Asymptotic results under multiway clustering

Laurent Davezies[*] Xavier D'Haultfœuille[†] Yannick Guyonvarch[‡]

June 25, 2018

PRELIMINARY AND INCOMPLETE

Clustering has become pervasive practice in applied economics, but somewhat surprisingly, theoretical results still lag behind. This paper aims to fill this gap, by developing a general asymptotic theory with multiway clustering. For that purpose, we first prove results on empirical processes in the presence of clustered data. Such results does not impose functional form restrictions on the dependence within a given cluster and allows for random, possibly unbounded cluster sizes. We show weak convergence of the empirical process provided that the number of clusters tends to infinity in each dimension and under conditions on the class of functions that are very similar to the i.i.d. case. This result allows us to directly extend standard asymptotic results for general classes of estimators in the i.i.d. case to this general set-up with clustering. We propose estimators of the asymptotic variance of such estimators, and exhibit conditions for their consistency. We also propose bootstrap procedure. As a proof of concept, we show convergence and asymptotic normality of GMM estimators and nonlinear functionals of the empirical cumulative distribution function (like the Gini coefficient). Finite sample properties of the estimators are explored throught few Monte-Carlo simulations.

Keyword: Multiway clustering, Empirical Process, Exchangeable Array

[*]CREST. laurent.davezies@ensae.fr
[†]CREST. xavier.dhaultfoeuille@ensae.fr
[‡]CREST. yannick.guyonvarch@ensae.fr.

1 Introduction

Supposing that samples are obtained from the realization of independent random variables is often unreasonable, because units of interest such as individuals or firms usually face common shocks. Failing to take such dependence into account generally leads to overly optimistic inference, because shocks tend to correlate observations positively (Bertrand et al., 2004). As a result, estimation of standard errors robust to clustering has become pervasive in applied economics. In particular, following Cameron et al. (2011), empirical studies now routinely report standard errors accounting for multiway clustering.

Perhaps surprisingly, however, econometric theory has somewhat lagged behind this practice. Most of the research, following the early contributions of Pfeffermann and Nathan (1981), Moulton (1986, 1990) Liang and Zeger (1986) and Arellano (1987), has focused on linear models with one-way clustering (see in particular Hansen, 2007; Bester et al., 2011; Ibragimov and Müller, 2016; Carter et al., 2017). The only paper we are aware of that provides a general treatment of clustering is the recent independent work of Hansen and Lee (2017). They show in particular weak law of large numbers, central limit theorems and consistency of variance matrix estimators under moment conditions and restrictions on the clusters. Their paper is very useful but does not include multiway clustering. This case entails substantial complications, as all units sharing at least one cluster are dependent. And while Cameron et al. (2011) and Cameron and Miller (2015) study such a case, they do not provide any convergence results. They rather focus on the estimation of the asymptotic variance matrix, assuming that the estimators are consistent and asymptotically normal.

In this paper, we consider a fairly general set-up of multiway clustering. We do not impose functional form restrictions on the dependence within a given cluster and we allow for random, possibly unbounded cluster sizes. Cluster size may also depend on the data themselves. These two latter features are important to account for cluster heterogeneity, following the terminology of Carter et al. (2017). We do maintain, on the other hand and contrary to a strand in the literature that has considered the case of a fixed number of large clusters, that the number of clusters tends to infinity in each dimension of clustering.

Given this set-up, our main result is the weak convergence of the empirical process under conditions on the class of functions that are very similar to those used in the i.i.d. case. In particular, when cluster sizes are bounded, the conditions turn out to be exactly the same. If not, we have to slightly reinforce some moment conditions. This reinforcement appears to be necessary, if only for establishing pointwise asymptotic normality. This latter point is also shown by Hansen and Lee (2017) when discussing pointwise central limit theorem with one-way clustering. On the other hand, to the best of our knowledge we are the first to establish weak convergence of the empirical process with either one-way or multiway clustering. We are also the first to show a pointwise central limit theorem for multiway clustering. Our proof of pointwise asymptotic normality relies on a Hájek projection, in a similar way as for U-statistics though our object of interest is not a usual U-statistic. To prove the weak convergence of the empirical process, we rely on symmetrization arguments akin to those developed in Arcones and Giné (1993).

A byproduct of our main result is to exhibit a new form of the asymptotic variance of means under multiway clustering. This asymptotic variance takes a somewhat complicated form because of the particular dependence created by such clustering. We show that asymptotically, only some covariances between cells, defined as intersections of clusters along all the dimensions, matter. These are the covariances between cells sharing exactly one common cluster. As other byproducts of our result on the empirical process, we show the asymptotic normality of GMM estimators and smooth functionals of the empirical distribution function. We also provide weak conditions under which the estimator of their asymptotic variance is consistent, thus making it possible to draw asymptotically valid inference.

Our paper is organized as follows. Section 2 describes the set-up of multiway clustering, and parameters of interest we consider afterwards. Section 3.1 provides our main result on the convergence of empirical processes in this set-up. We also discuss therein asymptotic variances of means. Section 4 describes two applications of this result, to GMM estimators and smooth functionals of the empirical distribution function.

2 The set up

We consider a framework where the goal is to draw inference on some theoretical parameters of the distribution of $Y \in \mathcal{Y} \subset \mathbb{R}^\ell$ in the population. For that purpose, we use a sample of identically distributed copies of this variable, but these copies are not independent, because of clustering. Specifically, we suppose to have k non-nested partitions of the population, which correspond to the different dimensions of clustering. For instance, when considering wages of individuals, such dimensions may be areas of residence and sectors of activity, in which case $k = 2$. We then denote the index of the first dimension of clustering (e.g. area of residence) by j_1, the second (e.g. sector of activity) by j_2 etc. Hereafter, the intersection of k given clusters in the different dimensions (e.g., the second area of residence and the first sector of activity) is called a cell. Cells are indexed by the k-tuple $\boldsymbol{j} = (j_1, ..., j_k)$ for $j_i = 1, ..., C_i$, where C_i denotes the number of clusters in the sample for dimension i. We let $\boldsymbol{j} \geq \boldsymbol{j}'$ to mean that $j_i \geq j'_i$ for all $i = 1, ..., k$. In the following, $\mathbf{1}$ (respectively $\mathbf{2}_i$) denotes the k-tuple $(1, ..., 1)$ (respectively the k-tuple with 2 in each entry but 1 in entry i). We also let $\boldsymbol{C} = (C_1, ..., C_k)$ and again, the numbers of observations within each cell is denoted by $N_{\boldsymbol{j}}$. The random variable corresponding to unit $\ell = 1, ..., N_{\boldsymbol{j}}$ in cell $\boldsymbol{j}$ (with $\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}$) is then denoted $Y_{\ell,j.}$.

The key properties of this k-way clustering are the following. First, for a given $\boldsymbol{j}$, the $(Y_{\ell,\boldsymbol{j}})_{\ell=1,...,N_{\boldsymbol{j}}}$ are identically distributed, but not necessarily independent across $\boldsymbol{j}$. Second,

$Y_{\ell,\boldsymbol{j}}$ and $Y_{\ell',\boldsymbol{j}'}$ are independent if $j_i \neq j'_i$ for all $i = 1, ..., k$. Our goal is to provide some tools to perform asymptotic inference on statistics based on the sample $(Y_{1,\boldsymbol{j}}, ..., Y_{N_{\boldsymbol{j}},\boldsymbol{j}})_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}$ when $\underline{C} = \min_{i\in\{1,...,k\}} C_i$ tends to infinity. We make for that purpose the following assumption, which formalizes the dependence structure we described in this set-up.

Assumption 1 (DGP)

1. *$(N_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{l\geq1})_{\boldsymbol{j}}$ are separably exchangeable. Namely, for any $(\pi_1, ..., \pi_k)$ k-tuple of permutations of $\mathbb{N}$,*

$$(N_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{\ell\geq1})_{\boldsymbol{j}\geq\boldsymbol{1}} \stackrel{d}{=} (N_{\pi_1(j_1),...,\pi_k(j_k)}, (Y_{\ell,\pi_1(j_1),...,\pi_k(j_k)})_{l\geq1})_{\boldsymbol{j}\geq\boldsymbol{1}}.$$

2. *For any $\boldsymbol{c} \in \mathbb{N}^{*k}$, $(N_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{\ell\geq1})_{\boldsymbol{j}\leq\boldsymbol{c}}$ is independent of $(N_{\boldsymbol{j}'}, (Y_{\ell,\boldsymbol{j}'})_{\ell\geq1})_{\boldsymbol{j}'\geq\boldsymbol{c}+\boldsymbol{1}}$.*

3. *$E(N_{\boldsymbol{1}}) > 0$.*

4. *$\underline{C} \to \infty$ and for all $i = 1...k$, $\underline{C}/C_i \to \lambda_i \geq 0$.*

Separate exchangeability imposes for instance, in two-way clustering, that $(N_{(1,1)}, N_{(1,2)})$ has the same distribution as $(N_{(2,1)}, N_{(2,2)})$, since data of all rows have the same distribution. But importantly, it does not impose that $(N_{(1,1)}, N_{(1,2)})$ has the same distribution as $(N_{(1,1)}, N_{(2,2)})$. This would indeed amount to neglect possible dependence along rows. The second condition imposes that any collection of cells sharing no cluster are mutually independent. Importantly, it does not impose any restriction on the distribution of $(N_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{\ell\geq1})$. First, we leave the dependence between $N_{\boldsymbol{j}}$ and the $(Y_{\ell,\boldsymbol{j}})_{\ell\geq1}$ unrestricted. Second, the dependence between the $(Y_{\ell,\boldsymbol{j}})_{\ell\geq1}$ is also unrestricted. Note also that we can allow $Y_{\ell,\boldsymbol{j}}$ to have a different distribution from $Y_{\ell',\boldsymbol{j}}$, $\ell \neq \ell'$. Finally, the condition that $\underline{C}/C_i$ tends to $\lambda_i \geq 0$ is very mild since it allows for different rates of convergence on the different dimensions of clusters.

Importantly, whereas the data generating process is defined at the cell level here, parameters of interest are rather defined at the unit level. Consider for instance that we are interested in the average size of US firms, with US states and sectors corresponding to the two dimensions of clustering. Then cells correspond to all firms within a given state and a given sector. We would typically estimate the average size by

$$\widehat{\theta} = \frac{\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} Y_{\ell,\boldsymbol{j}}}{\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}} N_{\boldsymbol{j}}}, \tag{1}$$

with $Y_{\ell,\boldsymbol{j}}$ the size of firm l in cell $\boldsymbol{j}$. This estimator is the empirical counterpart of

$$\theta_0 = \frac{\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} Y_{\ell,\boldsymbol{1}}\right)}{\mathbb{E}\left(N_{\boldsymbol{1}}\right)}, \tag{2}$$

rather than of, e.g., $\mathbb{E}\left(Y_{\ell,\boldsymbol{1}}\right)$. We study how to conduct inference on $\widehat{\theta}$ in Section 4.1.

An equivalent but more abstract way to define θ_0 is through the following equivalence:

$$\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}}(Y_{\ell,\mathbf{1}}-\theta)\right)=0 \Longleftrightarrow \theta=\theta_0.$$

This example generalizes to GMM. Let $\theta_0 \in \Theta \subset \mathbb{R}^p$ be such that for a vector-valued function $m(y,\theta) \in \mathbb{R}^L$, we have

$$\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right)=0 \Longleftrightarrow \theta=\theta_0.$$

For any symmetric random matrix $\widehat{\Xi}$ tending in probability to Ξ symmetric positive, the GMM estimator $\widehat{\theta}$ of θ_0 is then defined as:

$$\widehat{\theta}=\arg\min_{\theta\in\Theta}\left(\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} m(Y_{\ell,\boldsymbol{j}},\theta)\right)'\widehat{\Xi}\left(\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} m(Y_{\ell,\boldsymbol{j}},\theta)\right), \quad (3)$$

where the transpose of a matrix is denoted by the superscript T. We study inference on θ_0 based on $\widehat{\theta}$ in Section 4.2.

Finally, if one is interested in, e.g., the median instead of the mean of firms size, θ_0 can be defined as $F^{-1}(1/2)$, with

$$F(y)=\frac{\mathbb{E}\left[\sum_{\ell=1}^{N_1}\mathbb{1}\{Y_{\ell,\mathbf{1}}\leq y\}\right]}{\mathbb{E}(N_1)}, \quad (4)$$

More generally, we study inference on smooth functionals of F in Section 4.3.

Before considering these different classes of estimators, we provide in Section 3 some general results that are key for analyzing their asymptotic behavior.

3 Weak convergence results

3.1 Empirical processes

Let P denotes the probability measure of $(N_{\mathbf{1}},(Y_{\ell,\mathbf{1}})_{l\geq 1})$ and $\mathcal{F}$ denote a class of real-valued functions in $L_2(P)$. Let also $\Pi_C=\prod_{i=1}^k C_i$ denote the total number of cells. In this section, we study the empirical process $\mathbb{G}_C$ defined on $\mathcal{F}$ by:

$$\mathbb{G}_C f=\sqrt{\underline{C}}\left\{\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})-\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{1,\mathbf{1}})\right]\right\}.$$

Specifically, we prove that as $\underline{C}$ tends to infinity, $\mathbb{G}_C$ converges weakly to a gaussian process. While we refer to, e.g., van der Vaart and Wellner (1996) for a formal definition of weak convergence of processes, we recall that this result is stronger than pointwise asymptotic normality of $\mathbb{G}_C f$. Our result below will therefore entail central limit theorems for means of the form

$$\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}),$$

and therefore, by the delta method (considering $f(y)=y$ and $f(y)=1$), for sample averages defined by (1). But such a result is not sufficient for, e.g., asymptotic normality of GMM estimators or smooth functionals of the empirical cdf. Convergence of the whole process, on the other hand, allows one to establish such results, as shown below. Note finally that we cannot apply standard results on the empirical process for two reasons. First, the different cells are potentially dependent rather than i.i.d. Second, even if this were the case, we do not consider the usual empirical process over cells, because we also sum over a random number of units within each cell.

Before giving our main asymptotic result on $\mathbb{G}_C$, we introduce additional notation related to the class $\mathcal{F}$. An envelope of $\mathcal{F}$ is a function F satisfying $F(u)\geq \sup_{f\in\mathcal{F}}|f(u)|$. For any $\varepsilon>0$ and any norm $||.||$ on a space containing $\mathcal{F}$, $N(\varepsilon,\mathcal{F},||.||)$ denotes the minimal number of $||.||$-closed balls of radius ε needed to cover $\mathcal{F}$.[1] The norms we consider hereafter are $\|f\|_{\mu,r}=(\int|f|^r d\mu)^{1/r}$ for any $r\geq 1$ and probability measure μ on $\mathcal{Y}$. Finally, a class of measurable functions $\mathcal{F}$ is pointwise measurable if there exists a countable subclass $\mathcal{H}\subset\mathcal{F}$ such that elements of $\mathcal{F}$ are pointwise limit of elements of $\mathcal{H}$.

We consider the following standard assumptions on the class $\mathcal{F}$.

Assumption 2 *$\mathcal{F}$ is a pointwise measurable class of functions.*

Assumption 3 *The class $\mathcal{F}$ admits a P-measurable envelope F with either:*
- $\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}} F\left(Y_{\ell,\mathbf{1}}\right)\right)^2\right]<+\infty$ *and $\mathcal{F}$ is finite;*
- *or* $\mathbb{E}\left[N_{\mathbf{1}}^2\right]<+\infty$, $\mathbb{E}\left[N_{\mathbf{1}}\sum_{\ell=1}^{N_{\mathbf{1}}} F\left(Y_{\ell,\mathbf{1}}\right)^2\right]<+\infty$ *and*

$$\int_0^{+\infty}\sup_Q\sqrt{\log N\left(\varepsilon||F||_{Q,2},\mathcal{F},||.||_{Q,2}\right)}d\varepsilon<+\infty,$$

where the supremum is taken over the set of probability measures with finite support on $\mathcal{Y}$.

