \\ \Phi_{PF}^{mix} : (x^{(1)}, x^{(2)}) &\longrightarrow f(x^{(1)})f(x_u^{(1)}, x_{-u}^{(2)}) - \mathbb{E}(Y)^2, \\ \Phi_{PF}^{knn} : (x^{(1)}, x^{(2)}) &\longrightarrow f(x^{(1)})f(x^{(2)}) - \mathbb{E}(Y)^2. \end{aligned}$$

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 , $\text{Var}(Y|X_{-u})$ by $\mathbb{E}(Y|X_u)^2 - \mathbb{E}(Y)^2$ 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.

Remark 13. In the definition of $\widehat{V}_{u,PF}$ given in Section 5.1.2, the random variables $\widehat{V}_{u,l,PF}$ depend on $X^{(l)}$ and $X^{k(l,2)}$, i.e.

$$\widehat{V}_{u,l,PF} = \Phi_{PF}(X^{(l)}, X^{k(l,2)}).$$

We could choose them to depend on $X^{(k(l,1))}$ and $X^{(k(l,2))}$, i.e. to replace $X^{(l)}$ by $X^{(k_N(l,1))}$. These two definitions are equivalent under the assumptions of Theorem 4 but can be different under the assumptions of Theorem 3 (more particularly in the case where $X_u^{(l)}$ is not the only nearest neighbour of $X_u^{(l)}$, i.e. there is another $X_u^{(n)}$, $n \neq l$ such that $X_u^{(n)} = X_u^{(l)}$). However, the proof of Theorem 3 works in both cases. To unify the notations of Theorems 3 and 1, we will assume in the proof that $\widehat{V}_{u,l,PF}$ depends on $X^{(k(l,1))}$ and $X^{(k(l,2))}$, i.e.

$$\widehat{V}_{u,l,PF} = \Phi_{PF}(X^{(k(l,1))}, X^{(k(l,2))}).$$

A.0.1 Proof of consistency: Theorems 1 and 3

Recall that for all $i \in [1 : p]$, (E_i, d_i) is a Polish space. Then, for all $v \subset [1 : p]$, $E_v := \times_{i \in v} E_i$ is a Polish space for the distance $d_v := \max_{i \in v} d_i$. We will write $B_v(x_v, r)$ the open ball in E_v of radius r and center x_v . Recall that the choice of the N_I -nearest neighbours could be not unique. In this case, conditionally to $(X_{-u}^{(n)})_n$, the $(k_N(l, i))_{N,l,i}$ are random variables that we choose in the following way. Conditionally to $(X_{-u}^{(n)})_n$, we choose $k_N(l, i)$ uniformly over all the indices of the i -th nearest neighbours such that the $(k_N(l, i))_{i \leq N_I}$ are two by two distinct and, for all N , $(k_{N'}(l, i))_{i \leq N_I, N' \neq N} \perp\!\!\!\perp \sigma((X_u^{(n)})_n, (k_N(l, i))_{i \leq N_I})$ and $(k_N(l, i))_{i \leq N_I} \perp\!\!\!\perp \sigma((X_u^{(n)})_n, (k_N(l', i))_{i \leq N_I})$ for all $l \neq l'$.

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

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

Lemma 1. *If \mathcal{H} is independent of $\sigma(\sigma(X), \mathcal{G})$, then*

$$\mathbb{E}(X | \sigma(\mathcal{G}, \mathcal{H})) = \mathbb{E}(X | \mathcal{G}).$$

Lemma 2. *For all measurable ϕ ,*

$$\mathcal{L}(\phi(X, Y) | X = x) = \mathcal{L}(\phi(x, Y) | X = x)$$

and if Y is independent of X , then

$$\mathcal{L}(\phi(X, Y) | X = x) = \mathcal{L}(\phi(x, Y)).$$

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

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

$$X_{-u}^{(k_N(l))} \xrightarrow[N \rightarrow +\infty]{a.s.} X_{-u}^{(1)}. \quad (22)$$

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

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

because if $x_{-u} \in \text{supp}(E_{-u})$, then $B_{-u}(x_{-u}, \varepsilon) \not\subset \text{supp}(E_{-u})^c$ and $\mathbb{P}_{X_{-u}}(B_{-u}(x_{-u}, \varepsilon)) > 0$.

Next, remark that

$$X_{-u}^{(k_N^{(l)})} \xrightarrow[N \rightarrow +\infty]{a.s.} X_{-u}^{(1)} \iff X_{-u}^{(k_N^{(2)})} \xrightarrow[N \rightarrow +\infty]{a.s.} X_{-u}^{(1)}$$

and,

$$\begin{aligned} \mathbb{P}\left(\left\{X_{-u}^{(k_N^{(2)})} \xrightarrow[N \rightarrow +\infty]{} X_{-u}^{(1)}\right\}^c\right) &= \mathbb{P}\left(\bigcup_{k \geq 1} \bigcap_{n \geq 1} d_{-u}(X_{-u}^{(n)}, X_{-u}^{(1)}) \geq \frac{1}{k}\right) \\ &\leq \sum_{k \geq 1} \mathbb{E}\left(\mathbb{P}\left(\bigcap_{n \geq 1} d_{-u}(X_{-u}^{(n)}, X_{-u}^{(1)}) \geq \frac{1}{k} \mid X_{-u}^{(1)}\right)\right) \\ &= \sum_{k \geq 1} \mathbb{E}\left(\lim_{N \rightarrow +\infty} \mathbb{P}\left(d_{-u}(X_{-u}^{(2)}, X_{-u}^{(1)}) \geq \frac{1}{k}\right)^N\right) \\ &= \sum_{k \geq 1} \mathbb{E}\left(\lim_{N \rightarrow +\infty} \left[1 - \mathbb{P}\left(d_{-u}(X_{-u}^{(2)}, X_{-u}^{(1)}) < \frac{1}{k}\right)\right]^N\right) \\ &= \sum_{k \geq 1} 0 \\ &= 0. \end{aligned}$$

□

Lemma 4. *There exists a continuous version of*

$$\mathcal{L}(X_u | X_{-u} = \cdot) : (\{f_{X_{-u}} > 0\}, d_{-u}) \longrightarrow (\mathcal{M}_1(E_u), \mathcal{T}(\text{weak}))$$

(where $\mathcal{T}(\text{weak})$ is the topology of weak convergence).

Proof. Let $x_u \in E_u$. For all $x_{-u} \in E_{-u}$ such that $f_{X_{-u}}(x_{-u}) > 0$, we have

$$f_{X_u | X_{-u} = x_{-u}}(x_u) = \frac{f(x_u, x_{-u})}{f_{X_{-u}}(x_{-u})}.$$

Let $(x_{-u}^{(n)})$ be a sequence converging to x_{-u} with $f_{X_{-u}}(x_{-u}) > 0$. There exists n_0 such that for all $n \geq n_0$, $f_{X_{-u}}(x_{-u}^{(n)}) > 0$. Thus, by continuity of f which respect to x_{-u} and $f_{X_{-u}}$, we have $f_{X_u|X_{-u}=x_{-u}}(x_u) = \lim_{n \rightarrow +\infty} f_{X_u|X_{-u}=x_{-u}^{(n)}}(x_u)$. Then, using the dominated converging Theorem,

$$\mathcal{L}(X_u|X_{-u} = x_{-u}^{(n)}) \xrightarrow[N \rightarrow +\infty]{\text{weakly}} \mathcal{L}(X_u|X_{-u} = x_{-u}).$$

□

Remark 14. The assumption "X = (X_u, X_{-u}) has a continuous density f_X with respect to a finite measure μ = μ_u ⊗ μ_{-u}" is only used in the proof of Lemma 4.

Remark 15. There exists a different proof of Lemma 4 if we assume that μ is regular. Theorem 8.1 of [19] ensures that the conditional distribution in the sense of Tjur is defined for all x_{-u} such that f_{X_{-u}} > 0 (and not only for almost all x_{-u}) and the continuity of f_{X_u|X_{-u}=x_{-u}}(x_u) with respect to x_{-u} comes from Theorem 22.1 of [19].

Remark 16. To avoid confusion, we can now define $\mathcal{L}(X_u|X_{-u} = x_{-u})$ as the probability measure of density $\frac{f(\cdot, x_{-u})}{f_{X_{-u}}(x_{-u})}$, which is defined for all (and not "almost all") x_u in {f_{X_{-u}} > 0}.

Proposition 9. If

$$\mathcal{L}(X_u|X_{-u} = \cdot) : (E_{-u}, d_{-u}) \longrightarrow (\mathcal{M}_1(E_u), \mathcal{T}(\text{weak}))$$

is continuous (where $\mathcal{T}(\text{weak})$ is the topology of weak convergence), then, for almost all $((x_{-u}^{(n)})_n, h)$, we have

$$\mathbb{E}(\widehat{E}_{u,1} \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h) \xrightarrow[N \rightarrow +\infty]{} \text{Var}(Y|X_{-u} = x_{-u}^{(1)}) \quad (23)$$

and,

$$\mathbb{E}(\widehat{E}_{u,1}) \xrightarrow[N \rightarrow +\infty]{} E_u. \quad (24)$$

Proof. Let $Z = (Z_1, \dots, Z_{N_I}) : (\Omega, \mathcal{A}) \rightarrow (E^{N_I}, \mathcal{E}^{\otimes N_I})$ measurable such that for almost all $((x_{-u}^{(n)})_n, h)$, we have

$$\mathcal{L}(Z \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h) = \bigotimes_{i=1}^{N_I} \mathcal{L}(X^{(1)} \mid X_{-u}^{(1)} = x_{-u}^{(1)}).$$

It suffices to show that, for almost all $((x_{-u}^{(n)})_n, h)$,

$$(X^{(k_N(i))})_{i \leq N_I} \xrightarrow[N \rightarrow +\infty]{\mathcal{L} \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k=h} Z. \quad (25)$$

Indeed, if (25) is true, then, using that Φ is bounded,

$$\begin{aligned}
& \mathbb{E} \left(\widehat{E}_{u,1} \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h \right) \\
&= \mathbb{E} \left[\Phi \left((X^{(k_N(i))})_{i \leq N_I} \right) \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h \right] \\
&\xrightarrow{N \rightarrow +\infty} \mathbb{E}(\Phi(Z) \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h) \\
&= \mathbb{E}(\Phi(Z) \mid X_{-u}^{(1)} = x_{-u}^{(1)}) \\
&= \text{Var}(Y \mid X_{-u} = x_{-u}^{(1)}).
\end{aligned}$$

Thus, we have (23). Furthermore, using dominated convergence theorem, integrating on $((x_{-u}^{(n)})_n, h)$, we obtain (24).