Assumption 2 is not necessary but usually imposed (see, e.g. Chernozhukov et al., 2014; Kato, 2017) to avoid as:measurability issues and the use of outer expectations. For further discussion about these classes, we refer to Kosorok (2006, pp.137-140). Assumptions 3 imposes a condition on what is usually referred to as the uniform entropy integral, see, e.g., van der Vaart and

[1] With a slight abuse of language, we use here the term norms in lieu of seminorms. For instance, Assumption 3 involves seminorms rather than norms. Also, we use $\|.\|$ for (semi)norms on functions and $|.|$ for norms on finite-dimensional objects. Specifically, for any vector b, $|b|$ denotes the Euclidean norm of b; and for any matrix A, $|A|$ denotes the Frobenius norm of A.

Wellner (1996). Finiteness of the uniform entropy integral is satisfied by any VC-type class of functions (see Chernozhukov et al., 2014, for a definition), or by the convex hull of such classes under some restrictions. These conditions are nearly the same as those used with i.i.d. data. The only difference is that we require $\mathbb{E}\left[N_{\mathbf{1}}^2\right] < +\infty$ and $\mathbb{E}\left[\left(N_1 \sum_{\ell=1}^{N_{\mathbf{1}}} F^2\left(Y_{\ell,\mathbf{1}}\right)\right)\right] < +\infty$ instead of $\mathbb{E}\left[F\left(Y_{1,\mathbf{1}}\right)^2\right] < +\infty$, to account for multiple units within each cell. But note that the two conditions coincide whenever $N_{\mathbf{1}}$ is bounded and the $(Y_{\ell,\mathbf{1}})_{l\geq 1}$ are identically distributed.

Theorem 3.1 *Suppose that Assumptions 1-3 hold. Then the process $\mathbb{G}_C$ converges weakly to a centered Gaussian process $\mathbb{G}$ on $\mathcal{F}$ as $\underline{C}$ tends to infinity. Moreover, the covariance kernel K of $\mathbb{G}$ satisfies:*

$$K(f_1, f_2) = \sum_{i=1}^{k} \lambda_i \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f_1(Y_{\ell,\mathbf{1}}), \sum_{\ell=1}^{N_{\mathbf{2}_i}} f_2(Y_{\ell,\mathbf{2}_i})\right).$$

Let us first summarize the proof of Theorem 3.1. Weak convergence of $\mathbb{G}_C$ holds under two main conditions. First, $(\mathbb{G}_C f_1, ..., \mathbb{G}_C f_m)$ should be asymptotically normal for any $m \in \mathbb{N}^*$ and $(f_1, ..., f_m)$ in $\mathcal{F}$. Second, one should establish asymptotic equicontinuity. Regarding finite-dimensional convergence, we proceed in several steps. To simplify the discussion, we consider here the case where $m = 1$ and two-way clustering. We first prove that there exist mutually independant variables $(U_{j_1,0}, U_{0,j_2}, U_{\boldsymbol{j}})_{j_1\geq 1, j_2\geq 1, \boldsymbol{j}\geq \mathbf{1}}$ are mutually independent such that for all $\boldsymbol{j}$,

$$(N_{\boldsymbol{j}}, (Y_{\ell,j})_{\ell\geq 1}) = \tau(U_{j_1,0}, U_{0,j_2}, U_{\boldsymbol{j}}).$$

The variable $U_{j_1,0}$ (resp. U_{0,j_i}) may be seen as a shock specific to cluster 1 (resp. 2), while $U_{\boldsymbol{j}}$ can be interpreted as a shock specific to Cell $\boldsymbol{j}$.[2] Our representation result is based on the Aldous-Hoover representation theorem (Aldous, 1981; Hoover, 1979), which itself extends de Finetti's theorem to separately exchangeable random sequences. We can improve the Aldous-Hoover representation in our context, using the fact that cells sharing no common clusters are independent.

In the second step, we consider the Hájek projection of $\mathbb{G}_C f_1$ on the set of random variables depending only on the marginal cluster specific factors:

$$\left\{\sum_{j_1=1}^{C_1} g_{j_1,0}(U_{j_1,0}) + \sum_{j_2=1}^{C_2} g_{0,j_2}(U_{0,j_2}),\ g_{j_1,0} \in L^2(U_{j_1,0}), g_{0,j_2} \in L^2(U_{0,j_2})\right\}.$$

We prove that $\mathbb{G}_C(f_1)$ gets close (in L^2) to its Hájek projection as $\underline{C} \to \infty$. Asymptotic normality then follows by a simple CLT on the Hájek projection.

[2] With more than two dimensions of clustering, the representation is similar but we have to include shocks specific to each subset of $(j_1, ..., j_k)$. For instance, with $k = 3$, we have to consider shocks such as $U_{j_1,j_2,0}$ in addition to U_{j_i} and $U_{\boldsymbol{j}}$.

To complete the proof of the theorem, we have to establish asymptotic equicontinuity. Roughly speaking, this means that whenever f_1 and f_2 are close to each other, $\mathbb{G}_C f_1 - \mathbb{G}_C f_2$ is close to zero (see, e.g., van der Vaart and Wellner, 1996, Section 2.1.2, for a formal definition). For that purpose, we prove a symmetrization lemma similar to Lemma 2.3.1 in van der Vaart and Wellner (1996). For that purpose, we use a decomposition similar to Arcones and Giné (1993) where independent copies of the $(U_{\boldsymbol{j}})_{\boldsymbol{j}}$ are introduced. With this lemma in hand, we use Assumption 3 or 3' to control the complexity of the class

$$\{g(n, y_1, ..., y_n) = \sum_{i=1}^{n} f(y_i) : n \in \mathbb{N}, (y_1, ..., y_n) \in \mathcal{Y}^n; f \in \mathcal{F}\}$$

and show at the end asymptotic equicontinuity.

We now comment on the asymptotic kernel of $\mathbb{G}_C$. For simplicity, let $f_1(y) = f_2(y) = y$ and define $S_{\boldsymbol{j}} = \sum_{\ell=1}^{N_{\boldsymbol{1}}} Y_{\ell,\boldsymbol{j}}$. Theorem 3.1 implies that the asymptotic variance of $\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}} S_{\boldsymbol{j}}/\Pi_C$ is $\sum_{i=1}^{k} \lambda_i \mathbb{C}ov\,(S_{\boldsymbol{1}}, S_{\boldsymbol{2}_i}))$. This formula involves cells that share exactly one common cluster, namely cluster 1 in dimension i. To better understand this, consider

$$\mathbb{V}\left(\frac{\sqrt{\underline{C}}}{\Pi_C} \sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}} S_{\boldsymbol{j}}\right).$$

This variance is complicated because of the particular dependence structure due to multiway clustering. To simplify it, we can write it as the sum of covariances between cells sharing no common cluster, cells sharing one common cluster... and finally the covariances of cells with themselves. The number of couples of cells sharing no common cluster is $\Pi_C \times \prod_{i=1}^{k}(C_i - 1)$, which is of the order Π_C^2 as $\underline{C}$ tends to infinity. The number of couples of cells sharing one common cluster is

$$\Pi_C \sum_{i=1}^{k} \prod_{j\neq i} (C_j - 1),$$

which is smaller than $k\Pi_C^2/\underline{C}$. Hence, the number of such couples of cells is negligible compared to the number of couples of cells sharing no common cluster. Similarly, we can prove that the number of cells sharing more than one common cluster is negligible compared to the number of cells sharing one common cluster. Hence, intuitively, the variance will be equivalent to the sum of only covariances between cells sharing either no or just one common cluster. But by independence, the covariance between cells sharing no common cluster is actually zero. Hence, at the end of the day, we only get covariances between cells sharing just one common cluster.

3.2 Bootstrap processes

We now consider the bootstrap counterpart of the weak convergence result in Theorem 3.1. Bootstrap offers several advantages over usual inference based on asymptotic normality. First, it enables one to avoid asymptotic variance calculations, which can be particularly useful with, e.g., multistep estimators. Second, it often exhibits a better behavior than normal approximations in finite samples. However, a consistent bootstrap in our clustering setting needs to reproduce the dependence between cells. We consider for that purpose the "pigeonhole bootstrap", suggested by McCullagh et al. (2000) and studied, in the case of the sample mean and for particular models, by Owen et al. (2007). We are, however, not aware of any result concerning the asymptotic validity of the pigeonhole bootstrap for inference. Theorem 3.2 below aims to fill this gap.

First, let us recall the principles of the pigeonhole bootstrap. Instead of drawing cells uniformly and independently from each other, they are drawn in the following way:

1. For each $i = 1, ..., k$ sample C_i elements with replacement and equal probability in the set $\{1, ..., C_i\}$. For each j_i in this set, let $W^i_{j_i}$ denote the number of times j_i is selected this way.

2. Cell $\boldsymbol{j} = (j_1, ..., j_k)$ is then selected $W_{\boldsymbol{j}} = \prod_{i=1}^k W^i_{j_i}$ times in the bootstrap sample.

Note that by construction, any bootstrap sample consists of exactly Π_C cells. Also, dependence between cells sharing cluster i is achieved through the term $W^i_{j_i}$. Actually, one can check that conditional on the data $(N_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{l\geq 1})_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}$, the bootstrap weights $(W_{\boldsymbol{j}})_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}$ satisfy the first two conditions in Assumption 1. This suggests that the pigeonhole bootstrap could be asymptotically valid.

We now consider the bootstrap counterpart of the empirical process $\mathbb{G}_C$, by considering, for any $f \in \mathcal{F}$,

$$\mathbb{G}^*_C(f) = \frac{\sqrt{C}}{\Pi_C} \sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}} \left(W^{\boldsymbol{C}}_{\boldsymbol{j}} - 1\right) \sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}).$$

The asymptotic validity of the pigeonhole bootstrap amounts to show that conditional on the data $\{N_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{l\geq 1}\}_{\boldsymbol{j}\geq\boldsymbol{1}}$, $\mathbb{G}^*_C$ converges weakly in probability to the process $\mathbb{G}$ defined in Theorem 3.1. Note that we study the behavior of $\mathbb{G}^*_C$ conditional on the data because the variations we rely on to draw bootsrap based inference are only due to the randomness of the weights $(W_{\boldsymbol{j}})_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}$. As discussed in, e.g., van der Vaart and Wellner (1996, Chapter 3.6), conditional weak convergence in probability amounts to prove

$$\sup_{h\in \text{BL}_1} \left|E\left(h(\mathbb{G}^*_C)|\{N_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{\ell\geq 1}\}_{\boldsymbol{j}\geq\boldsymbol{1}}\right) - E\left(h(\mathbb{G})\right)\right| \xrightarrow{\mathbb{P}} 0, \tag{5}$$

where BL_1 is the set of bounded and Lipschitz functions from $\ell^\infty(\mathcal{F})$ to $\mathbb{R}$.

Theorem 3.2 *Suppose that Assumptions 1-2 and 3 or 3' hold. Then $\mathbb{G}^*_C$ converges weakly to $\mathbb{G}$ in probability, namely (5) holds.*

As we shall see below, this theorem ensures the asymptotic validity of the pigeonhole bootstrap not ony for sample means, but also for GMM estimators or smooth functionals of the empirical cdf.

4 Applications

4.1 Inference on simple averages

As before, let $S_{\boldsymbol{j}} = \sum_{\ell=1}^{N_{\boldsymbol{j}}} Y_{\ell,\boldsymbol{j}}$. We investigate here how inference can be conducted on $\theta_0 = E(S_{\boldsymbol{1}})$ based on the sample average

$$\widehat{\theta} = \frac{1}{\Pi_C} \sum_{\boldsymbol{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} S_{\boldsymbol{j}}. \tag{6}$$

We consider $\widehat{\theta}$ rather than the sample average defined in (1) because it is slightly simple, and nonlinear estimators, including that defined in (1), which is a ratio of two averages, can typically be shown to be asymptotically equivalent to such averages. For instance:

$$\sqrt{\underline{C}}\left(\frac{\sum_{\boldsymbol{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} Y_{\ell,\boldsymbol{j}}}{\sum_{\boldsymbol{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} N_{\boldsymbol{j}}} - \frac{\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} Y_{\ell,\boldsymbol{1}}\right)}{\mathbb{E}\left(N_{\boldsymbol{1}}\right)}\right) = \sqrt{\underline{C}}\left(\sum_{\boldsymbol{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} Y'_{\ell,\boldsymbol{j}} - \mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} Y'_{\ell,\boldsymbol{1}}\right)\right) + o_p(1),$$

with $Y'_{\ell,\boldsymbol{j}} = Y_{\ell,\boldsymbol{j}}/\mathbb{E}(N_{\boldsymbol{1}})$.
Theorem 3.1 ensures that provided that $E(S_{\boldsymbol{j}}^2) < +\infty$,

$$\sqrt{\underline{C}}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left\{0, \sum_{i=1}^{k} \lambda_i \mathbb{C}ov\left(S_{\boldsymbol{1}}, S_{\boldsymbol{2}_i}\right)\right\}. \tag{7}$$

A first strategy to make inference on θ_0 is therefore to use the normal approximation and a consistent estimator of the asymptotic variance. A second strategy is to rely on the pigeonhole bootstrap.[3]

Let us consider first inference based on asymptotic normality. The asymptotic variance V depends on λ_i, which can be simply approximated by $\underline{C}/C_i$, and on $\mathbb{C}ov\left(S_{\boldsymbol{1}}, S_{\boldsymbol{2}_i}\right) = \mathbb{E}\left((S_{\boldsymbol{1}} - \theta_0)(S_{\boldsymbol{2}_i} - \theta_0)\right)$ for $i = 1...k$. Because $\boldsymbol{1}$ and $\boldsymbol{2}_i$ share exactly one cluster in common, it is natural to consider the following estimator for this covariance:

$$\widehat{\mathbb{C}ov}\left(S_{\boldsymbol{1}}, S_{\boldsymbol{2}_i}\right) = \frac{1}{C_i \prod_{s \neq i} C_s(C_s - 1)} \sum_{(\boldsymbol{j}, \boldsymbol{j}') \in \mathcal{A}_i} \left(S_{\boldsymbol{j}} - \widehat{\theta}\right)\left(S_{\boldsymbol{j}'} - \widehat{\theta}\right)',$$

where $\mathcal{A}_i := \{(\boldsymbol{j}, \boldsymbol{j}') : j_i = j'_i, \quad j_s \neq j'_s \quad \forall s \neq i\}$. This estimator is the average of cross products between clusters sharing just one common cluster (the denominator $C_i \prod_{s \neq i} C_s(C_s - 1)$ corresponding to the number of such pairs). This leads to the following estimator for V:

$$\widehat{V}_2 = \sum_{i=1}^{k} \frac{\underline{C}}{C_i} \frac{1}{C_i \prod_{s \neq i} C_s(C_s - 1)} \sum_{(\boldsymbol{j}, \boldsymbol{j}') \in \mathcal{A}_i} \left(S_{\boldsymbol{j}} - \widehat{\theta}\right)\left(S_{\boldsymbol{j}'} - \widehat{\theta}\right)'.$$

[3]Still another strategy is to rely on other bootstrap schemes. We refer to Menzel (2017) for the construction and analysis of a wild bootstrap for sample averages on such clustered data.

We will show that $\widehat{V}_2$ is consistent for V. An important drawback of this estimator, however, is that it is not necessarily positive. Note also that V is the variance of a Hájek projection, as explained above. As such, it is likely to underestimate the finite distance variance of $\widehat{\theta}$. Because $\widehat{\mathbb{C}ov}\left(S_{\mathbf{1}}, S_{\mathbf{2}_i}\right)$ itself slightly underestimates $\mathbb{C}ov\left(S_{\mathbf{1}}, S_{\mathbf{2}_i}\right)$ ($\mathbb{E}[\widehat{\mathbb{C}ov}\left(S_{\mathbf{1}}, S_{\mathbf{2}_i}\right)] = \mathbb{C}ov\left(S_{\mathbf{1}}, S_{\mathbf{2}_i}\right) - \mathbb{V}(\widehat{\theta})$), we can expect the corresponding confidence regions to undercover in practice.

To avoid these issues, we suggest to simply add to $\widehat{V}_2$ pairs sharing more than one cluster. Specifically, we consider

$$\begin{aligned}\widehat{V}_1 &= \sum_{i=1}^{k} \frac{\underline{C}}{C_i} \frac{1}{C_i \prod_{s \neq i} C_s^2} \sum_{(\boldsymbol{j},\boldsymbol{j}'):\boldsymbol{j}_i=\boldsymbol{j}'_i} \left(S_{\boldsymbol{j}} - \widehat{\theta}\right)\left(S_{\boldsymbol{j}'} - \widehat{\theta}\right)' \\ &= \sum_{i=1}^{k} \frac{\underline{C}}{C_i} \frac{1}{C_i} \sum_{j'_i=1}^{C_i} \left(\frac{1}{\prod_{s \neq i} C_s} \sum_{\boldsymbol{j}:j_i=j'_i} S_{\boldsymbol{j}} - \widehat{\theta}\right)\left(\frac{1}{\prod_{s \neq i} C_s} \sum_{\boldsymbol{j}:j_i=j'_i} S_{\boldsymbol{j}} - \widehat{\theta}\right)'.\end{aligned}$$

This estimator solves the positivity problem, as the second line shows. Also, it is likely to overestimate V, but this may somewhat compensate the fact that V itself underestimates $\mathbb{V}(\widehat{\theta})$. From an asymptotic point of view, it turns out that the additional terms corresponding to pairs sharing more than one cluster are negligible, implying that $\widehat{V}_1$ is consistent just as $\widehat{V}_2$. Finally, in the simulations considered in Section 5 below, inference is more accurate when using $\widehat{V}_1$ rather than $\widehat{V}_2$. We therefore recommend to use $\widehat{V}_1$ rather than $\widehat{V}_2$ in practice.

Finally, we compare our two estimators with that proposed by Cameron et al. (2011). Their estimator relies on a reformulation of $\mathbb{V}(\widehat{\theta})$. For any $m \in \{1, ..., k\}$ and $1 \leq i_1 < ... < i_m \leq k$, let $\mathcal{B}_{i_1,...,i_m} = \left\{(\boldsymbol{j}, \boldsymbol{j}') : j_{i_1} = j'_{i_1}, ..., j_{i_m} = j'_{i_m}\right\}$. Then,

$$\begin{aligned}\mathbb{V}(\widehat{\theta}) &= \frac{1}{\Pi_C^2} \sum_{(\boldsymbol{j},\boldsymbol{j}') \in \cup_{i=1}^k \mathcal{B}_i} \mathbb{C}ov\left(S_{\boldsymbol{j}}, S_{\boldsymbol{j}'}\right) \\ &= \sum_{m=1}^{k} (-1)^m \sum_{1 \leq i_1 < ... < i_m \leq k} \frac{1}{\Pi_C^2} \sum_{(\boldsymbol{j},\boldsymbol{j}') \in \mathcal{B}_{i_1,...,i_m}} \mathbb{C}ov\left(S_{\boldsymbol{j}}, S_{\boldsymbol{j}'}\right).\end{aligned}$$

The first line follows because if $(\boldsymbol{j}, \boldsymbol{j}')$ share no cluster, $\mathbb{C}ov\left(S_{\boldsymbol{j}}, S_{\boldsymbol{j}'}\right) = 0$. The second line follows from the inclusion-exclusion principle. This leads to the following estimator for the asymptotic variance of $\widehat{\theta}$:

$$\widehat{V}_{cgm} = \underline{C} \sum_{m=1}^{k} (-1)^m \sum_{1 \leq i_1 < ... < i_m \leq k} \frac{1}{\Pi_C^2} \sum_{(\boldsymbol{j},\boldsymbol{j}') \in \mathcal{B}_{i_1,...,i_m}} \left(S_{\boldsymbol{j}} - \widehat{\theta}\right)\left(S_{\boldsymbol{j}'} - \widehat{\theta}\right)'. \tag{8}$$

Note that we can consider various finite sample adjustments where $1/(\Pi_C)^2$ is replaced by $c_{i_1,...,i_m}/(\Pi_C)^2$, with $c_{i_1,...,i_m}$ tending to one as $\underline{C}$ tends to infinity. We refer to Cameron et al. (2011) for more details. The appeal of Formula (8) is that we can compute this estimator using variance estimators of $\widehat{\theta}$ assuming only one-way clustering along dimensions $(i_1, ..., i_m)$, for all $1 \leq i_1 < ... < i_m \leq k$.

To understand the link and differences between $\widehat{V}_1$ and $\widehat{V}_{cgm}$, it is instructive to consider the case $k = 2$. Then the formulas simplify into:

$$\widehat{V}_1 = \underline{C} \sum_{i_1=1}^{2} \frac{1}{\Pi_C^2} \sum_{(\boldsymbol{j},\boldsymbol{j}') \in \mathcal{B}_{i_1}} \left(S_{\boldsymbol{j}} - \widehat{\theta}\right) \left(S_{\boldsymbol{j}'} - \widehat{\theta}\right)',$$

$$\widehat{V}_{cgm} = \widehat{V}_1 - \frac{\underline{C}}{\Pi_C^2} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \left(S_{\boldsymbol{j}} - \widehat{\theta}\right) \left(S_{\boldsymbol{j}} - \widehat{\theta}\right)'. \tag{9}$$

In other words, $\widehat{V}_1$ estimates V by counting twice the pairs $(\boldsymbol{j}, \boldsymbol{j}')$ sharing two clusters, or equivalently the pairs $(\boldsymbol{j}, \boldsymbol{j})$, $\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}$. $\widehat{V}_{cgm}$ counts such pairs only once, whence the correction in (9). The cost of this correction is that $\widehat{V}_{cgm}$ is not always positive. Finally, note that there are only Π_C pairs $(\boldsymbol{j}, \boldsymbol{j})$. Thus, the second term in (9) is of order $\underline{C}/\Pi_C$ and tends to 0 as $\underline{C}$ tends to infinity. We can therefore expect $\widehat{V}_1$ and $\widehat{V}_{cgm}$ to be asymptotically equivalent. Proposition 4.1 shows indeed that $\widehat{V}_1$, $\widehat{V}_2$ and $\widehat{V}_{cgm}$ are all consistent estimators of V.

Proposition 4.1 *Suppose that Assumption 1 holds and $\mathbb{E}\left[S_{\mathbf{1}}^2\right] < +\infty$. Then $\widehat{V}_1$, $\widehat{V}_2$ and $\widehat{V}_{cgm}$ are consistent for V. Hence, confidence regions and tests on θ_0 based on asymptotic normality and any of these estimators are valid as long as V is symmetric positive.*

Proposition 4.1 shows that valid inference can be obtained using asymptotic normality and either $\widehat{V}_1$, $\widehat{V}_2$ or $\widehat{V}_{cgm}$, provided that V is positive definite. By the delta method, asymptotically valid inference can also be conducted on $\theta_0 = E(S_{\mathbf{1}})/E(N_{\mathbf{1}})$. More generally, we can use this strategy whenever the asymptotic linear approximation

$$\sqrt{\underline{C}} \left(\widehat{\theta} - \theta_0\right) = \frac{1}{\sqrt{\underline{C}}} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \psi_{\boldsymbol{j}} + o_P(1),$$

holds and provided that the $(\psi_{\boldsymbol{j}})_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}}$ are known. When the $(\psi_{\boldsymbol{j}})_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}}$ must be estimated, additional restrictions should be considered, as we will see below when considering GMM estimators.

The condition that V is positive definite basically states that at least one of the dimension of clustering matters, in the sense that for at least one $i \in \{1, ..., k\}$, $\mathbb{C}ov(S_{\mathbf{1}}, S_{\mathbf{2}_i})$ is positive definite. Note that under Assumption 1, $\mathbb{C}ov(S_{\mathbf{1}}, S_{\mathbf{2}_i})$ is necessarily positive; but it may not be definite. For instance, consider two-way clustering and $S_{\boldsymbol{j}} = U_{j_1,0} + U_{0,j_2} + U_{\boldsymbol{j}} \in \mathbb{R}$ and the $(U_{j_1,0})_{j_1}$, $(U_{0,j_2})_{j_2}$ and $(U_{\boldsymbol{j}})_{\boldsymbol{j}}$ all mutually independent. Then this condition holds if and only if $\mathbb{V}(U_{j_1,0}) + \mathbb{V}(U_{0,j_2}) > 0$. As discussed in Menzel (2017), there may be dependence between $S_{\mathbf{1}}$ and $S_{\mathbf{2}_i}$ but still $\mathbb{C}ov(S_{\mathbf{1}}, S_{\mathbf{2}_i})$ is not definite. This is the case if in the example above, we assume now that $S_{\boldsymbol{j}}$ satisifes

$$S_{\boldsymbol{j}} = (U_{j_1,0} - \mathbb{E}(U_{j_1,0}))(U_{0,j_2} - \mathbb{E}(U_{0,j_2})) + U_{\boldsymbol{j}}.$$

If V is not positive definite, standard tests and confidence regions, which rely on inverses of its estimators, are not valid in general. When $V = 0$, $\widehat{\theta}$ actually converges at a rate of convergence that is faster than $1/\sqrt{C}$. Its asymptotic distribution may also be non-normal, as in the example above. We refer to Menzel (2017), Example 1.6, for more details.

We now turn to the pigeonhole bootstrap. Let

$$\widehat{\theta}^* = \frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} W_{\boldsymbol{j}} S_{\boldsymbol{j}},$$

where $W_{\boldsymbol{j}}$ is defined in Section 3.2 above. Then let $q^*_{1-\alpha}$ denote the quantile of order $1-\alpha$ of the distribution of $|\widehat{\theta}^* - \widehat{\theta}|$ conditional on the data. We consider the confidence region $R_{1-\alpha}$ for θ_0 defined by

$$R_{1-\alpha} = \left\{\theta : \, |\theta - \widehat{\theta}| \leq q^*_{1-\alpha}\right\}.$$

The asymptotic validity of $R_{1-\alpha}$ is an immediate consequence of Theorem 3.2.

Proposition 4.2 *Suppose that Assumption 1 holds, $\mathbb{E}\left[S_{\mathbf{1}}^2\right] < +\infty$ and V is symmetric positive. Then*

$$\lim_{n\to\infty} P\left(R_{1-\alpha} \ni \theta_0\right) = 1 - \alpha.$$

When $\theta_0 \in \mathbb{R}$, an alternative, popular confidence region is the percentile bootstrap. This amounts to consider $R'_{1-\alpha} = [q_{\alpha/2}(\widehat{\theta}^*), q_{1-\alpha/2}(\widehat{\theta}^*)]$. This interval is also valid asymptotically, since the asymptotic distribution of $\widehat{\theta} - \theta_0$ is normal, and therefore symmetric.

4.2 GMM estimators

We consider in this section GMM estimators defined by (3), with possibly nonsmooth moments. We suppose that $m(y,\theta) \in \mathbb{R}^L$, with $m(y,\theta) = (m_1(y,\theta), ..., m_L(y,\theta))'$. We show the asymptotic normality of $\widehat{\theta}$ under the following condition.

Assumption 4 *We have:*

1. *θ_0 belongs to the interior of Θ, a compact subset of $\mathbb{R}^p$.*
2. *$\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right] = 0$ if and only if $\theta = \theta_0$.*
3. *For any $\theta \in \Theta$ we have $\lim_{\theta' \to \theta} \mathbb{E}\left[\left|\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta') - \sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right|^2\right] = 0.$*
4. *$\theta \mapsto \mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right)$ is differentiable at θ_0 with a jacobian matrix J of rank p.*
5. *For all $s = 1, ..., L$ the class $\mathcal{F}_s = \{y \mapsto m_s(y,\theta) : \theta \in \Theta\}$ fulfills Assumptions 2-3.*
6. *$\widehat{\Xi}$ is a sequence of random symmetric matrices of size L tending in probability to Ξ, a positive square matrix.*

Assumption 4.1, 4.2 and 4.6 are standard. Assumption 4.3, combined with 4.1 and 4.2, ensures that the minimum of

$$\theta \mapsto \mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right)' \Xi\, \mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right)$$

is well-separated on Θ, thus ruling out possible inconsistency of the GMM estimator. Note that Assumption 4.3 is weaker than the standard continuity assumption of $\theta \mapsto m(Y_{\ell,\mathbf{1}},\theta)$, which may fail if m includes for instance indicator functions. Assumption 4.4 is standard in GMM with nonsmooth moments where $\theta \mapsto m(Y_{\ell,\mathbf{1}},\theta)$ is not differentiable, see e.g. Condition (ii) in Theorem 7.2 of Newey and McFadden (1994). Finally, by Theorem 3.1, Assumption 4.5 ensures the stochastic equicontinuity condition (e.g., Condition (ii) in Theorem 7.2 of Newey and McFadden, 1994), which together with 4.4, is key to obtain $\sqrt{\underline{C}}$-asymptotic normality of $\widehat{\theta}$.

To illustrate that Assumption 4 can handle nonsmooth moments, let us consider the example of quantile IV regressions. Let $Y_{\ell,\boldsymbol{j}} = (W_{\ell,\boldsymbol{j}}, X'_{\ell,\boldsymbol{j}}, Z'_{\ell,\boldsymbol{j}})'$, where $W_{\ell,\boldsymbol{j}} \in \mathbb{R}$ denotes the outcome variable, $X_{\ell,\boldsymbol{j}} \in \mathbb{R}^p$ denotes the explanatory, potentially endogenous variable and $Z_{\ell,\boldsymbol{j}} \in \mathbb{R}^L$ denotes the set of instruments (where $Z_{\ell,\boldsymbol{j}}$ may include some components of $X_{\ell,\boldsymbol{j}}$). The moment conditions are then

$$m(Y_{\ell,\mathbf{1}},\theta) = Z_{\ell,\boldsymbol{j}}\left(\tau - \mathbb{1}\{W_{\ell,\boldsymbol{j}} - X'_{\ell,\boldsymbol{j}}\theta \le 0\}\right).$$

Let us assume for simplicity that the $(Y_{\ell,\mathbf{1}})_{\ell\ge 1}$ are identically distributed. We show in Appendix B.6 that Assumption 4.3-4.5 hold if, basically, $\mathbb{E}[N_{\mathbf{1}}^2|Z_{1,\mathbf{1}}|^2] < +\infty$, X is in a compact set, the conditional cdf $F_{W_{1,\mathbf{1}}|X_{1,\mathbf{1}},Z_{1,\mathbf{1}}}(\cdot|X_{1,\mathbf{1}},Z_{1,\mathbf{1}})$ is continuous everywhere and admits a bounded derivative $f_{W_{1,\mathbf{1}}|X_{1,\mathbf{1}},Z_{1,\mathbf{1}}}(\cdot|X_{1,\mathbf{1}},Z_{1,\mathbf{1}})$ in a neighborhood of $X'_{1,\mathbf{1}}\theta_0$ and the rank of $\mathbb{E}\left[N_{\mathbf{1}}X_{1,\mathbf{1}}Z'_{1,\mathbf{1}}f_{W_{1,\mathbf{1}}|X_{1,\mathbf{1}},Z_{1,\mathbf{1}}}(X'_{\ell,\boldsymbol{j}}\theta_0|X_{1,\mathbf{1}},Z_{1,\mathbf{1}})\right]$ is equal to p.[4]

Theorem 4.3 *Suppose that Assumption 1 and 4 hold. Then $\widehat{\theta}$ is well-defined with probability approaching one and*

$$\sqrt{\underline{C}}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V_0),$$

where $V_0 = (J'\Xi J)^{-1}J'\Xi H \Xi J(J'\Xi J)^{-1}$ and

$$H = \sum_{i=1}^{k} \lambda_i \mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta_0)\right)\left(\sum_{\ell=1}^{N_{\mathbf{2}_i}} m(Y_{\ell,\mathbf{2}_i},\theta_0)\right)'\right].$$

Besides showing the asymptotic normality of GMM estimators, Theorem 4.3 specifies the expression of its asymptotic variance V_0. This matrix takes the usual form, except that the matrix H, which would simply be $E[m(Y_{\ell,\mathbf{1}},\theta_0)m(Y_{\ell,\mathbf{1}},\theta_0)']$ without clustering, takes a more

[4]For the exact conditions, see Assumption 7 in Appendix B.6.

complicated form here. This form is in line with our result on the covariance kernel of the empirical process considered above.