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

Lemma 5. *For almost all $(x_{-u}^{(n)})_n$,*

$$\mathcal{L}((X_u^{(n)})_n \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n) = \bigotimes_{n \geq 1} \mathcal{L}(X_u \mid X_{-u} = x_{-u}^{(n)}).$$

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

$$\mathcal{L}((\tilde{X}_u^{(n)})_n \mid (\tilde{X}_{-u}^{(n)})_n = (x_{-u}^{(n)})_n) = \bigotimes_{n \geq 1} \mathcal{L}(X_u \mid X_{-u} = x_{-u}^{(n)}).$$

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

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

$$\begin{aligned}
\mathbb{E}(\phi(\tilde{X}^{(n)})) &= \int_{\Omega} \phi(\tilde{X}^{(n)}(\omega)) d\mathbb{P}(\omega) \\
&= \int_{E_u \times E_{-u}} \phi(x_u, x_{-u}) d\mathbb{P}_{(X_u, X_{-u})}(x_u, x_{-u}) \\
&= \int_{E_{-u}} \left(\int_{E_u} \phi(x_u, x_{-u}) d\mathbb{P}_{X_u \mid X_{-u} = x_{-u}}(x_u) \right) d\mathbb{P}_{X_{-u}}(x_{-u}) \\
&= \int_E \phi(x) d\mathbb{P}_X(x).
\end{aligned}$$

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

Borel functions ϕ_1 and ϕ_2 , we have:

$$\begin{aligned}
& \mathbb{E}(\phi_1(X^{(n)})\phi_2(X^{(m)})) \\
&= \int_{E_u^2 \times E_{-u}^2} \phi_1(x_u^{(n)}, x_{-u}^{(n)})\phi_2(x_u^{(m)}, x_{-u}^{(m)})d\mathbb{P}_{(X_u^{(n)}, X_u^{(m)}, X_{-u}^{(n)}, X_{-u}^{(m)})}(x_u^{(n)}, x_u^{(m)}, x_{-u}^{(n)}, x_{-u}^{(m)}) \\
&= \int_{E_{-u}^2} \left(\int_{E_u^2} \phi_1(x_u^{(n)}, x_{-u}^{(n)})\phi_2(x_u^{(m)}, x_{-u}^{(m)})d\mathbb{P}_{(X_u^{(n)}, X_u^{(m)})|(X_{-u}^{(n)}, X_{-u}^{(m)})=(x_{-u}^{(n)}, x_{-u}^{(m)})}(x_u^{(n)}, x_u^{(m)}) \right) \\
&\quad d\mathbb{P}_{(X_{-u}^{(n)}, X_{-u}^{(m)})}(x_{-u}^{(n)}, x_{-u}^{(m)}) \\
&= \int_{E_{-u}^2} \left(\int_{E_u^2} \phi_1(x_u^{(n)}, x_{-u}^{(n)})\phi_2(x_u^{(m)}, x_{-u}^{(m)})d\mathbb{P}_{X_u|X_{-u}=x_{-u}^{(n)}} \otimes \mathbb{P}_{X_u|X_{-u}=x_{-u}^{(m)}}(x_u^{(n)}, x_u^{(m)}) \right) d\mathbb{P}_{X_{-u}}^{\otimes 2}(x_{-u}^{(n)}, x_{-u}^{(m)}) \\
&= \int_{E_{-u}^2} \left(\int_{E_u} \phi_1(x_u^{(n)}, x_{-u}^{(n)})d\mathbb{P}_{X_u|X_{-u}=x_{-u}^{(n)}}(x_u^{(n)}) \right) \\
&\quad \left(\int_{E_u} \phi_2(x_u^{(m)}, x_{-u}^{(m)})d\mathbb{P}_{X_u|X_{-u}=x_{-u}^{(m)}}(x_u^{(m)}) \right) d\mathbb{P}_{X_{-u}}^{\otimes 2}(x_{-u}^{(n)}, x_{-u}^{(m)}) \\
&= \int_{E_{-u}} \left(\int_{E_u} \phi_1(x_u^{(n)}, x_{-u}^{(n)})d\mathbb{P}_{X_u|X_{-u}=x_{-u}^{(n)}}(x_u^{(n)}) \right) d\mathbb{P}_{X_{-u}}(x_{-u}^{(n)}) \\
&\quad \left(\int_{E_u} \phi_2(x_u^{(m)}, x_{-u}^{(m)})d\mathbb{P}_{X_u|X_{-u}=x_{-u}^{(m)}}(x_u^{(m)}) \right) d\mathbb{P}_{X_{-u}}(x_{-u}^{(m)}) \\
&= \mathbb{E}(\phi_1(X^{(n)}))\mathbb{E}(\phi_2(X^{(m)})).
\end{aligned}$$

That concludes the proof of Lemma 5. \square

Lemma 6. For almost all $((x_{-u}^{(n)})_n, h)$, we have:

$$\mathcal{L}\left(\left(X_u^{(k_N(i))}\right)_{i \leq N_I} | (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h\right) = \bigotimes_{i=1}^{N_I} \mathcal{L}\left(X_u | X_{-u} = x_{-u}^{(h_N(i))}\right).$$

Proof. For all bounded Borel function ϕ ,

$$\begin{aligned}
& \mathbb{E}\left(\phi\left(\left(X_u^{(k_N(i))}\right)_{i \leq N_I} | (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h\right)\right) \\
&= \mathbb{E}\left(\phi\left(\left(X_u^{(k_N(i))}\right)_{i \leq N_I}\right) \middle| (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, (k_{N'}(i))_{i \leq N_I, N' \in \mathbb{N}^*} = (h_{N'}(i))_{i \leq N_I, N' \in \mathbb{N}^*}\right) \\
&= \mathbb{E}\left(\phi\left(\left(X_u^{(k_N(i))}\right)_{i \leq N_I}\right) \middle| (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, (k_N(i))_{i \leq N_I} = (h_N(i))_{i \leq N_I}\right) \\
&= \mathbb{E}\left(\phi\left(\left(X_u^{(h_N(i))}\right)_{i \leq N_I}\right) \middle| (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n\right),
\end{aligned}$$

using Lemmas 1 and 2 conditionally to $(X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n$. Then,

$$\begin{aligned}
& \mathbb{E} \left(\phi \left((X_u^{(h_N(i))})_{i \leq N_I} \right) \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n \right) \\
&= \int_{E_u^{N_I}} \phi(x_u^{(1)}, \dots, x_u^{(N_I)}) d\mathbb{P}_{(X_u^{(h_N(i))})_{i \leq N_I} \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n} (x_u^{(1)}, \dots, x_u^{(N_I)}) \\
&= \int_{E_u^{N_I}} \phi(x_u^{(1)}, \dots, x_u^{(N_I)}) d \bigotimes_{i=1}^{N_I} \mathbb{P}_{X_u \mid X_{-u} = x_{-u}^{(h_N(i))}} (x_u^{(1)}, \dots, x_u^{(N_I)}).
\end{aligned}$$

That concludes the proof of Lemma 6. \square

Recall that $X_{-u}^{(k_N(i))} \xrightarrow{N \rightarrow +\infty} X_{-u}^{(1)}$ \mathbb{P} -a.e., thus, for almost all $\left((x_{-u}^{(n)})_n, h \right)$,

$$x_{-u}^{(h_N(i))} \xrightarrow{N \rightarrow +\infty} x_{-u}^{(1)}.$$

Thus, using the continuity of the conditional distribution given by Lemma 4, for almost all $\left((x_{-u}^{(n)})_n, h \right)$, we have that $x_{-u}^{(1)} \in \{f_{X_{-u}} > 0\}$ and

$$\mathcal{L}(X_u \mid X_{-u} = x_{-u}^{(h_N(i))}) \xrightarrow{N \rightarrow +\infty, \text{ weakly}} \mathcal{L}(X_u \mid X_{-u} = x_{-u}^{(1)}).$$

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

$$\bigotimes_{i=1}^{N_I} \mathcal{L}(X_u \mid X_{-u} = x_{-u}^{(h_N(i))}) \xrightarrow{N \rightarrow +\infty, \text{ weakly}} \bigotimes_{i=1}^{N_I} \mathcal{L}(X_u \mid X_{-u} = x_{-u}^{(1)}) = \mathcal{L}(Z_u \mid X_{-u} = x_{-u}^{(1)}).$$

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

$$\mathcal{L} \left((X_u^{(k_N(i))})_{i \leq N_I} \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h \right) \xrightarrow{N \rightarrow +\infty, \text{ weakly}} \mathcal{L}(Z_u \mid X_{-u} = x_{-u}^{(1)}).$$

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

$$\mathcal{L} \left((X_u^{(k_N(i))})_{i \leq N_I} \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h \right) \xrightarrow{N \rightarrow +\infty, \text{ weakly}} \mathcal{L} \left(Z_u \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h \right).$$

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

$$\mathcal{L} \left((X_u^{(k_N(i))})_{i \leq N_I} \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h \right) \xrightarrow{N \rightarrow +\infty, \text{ weakly}} \mathcal{L} \left(Z \mid (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h \right),$$

that concludes the proof of Proposition 9. \square

Lemma 7. *The value of $\text{Var}(\widehat{E}_{u,1,MC})$ is bounded by $32 \frac{N_I^2}{(N_I-1)^2} \|f\|_\infty^4$.*

Proof. As f is bounded, Φ is bounded by $\frac{1}{N_I-1} \sum_{k=1}^{N_I} (2\|f\|_\infty)^2 = \frac{N_I}{N_I-1} 4\|f\|_\infty^2$ so $\text{Var}(\widehat{E}_{u,1})$ is bounded by $2\|\Phi\|_\infty^2 = 32\frac{N_I^2}{(N_I-1)^2}\|f\|_\infty^4$. \square

Proposition 10. *We have*

$$\text{Cov}(\widehat{E}_{u,1}, \widehat{E}_{u,2}) \xrightarrow{N \rightarrow +\infty} 0.$$

Proof. We use the law of total covariance

$$\text{cov}(\widehat{E}_{u,1}, \widehat{E}_{u,2}) = \text{E} \left(\text{cov} \left(\widehat{E}_{u,1}, \widehat{E}_{u,2} | X_{-u}^{(1)}, X_{-u}^{(2)} \right) \right) + \text{cov} \left(\text{E}(\widehat{E}_{u,1} | X_{-u}^{(1)}, X_{-u}^{(2)}), \text{E}(\widehat{E}_{u,2} | X_{-u}^{(1)}, X_{-u}^{(2)}) \right). \quad (26)$$

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

$$\begin{aligned} & \text{cov} \left(\text{E}(\widehat{E}_{u,1} | X_{-u}^{(1)}, X_{-u}^{(2)}), \text{E}(\widehat{E}_{u,2} | X_{-u}^{(1)}, X_{-u}^{(2)}) \right) \\ &= \text{E} \left(\text{E}(\widehat{E}_{u,1} | X_{-u}^{(1)}, X_{-u}^{(2)}) \text{E}(\widehat{E}_{u,2} | X_{-u}^{(1)}, X_{-u}^{(2)}) \right) - \text{E}(\widehat{E}_{u,1}) \text{E}(\widehat{E}_{u,2}) \\ &\xrightarrow{N \rightarrow +\infty} \text{E} \left(\text{Var}(Y | X_{-u} = X_{-u}^{(1)}) \text{Var}(Y | X_{-u} = X_{-u}^{(2)}) \right) - E_u^2 \\ &= 0. \end{aligned}$$

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

$$\text{cov} \left(\widehat{E}_{u,1}, \widehat{E}_{u,2} | X_{-u}^{(1)} = x_{-u}^{(1)}, X_{-u}^{(2)} = x_{-u}^{(2)} \right) \xrightarrow{N \rightarrow +\infty} 0. \quad (27)$$

From now on, we aim at proving (27).

Fist, we want to prove (27) for $x_{-u}^{(1)} \neq x_{-u}^{(2)}$. Using dominated convergence theorem and Proposition 9, it will suffice to show that (conditionally to $X_{-u}^{(1)} = x_{-u}^{(1)}, X_{-u}^{(2)} = x_{-u}^{(2)}$), for almost all $((x_{-u}^{(n)})_{n \geq 3}, h, h')$,

$$\text{E} \left(\widehat{E}_{u,1}, \widehat{E}_{u,2} | (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h, k' = h' \right) \xrightarrow{N \rightarrow +\infty} \text{E} \left(\widehat{E}_{u,1} | X_{-u}^{(1)} = x_{-u}^{(1)} \right) \text{E} \left(\widehat{E}_{u,2} | X_{-u}^{(2)} = x_{-u}^{(2)} \right).$$

Let

$$A := \left\{ \left((x_{-u}^{(n)})_n, h, h' \right) \mid x_{-u}^{h_N(N_I)} \xrightarrow{N \rightarrow +\infty} x_{-u}^{(1)}, x_{-u}^{h'_N(N_I)} \xrightarrow{N \rightarrow +\infty} x_{-u}^{(2)} \right\}.$$

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

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

Thus, for all $N \geq N_1$,

$$\begin{aligned}
& E(\widehat{E}_{u,1}\widehat{E}_{u,2}|(X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h, k' = h') \\
&= E\left[\Phi\left((X^{k_N(i)})_{i \leq N_I}\right)\Phi\left((X^{k'_N(i)})_{i \leq N_I}\right)\middle|(X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h, k' = h'\right] \\
&= E\left[\Phi\left((X^{k_N(i)})_{i \leq N_I}\right)\Phi\left((X^{k'_N(i)})_{i \leq N_I}\right)\middle|(X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k_N = h_N, k'_N = h'_N\right] \\
&= E\left[\Phi\left((X^{h_N(i)})_{i \leq N_I}\right)\Phi\left((X^{h'_N(i)})_{i \leq N_I}\right)\middle|(X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k'_N = h'_N\right] \\
&= E\left[\Phi\left((X^{h_N(i)})_{i \leq N_I}\right)\Phi\left((X^{h'_N(i)})_{i \leq N_I}\right)\middle|(X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n\right] \\
&= E\left[\Phi\left((x_{-u}^{h_N(i)})_{i \leq N_I}, (X_u^{h_N(i)})_{i \leq N_I}\right)h\left((x_{-u}^{h'_N(i)})_{i \leq N_I}, (X_u^{h'_N(i)})_{i \leq N_I}\right)\middle|(X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n\right] \\
&= E\left[\Phi\left((x_{-u}^{h_N(i)})_{i \leq N_I}, (X_u^{h_N(i)})_{i \leq N_I}\right)\middle|(X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n\right] \\
&\quad E\left[\Phi\left((x_{-u}^{h'_N(i)})_{i \leq N_I}, (X_u^{h'_N(i)})_{i \leq N_I}\right)\middle|(X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n\right] \\
&= E\left[\widehat{E}_{u,1}\middle|(X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k = h\right]E\left[\widehat{E}_{u,2}\middle|(X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n, k' = h'\right] \\
&\xrightarrow{N \rightarrow +\infty} E\left[\widehat{E}_{u,1}\middle|X_{-u}^{(1)} = x_{-u}^{(1)}\right]E\left[\widehat{E}_{u,2}\middle|X_{-u}^{(2)} = x_{-u}^{(2)}\right],
\end{aligned}$$

thanks to Proposition 9.

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

$$E\left(\widehat{E}_{u,1}\widehat{E}_{u,2}\middle|X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}\right) - E(\widehat{E}_{u,1}\middle|X_{-u}^{(1)} = x_{-u})E(\widehat{E}_{u,2}\middle|X_{-u}^{(2)} = x_{-u}) \xrightarrow{N \rightarrow +\infty} 0.$$

Let $\varepsilon > 0$.

Let M_N the number of observations which are equal to x_{-u} ,

$$M_N := \#\{n \leq N \mid X_{-u}^{(n)} = x_{-u}\},$$

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

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

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

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

$$\begin{aligned}
& \mathbb{P}(H_N = 0 | M_N = m, X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}) \\
&= \frac{\binom{m - N_I}{N_I}}{\binom{m}{N_I}} \\
&= \frac{(m - 2N_I + 1)(m - 2N_I + 2) \dots (m - N_I)}{(m - N_I + 1)(m - N_I + 2) \dots m} \\
&\xrightarrow{m \rightarrow +\infty} 1.
\end{aligned}$$

Thus, there exists m_1 such that

$$\alpha_{m_1} := \mathbb{P}(H_N = 0 | M_N \geq m_1, X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}) > 1 - \frac{\varepsilon}{5 \|\Phi\|_\infty^2}. \quad (28)$$

So,

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

Let

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

Conditionally to $X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}$, we know that $M_N - 2 \sim \mathcal{B}(N - 2, \mathbb{P}(X_{-u} = x_{-u}))$. Thus, there exists N_1 such that for all $N \geq N_1$,

$$\mathbb{P} \left(M_N < m_1 | X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u} \right) < \frac{\varepsilon}{5 \max(1, \|\Phi\|_\infty^2)}, \quad (29)$$

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

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

Let

$$\gamma_N := \mathbb{P} \left(M_N \geq m_1 | X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u} \right).$$

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

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

Proof. We know that, conditionally to $X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}$, $M_N \geq m_1$, $H_N = 0$, the vector $\left((X_{-u}^{(k_N(i))})_{i \leq N_I}, (X_{-u}^{(k'_N(i))})_{i \leq N_I} \right)$ is constant equal to $(x_{-u})_{i \leq 2N_I}$. It suffices to show that, conditionally to $X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}$, $M_N \geq m_1$, $H_N = 0$, the vector $\left((X_u^{(k_N(i))})_{i \leq N_I}, (X_u^{(k'_N(i))})_{i \leq N_I} \right)$ is composed of $2N_I$ i.i.d. random variables of distribution X conditionally to $X_{-u} = x_{-u}$. Let $((x_{-u}^{(n)})_n, h_N, h'_N)$ such that $X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}$, $M_N \geq m_1$ and $H_N = 0$. As $M_N \geq m_1 \geq N_I$, for all $i \leq N_I$, we have $x_{-u}^{(k_N(i))} = x_{-u} = x_{-u}^{(k'_N(i))}$. 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 ,

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

Thus,

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

that concludes the proof of Lemma 8. \square

Thus

$$\begin{aligned} & \mathbb{E} \left(\widehat{E}_{u,1} \widehat{E}_{u,2} \middle| X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}, M_N \geq m_1, H_N = 0 \right) \\ &= \mathbb{E} \left(\widehat{E}_{u,1} \middle| X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}, M_N \geq m_1, H_N = 0 \right)^2 \end{aligned}$$