We now turn to inference on θ_0. As for sample averages, we consider inference based on asymptotic normality and a consistent estimator of V_0, or the pigeonhole bootstrap. To ensure the consistency of our estimator of V_0, we impose the following additional regularity condition.

Assumption 5

1. *The jacobian matrix J of $\theta \mapsto \mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right)$ at θ_0 admits the following representation:*
$$J = \mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} d(Y_{\ell,\mathbf{1}},\theta)\right)$$
for some matrix-valued function $d(.,.) = (d_{r,s}(.,.))_{1\leq r\leq p, 1\leq s\leq L}$.

2. *For all $(r,s) \in \{1,...,p\} \times \{1,...,L\}$, the class $\mathcal{F}_{r,s} = \{y \mapsto d_{r,s}(y,\theta) : \theta \in \Theta\}$ fulfills Assumption 2 and admits an envelope $F_{r,s}$ such that $E[\sum_{\ell=1}^{N_{\mathbf{1}}} F_{r,s}(Y_{\ell,\mathbf{1}})] < +\infty$, and for any $\varepsilon > 0$ $\sup_Q N(\varepsilon||.||_{Q,1}, \mathcal{F}_{r,s}, ||.||_{Q,1})$ where the supremum is taken over the set of probability measures with finite support on $\mathcal{Y}$.*

3. $\lim_{\theta'\to\theta_0} \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} d(Y_{\ell,\mathbf{1}},\theta')\right] = \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} d(Y_{\ell,\mathbf{1}},\theta_0)\right]$.

4. *For every $i = 1,...,k$,*
$$\lim_{\theta'\to\theta_0} \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}}\sum_{\ell=1}^{N_{\mathbf{2}_i}} m(Y_{\ell,\mathbf{1}},\theta')m(Y_{\ell,\mathbf{2}_i},\theta')\right] = \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}}\sum_{\ell=1}^{N_{\mathbf{2}_i}} m(Y_{\ell,\mathbf{1}},\theta_0)m(Y_{\ell,\mathbf{2}_i},\theta_0)\right].$$

MODIFY. Assumption 5.1 imposes some regularity on the derivative of $\theta \mapsto \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right]$. Note that it holds in important cases where $m(Y_{\ell,\mathbf{1}},.)$ is not differentiable. In quantile or quantile IV regression for instance, $m(Y_{\ell,\mathbf{1}},.)$ is a step function but this assumption still holds if the residual of the model admits a conditional density in a neighborhood of 0. In Assumption 5.2 the condition on the class is a Glivenko-Cantelli type restriction on $\mathcal{F}$, which is weaker than Assumptions 3 or 3' considered before. Assumption 5.4 is the same as Assumption 4.5, except that if we assume that $\mathcal{F}$ satisfies a uniform entropy condition, we need an additional moment restriction on cluster sizes. This restriction naturally shows up by definition of H.

Now, we define our estimator of V_0. For that purpose, we estimate J and H. Given Assumption 5.1, $\widehat{J}$ is the simple plug-in estimator:
$$\widehat{J} = \frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} d(Y_{\ell,\boldsymbol{j}},\widehat{\theta}).$$

To estimate H, we adapt our previous estimator $\widehat{V}_1$ to this context by considering

$$\widehat{H} = \sum_{i=1}^{k} \frac{C}{C_i}\frac{1}{C_i}\sum_{j_i'=1}^{C_i}\left(\frac{1}{\prod_{s\neq i} C_s}\sum_{\boldsymbol{j}:j_i=j_i'}\sum_{\ell=1}^{N_{\boldsymbol{j}}} m\left(Y_{\ell,\boldsymbol{j}},\widehat{\theta}\right)\right)\left(\frac{1}{\prod_{s\neq i} C_s}\sum_{\boldsymbol{j}:j_i=j_i'}\sum_{\ell=1}^{N_{\boldsymbol{j}}} m\left(Y_{\ell,\boldsymbol{j}},\widehat{\theta}\right)\right)'.$$

Our estimator of variance is then $\widehat{V} = (\widehat{J}'\widehat{\Xi}\widehat{J})^{-1}\widehat{J}'\widehat{\Xi}\widehat{H}\widehat{\Xi}\widehat{J}(\widehat{J}'\widehat{\Xi}\widehat{J})^{-1}$.

Theorem 4.4 *Assume that Assumptions 1 and 4 hold and H is positive definite. Then:*

1. *If Assumption 5 holds as well, $\widehat{V} \xrightarrow{\mathbb{P}} V_0$ and confidence regions and tests on θ_0 based on asymptotic normality and $\widehat{V}$ are asymptotically valid.*

2. *Confidence regions and tests on θ_0 based on the pigeonhole bootstrap are asymptotically valid.*

Note that the pigeonhole bootstrap does not require any additional condition above that ensuring the $\sqrt{\underline{C}}$-asymptotic normality of the GMM estimator and the fact that H is positive definite. Hence, Theorem 4.4 implies for instance that under the conditions displayed above, the pigeonhole bootstrap is valid for quantile IV regressions under multiway clustering.

4.3 Nonlinear functionals of the distribution

The results in Section 3 can also be applied to obtain asymptotic normality of, and inference on, smooth, nonlinear functionals of the empirical distribution. Let F be defined as in (4) and let $\theta_0 = g(F)$. To take examples related to income inequalities (so that here the support of Y is $\mathbb{R}^+$), we may consider for instance quantiles, interquantile ratios, poverty rates or the Gini index, for which we have respectively $g(F) = F^{-1}(\tau)$ for any $\tau \in (0,1)$, $g(F) = F^{-1}(\tau)/F^{-1}(1-\tau)$ for $q \in (1/2,1)$, $g(F) = F(\alpha F^{-1}(\beta))$ for $(\alpha,\beta) \in (0,1)^2$ and $g(F) = \int_0^{+\infty}(1-F(y))^2 dy/\int_0^{+\infty} y dF(y)$. Other examples include the Kaplan-Meier functional (see, e.g., Example 20.15 in van der Vaart, 2000) or the nonlinear difference-in-difference estimand of Athey and Imbens (2006), for which $\theta_0 = \int y dF_1(y) - \int \left[F_2^{-1}\circ F_3(y)\right] dF_4(y)$, where $(F_1, ..., F_4)$ are the cdf's of Y on four distinct subpopulations..

We consider here the plug-in estimator $\widehat{\theta} = g(\widehat{F})$ of θ_0, with

$$\widehat{F}(y) = \frac{\sum_{\boldsymbol{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} \mathbb{1}\{Y_{\ell,\boldsymbol{j}} \leq y\}}{\sum_{\boldsymbol{1}\leq \boldsymbol{j}\leq \boldsymbol{C}} N_{\boldsymbol{j}}}. \tag{10}$$

To state the smoothness condition on g, we need additional notation and definition. Let $\mathbb{D}$ denote a subset of the set of all cumulative distribution functions on $\mathbb{R}^\ell$ and suppose that $g : \mathbb{D} \mapsto \mathbb{R}^r$. We consider for simplicity here vector-valued functions g, but could easily our result below to functions taking values in normed spaces. We say that g is Hadamard differentiable at F tangentially to $\mathbb{D}_0$ if there exists a continuous, linear map $g'_F : \mathbb{D} \mapsto \mathbb{R}^r$ such that for every $(h_t)_{t\in\mathbb{R}^+}$ such that $h_t \to h \in \mathbb{D}_0$ as $t \downarrow 0$,

$$\lim_{t\downarrow 0} \left| \frac{g(F + th_t) - g(F)}{t} - g'_F(h) \right| = 0.$$

Proposition 4.5 shows the asymptotic normality of plug-in estimators as long as, basically, g is Hadamard differentiable at F.

Proposition 4.5 *Suppose that $\theta_0 = g(F)$ and $\widehat{\theta} = g(\widehat{F})$, where F and $\widehat{F}$ are defined respectively by* (4) *and* (10). *Suppose also that Assumption 1 holds and $\mathbb{E}(N_{1,\mathbf{1}}^2) < +\infty$. Finally, suppose that g is Hadamard differentiable at F tangentially to $\mathbb{D}_0$. Then*

$$\sqrt{\underline{C}} \left(\widehat{\theta} - \theta_0 \right) \to \mathcal{N}(0, \mathbb{V}(g'_F(\mathbb{G}))),$$

where $\mathbb{G}$ is a centered gaussian process on $\mathbb{R}^r$ with covariance kernel

$$K(y_1, y_2) = \sum_{i=1}^{k} \lambda_i \mathbb{C}ov \left(\sum_{\ell=1}^{N_{\mathbf{1}}} \mathbb{1}\{Y_{\ell,\mathbf{1}} \leq y_1\}, \sum_{\ell=1}^{N_{\mathbf{2}_i}} \mathbb{1}\{Y_{\ell,\mathbf{2}_i} \leq y_2\} \right).$$

Moreover, confidence regions and tests on θ_0 based on the pigeonhole bootstrap are asymptotically valid. as soon as $\mathbb{V}(g'_F(\mathbb{G})$ is symmetric positive.

The first part follows from Theorem 3.1 and the functional delta method (see, e.g., van der Vaart, 2000, Theorem 20.8). The second part is a direct consequence of Theorem 3.2 and the functional delta method for the bootstrap (see, e.g., van der Vaart and Wellner, 1996, Theorem 3.9.11). It can then be applied to, e.g., the previous examples of quantiles or inequality indices, under suitable regularity conditions on F. For quantiles for instance, one has to assume that F is differentiable at $F^{-1}(\tau)$, with a strictly positive derivative. We refer to, e.g., van der Vaart (2000) for details on this example and on others.

5 Simulations and finite sample issues

One can wonder about the pratical relevance of the asymptotic results derive in the previous sections, in particular with respect to performance of the estimator on finite sample. Here, we want to perform some Monte-Carlo simulation to investigate the finite sample properties of the strategies we propose.

5.1 Baseline simulations

Our baseline senario is the simple estimation of the expectation of a variable Y, in a two-dimensional array with one observation per cell ($N_{\boldsymbol{j}} = 1$). Each $Y_{1,\boldsymbol{j}}$ is draw in a standard gaussian distribution, but the variance due to cell-shocks represent only 60% of the total variance, 20% of the variance is due to row shocks and the last 20% of the variance are due to the column shocks.

$$Y_{1,(j_1,j_2)} = \frac{1}{\sqrt{5}}\left(\sqrt{3}U_{j_1,j_2} + U_{j_1,0} + U_{0,j_2}\right), \text{ with } (U_{j_1,j_2}, U_{j_1,0}, U_{0,j_2}) \sim \mathcal{N}(0, I_3). \tag{11}$$

The estimator considered here is the sample average, and the Monte-Carlo exercice focus on inference. We focus on the building of confidence interval at 5%, using strategies presented in the previous sections. We compare inference based on asymptotic normality with various estimators of the asymptotic variance. Our prefered estimator presented in Section 4.1 of the variance based on the asymptotic approximation is $\underline{C}^{-1}\widehat{V}_1 = \sum_{i=1}^{k}\widehat{\Sigma}_i$, where $\widehat{\Sigma}_i$ are the clustered estimated variance (with respect to the i-th dimension of clustering). This appealing feature made this estimate quite easy to compute with usual econometric software like Stata or R. The econometrician has only to compute k variances under one-way cluster design and to add them to get a consistent estimate of variance under multiway clustering. In Stata, the small sample correction $\frac{C_i}{C_i-1}$ is used by default for the computation of the clustered variance $\widehat{\Sigma}_i$. To stick to the usual practice of applied econometricians, we use this small sample correction for each dimension of clustering. We procede similarly for the estimator suggested by Cameron et al. (2011) and for the estimator correponding to $\widehat{V}_2$ in Section 4.1. Based on this estimates we build a symmetric 95% confidence intervals: For $\widehat{\theta} = \frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}} Y_{1,\boldsymbol{j}}$ and $z_{0,025}$ the upper quantile of order 2,5% of the gaussian distribution, the confidence interval we compute are the following:

$$\mathcal{IC}_1 = \left[\widehat{\theta} \pm z_{0,025}\sqrt{\widehat{\Sigma}_1 + \widehat{\Sigma}_2}\right] \text{ with } \widehat{\Sigma}_1 = \frac{C_1}{C_1-1}\frac{1}{C_1^2C_2^2}\sum_{j_1=1}^{C_1}\left(\sum_{j_2=1}^{C_2}(Y_{1,(j_1,j_2)} - \widehat{\theta})\right)^2 \tag{12}$$

$$\widehat{\Sigma}_2 = \frac{C_2}{C_2-1}\frac{1}{C_1^2C_2^2}\sum_{j_2=1}^{C_2}\left(\sum_{j_1=1}^{C_2}(Y_{1,(j_1,j_2)} - \widehat{\theta})\right)^2,$$

$$\mathcal{IC}_{cgm} = \left[\widehat{\theta} \pm z_{0,025}\sqrt{\widehat{\Sigma}_1 + \widehat{\Sigma}_2 - \widehat{\Sigma}_{12}}\right] \text{ with } \quad \widehat{\Sigma}_{12} = \frac{C_1C_2}{C_1C_2-1}\frac{1}{C_1^2C_2^2}\sum_{j_1=1}^{C_1}\sum_{j_2=1}^{C_2}\left(Y_{1,(j_1,j_2)} - \widehat{\theta}\right)^2, \tag{13}$$

$$\mathcal{IC}_2 = \left[\widehat{\theta} \pm z_{0,025}\sqrt{\widetilde{\Sigma}_1 + \widetilde{\Sigma}_2}\right] \text{ with } \widetilde{\Sigma}_1 = \frac{C_1}{C_1-1}\frac{1}{C_1^2C_2^2}\sum_{j_1=1}^{C_1}\left(\sum_{\substack{1\leq j_2,j_2'\leq C_2\\ j_2'\neq j_2}}(Y_{1,(j_1,j_2)} - \widehat{\theta})(Y_{1,(j_1,j_2')} - \widehat{\theta})\right), \tag{14}$$

$$\widetilde{\Sigma}_2 = \frac{C_2}{C_2-1}\frac{1}{C_1^2C_2^2}\sum_{j_2=1}^{C_2}\left(\sum_{\substack{1\leq j_1,j_1'\leq C_1\\ j_1'\neq j_1}}(Y_{1,(j_1,j_2)} - \widehat{\theta})(Y_{1,(j_1',j_2)} - \widehat{\theta})\right).$$

Last, also compute the Efron's percentile bootstrap confidence interval based on the pigeonhole sampling presented in Section 3.2:

$$\mathcal{IC}_{boot} = \left[q^*_{0,025}; q^*_{0,975}\right] \text{ with } q^*_\alpha \text{ the quantile } \alpha \text{ of } \theta^*|(N_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{\ell \leq N_{\boldsymbol{j}}})_{\boldsymbol{1}\leq \boldsymbol{j} \leq \boldsymbol{C}} \tag{15}$$

Since the asymptotic distribution of $\widehat{\theta}$ is symmetric, this appraoch is valid. To simulate the distribution of $\theta^*|(N_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{\ell \leq N_{\boldsymbol{j}}})_{\boldsymbol{1}\leq \boldsymbol{j} \leq \boldsymbol{C}}$, we use 1000 bootstrap samples for each simulation. We simulate 1000 samples according to the data generating process 11, for designs such that $C_1 = C_2 = 5, 10, 30, 50, 100$. For each design and each estimated confidence interval (12)-(15), we compute the frequency of the event $\mathcal{IC} \ni \theta_0$. We then compare the actual coverage rate of the confidence intervals with the nominal one (95%). For intervals based on formula 13 and 14, there is no garanty that the estimated variance is positive. In that case, we consider that the variance is null (or arbitrary small), leading us to consider that θ_0 is outside the corresponding confidence interval. This problem of negative estimated variance only occurs for the smaller size of sample ($C_1 = C_2 = 5$) in approximatively 16% of simulations for 14 and 1% of for 13. The results are summarized in Figure 1.