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

$$\left| \mathbb{E} \left(\widehat{E}_{u,1} \widehat{E}_{u,2} \middle| X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}, M_N \geq m_1, H_N = 0 \right) - \mathbb{E} \left(\widehat{E}_{u,1} \middle| X_{-u}^{(1)} = x_{-u} \right)^2 \right| < \frac{\varepsilon}{5}. \quad (30)$$

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

$$\begin{aligned} & \left| \mathbb{E} \left(\widehat{E}_{u,1} \widehat{E}_{u,2} | X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u} \right) - \mathbb{E}(\widehat{E}_{u,1} | X_{-u}^{(1)} = x_{-u})^2 \right| \\ & \leq |\beta_N| + \left| \gamma_N \mathbb{E} \left(\widehat{E}_{u,1} \widehat{E}_{u,2} | X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}, M_N \geq m_1, H_N \geq 1 \right) (1 - \alpha_{m_1}) \right| \\ & \quad + \left| \gamma_N \alpha_{m_1} \mathbb{E} \left(\widehat{E}_{u,1} \widehat{E}_{u,2} | X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}, M_N \geq m_1, H_N = 0 \right) - \mathbb{E}(\widehat{E}_{u,1} | X_{-u}^{(1)} = x_{-u})^2 \right|. \end{aligned}$$

The upper-bound is a sum of three terms. The first one is bounded by $\varepsilon/5$ using (29) and the second one is bounded by $\varepsilon/5$ using (28). 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,

$$\begin{aligned} & \left| \mathbb{E} \left(\widehat{E}_{u,1} \widehat{E}_{u,2} | X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u} \right) - \mathbb{E}(\widehat{E}_{u,1} | X_{-u}^{(1)} = x_{-u})^2 \right| \\ & \leq \frac{\varepsilon}{5} + \frac{\varepsilon}{5} + |\gamma_N \alpha_{m_1} - 1| \|\Phi\|_\infty^2 \\ & \quad + \left| \mathbb{E} \left(\widehat{E}_{u,1} \widehat{E}_{u,2} | X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u}, M_N \geq m_1, H_N = 0 \right) - \mathbb{E}(\widehat{E}_{u,1} | X_{-u}^{(1)} = x_{-u})^2 \right| \\ & \leq \frac{3\varepsilon}{5} + (|\gamma_N - 1| \alpha_N + |\alpha_N - 1|) \|\Phi\|_\infty^2 \quad \text{using (30)} \\ & \leq \varepsilon, \end{aligned}$$

using (29) and (28). Finally, we proved that

$$\mathbb{E} \left(\widehat{E}_{u,1} \widehat{E}_{u,2} | X_{-u}^{(1)} = X_{-u}^{(2)} = x_{-u} \right) - \mathbb{E}(\widehat{E}_{u,1} | X_{-u}^{(1)} = x_{-u}) \mathbb{E}(\widehat{E}_{u,2} | X_{-u}^{(2)} = x_{-u}) \xrightarrow[N \rightarrow +\infty]{} 0.$$

Hence, (27) is proved and the proof of Proposition 10 is concluded. \square

Proposition 11. *We have*

$$\widehat{E}_u - \mathbb{E} \left(\widehat{E}_{u,1} \right) \xrightarrow[N \rightarrow +\infty, N_O \rightarrow +\infty]{\mathbb{P}} 0. \quad (31)$$

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

$$\mathbb{P} \left(\left| \widehat{E}_u - \mathbb{E} \left(\widehat{E}_u \right) \right| > \varepsilon \right) \leq \frac{\text{Var}(\widehat{E}_u)}{\varepsilon^2}. \quad (32)$$

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

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

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

thus

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

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

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

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

$$\mathbb{P} \left(\left| \widehat{E}_u - \mathbb{E} \left(\widehat{E}_u \right) \right| > \varepsilon \right) \xrightarrow[N_O \rightarrow +\infty]{N \rightarrow +\infty} 0.$$

□

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

A.0.2 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 notations. We will write C_{sup} for a generic non-negative finite constant (depending only on u , f and the distribution of 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 $C_{\text{sup}}(\varepsilon)$ for a generic non-negative finite constant depending only on ε , u , f and the distribution of X .

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

Note that with these assumptions, Φ is \mathcal{C}^1 on the compact set E and so Lipschitz continuous. For all n , we will write d for the euclidean distance on \mathbb{R}^n and $B(x, r)$ for the open ball of radius r and center x .

Remark that

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

because the Lebesgue measure of the sphere is zero. Thus, almost everywhere, for all l and all $i \neq j$,

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

Thus, the indices of the nearest neighbours $(k_N(l, i))_{l, i}$ are constant random variables conditionally to $(X_{-u}^{(n)})_n$ or to $(X_{-u}^{(n)})_{n \leq N}$. In particular, for all N and l , $k_N(l, 1) = l$. Thanks to Doob-Dynkin lemma, we can write, abusing notations, $k_N(l, i)(\omega) = k_N(l, i)[(X_{-u}^{(n)}(\omega))_n] = k_N(l, i)[(X_{-u}^{(n)}(\omega))_{n \leq N}]$. To simplify

notation, let us write $k_N(i) := k_N(1, i)$ (the index of one i -th neighbour of $X_{-u}^{(1)}$) and $k'_N(i) := k_N(2, i)$ (the index of one i -th neighbour of $X_{-u}^{(2)}$).

Remark 17. *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 E with a dominating finite measure $\mu = \bigotimes \mu_i$ such that for μ_i -almost all $x_i \in E_i$ and for all $\delta > 0$,*

$$C_{\text{inf}}\delta \leq \mu_i(B(x_i, \delta)) = \mu_i(\overline{B}(x_i, \delta)) \leq C_{\text{sup}}\delta.$$

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 , $|a_i| \leq M$, $|b_i| \leq M$ et $|a_i - b_i| \leq \varepsilon$. 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 $i \leq N$, $d(x_{-u}^{(i)}, y_{-u}^{(i)}) < \varepsilon$, then, for all $(a_{-u}^{(i)})_{i \leq N_I} \in E_{-u}^{N_I}$,*

$$\begin{aligned} & \left| \mathbb{E} \left[\Phi \left((a_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (x_{-u}^{(i)})_{i \leq N_I} \right] \right. \\ & - \left. \mathbb{E} \left[\Phi \left((a_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (y_{-u}^{(i)})_{i \leq N_I} \right] \right| \leq C_{\text{sup}}\varepsilon. \end{aligned}$$

Proof.

$$\begin{aligned} & \left| \mathbb{E} \left[\Phi \left((a_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (x_{-u}^{(i)})_{i \leq N_I} \right] \right. \\ & - \left. \mathbb{E} \left[\Phi \left((a_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (y_{-u}^{(i)})_{i \leq N_I} \right] \right| \\ & = \left| \int_{E_u^{N_I}} \Phi((a_{-u}^{(i)})_{i \leq N_I}, (x_u^{(i)})_{i \leq N_I}) \left(f_{(X_u^{(i)})_{i \leq N_I} | (X_{-u}^{(i)})_{i \leq N_I} = (x_{-u}^{(1)})_{i \leq N_I}}((x_u^{(i)})_{i \leq N_I}) \right. \right. \\ & \quad \left. \left. - f_{(X_u^{(i)})_{i \leq N_I} | (X_{-u}^{(i)})_{i \leq N_I} = (y_{-u}^{(1)})_{i \leq N_I}}((x_u^{(i)})_{i \leq N_I}) \right) d((x_u^{(i)})_{i \leq N_I}) \right| \\ & \leq C_{\text{sup}} \int_{E_u^{N_I}} \left| \prod_{i=1}^{N_I} f_{X_u | X_{-u} = x_{-u}^{(i)}}(x_u^{(i)}) - \prod_{i=1}^{N_I} f_{X_u | X_{-u} = y_{-u}^{(i)}}(x_u^{(i)}) \right| d((x_u^{(i)})_{i \leq N_I}). \end{aligned}$$

We know that,

$$\begin{aligned}
& \left| f_{X_u|X_{-u}=x_{-u}}(x_u) - f_{X_u|X_{-u}=y_{-u}}(x_u) \right| \\
& \leq \left| \frac{f_X(x_u, x_{-u})}{\int_{E_u} f_X(x'_u, x_{-u})d(x'_u)} - \frac{f_X(x_u, y_{-u})}{\int_{E_u} f_X(x'_u, y_{-u})d(x'_u)} \right| \\
& \leq \frac{1}{\int_{E_u} f_X(x'_u, x_{-u})d(x'_u)} |f_X(x_u, x_{-u}) - f_X(x_u, y_{-u})| \\
& \quad + f_X(x_u, y_{-u}) \left| \frac{1}{\int_{E_u} f_X(x'_u, x_{-u})d(x'_u)} - \frac{1}{\int_{E_u} f_X(x'_u, y_{-u})d(x'_u)} \right| \\
& \leq C_{\text{sup}} |f_X(x_u, x_{-u}) - f_X(x_u, y_{-u})| + C_{\text{sup}} |f_X(x_u, x_{-u}) - f_X(x_u, y_{-u})| \\
& \leq C_{\text{sup}} d(x_{-u}, y_{-u}).
\end{aligned}$$

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

$$\left| f_{X_u|X_{-u}=x_{-u}^{(i)}}(x_u^{(i)}) - f_{X_u|X_{-u}=y_{-u}^{(i)}}(x_u^{(i)}) \right| \leq C_{\text{sup}} \varepsilon.$$

Thus, using Lemma 9,

$$\begin{aligned}
& \left| \mathbb{E} \left[\Phi \left((a_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (x_{-u}^{(i)})_{i \leq N_I} \right] \right. \\
& \left. - \mathbb{E} \left[\Phi \left((a_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (y_{-u}^{(i)})_{i \leq N_I} \right] \right| \leq C_{\text{sup}} \varepsilon.
\end{aligned}$$