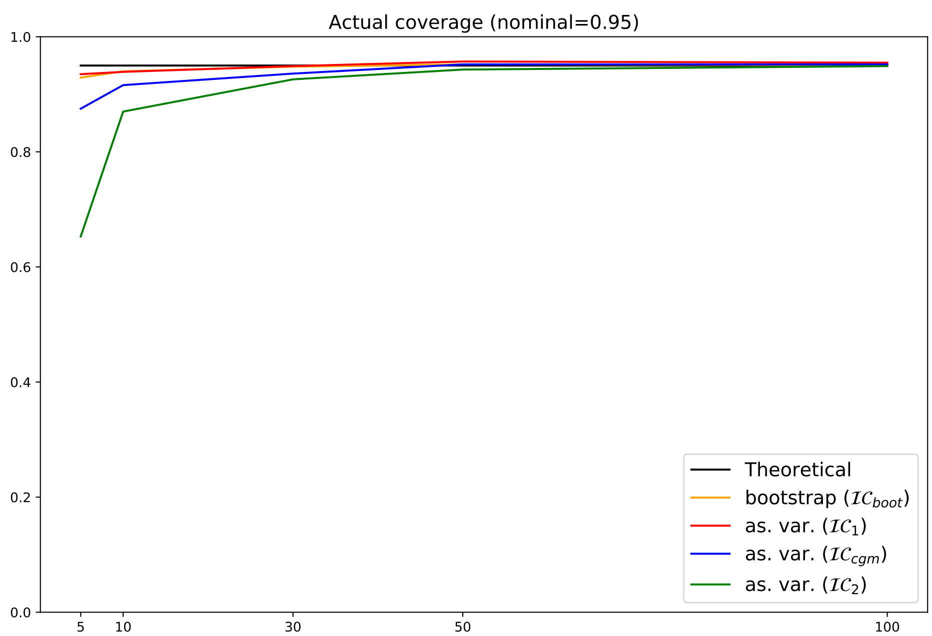

Figure 1: Balanced two-way clustering, one observation per cell, $Y_{1,(j_1,j_2)} = \frac{1}{\sqrt{5}}\left(\sqrt{3}U_{j_1,j_2} + U_{j_1,0} + U_{0,j_2}\right)$, $(U_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathbb{N}^2} \sim \mathcal{N}(0,1)^{\otimes \mathbb{N}^2}$, with small sample adjustement

For our prefered estimator of variance and for the pigeonhole bootstrap, the nominal coverage is always very close to the actual one, even for smaller samples (actual level between 93% and 96%). It is often considered that between 30 and 50 clusters are necessary in a one-way clustering to get reliable confidence intervals, (Cameron and Miller, 2015; Bertrand et al., 2004). Even with 40% of the variance of the cells related to clusters shocks, 25 cells resulting from a 5×5 design seems to be sufficient to get reliable inference at least with fixed size of cells and with a total at cell level following a standard gaussian. For the same design, inference proposed by Cameron et al. (2011) leads to an actual coverage of around 88% for a nominal coverage of 95%. Last the confidence intervals based on formula 14 perform poorly on small sample, in a 5×5 design we get an actual coverage around 62% far from the nominal one (remember that for 16% of the simulation we get a negative estimator of the variance leading to a decrease of the actual coverage in the same amount). According to the theory, when the number of clusters in each dimension increases, the actual level of the four confidence intervals converge to 95%. There are all between 92% and 95% for a design 30×30 and between 94% and 96% for a design 50×50.

5.2 Variation 1: without small sample corrections

The adjustement $\frac{C}{C-1}$ are usually motivated by small sample considerations in the case of one-way clustering. One can wonder in our framework how the inference is sensitive to such adjustement. As a robustness check, we also build the confidence interval omitting the factor $\frac{C_i}{C_i-1}$ in the definition of $\widehat{\Sigma}_i$ and $\widetilde{\Sigma}_i$ and omitting the factor $\frac{C_1 C_2}{C_1 C_2 - 1}$ in the definition of $\widehat{\Sigma}_1 2$. Figure 2 reports the result of this alternative estimated confidence intervals.
Such modification does not have a notable influence when $C_1, C_2 \geq 30$ for any of the confidence itervals. For the design 10×10 the coverage rate drops from 92% with adjustement to 90% for $\mathcal{IC}_{cgm}$ and from 87% to 85% for $\mathcal{IC}_2$, whereas the coverage rate of $\mathcal{IC}_2$ remains relativelly stable around $93,5\%$. For the design 5×5, the drop is more important for $\mathcal{IC}_{cgm}$ (from 87,5% to less than 82%) than for $\mathcal{IC}_1$, $\mathcal{IC}_2$ (with respective drop from 93,5% to 90,5% and from 66,5% to 62,5%). So, even without ajustement our prefered confidence intervals shows good properties and seems to be less sensible to adjustements than the confidence intervals proposed by Cameron et al. (2011).

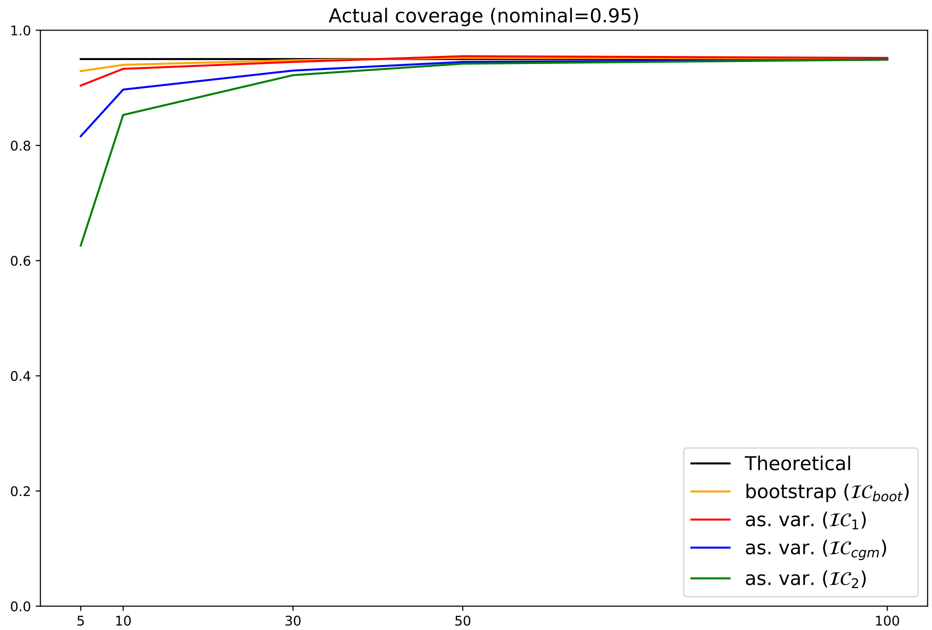

Figure 2: Balanced two-way clustering, one observation per cell, $Y_{1,(j_1,j_2)} = \frac{1}{\sqrt{5}}\left(\sqrt{3}U_{j_1,j_2} + U_{j_1,0} + U_{0,j_2}\right)$, $(U_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathbb{N}^2} \sim \mathcal{N}(0,1)^{\otimes\mathbb{N}^2}$, without small sample adjustement

5.3 Variation 2: mean of binary variable

In the previous simulations the variables Y are drawn in gaussian distributions, leading to gaussian distribution of the average on finite sample. One can wonder that the good properties of the $\mathcal{IC}_{boot}$ and $\mathcal{IC}_1$ are driven by the gaussianity of $\widehat{\theta}$ on finite sample. To investigate this issue, we run the same simulations tring to perform inference on the expectation of $\mathbb{1}\{Y > 0\}$. The results are qualitativelly the same as our baseline simulation. Quantitativelly the coverage rate of $\mathcal{IC}_2$ and $\mathcal{IC}_{cgm}$ are farer to from the nominal rate (falling respectivelly under 57% and 84% in the 5×5 design), while the coverage rate of $\mathcal{IC}_1$ and $\mathcal{IC}_{boot}$ are not affected (93,5% and 95,2% in the 5×5 design).

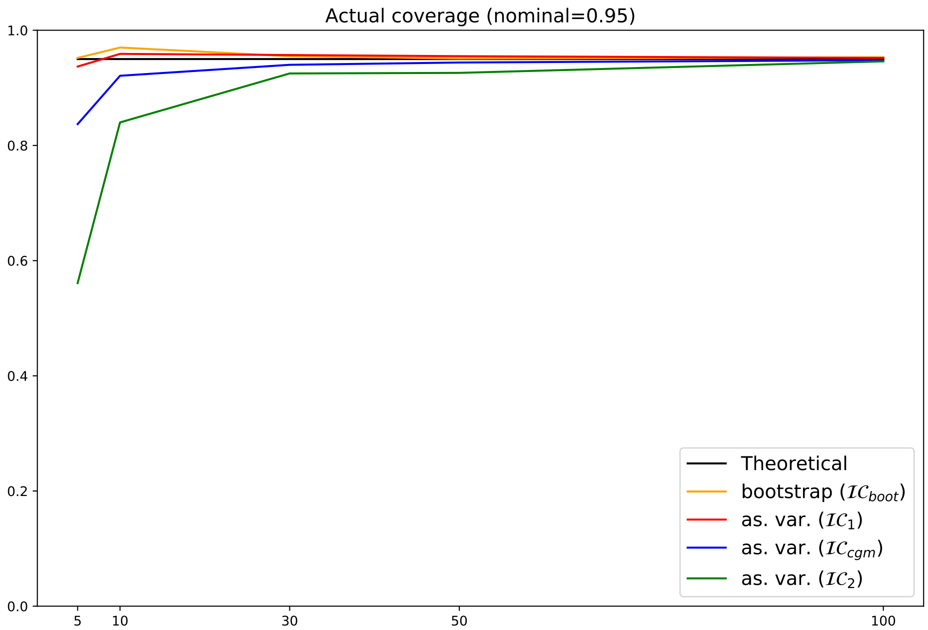

Figure 3: Balanced two-way clustering, one observation per cell, $Y_{1,(j_1,j_2)} = \mathbb{1}\left\{\frac{1}{\sqrt{5}}\left(\sqrt{3}U_{j_1,j_2} + U_{j_1,0} + U_{0,j_2}\right) > 0\right\}$, $(U_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathbb{N}^2} \sim \mathcal{N}(0,1)^{\otimes\mathbb{N}^2}$, with small sample adjustement

5.4 Variation 3: Probit model with heterogeneous cells size

In the previous simulations, we have considered linear estimators (namely average) of parameters of interest. But for many parameters of interest, only asymptotically linear estimator are available. In that case, one can wonder that non linearity of the estimator on finite sample generates less reliable inference. Moreover, until now, the number of observation par cells were fixed to 1.

In another set of simulation, we have considered a simple probit model, with heterogeneous cells size. We consider the following DGP:

$$\begin{aligned} N_{(j_1,j_2)} &\sim 1 + \mathcal{P}(5) \\ (X_{\ell,\boldsymbol{j}}, U_{\boldsymbol{j}}, \varepsilon_\ell)_{\ell,\boldsymbol{j}\in\mathbb{N}^3} | (N_{\boldsymbol{c}})_{\boldsymbol{c}\in\mathbb{N}^2} &\sim \mathcal{N}(0,1)^{\otimes\mathbb{N}^3} \otimes \mathcal{N}(0,1)^{\otimes\mathbb{N}^2} \otimes \mathcal{N}(0,1)^{\otimes\mathbb{N}} \\ Y_{\ell,(j_1,j_2)} &= \mathbb{1}\left\{\beta_0 + \beta_1 X_{\ell,(j_1,j_2)} + \frac{1}{\sqrt{6}}\left(U_{(j_1,0)} + U_{(0,j_2)} + U_{(j_1,j_2)}\right) + \frac{\varepsilon_\ell}{\sqrt{2}} > 0\right\} \\ \beta_0 &= 0 \\ \beta_1 &= 1 \end{aligned}$$

The parameter of interest is β_1 and we consider that the econometrician observe $(N_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}}, X_{\ell,\boldsymbol{j}})_{1\leq\ell\leq N_{\boldsymbol{j}}})_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}$. In these simulations $\widehat{\theta}$ is the (pseudo-)maximum likelyhood estimator of β_1, ie the estimator obtained with an usual probit estimation on the pooled sample. We know from Theorems 4.3

and 4.4 that $\widehat{\theta}$ is asymptotically normal and that confidence intervals based on estimation of the sandwich formula and/or on Efron's percentile pigeonhole bootstrap are consistent.
Following the default option in usual software like STATA, we also use adjustments of type $\frac{C}{C-1}$ for the estimation of terms $\widehat{\Sigma}_i$, $\widetilde{\Sigma}_i$ and $\widehat{\Sigma}_{12}$. More precisely, we apply formula (12), (13) and (14), adapting $\widehat{\Sigma}_i$, $\widetilde{\Sigma}_i$ and $\widehat{\Sigma}_{12}$ as below:

$$\widehat{\Sigma}_1 = \frac{C_1}{C_1-1} J^{-1} \left[\sum_{j_1=1}^{C_1} \left(\sum_{j_2=1}^{C_2} \sum_{\ell=1}^{N_{(j_1,j_2)}} s_{\ell,(j_1,j_2)}(\widehat{\beta}) \right) \left(\sum_{j_2=1}^{C_2} \sum_{\ell=1}^{N_{(j_1,j_2)}} s'_{\ell,(j_1,j_2)}(\widehat{\beta}) \right) \right] J^{-1}$$

$$\widehat{\Sigma}_{12} = \frac{C_1C_2}{C_1C_2-1} J^{-1} \sum_{\mathbf{1}\leq \boldsymbol{j} \leq \boldsymbol{C}} \left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} s_{\ell,\boldsymbol{j}}(\widehat{\beta}) \sum_{\ell=1}^{N_{\boldsymbol{j}}} s'_{\ell,\boldsymbol{j}}(\widehat{\beta}) \right] J^{-1}$$

$$\widetilde{\Sigma}_1 = \frac{C_1}{C_1-1} J^{-1} \left[\sum_{j_1=1}^{C_1} \sum_{\substack{1\leq j_2, j_2' \leq C_2 \\ j_2' \neq j_2}} \left(\sum_{\ell=1}^{N_{(j_1,j_2)}} s_{\ell,(j_1,j_2)}(\widehat{\beta}) \sum_{\ell=1}^{N_{(j_1,j_2')}} s'_{\ell,(j_1,j_2')}(\widehat{\beta}) \right) \right] J^{-1}$$

$$s_{\ell,\boldsymbol{j}}(\widehat{\beta}) = \lambda_{\ell,\boldsymbol{j}}(\widehat{\beta}) \begin{pmatrix} 1 \\ X_{\ell,\boldsymbol{j}} \end{pmatrix}, \qquad \lambda_{\ell,\boldsymbol{j}}(\widehat{\beta}) = (2Y_{\ell,\boldsymbol{j}}-1) \frac{\phi\left((2Y_{\ell,\boldsymbol{j}}-1)(\widehat{\beta}_0 + \widehat{\beta}_1 X_{\ell,\boldsymbol{j}})\right)}{\Phi\left((2Y_{\ell,\boldsymbol{j}}-1)(\widehat{\beta}_0 + \widehat{\beta}_1 X_{\ell,\boldsymbol{j}})\right)}$$

$$J = \sum_{\mathbf{1}\leq \boldsymbol{j} \leq \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} \nabla s_{\ell,\boldsymbol{j}}(\widehat{\beta}) = \sum_{\mathbf{1}\leq \boldsymbol{j} \leq \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} \lambda_{\ell,\boldsymbol{j}}(\widehat{\beta}) \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_{\ell,\boldsymbol{j}} + \lambda_{\ell,\boldsymbol{j}}(\widehat{\beta})\right) \begin{pmatrix} 1 & X_{\ell,\boldsymbol{j}} \\ X_{\ell,\boldsymbol{j}} & X_{\ell,\boldsymbol{j}}^2 \end{pmatrix}$$

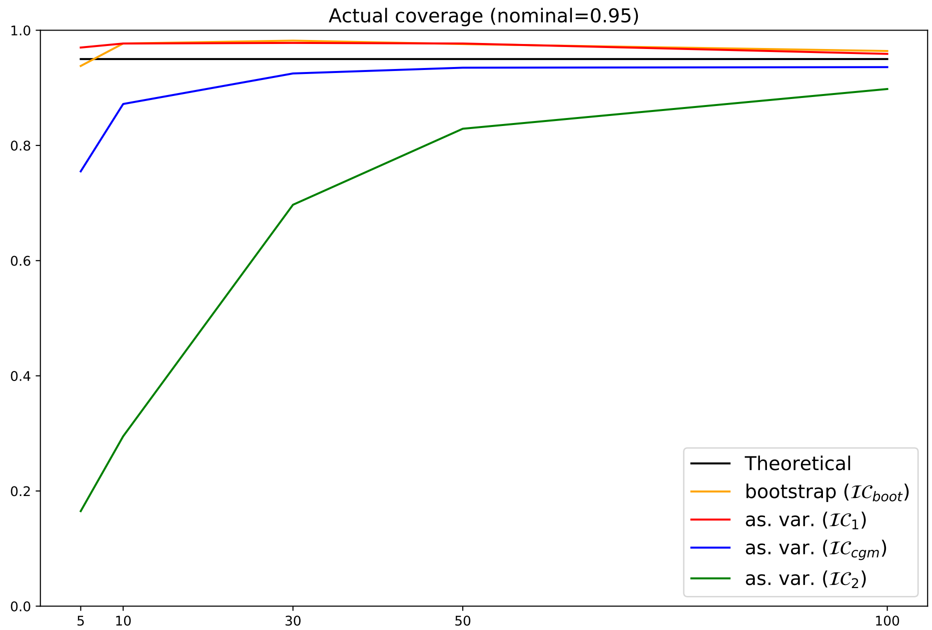

Figure 4: Balanced two-way clustering, one observation per cell, $Y_{\ell,(j_1,j_2)} = \mathbb{1}\{\beta_0 + \beta_1 X_{\ell,(j_1,j_2)} + \frac{1}{\sqrt{6}}\left(U_{(j_1,0)} + U_{(0,j_2)} + U_{(j_1,j_2)}\right) + \frac{\varepsilon_\ell}{\sqrt{2}} > 0\}$, $N_{(j_1,j_2)} \sim 1+\mathcal{P}(5)$, $(X_{\ell,\boldsymbol{j}}, U_{\boldsymbol{j}}, \varepsilon_\ell)_{\ell,\boldsymbol{j}\in\mathbb{N}^3}\,|(U_{\boldsymbol{c}})_{\boldsymbol{c}\in\mathbb{N}^2} \sim \mathcal{N}(0,1)^{\otimes\mathbb{N}^3} \otimes \mathcal{N}(0,1)^{\otimes\mathbb{N}^2} \otimes \mathcal{N}(0,1)^{\otimes\mathbb{N}}$, with small sample adjustement

With this data generating process, confidence intervals based on formula (14) perform dramatically on finite sample: in the 5×5 design, more than 75% of the variance estimates are non positive, leading to an actual rate lower than 20%.
Actual rates of confidence estimates based on (13) and proposed by Cameron et al. (2011) are quite below the nominal rate: below 76% for the 5×5 design and below 88% for the 10×10 design. For this design estimators of variance are sometimes non positive (15% of the simulations for the 5×5 design). Actual rates of confidence intervals based on 13 approximate reasonably well the nominal rate only for designs such that $C_1, C_2 \geq 30$.
Concerning the confidence intervals based on the bootstrap or on formula (12), actual rates are close to the nominal one, even if the confidence intervals tend to be slightly conservative (for this confidence intervals all the actual rates are between 93,8% and 98,2%).

5.5 Variation 4: 3-way clustering

The last simulations we present in details here investigate how the nominal rates evolves when the dimension of clustering increases. We have simulated a data generating process in three-ways clustering design. In order to have relevant comparison with our baseline simulations, we have draw Y in a gaussian distribution with unit variance. As in our baseline simulations, 60% of the variance is due to cell-shocks. 6,67% of the variance is due to shocks specific to the dimension 1 of the clustering, the same for the dimensions 2 et 3. The remaining variance is due to shocks common to dimensions 1 and 2, (respectively to dimensions 2 and 3, and to dimensions 1 and 3):

$$Y_{1,\boldsymbol{j}} = \frac{1}{\sqrt{15}} \left(3U_{\boldsymbol{j}} + U_{(j_1,j_2,0)} + U_{(0,j_2,j_3)} + U_{(j_1,0,j_3)} + U_{(j_1,0,0)} + U_{(0,j_2,0)} + U_{(0,0,j_3)} \right),$$
$$(U_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathbb{N}^3} \sim \mathcal{N}(0,1)^{\otimes \mathbb{N}^3}$$

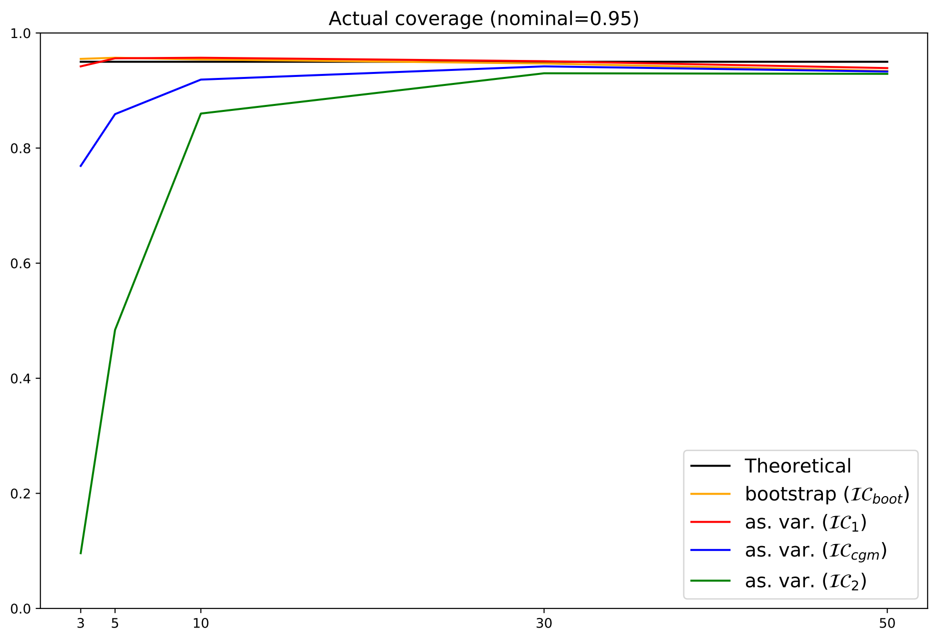

Figure 5: Balanced three-way clustering, one observation per cell, $Y_{1,\boldsymbol{j}} = \frac{1}{\sqrt{15}} \left(3U_{\boldsymbol{j}} + U_{(j_1,j_2,0)} + U_{(0,j_2,j_3)} + U_{(j_1,0,j_3)} + U_{(j_1,0,0)} + U_{(0,j_2,0)} + U_{(0,0,j_3)} \right)$, $(U_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathbb{N}^3} \sim \mathcal{N}(0,1)^{\otimes \mathbb{N}^3}$, with small sample adjustement

We consider $3 \times 3 \times 3$ design corresponding to 27 cells, a number close to the one of the 5×5 design of our baseline simulations. And because $5^3 \approx 10^2$ and $10^3 \approx 30^2$, so we also consider $5 \times 5 \times 5$ and $10 \times 10 \times 10$ design. On the top of that, to complete this simulation exercice, we also consider $30 \times 30 \times 30$ and $50 \times 50 \times 50$ designs.
TBC

5.6 Additional remarks

TBC

6 Conclusion

TBC

References

Aldous, D. J. (1981), 'Representations for partially exchangeable arrays of random variables', *Journal of Multivariate Analysis* **11**(4), pp. 581–598.

Aldous, D. J. (1985), Exchangeability and related topics, *in* 'École d'Été St Flour 1983', Springer-Verlag, pp. 1–198. Lecture Notes in Math. 1117.

Arcones, M. and Giné, E. (1993), 'Limit theorems for u-processes', *The Annals of Probability* **21**(3), pp. 1494–1542.

Arellano, M. (1987), 'Computing robust standard errors for within-groups estimators', *Oxford bulletin of Economics and Statistics* **49**(4), 431–434.

Athey, S. and Imbens, G. W. (2006), 'Identification and inference in nonlinear difference-in-differences models', *Econometrica* **74**(2), 431–497.

Bertrand, M., Duflo, E. and Mullainathan, S. (2004), 'How much should we trust differences-in-differences estimates?', *The Quarterly journal of economics* **119**(1), 249–275.

Bester, C. A., Conley, T. G. and Hansen, C. B. (2011), 'Inference with dependent data using cluster covariance estimators', *Journal of Econometrics* **165**(2), 137–151.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2011), 'Robust inference with multiway clustering', *Journal of Business & Economic Statistics* **29**(2), 238–249.

Cameron, A. C. and Miller, D. L. (2015), 'A practitioner's guide to cluster-robust inference', *Journal of Human Resources* **50**(2), 317–372.

Carter, A. V., Schnepel, K. T. and Steigerwald, D. G. (2017), 'Asymptotic behavior of at-test robust to cluster heterogeneity', *Review of Economics and Statistics* **99**(4), 698–709.

Chen, X. (2007), 'Large sample sieve estimation of semi-nonparametric models', *Handbook of econometrics* **6**, 5549–5632.

Chen, X. and Pouzo, D. (2015), 'Sieve wald and qlr inferences on semi/nonparametric conditional moment models', *Econometrica* **83**(3), 1013–1079.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014), 'Gaussian approximation of suprema of empirical processes', *Annals of Statistics* **42**(4), 1564–1597.

Freyberger, J. and Masten, M. (2015), Compactness of infinite dimensional parameter spaces, Technical report, cemmap working paper, Centre for Microdata Methods and Practice.

Gallant, A. R. and Nychka, D. W. (1987), 'Semi-nonparametric maximum likelihood estimation', *Econometrica: Journal of the Econometric Society* pp. 363–390.

Giné, E. and Nickl, R. (2015), *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Hahn, J. (1996), 'A note on bootstrapping generalized method of moments estimators', *Econometric Theory* **12**(1), 187–197.

Hansen, B. and Lee, S. (2017), Asymptotic theory for clustered samples. Working paper.

Hansen, C. B. (2007), 'Asymptotic properties of a robust variance matrix estimator for panel data when t is large', *Journal of Econometrics* **141**(2), 597–620.

Hoover, D. N. (1979), Relations on probability spaces and arrays of random variables, Technical report, Institute of Advanced Study, Princeton.

Ibragimov, R. and Müller, U. K. (2016), 'Inference with few heterogeneous clusters', *Review of Economics and Statistics* **98**(1), 83–96.

Kallenberg, O. (2005), *Probabilistic Symmetries and Invariance Principles*, number vol. 10 *in* 'Applied probability', Springer.

Kato, K. (2017), Lecture notes on empirical process theory, Technical report, University of Tokyo.

Kosorok, M. (2006), *Introduction to Empirical Processes and Semiparametric Inference*, Springer Verlag New York.

Liang, K.-Y. and Zeger, S. L. (1986), 'Longitudinal data analysis using generalized linear models', *Biometrika* **73**(1), 13–22.

McCullagh, P. et al. (2000), 'Resampling and exchangeable arrays', *Bernoulli* **6**(2), 285–301.

Menzel, K. (2017), Bootstrap with clustering in two or more dimensions. Working paper.

Moulton, B. R. (1986), 'Random group effects and the precision of regression estimates', *Journal of econometrics* **32**(3), 385–397.

Moulton, B. R. (1990), 'An illustration of a pitfall in estimating the effects of aggregate variables on micro units', *The review of Economics and Statistics* pp. 334–338.

Newey, W. K. and McFadden, D. (1994), 'Large sample estimation and hypothesis testing', *Handbook of econometrics* **4**, 2111–2245.

Newey, W. K. and Powell, J. L. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**(5), 1565–1578.

Nickl, R. and Pötscher, B. M. (2007), 'Bracketing metric entropy rates and empirical central limit theorems for function classes of besov-and sobolev-type', *Journal of Theoretical Probability* **20**(2), 177–199.

Owen, A. B. et al. (2007), 'The pigeonhole bootstrap', *The Annals of Applied Statistics* **1**(2), 386–411.

Pakes, A. and Pollard, D. (1989), 'Simulation and the asymptotics of optimization estimators', *Econometrica* **57**(5), 1027–1057.

Pfeffermann, D. and Nathan, G. (1981), 'Regression analysis of data from a cluster sample', *Journal of the American Statistical Association* **76**, 681–689.

van der Vaart, A. (2000), *Asymptotics Statistics*, Cambridge University Press.

van der Vaart, A. and Wellner, J. (1996), *Weak Convergence of Empirical Processes: with Applications to Statistics*, Springer-Verlag New York.

A Weak convergence under another restriction on $\mathcal{F}$

In this appendix, we discuss another set of results that can be obtained under another restriction on the class $\mathcal{F}$. For that purpose, let us consider the norm $||\cdot||_{\infty,\beta}$ defined by

$$||f||_{\infty,\beta} = \sup_{y \in \mathcal{Y}} \left| f(y)(1+|y|^2)^{\beta/2} \right|$$

for $\beta \in \mathbb{R}$. When $\beta = 0$, $||\cdot||_{\infty,\beta}$ corresponds to the standard supremum norm. When $\beta < 0$, $||\cdot||_{\infty,\beta}$ allows to deal with classes of smooth but unbounded functions that diverge in the tails at an appropriate rate. When $\beta > 0$, $||\cdot||_{\infty,\beta}$ is useful when the class $\mathcal{F}$ consists of uniformly bounded functions decaying sufficiently fast in the tails (see e.g. Freyberger and Masten, 2015, for more details).

We then consider the following restriction on $\mathcal{F}$.