□

Lemma 11. *If for all i , $d(x_{-u}^{(i)}, y_{-u}^{(i)}) < \varepsilon$, then*

$$\begin{aligned}
& \left| \mathbb{E} \left[\Phi \left((x_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (x_{-u}^{(i)})_{i \leq N_I} \right] \right. \\
& \left. - \mathbb{E} \left[\Phi \left((y_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (y_{-u}^{(i)})_{i \leq N_I} \right] \right| \leq C_{\text{sup}} \varepsilon.
\end{aligned}$$

Proof.

$$\begin{aligned}
& \left| \mathbb{E} \left[\Phi \left((x_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (x_{-u}^{(i)})_{i \leq N_I} \right] \right. \\
& \left. - \mathbb{E} \left[\Phi \left((y_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (y_{-u}^{(i)})_{i \leq N_I} \right] \right| \\
& \leq \left| \mathbb{E} \left[\Phi \left((x_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (x_{-u}^{(i)})_{i \leq N_I} \right] \right. \\
& \left. - \mathbb{E} \left[\Phi \left((x_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (y_{-u}^{(i)})_{i \leq N_I} \right] \right| \\
& \quad + \left| \mathbb{E} \left[\Phi \left((x_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) - \Phi \left((y_{-u}^{(i)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (y_{-u}^{(i)})_{i \leq N_I} \right] \right| \\
& \leq C_{\text{sup}} \varepsilon + C_{\text{sup}} \varepsilon,
\end{aligned}$$

using Lemma 10 and using that Φ is Lipschitz continuous on E .

□

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

$$\mathbb{P}\left(d\left(X_{-u}^{(1)}, X_{-u}^{(k_N(N_I))}\right) \geq a \mid X_{-u}^{(1)}\right) \leq C_{\text{sup}} N^{N_I} (1 - C_{\text{inf}} a^{|-u|})^{N - N_I}. \quad (33)$$

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

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

We know that

$$\begin{aligned} p(a, X_{-u}^{(1)}) &= \int_{B_{-u}(X_{-u}^{(1)}, a)} f_{X_{-u}}(x_{-u}) dx_{-u} \\ &\geq C_{\text{inf}} \lambda_{|-u|} \left(B(X_{-u}^{(1)}, a)\right) \\ &\geq C_{\text{inf}} a^{|-u|}. \end{aligned}$$

Thus

$$\mathbb{P}\left(d\left(X_{-u}^{(1)}, X_{-u}^{(k_N(N_I))}\right) \geq a \mid X_{-u}^{(1)}\right) \leq C_{\text{sup}} N^{N_I} (1 - C_{\text{inf}} a^{|-u|})^{N - N_I}. \quad (34)$$

□

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

$$\mathbb{E}\left(d\left(X_{-u}^{(1)}, X_{-u}^{(k_N(N_I))}\right)\right) \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{\frac{1}{p-|u|} - \varepsilon}}, \quad (35)$$

and for all $x_{-u}^{(1)}$,

$$\mathbb{E}\left(d\left(X_{-u}^{(1)}, X_{-u}^{(k_N(N_I))}\right) \mid X_{-u}^{(1)} = x_{-u}^{(1)}\right) \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{\frac{1}{p-|u|} - \varepsilon}}. \quad (36)$$

Proof. Using Lemma 12, we have

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

and

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

Thus,

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

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

$$\mathbb{E} \left(d \left(X_{-u}^{(1)}, X_{-u}^{(k_N(N_I))} \right) \middle| X_{-u}^{(1)} \right) \leq \frac{1 + C_{\text{sup}}(\varepsilon)}{(N - N_I)^{\frac{1}{p-|\cdot|} - \varepsilon}} \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{\frac{1}{p-|\cdot|} - \varepsilon}}.$$

That concludes the proof of Lemma 13. \square

Remark 18. For the estimators $\widehat{V}_{u,PF}$, we choose only one nearest neighbour different from $X_u^{(1)}$ in $\widehat{V}_{u,1,PF}$, which is $X_u^{(k(2))}$. Thus, in the previous computation, we do not have the N^{N_I} . Thus, we can choose $\varepsilon = 0$ up to Proposition 12. Remark that this is also true for $\widehat{E}_{u,MC}$ taking $N_I = 2$ (because we always have $k(l, 1) = l$ under the assumptions of Theorem 2).

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

$$\left| \mathbb{E} \left(\widehat{E}_u \right) - E_u \right| \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{\frac{1}{p-|\cdot|} - \varepsilon}} \quad (37)$$

and for almost all $x_{-u}^{(1)}$,

$$\left| \mathbb{E} \left(\widehat{E}_{u,1} | X_{-u}^{(1)} = x_{-u}^{(1)} \right) - \text{Var}(Y | X_{-u} = x_{-u}^{(1)}) \right| \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{\frac{1}{p-|u|} - \varepsilon}}. \quad (38)$$

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

$$\begin{aligned} & \left| \mathbb{E} \left(\Phi \left((X_{-u}^{(k_N(i)[(X_{-u}^{(n)})_n]})_{i \leq N_I}, (X_u^{(k_N(i)[(X_{-u}^{(n)})_n]})_{i \leq N_I} \right) \middle| (X_{-u}^{(n)})_n = (x_{-u}^{(n)})_n \right) \right. \\ & \quad \left. - \mathbb{E} \left(\Phi(Z) | X_{-u}^{(1)} = x_{-u}^{(1)} \right) \right| \\ &= \left| \mathbb{E} \left(\Phi \left((x_{-u}^{(k_N(i)[(x_{-u}^{(n)})_n]})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (x_{-u}^{(k_N(i)[(x_{-u}^{(n)})_n]})_{i \leq N_I} \right) \right. \\ & \quad \left. - \mathbb{E} \left(\Phi \left((x_{-u}^{(1)})_{i \leq N_I}, (X_u^{(i)})_{i \leq N_I} \right) \middle| (X_{-u}^{(i)})_{i \leq N_I} = (x_{-u}^{(1)})_{i \leq N_I} \right) \right| \\ &\leq C_{\text{sup}} d \left(x_{-u}^{(k_N(N_I)[(x_{-u}^{(n)})_n]}, x_{-u}^{(1)} \right), \end{aligned}$$

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

$$\begin{aligned} \left| \mathbb{E} \left(\widehat{E}_{u,1} | X_{-u}^{(1)} = x_{-u}^{(1)} \right) - \text{Var}(Y | X_{-u} = x_{-u}^{(1)}) \right| &\leq C_{\text{sup}} \mathbb{E} \left(d \left(X_{-u}^{(1)}, X_{-u}^{(k_N(N_I))} \right) \middle| X_{-u}^{(1)} = x_{-u}^{(1)} \right) \\ &\leq C_{\text{sup}} \frac{C_{\text{sup}}(\varepsilon)}{N^{\frac{1}{p-|u|} - \varepsilon}}. \end{aligned}$$

□

In the following, to simplify notation, we could write " $X_{-u}^{(1,2)} = x_{-u}^{(1,2)}$ " for " $X_{-u}^{(1)} = x_{-u}^{(1)}$ and $X_{-u}^{(2)} = x_{-u}^{(2)}$ ".

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

$$\mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(k_N(N_I))}) \geq a \middle| X_{-u}^{(1,2)} = x_{-u}^{(1,2)} \right) \leq \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(k_{N-1}(N_I))}) \geq a \middle| X_{-u}^{(1)} = x_{-u}^{(1)} \right),$$

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

$$\mathbb{E} \left(d \left(X_{-u}^{(k_N(N_I))}, X_{-u}^{(1)} \right) \middle| X_{-u}^{(1,2)} = x_{-u}^{(1,2)} \right) \leq \mathbb{E} \left(d \left(X_{-u}^{(k_{N-1}(N_I))}, X_{-u}^{(1)} \right) \middle| X_{-u}^{(1)} = x_{-u}^{(1)} \right).$$

Proof. Let $g_N(i)$ be the index of the i -th nearest neighbour of $X_{-u}^{(1)}$ in $(X_{-u}^{(n)})_{n \in [1:N] \setminus \{2\}}$.

For almost all $(x_{-u}^{(1)}, x_{-u}^{(2)})$, we have

$$\begin{aligned}
& \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(k_N(N_I))}) \geq a \mid X_{-u}^{(1,2)} = x_{-u}^{(1,2)} \right) \\
= & \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(k_N(N_I))}) \geq a \mid X_{-u}^{(1,2)} = x_{-u}^{(1,2)}, d(x_{-u}^{(1)}, x_{-u}^{(2)}) > d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \right) \\
& \mathbb{P} \left(d(x_{-u}^{(1)}, x_{-u}^{(2)}) > d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \mid X_{-u}^{(1,2)} = x_{-u}^{(1,2)} \right) \\
& + \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(k_N(N_I))}) \geq a \mid X_{-u}^{(1,2)} = x_{-u}^{(1,2)}, d(x_{-u}^{(1)}, x_{-u}^{(2)}) \leq d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \right) \\
& \mathbb{P} \left(d(x_{-u}^{(1)}, x_{-u}^{(2)}) \leq d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \mid X_{-u}^{(1,2)} = x_{-u}^{(1,2)} \right).
\end{aligned}$$

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

$$\begin{aligned}
& \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(k_N(N_I))}) \geq a \mid X_{-u}^{(1,2)} = x_{-u}^{(1,2)}, d(x_{-u}^{(1)}, x_{-u}^{(2)}) > d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \right) \\
= & \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \geq a \mid X_{-u}^{(1,2)} = x_{-u}^{(1,2)}, d(x_{-u}^{(1)}, x_{-u}^{(2)}) > d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \right) \\
= & \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \geq a \mid X_{-u}^{(1)} = x_{-u}^{(1)}, d(x_{-u}^{(1)}, x_{-u}^{(2)}) > d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \right).
\end{aligned}$$

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

$$\begin{aligned}
& \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(k_N(N_I))}) \geq a \mid X_{-u}^{(1,2)} = x_{-u}^{(1,2)}, d(x_{-u}^{(1)}, x_{-u}^{(2)}) \leq d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \right) \\
\leq & \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \geq a \mid X_{-u}^{(1,2)} = x_{-u}^{(1,2)}, d(x_{-u}^{(1)}, x_{-u}^{(2)}) \leq d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \right) \\
= & \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \geq a \mid X_{-u}^{(1)} = x_{-u}^{(1)}, d(x_{-u}^{(1)}, x_{-u}^{(2)}) \leq d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \right).
\end{aligned}$$

Finally,

$$\begin{aligned}
& \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(k_N(N_I))}) \geq a \mid X_{-u}^{(1,2)} = x_{-u}^{(1,2)} \right) \\
\leq & \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \geq a \mid X_{-u}^{(1)} = x_{-u}^{(1)}, d(x_{-u}^{(1)}, x_{-u}^{(2)}) > d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \right) \\
& \mathbb{P} \left(d(x_{-u}^{(1)}, x_{-u}^{(2)}) > d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \mid X_{-u}^{(1)} = x_{-u}^{(1)} \right) \\
& + \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \geq a \mid X_{-u}^{(1)} = x_{-u}^{(1)}, d(x_{-u}^{(1)}, x_{-u}^{(2)}) \leq d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \right) \\
& \mathbb{P} \left(d(x_{-u}^{(1)}, x_{-u}^{(2)}) \leq d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \mid X_{-u}^{(1)} = x_{-u}^{(1)} \right) \\
= & \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(g_N(N_I))}) \geq a \mid X_{-u}^{(1)} = x_{-u}^{(1)} \right) \\
= & \mathbb{P} \left(d(x_{-u}^{(1)}, X_{-u}^{(k_{N-1}(N_I))}) \geq a \mid X_{-u}^{(1)} = x_{-u}^{(1)} \right),
\end{aligned}$$

and we proved Lemma 14. \square

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

$$\left| \text{Cov}(\widehat{E}_{u,1}, \widehat{E}_{u,2}) \right| \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}. \quad (39)$$

Proof. We use the law of total covariance,

$$\text{cov}(\widehat{E}_1, \widehat{E}_2) = \mathbb{E} \left[\text{cov} \left(\widehat{E}_1, \widehat{E}_2 \mid X_{-u}^{(1)}, X_{-u}^{(2)} \right) \right] + \text{cov} \left[\mathbb{E} \left(\widehat{E}_{u,1} \mid X_{-u}^{(1)}, X_{-u}^{(2)} \right), \mathbb{E} \left(\widehat{E}_{u,2} \mid X_{-u}^{(1)}, X_{-u}^{(2)} \right) \right]. \quad (40)$$

Part 1: First, we will bound the second term of (40). Thanks to Lemma 11, we have

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

Similarly,

$$\left| \mathbb{E} \left(\widehat{E}_{u,2} \mid X_{-u}^{(1)} = x_{-u}^{(1)}, X_{-u}^{(2)} = x_{-u}^{(2)} \right) - \text{Var} \left(Y \mid X_{-u} = x_{-u}^{(2)} \right) \right| \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}.$$

Thus, using that Φ is bounded,

$$\begin{aligned} & \left| \mathbb{E} \left(\widehat{E}_{u,1} \mid X_{-u}^{(1)} = x_{-u}^{(1)}, X_{-u}^{(2)} = x_{-u}^{(2)} \right) \mathbb{E} \left(\widehat{E}_{u,2} \mid X_{-u}^{(1)} = x_{-u}^{(1)}, X_{-u}^{(2)} = x_{-u}^{(2)} \right) \right. \\ & \quad \left. - \text{Var} \left(Y \mid X_{-u} = x_{-u}^{(1)} \right) \text{Var} \left(Y \mid X_{-u} = x_{-u}^{(2)} \right) \right| \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}. \end{aligned}$$

Moreover, using Proposition 12, we have

$$\begin{aligned} & \left| \mathbb{E} \left(\widehat{E}_{u,1} \mid X_{-u}^{(1)} = x_{-u}^{(1)} \right) \mathbb{E} \left(\widehat{E}_{u,2} \mid X_{-u}^{(2)} = x_{-u}^{(2)} \right) \right. \\ & \quad \left. - \text{Var} \left(Y \mid X_{-u} = x_{-u}^{(1)} \right) \text{Var} \left(Y \mid X_{-u} = x_{-u}^{(2)} \right) \right| \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}. \end{aligned}$$

Thus,

$$\begin{aligned} & \left| \mathbb{E} \left(\widehat{E}_{u,1} \mid X_{-u}^{(1)} = x_{-u}^{(1)}, X_{-u}^{(2)} = x_{-u}^{(2)} \right) \mathbb{E} \left(\widehat{E}_{u,2} \mid X_{-u}^{(1)} = x_{-u}^{(1)}, X_{-u}^{(2)} = x_{-u}^{(2)} \right) \right. \\ & \quad \left. - \mathbb{E} \left(\widehat{E}_{u,1} \mid X_{-u}^{(1)} = x_{-u}^{(1)} \right) \mathbb{E} \left(\widehat{E}_{u,2} \mid X_{-u}^{(2)} = x_{-u}^{(2)} \right) \right| \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{\frac{1}{p-|u|}-\varepsilon}}. \end{aligned}$$

Finally,

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

Remark 19. 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 (40): $\mathbb{E} \left[\text{cov} \left(\widehat{E}_1, \widehat{E}_2 \mid X_{-u}^{(1)}, X_{-u}^{(2)} \right) \right]$. We want to prove that

$$\begin{aligned}
& \left| \int_{E_{-u}^2} \mathbb{E} \left(\widehat{E}_{u,1} \widehat{E}_{u,2} | X_{-u}^{(1,2)} = x_{-u}^{(1,2)} \right) \right. \\
& \left. - \mathbb{E}(\widehat{E}_{u,1} | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \mathbb{E}(\widehat{E}_{u,2} | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) d\mathbb{P}_{X_{-u}}^{\otimes 2} (x_{-u}^{(1)}, x_{-u}^{(2)}) \right| \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{1-\varepsilon}}.
\end{aligned}$$

Let us write

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

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

$$\begin{aligned}
G(x_{-u}^{(1)}, x_{-u}^{(2)}) := & \left\{ (x_{-u}^{(n)})_{n \in [3:N]} \mid d(x_{-u}^{(1)}, x_{-u}^{(k_N(N_I)[(x_{-u}^{(n)})_{n \leq N}]}) < l(x_{-u}^{(1)}, x_{-u}^{(2)}), \right. \\
& \left. d(x_{-u}^{(2)}, x_{-u}^{(k'_N(N_I)[(x_{-u}^{(n)})_{n \leq N}]}) < l(x_{-u}^{(1)}, x_{-u}^{(2)}) \right\}.
\end{aligned}$$

Part 2.A: We prove the following lemmas.

Lemma 15. *There exists $C_{\text{sup}}(\varepsilon)$ such that ,*

$$\int_{E_{-u}^2} \mathbb{P} \left(d(X_{-u}^{(1)}, X_{-u}^{(k_{N-1}(N_I))}) \geq d(x_{-u}^{(1)}, x_{-u}^{(2)})/2 \mid X_{-u}^{(1)} = x_{-u}^{(1)} \right) d\mathbb{P}_{X_{-u}}^{\otimes 2} (x_{-u}^{(1)}, x_{-u}^{(2)}) \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{1-\varepsilon}}. \quad (41)$$

Proof. We divide E_{-u}^2 in $F_1 := \{(x_{-u}^{(1)}, x_{-u}^{(2)}) \in E_{-u}^2, d(x_{-u}^{(1)}, x_{-u}^{(2)}) < (N - N_I -$

1) $\frac{-1+\varepsilon}{|-u|}$ } and $F_2 := \{(x_{-u}^{(1)}, x_{-u}^{(2)}) \in E_{-u}^2, d(x_{-u}^{(1)}, x_{-u}^{(2)}) \geq (N - N_I - 1) \frac{-1+\varepsilon}{|-u|}\}$.

$$\begin{aligned}
& \int_{F_1} \mathbb{P} \left(d(X_{-u}^{(1)}, X_{-u}^{(k_{N-1}(N_I))}) \geq d(x_{-u}^{(1)}, x_{-u}^{(2)})/2 \mid X_{-u}^{(1)} = x_{-u}^{(1)} \right) d\mathbb{P}_{X_{-u}}^{\otimes 2}(x_{-u}^{(1)}, x_{-u}^{(2)}) \\
& \leq C_{\text{sup}} \lambda_{|-u|}^{\otimes 2}(F_1) \\
& \leq C_{\text{sup}} \int_{E_{-u}} \lambda_{|-u|} \left(B_{-u} \left[x_{-u}, (N - N_I - 1) \frac{-1+\varepsilon}{|-u|} \right] \right) dx_{-u} \\
& \leq C_{\text{sup}} \int_{E_{-u}} (N - N_I - 1)^{\frac{-1+\varepsilon}{|-u|}|-u|} dx_{-u} \\
& \leq C_{\text{sup}} (N - N_I - 1)^{-1+\varepsilon} \\
& \leq \frac{C_{\text{sup}}}{N^{1-\varepsilon}}.
\end{aligned}$$

Furthermore, using Lemma 12, we have

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

□

Remark 20. 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. There exists $C_{\text{sup}}(\varepsilon)$ such that

$$\int_{E_{-u}^2} \mathbb{P}(G(x_{-u}^{(1)}, x_{-u}^{(2)})^c) d\mathbb{P}_{X_{-u}}^{\otimes 2}(x_{-u}^{(1)}, x_{-u}^{(2)}) \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{1-\varepsilon}}. \quad (42)$$

Proof. Using Lemma 12, we have

$$\mathbb{P} \left(d(X_{-u}^{(k_{N-1}(N_I))}, x_{-u}^{(1)}) \geq N^{-\frac{1}{|-u|}+\delta} \mid X_{-u}^{(1)} \right) \leq C_{\text{sup}} (N - 1)^{N_I} (1 - C_{\text{inf}} N^{-1+\delta|-u|})^{N-1-N_I},$$

so

$$\mathbb{P} \left(d(X_{-u}^{(k_{N-1}(N_I))}, x_{-u}^{(1)}) \geq N^{-\frac{1}{|-u|}+\delta} \mid X_{-u}^{(1)} \right) \leq \frac{C_{\text{sup}}(\varepsilon)}{N}. \quad (43)$$

Thus, we have

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

and we conclude the proof of Lemma 16 using Lemma 15 and Equation (43). \square

For $i = 1, 2$, let B_i be the ball of center $x_{-u}^{(i)}$ and of rayon $l(x_{-u}^{(1)}, x_{-u}^{(2)})$, let p_i be the probability of B_i and N_i be the number of observations $(X_{-u}^{(n)})_{n \leq N}$ in the ball B_i . Remark that

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

We have the two following lemmas.