Assumption 3' *$\mathcal{F}$ has a P-measurable envelope F with $||F||_{\infty,\beta} < +\infty$ and*

$$\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}(1+|Y_{\ell,\mathbf{1}}|^2)^{-\beta/2}\right)^2\right] < +\infty,$$

$$\int_0^{+\infty} \sqrt{\log N(\epsilon||F||_{\infty,\beta}, \mathcal{F}, ||.||_{\infty,\beta})} d\epsilon < +\infty.$$

Compared to Assumption 3, Assumption 3' is useful to establish asymptotic normality in models involving infinite-dimensional but smooth parameters. In the i.i.d. setup, Nickl and Pötscher (2007) show that under Assumption 3' and a moment condition, the $L_2(P)$ bracketing integral of many well-known classes of smooth functions is finite, which is a key ingredient in proving asymptotic normality results for estimators of smooth functional parameters.[5] The two main types of smoothness classes of functions used in practice are Sobolev classes and Hölder classes. The former have been used for instance by Newey and Powell (2003) in the nonparametric instrumental variable model and Gallant and Nychka (1987) in a semi-nonparametric maximum likelihood estimation framework. The Hölder classes have been used for instance by Chen and Pouzo (2015) in the context of nonparametric quantile instrumental variable models. For more details on the use of (weighted-)nonparametric smoothness classes in econometrics, we refer to Chen (2007) and Freyberger and Masten (2015).

We obtain the same result as Theorem 3.1 when replacing Assumption 3 by Assumption 3'.

Theorem A.1 *Suppose that Assumptions 1-2 and 3' hold. Then the process $\mathbb{G}_C$ converges weakly to a centered Gaussian process $\mathbb{G}$ on $\mathcal{F}$ as $\underline{C}$ tends to infinity. Moreover, the covariance kernel K of $\mathbb{G}$ satisfies:*

$$K(f_1, f_2) = \sum_{i=1}^{k} \lambda_i \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f_1(Y_{\ell,\mathbf{1}}), \sum_{\ell=1}^{N_{\mathbf{2}_i}} f_2(Y_{\ell,\mathbf{2}_i})\right).$$

[5] We refer to, e.g., van der Vaart and Wellner (1996) for a definition of bracketing integrals.

TO DO: discuss how the results on the GMM can be modified.

B Proofs of the main results

B.1 Notations

B.1.1 Notations relative to the processes

To simplify the exposition hereafter, we introduce the following notations, which hold for any vector-valued function f on $\mathcal{Y}$ such that $\mathbb{E}\left(\left|\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right|\right) < \infty$:

$$\mathbb{P}f = \mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right)$$

$$\mathbb{P}_C f = \frac{1}{\Pi_C} \sum_{\mathbf{1}\leq \boldsymbol{j} \leq \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}),$$

as a consequence, that $\mathbb{G}_C(f) = \sqrt{\underline{C}}\,(\mathbb{P}_C f - \mathbb{P}f)$.

B.1.2 Notations relative to algebra in $\mathbb{N}^k$

$\boldsymbol{j}, \boldsymbol{j}', \boldsymbol{e}, \boldsymbol{C}$...	denote elements of $\mathbb{N}^k$, with respective component $(j_1, ..., j_k)$, $(j'_1, ..., j'_k)$, $(e_1, ..., e_k)$, $(C_1, ..., C_k)$ ...
$\mathbf{0}, \mathbf{1}, \mathbf{2}$	denote respectivelly $(0, ..., 0)$, $(1, ..., 1)$, $(2, ..., 2)$
$\boldsymbol{j} \leq \boldsymbol{j}'$	means that for any $i = 1, ..., k$, $j_i \leq j'_i$
$\boldsymbol{j} < \boldsymbol{j}'$	means that $\boldsymbol{j} \leq \boldsymbol{j}'$ and $\boldsymbol{j} \neq \boldsymbol{j}'$
$\odot$	denotes the Hadamard product, ie $\boldsymbol{j} \odot \boldsymbol{j}' = (j_1 j'_1, ..., j_k j'_k)$
$\vee$ and $\wedge$	denote respectivelly the componentwise maximum and minimum: $\boldsymbol{j} \vee \boldsymbol{j}' = (\max(j_1, j'_1), ..., \max(j_k, j'_k))$
$\mathcal{E}_i$	for $i = 1, ..., k$ is the set $\{\boldsymbol{e} \in \{0;1\}^k : \sum_{i'=1}^k e_{i'} = i\}$
$\mathcal{I}_i$	for $i = 1, ..., k$ is the set $\{\boldsymbol{c} = \boldsymbol{j} \odot \boldsymbol{e} : \mathbf{1} \leq \boldsymbol{j}, \boldsymbol{e} \in \mathcal{E}_i\}$
$\mathcal{I}_i(\boldsymbol{C})$	for $i = 1, ..., k$ is the set $\{\boldsymbol{c} = \boldsymbol{j} \odot \boldsymbol{e} : \mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}, \boldsymbol{e} \in \mathcal{E}_i\}$
$\boldsymbol{e} \preceq \boldsymbol{e}'$	for $(\boldsymbol{e}, \boldsymbol{e}') \in (\{0;1\}^k)^2$ means that 1. either the number of null components of e is greater than the number of null component of $\boldsymbol{e}'$, 2. or either the number of null components of e is equal to the number of null component of $\boldsymbol{e}'$ and $\sum_{i=1}^k e_i \times 10^i \leq \sum_{i=1}^k e'_i \times 10^i$.
$\boldsymbol{e} \prec \boldsymbol{e}'$	for $(\boldsymbol{e}, \boldsymbol{e}') \in (\{0;1\}^k)^2$ means that $\boldsymbol{e} \preceq \boldsymbol{e}'$ and $\boldsymbol{e} \neq \boldsymbol{e}'$.
$\mathcal{A}_{\boldsymbol{e}}$	for $\boldsymbol{e} \in \{0;1\}^k$, the set of couple $(\boldsymbol{j}, \boldsymbol{j}')$ of cells such that $j_i = j'_i$ if and only if $e_i = 1$: $\{(\boldsymbol{j}, \boldsymbol{j}') : \mathbf{1} \leq \boldsymbol{j}, \boldsymbol{j}' \leq \boldsymbol{C} : \forall i = 1, ..., k,\ e_i = 1 \Leftrightarrow j_i = j'_i\}$

$\mathcal{B}_{\boldsymbol{e}}$	for $\boldsymbol{e} \in \{0;1\}^k$, the set of couple $(\boldsymbol{j}, \boldsymbol{j}')$ of cells such that $j_i = j'_i$ if $e_i = 1$: $\{(\boldsymbol{j}, \boldsymbol{j}') : \mathbf{1} \le \boldsymbol{j}, \boldsymbol{j}' \le \boldsymbol{C} : \forall i = 1, ..., k,\ e_i = 1 \Rightarrow j_i = j'_i\}$
$\mathcal{A}_i, \mathcal{B}_i$	$\mathcal{A}_{\boldsymbol{e}}$ (resp. $\mathcal{B}_{\boldsymbol{e}}$) for $\boldsymbol{e} = (0, ..., 0, 1, 0, ..., 0)$ where 1 is located at component i.

B.1.3 Notations relative to the classes of functions

We introduce additional classes of functions, related to any initial class $\mathcal{F} \subset L^1(P)$ or $L^2(P)$ (depending on the context).

$\mathcal{F}^2$	the set of squared functions in $\mathcal{F}$: $\{f^2 : f \in \mathcal{F}\}$
$\mathcal{F} \times \mathcal{F}$ or $(\mathcal{F})^2$	the cartesian product of $\mathcal{F}$ with itself: $\{(f_1, f_2) : f_1 \in \mathcal{F}, f_2 \in \mathcal{F}\}$
$\mathcal{F}_\delta$	the class of differences smaller than δ for the norm $\|\|.\|\|_{P,2}$: $\left\{h = f_1 - f_2 : (f_1, f_2) \in \mathcal{F} \times \mathcal{F}, \left\|\left\|\sum_{\ell=1}^{N_{\mathbf{1}}} \left(f_1(Y_{\ell,\mathbf{1}}) - f_2(Y_{\ell,\mathbf{1}})\right)\right\|\right\|_{P,2} \le \delta\right\}$.
$\mathcal{F}_\infty$	the class of differences: $\{h = f_1 - f_2 : (f_1, f_2) \in \mathcal{F} \times \mathcal{F}\}$.
$\mathcal{F}_\infty^2$	the class of the square differences: $\left\{(f_1 - f_2)^2 : (f_1, f_2) \in \mathcal{F} \times \mathcal{F}\right\}$
$\pi \circ \mathcal{F}$	the class $\left\{(n, (y_\ell)_{\ell \in \mathbb{N}}) \in \mathcal{N} \times \mathcal{Y}^{\mathbb{N}} \mapsto \sum_{\ell=1}^n f(y_\ell) : f \in \mathcal{F}\right\}$
$\pi^2 \circ \mathcal{F}$	the class $\left\{(n, (y_\ell)_{\ell \in \mathbb{N}}) \in \mathcal{N} \times \mathcal{Y}^{\mathbb{N}} \mapsto \left(\sum_{\ell=1}^n f(y_\ell)\right)^2 : f \in \mathcal{F}\right\}$

If $\mathcal{F}$ is pointwise measurable, then so are $\mathcal{F}_\infty$ and $\mathcal{F}_\infty^2$. If F is an envelope of $\mathcal{F}$ such that $\|\|F\|\|_{2,P} < +\infty$, pointwise as:measurability of $\mathcal{F}$ also extends to $\mathcal{F}_\delta$ for any $\delta \in]0, +\infty[$. We refer to, e.g., Kosorok (2006) for a proof of these statements. We also have the pointwise measurability of $\pi \circ \mathcal{F}$ and $\pi \circ \mathcal{F}$.

Finally, we introduce the following notations to simplify somewhat our formulas below:

$\pi \circ f$	the function defined at the cell level $\left(n, (y_l)_{l \ge 1}\right) \mapsto \sum_{\ell=1}^n f(y_l)$
$\pi^2 \circ f$	the function defined at the cell level $\left(n, (y_l)_{l \ge 1}\right) \mapsto \left(\sum_{\ell=1}^n f(y_l)\right)^2$
$\vec{Y}_{\boldsymbol{j}}$	the sequence of the Y within the cell $\boldsymbol{j}$: $\left(Y_{\ell,\boldsymbol{j}}\right)_{\ell \ge 1}$
$\langle y \rangle$	the quantity $(1 + \|y\|^2)^{1/2}$
$\|\|g\|\|_{\mu_C,r}^r$	the random variable $\frac{1}{\Pi_C} \sum_{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C}} \left\|g(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}})\right\|^r$, for g a function from $\mathbb{N} \times \mathcal{Y}^{\mathbb{N}}$ to $\mathbb{R}$ and for $r \in \mathbb{N}$
$\|\|g\|\|_{\mu_C(\omega),r}^r$	is $\frac{1}{\Pi_C} \sum_{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C}} \left\|g(N_{\boldsymbol{j}}(\omega), \vec{Y}_{\boldsymbol{j}}(\omega))\right\|^r$, for a given $\omega \in \Omega$,
$\|\|g\|\|_{\mu_C^{\boldsymbol{e}},r}^r$	the random variable $\frac{1}{\prod_{s:\boldsymbol{e}_s=1} C_s} \sum_{\boldsymbol{e} \le \boldsymbol{c} \le \boldsymbol{C} \odot \boldsymbol{e}} \left\|\frac{1}{\prod_{s:\boldsymbol{e}_s=0} C_s} \sum_{(\mathbf{1}-\boldsymbol{e}) \le \boldsymbol{c}' \le \boldsymbol{C} \odot (\mathbf{1}-\boldsymbol{e})} g(N_{\boldsymbol{c}+\boldsymbol{c}'}, \vec{Y}_{\boldsymbol{c}+\boldsymbol{c}'})\right\|^r$, note that $\|\|g\|\|_{\mu_C,r} = \|\|g\|\|_{\mu_C^{\mathbf{1}},r}$
A_F	$= N_{\mathbf{1}} \sum_{\ell=1}^{N_{\mathbf{1}}} F\left(Y_{\ell,\mathbf{1}}\right)^2$
$A_\beta(\boldsymbol{j})$	$= \sum_{\ell=1}^{N_{\boldsymbol{j}}} \langle Y_{\ell,\boldsymbol{j}} \rangle^{-\beta}$
$\overline{N}_r$	$= \frac{1}{\Pi_C} \sum_{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C}} N_{\boldsymbol{j}}^r$ for $r \in \mathbb{N}$
$\overline{N}_r(\omega)$	$= \frac{1}{\Pi_C} \sum_{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C}} N_{\boldsymbol{j}}^r(\omega)$ for $r \in \mathbb{N}$
A_r	$= \frac{1}{\Pi_C} \sum_{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C}} \|A_\beta(\boldsymbol{j})\|^r = \frac{1}{\Pi_C} \sum_{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C}} \left\|\sum_{\ell=1}^{N_{\boldsymbol{j}}} \langle Y_{\ell,\boldsymbol{j}} \rangle^{-\beta}\right\|^r$ for $r \in \mathbb{N}$
$\mathbb{Q}_C^r$	$= \frac{1}{\overline{N}_r \Pi_C} \sum_{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C}} N_{\boldsymbol{j}}^{r-1} \sum_{\ell=1}^{N_{\boldsymbol{j}}} \delta_{Y_{\ell,\boldsymbol{j}}}$, with δ_y the Dirac measure at y and $r \in \mathbb{N}$

$||f||^2_{\mathbb{Q}^2_C,2}$ $= \frac{1}{\overline{N}_2\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}} N_{\boldsymbol{j}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})^2$, for f a function from $\mathcal{Y}$ to $\mathbb{R}$

$||f||^2_{\mathbb{Q}^2_C(\omega),2}$ $= \frac{1}{\overline{N}_2(\omega)\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}} N_{\boldsymbol{j}}(\omega)\sum_{\ell=1}^{N_{\boldsymbol{j}}(\omega)} f(Y_{\ell,\boldsymbol{j}}(\omega))^2$ for f a function from $\mathcal{Y}$ to $\mathbb{R}$

σ^2_C $= \sup_{\mathcal{F}_\delta}\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right)^2 = \sup_{\mathcal{F}_\delta}||\pi\circ f||^2_{\mu_C,2}$ for $\mathcal{F}$ a class of functions from $\mathcal{Y}$ to $\mathbb{R}$ and $\delta > 0$

If $N_{\boldsymbol{j}} = 0$ for any $\boldsymbol{j}\leq \boldsymbol{C}$ the random variables A_F, A_β, $\overline{N}_r$, A_r, $||f||^2_{\mathbb{Q}^2_C,2}$, σ^2_C are equal to 0, and the random measure $\mathbb{Q}^r_C$ is the null measure.

B.1.4 Notation and definition related to the covering numbers

$N(\varepsilon,\mathcal{F},||.||)$ the covering numbers, ie the minimal number of closed balls for the semi-norm $||.||$ of radius ε with center in $\mathcal{F}$ that is necessary to cover $\mathcal{F}$. We follow the convention adopted by Giné and Nickl (2015) or Kato (2017). This convention alleviates discussion of some degenerate cases. In particular if the semi-norm $||.||$ is null on $\mathcal{F}$ then $N(\epsilon,\mathcal{F},||.||) = 1$ for any $\varepsilon\geq 0$. Note that some authors do not consider closed balls but open balls and do not impose that centers of the balls belong to $\mathcal{F}$ (for instance in Definition 2.1.5 in van der Vaart and Wellner, 1996), resulting in slight variations of numerical constantes for some majorizations.

$J_{p,\mathcal{F}}(u)$ $= \int_0^u \sup_Q\sqrt{\log N\left(\varepsilon\,||F||_{Q,p},\mathcal{F},||.||_{Q,p}\right)}d\varepsilon$ where the supremum is taken on the set of measures on $\mathcal{Y}$ with finite support (including the null measure).

$J_{\infty,\beta,\mathcal{F}}(u)$ $= \int_0^u\sqrt{\log N(\varepsilon||F||_{\infty,\beta},\mathcal{F},||.||_{\infty,\beta})}d\varepsilon$.