Lemma 17. *Conditionally to $X_{-u}^{(1,2)} = x_{-u}^{(1,2)}$, the random variable N_i is binomial $\mathcal{B}(N, p_i)$.*

Conditionally to $X_{-u}^{(1,2)} = x_{-u}^{(1,2)}$, $N_j = n_j$, the random variable N_i is binomial $\mathcal{B}(N - n_j, p_i(1 - p_j)^{-1})$.

Proof. For the first assertion, we use that the $(X_{-u}^{(n)})_n$ are i.i.d. For the second assertion, we compute $\mathbb{P}(N_i = n_i \mid X_{-u}^{(1,2)} = x_{-u}^{(1,2)}, N_j = n_j)$ with Bayes' Theorem. \square

Lemma 18. *If $N_i = n_i$, let $X_{-u}^{(M_i)}$ be the random vector composed of the n_i observations in B^i . We have:*

$$\begin{aligned} & \mathcal{L}\left(X^{(M_1)}, X^{(M_2)} | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}, N_1 = n_1, N_2 = n_2\right) \\ &= \mathcal{L}\left(X^{(M_1)} | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}, N_1 = n_1\right) \otimes \mathcal{L}\left(X^{(M_2)} | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}, N_2 = n_2\right). \end{aligned}$$

Proof. For any bounded Borell functions ϕ_1, ϕ_2 , we have

$$\begin{aligned} & \mathbb{E}\left(\phi_1(X^{(M_1)})\phi_2(X^{(M_2)}) | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}, N_1 = n_1, N_2 = n_2\right) \\ &= \frac{\mathbb{E}\left(\phi_1(X^{(M_1)})\phi_2(X^{(M_2)})\mathbb{1}_{N_1=n_1}\mathbb{1}_{N_2=n_2} | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}\right)}{\mathbb{P}(N_1 = n_1, N_2 = n_2 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)})}. \end{aligned}$$

Then, decomposing the cases and using Lemma 17, we have,

$$\begin{aligned} & \frac{\mathbb{E}\left(\phi_1(X^{(M_1)})\phi_2(X^{(M_2)})\mathbb{1}_{N_1=n_1}\mathbb{1}_{N_2=n_2} | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}\right)}{\mathbb{P}(N_1 = n_1, N_2 = n_2 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)})} \\ &= \left[\frac{1}{p_1^{n_1}}\mathbb{E}\left(\phi_1((X^{(i)})_{i \leq n_1})\mathbb{1}_{X_{-u}^{(i)} \in B_1, \forall i \leq n_1}\right)\right] \left[\frac{1}{p_2^{n_2}}\mathbb{E}\left(\phi_2((X^{(i)})_{i \leq n_2})\mathbb{1}_{X_{-u}^{(i)} \in B_2, \forall i \leq n_2}\right)\right]. \end{aligned}$$

That concludes the proof of Lemma 18. \square

Part 2.B: We aim at proving that

$$\begin{aligned} & \left| \int_{E_{-u}^2} \mathbb{E}\left(\widehat{E}_{u,1}\widehat{E}_{u,2} | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}\right) \right. \\ & \left. - \mathbb{E}(\widehat{E}_{u,1} | X_{-u}^{(1,2)} = x_{-u}^{(1,2)})\mathbb{E}(\widehat{E}_{u,2} | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) d\mathbb{P}_{X_{-u}}^{\otimes 2}(x_{-u}^{(1)}, x_{-u}^{(2)}) \right| \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{1-\varepsilon}}. \end{aligned}$$

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

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

On the other hand, we have

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

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

$$\begin{aligned} \sum_{n_1, n_2 = N_I}^N & \left| \mathbb{P}(N_1 = n_1, N_2 = n_2 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \right. \\ & \left. - \mathbb{P}(N_1 = n_1 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \mathbb{P}(N_2 = n_2 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \right| \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{1-\varepsilon}}. \end{aligned}$$

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

$$\begin{aligned} A(x_{-u}^{(1)}, x_{-u}^{(2)}) & := \sum_{n_1, n_2 = N_I}^{K_N} \left| \mathbb{P}(N_1 = n_1, N_2 = n_2 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \right. \\ & \quad \left. - \mathbb{P}(N_1 = n_1 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \mathbb{P}(N_2 = n_2 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \right|, \\ B(x_{-u}^{(1)}, x_{-u}^{(2)}) & := \sum_{\substack{n_1, n_2 = N_I, \\ n_1 > K_N \text{ or } n_2 > K_N}}^N \left| \mathbb{P}(N_1 = n_1, N_2 = n_2 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \right. \\ & \quad \left. - \mathbb{P}(N_1 = n_1 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \mathbb{P}(N_2 = n_2 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \right|. \end{aligned}$$

Let us bound these two terms.

First, we have

$$\begin{aligned} A(x_{-u}^{(1)}, x_{-u}^{(2)}) & = \sum_{n_1, n_2 = N_I}^{K_N} \mathbb{P}(N_1 = n_1 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \mathbb{P}(N_2 = n_2 | N_1 = n_1, X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \\ & \quad \times \left| 1 - \frac{\mathbb{P}(N_2 = n_2 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)})}{\mathbb{P}(N_2 = n_2 | N_1 = n_1, X_{-u}^{(1,2)} = x_{-u}^{(1,2)})} \right|. \end{aligned}$$

Thus, it suffices to bound

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

Thus, it suffices to show

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

Thanks to Lemma 17, we have,

$$\begin{aligned}
& \log \left(\frac{\mathbb{P}(N_2 = n_2 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)})}{\mathbb{P}(N_2 = n_2 | N_1 = n_1, X_{-u}^{(1,2)} = x_{-u}^{(1,2)})} \right) \\
&= \log \left(\frac{N(N-1)\dots(N-n_1+1)}{(N-n_2)(N-n_2-1)\dots(N-n_2-n_1+1)} \frac{(1-p_1)^{N-n_1}(1-p_2)^{N-n_2}}{(1-p_1-p_2)^{N-n_1-n_2}} \right) \\
&= \log \left(1 \left(1 - \frac{1}{N}\right) \dots \left(1 - \frac{n_1-1}{N}\right) \right) - \log \left(\left(1 - \frac{n_2}{N}\right) \left(1 - \frac{n_2+1}{N}\right) \dots \left(1 - \frac{n_2+n_1-1}{N}\right) \right) \\
&\quad (N-n_1) \log(1-p_1) + (N-n_2) \log(1-p_2) - (N-n_1-n_2) \log(1-p_1-p_2) \\
&= -\frac{n_1(n_1-1)}{2N} + n_1 O\left(\frac{n_1^2}{N^2}\right) + \frac{n_1(n_1+2n_2-1)}{2N} + n_1 O\left(\frac{(n_1+n_2)^2}{N^2}\right) \\
&\quad -(N-n_2)p_2 + (N-n_2)O(p_2^2) - (N-n_1)p_1 + (N-n_1)O(p_1^2) \\
&\quad + (N-n_1-n_2)(p_1+p_2) + (N-n_1-n_2)O((p_1+p_2)^2) \\
&= \frac{n_1 n_2}{N} + O\left(\frac{n_1^3}{N}\right) + O\left(\frac{n_1(n_1+n_2)^2}{N^2}\right) - n_2 p_1 - n_1 p_2 \\
&\quad + (N-n_2)O(p_1^2) + (N-n_1)O(p_2^2) + (N-n_1-n_2)O((p_1+p_2)^2).
\end{aligned}$$

Yet, we have

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

So,

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

Thus, we have shown that we have

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

Now, let us bound $B(x_{-u}^{(1)}, x_{-u}^{(2)})$. Remark that

$$\{(n_1, n_2) \in [N_I : N] \mid n_1 > K_N \text{ or } n_2 > K_N\} \subset ([K_N + 1 : N] \times [N_I : N]) \cup ([N_I : N] \cup [K_N + 1 : N]).$$

Thus, it suffices to bound

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

Thus, it suffices to bound

$$\sum_{n_1=K_N+1}^N \mathbb{P}(N_1 = n_1 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}).$$

We know that N_1 has a binomial distribution with parameters N and p_1 . Thus,

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

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

$$\mathbb{P}(N_1 = n_1 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \leq \mathbb{P}(N_1 = K_N + 1 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}).$$

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

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

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

$$(N - K_N)^{-N} N^N \leq C_{\text{sup}} \exp(K_N) \leq C_{\text{sup}}^{K_N}.$$

Moreover, we have

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

and so

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

Thus, we have

$$\begin{aligned}
& \sum_{n_1=K_N+1}^N \mathbb{P}(N_1 = n_1 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \\
& \leq C_{\text{sup}}(\varepsilon) e^{-K_N} (N - K_N)^{\frac{3}{2}} (K_N + 1)^{\frac{3}{2}} N^{-\frac{1}{2} + \delta | -u|} \\
& \leq \frac{C_{\text{sup}}(\varepsilon)}{N}.
\end{aligned}$$

Finally, we have

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

thus

$$\begin{aligned}
& \sum_{n_1, n_2=N_I}^N |\mathbb{P}(N_1 = n_1, N_2 = n_2 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \\
& \quad - \mathbb{P}(N_1 = n_1 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)}) \mathbb{P}(N_2 = n_2 | X_{-u}^{(1,2)} = x_{-u}^{(1,2)})| \leq \frac{C_{\text{sup}}(\varepsilon)}{N^{1-\varepsilon}}.
\end{aligned}$$

So, we have proved Proposition 13. \square

We conclude by the proof of Theorem 2.

Proof.

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

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

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

for all $\varepsilon' > 0$, thanks to Proposition 13. If $(s(l))_{l \leq N_O}$ is an sample uniform on $[1 : N]$ without replacement, then for all $\varepsilon > 0$,

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

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

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

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

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

That concludes the proof. \square

Appendix B Other proofs

Proof of Proposition 1 This computation is not new, but we write it to be self-consistent.

Proof. We know that $|\{\pi \in S_p, P_i(\pi) = u\}| = |u|(p-1-|u|)!$, so

$$\frac{1}{p} \binom{p-1}{|u|}^{-1} (V_{u+i} - V_u) = \frac{1}{p!} \sum_{\substack{\pi \in S_p, \\ P_i(\pi)=u}} (V_{P_i(\pi)+i} - V_{P_i(\pi)}).$$

Then, we have

$$\begin{aligned} \eta_i &= \frac{1}{p \text{Var}(Y)} \sum_{u \subset -i} \binom{p-1}{|u|}^{-1} (V_{u+i} - V_u) \\ &= \frac{1}{p! \text{Var}(Y)} \sum_{u \subset -i} \sum_{\substack{\pi \in S_p, \\ P_i(\pi)=u}} (V_{P_i(\pi)+i} - V_{P_i(\pi)}) \\ &= \frac{1}{p! \text{Var}(Y)} \sum_{\pi \in S_p} (V_{P_i(\pi)+i} - V_{P_i(\pi)}). \end{aligned}$$

\square

Proof of Proposition 2

Proof. Let us write $V := \text{Var}(\widehat{W}_u^{(1)})$ that does not depend on u by assumption. We find the relaxed problem

$$\min_{(N_u)_u} \text{Var}(\widehat{\eta}_i) = \frac{V}{\text{Var}(Y)^2} \sum_{u \subset -i} p_u^2 \left(\frac{1}{N_u} + \frac{1}{N_{u+i}} \right) = \frac{V}{\text{Var}(Y)^2} \sum_{u \subset [1:p]} \frac{p_u^2}{N_u},$$

subject to $\sum_u N_u = 2MN_O$. Let $U = (\mathbb{R}_+^*)^{2^p}$. Let f be the C^1 function on U defined by $f(x) = \sum_u \frac{p_u^2}{x_u}$, g be the C^1 function on U defined by $g(x) = (\sum_u x_u) - 2MN_O$. Finally, let $A = g^{-1}(\{0\})$. Using the method of Lagrange multipliers, if $f|_A$ has a local minimum in a , there exists c such that $Df(a) =$

$cDg(a)$, i.e. $\nabla f(a) = \nabla g(a)$ i.e. $\forall u, -\frac{p_u}{a_u^2} = c$ i.e. $a_u = c'p_u$. To sum up, if $f|_A$ has a local minimum, it is in a defined by

$$a_u = N_O M p_u.$$

Moreover, note that f is strictly convex and the set A is convex, thus $f|_A$ is strictly convex. Thus a is the strict global minimum point of $f|_A$. \square

Proof of Proposition 3 This proof totally arises from the appendix of [18]. The computations are the same.

Proof. Under Assumption 1, we have

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

Thus, the minimum is with $N_O = 1$. \square

Proof of Proposition 4

Proof. We only prove the second item. The first one is easier and uses the same idea. Recall that (see Section 3.1.2)

$$\hat{\eta}_i = \frac{1}{p\text{Var}(Y)} \sum_{u \subset C-i} \binom{p-1}{|u|}^{-1} (\tilde{W}_{u+i} - \tilde{W}_u)$$

with

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

Thus, for all u ,

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

where $\tilde{N}_{u,M} = \tilde{N}_{u+i,M} \sim \mathcal{B}(M, \frac{|u|!(p-1-|u|)!}{p!})$. First, remark that M goes to $+\infty$ when N_{tot} goes to $+\infty$ (recall that $N_{tot} = M(p-1)$). It suffices to show

that for all $u \subset [1 : p]$, the estimator $\omega \mapsto \widehat{W}_u^{\tilde{N}_{u,M}(\omega)}$ is consistent and we could conclude saying that for all i ,

$$\begin{aligned}\hat{\eta}_i &= \frac{1}{p\text{Var}(Y)} \sum_{u \subset -i} \binom{p-1}{|u|}^{-1} (\tilde{W}_{u \cup \{i\}} - \tilde{W}_u) \\ &\xrightarrow[N \rightarrow +\infty]{\mathbb{P}} \frac{1}{p\text{Var}(Y)} \sum_{u \subset -i} \binom{p-1}{|u|}^{-1} (W_{u \cup \{i\}} - W_u) \\ &= \eta_i.\end{aligned}$$

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,

$$\mathbb{P}(\tilde{N}_{u,M} \leq N_{O1}) = \sum_{k=0}^{N_{O1}} \binom{M}{k} p_{u-i}^k (1 - p_{u-i})^{M-k} \xrightarrow[M \rightarrow +\infty]{} 0.$$

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

$$\mathbb{P}(\tilde{N}_{u,M} \leq N_1) < \frac{\varepsilon}{2}.$$

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

$$\mathbb{P}(\tilde{N}_{u,M} \leq N_1) < \frac{\varepsilon}{2}.$$

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

$$\begin{aligned}\mathbb{P}\left(\left|\widehat{W}_u^{\tilde{N}_{u,M}} - W_u\right| > \delta\right) &\leq \mathbb{P}\left(\left|\widehat{W}_u^{\tilde{N}_{u,M}} - W_u\right| > \delta, \tilde{N}_{u,M} \geq N_1\right) + \mathbb{P}(\tilde{N}_{u,M} \leq N_1) \\ &< \varepsilon.\end{aligned}$$

□

Proof of Proposition 5

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 2, we have

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

Moreover,

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

Thus, we want to minimize

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

subject to

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

As in the proof of Proposition 2, we use the method of Lagrange multiplier. Let $U = (\mathbb{R}_+^*)^{2^p-2}$. Let f be the C^1 function on U defined by $f(x) = \sum_{\emptyset \subsetneq u \subsetneq [1:p]} \frac{C(p,u) \text{Var}(\widehat{W}_u^{(1)})}{x_u}$, let g be the C^1 function on U defined by $(\sum_{\emptyset \subsetneq u \subsetneq [1:p]} x_u) - N_{tot}$ and let $A = g^{-1}(\{0\})$. As in the proof of Proposition 2, $f|_A$ has a strict global minimum in a defined by

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

□

Proof of Proposition 7

Proof.

$$\begin{aligned}
& \mathbb{E}(f(X)f(X^u)) \\
&= \mathbb{E}(\mathbb{E}(f(X)f(X^u)|X_u)) \\
&= \mathbb{E}\left(\int_{E_{-u}^2} f(X_u, x_{-u})f(X_u, x'_{-u})d\mathbb{P}_{X_{-u}|X_u} \otimes \mathbb{P}_{X_{-u}|X_u}(x_{-u}, x'_{-u})\right) \\
&= \mathbb{E}\left(\int_{E_{-u}} f(X_u, x_{-u})d\mathbb{P}_{X_{-u}|X_u}(x_{-u}) \int_{E_{-u}} f(X_u, x'_{-u})d\mathbb{P}_{X_{-u}|X_u}(x'_{-u})\right) \\
&= \mathbb{E}(\mathbb{E}(f(X)|X_u)^2).
\end{aligned}$$

That concludes the proof of Proposition 7. \square

Proof of Corollaries 1 and 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

$$\mathbb{P}\left(N^{\frac{1}{2(p-|u|)}}^{-\delta} \left|\widehat{E}_{u,MC} - E_u\right| > \varepsilon\right) \leq \frac{C_{\text{sup}}(\delta/4)N^{\frac{1}{p-|u|}-2\delta}}{\varepsilon^2 N^{\frac{1}{p-|u|}-\delta}} \xrightarrow{N \rightarrow +\infty} 0.$$

That concludes the proof of Corollary 1. \square

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 use Proposition 6.

If we use the subset W -aggregation procedure, the consistency of the estimators of the Shapley effects comes from the second part of Proposition 4. We just have to verify the assumptions. Let $\widehat{W}_u^{(i)}$ of Proposition 4 be $\widehat{E}_{u,s(i),MC}$ or $\widehat{V}_{u,s(i),PF}$ defined in Section 5.1. Then, following the end of the proof of Theorems 1 and 3, we obtain

$$\frac{1}{N_O^2} \sum_{i,j=1}^{N_O} \text{cov}\left(\widehat{W}_u^{(i)}, \widehat{W}_u^{(j)}\right) \xrightarrow{N, N_O \rightarrow +\infty} 0,$$

and, by Proposition 9, we have

$$\mathbb{E}\left(\widehat{W}_u^{(1)}\right) \xrightarrow{N \rightarrow +\infty} W_u.$$

All the assumptions of Proposition 4 are verified. \square