B.1.5 Notation relative to the bootstrap

$W^i_{j_i}$ the bootstrap weight for the cluster j_i corresponding to the number of time where j_i is sampled during the sampling in dimension i of clustering

$W^{\boldsymbol{C}}_{\boldsymbol{j}}$ the bootstrap weight of cells, $W^{\boldsymbol{C}}_{\boldsymbol{j}} = \prod_{i=1}^k W^i_{j_i}$

$N^*_{\boldsymbol{j}},\vec{Y}^*_{\boldsymbol{j}}$ the bootstraped cell, corresponding to the intersection of the j_1-th draw in dimension $i=1$, j_2-th draw in dimension $i=2$,... In particular

$\mathbb{G}^*_C(f)$ the bootstrap process, $\mathbb{G}^*_C(f) = \frac{\sqrt{\underline{C}}}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}(W^{\boldsymbol{C}}_{\boldsymbol{j}}-1)\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) = \frac{\sqrt{\underline{C}}}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\left[\sum_{\ell=1}^{N^*_{\boldsymbol{j}}} f(Y^*_{\ell,\boldsymbol{j}}) - \sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right]$

$N^{\boldsymbol{e}*}_{\boldsymbol{j}},\vec{Y}^{\boldsymbol{e}*}_{\boldsymbol{j}}$ for $\mathbf{0}\leq \boldsymbol{e}\leq \mathbf{1}$, the cell whose component j_i corresponds to the component of the j_i-th draw in dimension i if $e_i = 1$ and with component j_i if $e_i = 0$. In particular: $(N^{\mathbf{1}*}_{\boldsymbol{j}},\vec{Y}^{\mathbf{1}*}_{\boldsymbol{j}}) = (N^*_{\boldsymbol{j}},\vec{Y}^*_{\boldsymbol{j}})$ and $(N^{\mathbf{0}*}_{\boldsymbol{j}},\vec{Y}^{\mathbf{0}*}_{\boldsymbol{j}}) = (N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}})$

A^*_F $= N^*_{\mathbf{1}}\sum_{\ell=1}^{N^*_{\mathbf{1}}} F^*\left(Y_{\ell,\mathbf{1}}\right)^2$

$A^*_\beta(\boldsymbol{j})$ $= \sum_{\ell=1}^{N^*_{\boldsymbol{j}}}\langle Y^*_{\ell,\boldsymbol{j}}\rangle^{-\beta}$

σ_C^{2*} $= \sup_{\mathcal{F}_\delta} \frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}^*} f(Y_{\ell,\boldsymbol{j}}^*) \right)^2$ for $\mathcal{F}$ a class of functions from $\mathcal{Y}$ to $\mathbb{R}$ and $\delta > 0$

B.2 Representation Lemma

Lemma B.1 *Let Assumptions 1.1 and 1.2 hold. Then there exists a measurable function τ such that*

$$\left\{ N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}} \right\}_{\boldsymbol{j} \geq \mathbf{1}} \overset{a.s.}{=} \left\{ \tau \left((U_{\boldsymbol{j} \odot \boldsymbol{e}})_{\mathbf{0} \prec \boldsymbol{e} \preceq \mathbf{1}} \right) \right\}_{\boldsymbol{j} \geq \mathbf{1}},$$

where $(U_{\boldsymbol{j}})_{\boldsymbol{j} > \mathbf{0}}$ is a family of mutually independent Uniform-$(0,1)$ random variables.

Proof
This representation follows from the equivalence of Condition (ii) and (iii) in Lemma 7.35 of Kallenberg (2005).

B.3 Proof of Theorem 3.1

The proof consists in three important steps. We first prove the asymptotic normality of $(\mathbb{G}_C(f_1), ..., \mathbb{G}_C(f_m))$ for any $m \geq 1$ and $(f_1, ..., f_m) \in (\mathcal{F})^m$. Second, we prove the asymptotic equicontinuity of $\mathbb{G}_C$. Third, and total boundedness of $\mathcal{F}$ for the norm $||.||_{P,2}$. By Theorems 1.5.4 and 1.5.7 in van der Vaart and Wellner (1996), this shows the weak convergence of $\mathbb{G}_C$ towards a centered gaussian process. The first and second step are themselves divided in several substeps.

B.3.1 Asymptotic normality of $(\mathbb{G}_C(f_1), ..., \mathbb{G}_C(f_m))$

By the Cramer-Wold device, it is sufficient to prove the asymptotic normality for any single function f of the form $f = \sum_{s=1}^m t_s f_s$, with $(t_s)_{s=1,...,m} \in \mathbb{R}^m$. Note for such a f, we have $\mathbb{E}\left(\left| \sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}) \right|^2 \right) \leq \sum_{s=1}^m |t_s| \mathbb{E}\left(\left| \sum_{\ell=1}^{N_{\mathbf{1}}} F(Y_{\ell,\mathbf{1}}) \right|^2 \right) \leq \sum_{s=1}^m |t_s| \mathbb{E}\left(N_{\mathbf{1}} \sum_{\ell=1}^{N_{\mathbf{1}}} F^2(Y_{\ell,\mathbf{1}}) \right) < \infty$.
We have to derive the weak limit of $\mathbb{G}_C(f) = \sqrt{\underline{C}}\left(\mathbb{P}_C(f) - \mathbb{P}(f)\right)$. The proof consists in four steps:

1. we introduce a Hájek projection $H_1(f)$ of $\mathbb{P}_C(f) - \mathbb{P}(f)$ on the of random variables $\sum_{\boldsymbol{c} \in \mathcal{I}_1(\boldsymbol{C})} g_{\boldsymbol{c}}(U_{\boldsymbol{c}})$ for $g_{\boldsymbol{c}} \in L^2([0,1])$ and derive the variance and the limit distribution of $\frac{\sqrt{\underline{C}} H_1(f)}{\mathbb{V}(\sqrt{\underline{C}} H_1(f))^{1/2}}$,
2. show that the variance of $\mathbb{V}(\mathbb{G}_C f) = \mathbb{V}(\sqrt{\underline{C}} H_1 f)(1 + o(1))$ when $\underline{C} \to \infty$,
3. conclude $\mathbb{G}_C f$ is asymptotically equivalent to $\sqrt{\underline{C}} H_1 f$.

1. **Hájek Projection** The Lemma C.2 ensures:

$$\sqrt{\underline{C}}H_1(f) \xrightarrow{d} \mathcal{N}\left(0, \sum_{i=1}^{k} \lambda_i \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}), \sum_{\ell=1}^{N_{\mathbf{2}_i}} f(Y_{\ell,\mathbf{2}_i})\right)\right),$$

$$\mathbb{V}(\sqrt{\underline{C}}H_1(f)) = \sum_{i=1}^{k} \frac{\underline{C}}{C_i} \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}), \sum_{\ell=1}^{N_{\mathbf{2}_i}} f(Y_{\ell,\mathbf{2}_i})\right),$$

with,

$$\sum_{i=1}^{k} \frac{\underline{C}}{C_i} \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}), \sum_{\ell=1}^{N_{\mathbf{2}_i}} f(Y_{\ell,\mathbf{2}_i})\right) = \sum_{e \in \mathcal{E}_1} \frac{\underline{C}}{\prod_{i:e_i=1} C_i} \mathbb{V}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}) \middle| U_e\right)\right).$$

2. **Variances comparison** We first expand $\mathbb{V}(\mathbb{P}_C f)$ using the law of total covariance:

$$\begin{aligned}
\mathbb{V}(\mathbb{P}_C f) &= \frac{1}{(\Pi_C)^2} \sum_{\substack{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C} \\ \mathbf{1} \le \boldsymbol{j}' \le \boldsymbol{C}}} \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}), \sum_{\ell=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell,\boldsymbol{j}'})\right) \\
&= \frac{1}{(\Pi_C)^2} \sum_{\substack{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C} \\ \mathbf{1} \le \boldsymbol{j}' \le \boldsymbol{C}}} \mathbb{C}ov\left(\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \middle| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}, U_{\boldsymbol{j}'\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}\right], \mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell,\boldsymbol{j}'}) \middle| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}, U_{\boldsymbol{j}'\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}\right]\right) \\
&+ \frac{1}{(\Pi_C)^2} \sum_{\substack{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C} \\ \mathbf{1} \le \boldsymbol{j}' \le \boldsymbol{C}}} \mathbb{E}\left[\mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}), \sum_{\ell=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell,\boldsymbol{j}'}) \middle| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}, U_{\boldsymbol{j}'\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}\right)\right]
\end{aligned}$$

For any $\boldsymbol{j}, \boldsymbol{j}'$, Assumption 1 ensures $\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \perp\!\!\!\perp \left\{U_{\boldsymbol{j}'\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1} \mid \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}$ and next:

$$\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \middle| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}, U_{\boldsymbol{j}'\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}\right] = \mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \middle| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}\right].$$

It follows that

$$\mathbb{V}(\mathbb{P}_C f) = A + B,$$

with $A = \frac{1}{(\Pi_C)^2} \sum_{\substack{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C} \\ \mathbf{1} \le \boldsymbol{j}' \le \boldsymbol{C}}} \mathbb{C}ov\left(\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \middle| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}\right], \mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell,\boldsymbol{j}'}) \middle| \left\{U_{\boldsymbol{j}'\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}\right]\right)$ and $B = \frac{1}{(\Pi_C)^2} \sum_{\substack{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C} \\ \mathbf{1} \le \boldsymbol{j}' \le \boldsymbol{C}}} \mathbb{E}\left[\mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}), \sum_{\ell=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell,\boldsymbol{j}'}) \middle| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}, U_{\boldsymbol{j}'\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}\right)\right]$. We will control these two terms.

   (a) $A = \mathbb{V}(H_1 f) + O(\underline{C}^{-2})$

   We can notice that $A = \mathbb{V}(S)$ with

$$S = \frac{1}{\Pi_C} \sum_{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C}} \mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \middle| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}\right].$$

S is a k-sample U-statistic of order 1 in each sample whose Hájek projection on $\sum_{\boldsymbol{c}\in\mathcal{I}_1(\boldsymbol{C})} g_{\boldsymbol{c}}(U_{\boldsymbol{c}})$ for $g_c \in L^2([0,1])$ is :

$$\sum_{\boldsymbol{c}\in\mathcal{I}_1(\boldsymbol{C})} \mathbb{E}(S|U_{\boldsymbol{c}}) = \sum_{\boldsymbol{c}\in\mathcal{I}_1(\boldsymbol{C})} \frac{1}{\Pi_C} \sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}} \mathbb{E}\left\{\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}f \middle| \{U_{\boldsymbol{j}\odot\boldsymbol{e}}\}_{\boldsymbol{e}\in\mathcal{E}_1}\right] \middle| U_{\boldsymbol{c}}\right\}$$

For $\boldsymbol{j}$ and $\boldsymbol{c}$ such that for any $\boldsymbol{e}\in\mathcal{E}_1$ we $\boldsymbol{j}\odot\boldsymbol{e}\neq\boldsymbol{c}$, independence of the U ensures that:

$$\mathbb{E}\left\{\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}f \middle| \{U_{\boldsymbol{j}\odot\boldsymbol{e}}\}_{\boldsymbol{e}\in\mathcal{E}_1}\right] \middle| U_{\boldsymbol{c}}\right\} = \mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}f\right] = 0.$$

For $\boldsymbol{j}$ and $\boldsymbol{c}$ such that for one $\boldsymbol{e}\in\mathcal{E}_1$ $\boldsymbol{j}\odot\boldsymbol{e}=\boldsymbol{c}$, the law of iterated expectation ensures:

$$\mathbb{E}\left\{\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}f \middle| \{U_{\boldsymbol{j}\odot\boldsymbol{e}}\}_{\boldsymbol{e}\in\mathcal{E}_1}\right] \middle| U_{\boldsymbol{c}}\right\} = \mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}f \middle| U_{\boldsymbol{c}}\right].$$

Because the U are iid, we also have:

$$\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}f \middle| U_{\boldsymbol{c}}\right] = \mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}}) - \mathbb{P}f \middle| U_{\boldsymbol{c}}\right],$$

and next,

$$\mathbb{E}\left\{\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}f \middle| \{U_{\boldsymbol{j}\odot\boldsymbol{e}}\}_{\boldsymbol{e}\in\mathcal{E}_1}\right] \middle| U_{\boldsymbol{c}}\right\} = \mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}}) - \mathbb{P}f \middle| U_{\boldsymbol{c}}\right].$$

For a $\boldsymbol{c}\in\mathcal{I}_1(\boldsymbol{C})$, there exist $\prod_{i:c_i=0} C_i$ tuples $\boldsymbol{j}$ such that $\boldsymbol{j}\odot\boldsymbol{e}=\boldsymbol{c}$ for some $\boldsymbol{e}\in\mathcal{E}_1$. It follows:

$$\begin{aligned}\sum_{\boldsymbol{c}\in\mathcal{I}_1(\boldsymbol{C})} \mathbb{E}(S|U_{\boldsymbol{c}}) &= \sum_{\boldsymbol{c}\in\mathcal{I}_1(\boldsymbol{C})} \frac{1}{\prod_{i:\boldsymbol{c}_i\neq 0} C_i}\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}}) - \mathbb{P}f \middle| U_{\boldsymbol{c}}\right]\\ &= H_1 f.\end{aligned}$$

We can therefore generalize ideas in the proof of Theorem 12.6 in van der Vaart (2000). As the Us are iid , we obtain:

$$A = \frac{1}{\Pi_C} \sum_{\boldsymbol{0}\leq\boldsymbol{e}\leq\boldsymbol{1}} \eta_{\boldsymbol{e}} \times \prod_{i=1}^{k} \binom{C_i - 1}{1 - e_i},$$

where

$$\eta_{\boldsymbol{e}} = \mathbb{C}ov\left(\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}) \middle| \{U_{\boldsymbol{e}'}\}_{\boldsymbol{e}'\in\mathcal{E}_1}\right], \mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{2}-\boldsymbol{e}}} f(Y_{\ell,\boldsymbol{2}-\boldsymbol{e}}) \middle| \{U_{(\boldsymbol{2}-\boldsymbol{e})\odot\boldsymbol{e}'}\}_{\boldsymbol{e}'\in\mathcal{E}_1}\right]\right).$$

For $\boldsymbol{e} \in \mathcal{E}_l$, we have $\prod_{i=1}^k \binom{C_i-1}{1-e_i} = O(\Pi_C \underline{C}^{-l})$ and since $\eta_{\mathbf{0}} = 0$ and $\sup_{\boldsymbol{e}\leq \mathbf{1}} |\eta_{\boldsymbol{e}}| \leq \mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right)^2\right] < \infty$, we have:

$$A = \frac{1}{\Pi_C}\left(\eta_0 + \sum_{\boldsymbol{e}\in\mathcal{E}_1} \eta_e \left(\prod_{i:e_i=0} C_i\right) + \sum_{\boldsymbol{e}\in\cup_{r=2}^k \mathcal{E}_r} O\left(\Pi_C \underline{C}^{-2}\right)\right)$$
$$= \sum_{\boldsymbol{e}\in\mathcal{E}_1} \frac{\eta_{\boldsymbol{e}}}{\prod_{i:e_i=1} C_i} + O\left(\underline{C}^{-2}\right)$$

Now observe that for any $\boldsymbol{e}$:

$$\eta_{\boldsymbol{e}} = \mathbb{E}\left[\mathbb{C}ov\left(\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\middle| \{U_{\boldsymbol{e}'}\}_{\boldsymbol{e}'\in\mathcal{E}_1}\right], \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{2}-\boldsymbol{e}}} f(Y_{\ell,\mathbf{2}-\boldsymbol{e}})\middle| \left\{U_{(\mathbf{2}-\boldsymbol{e})\odot\boldsymbol{e}'}\right\}_{\boldsymbol{e}'\in\mathcal{E}_1}\right]\middle| U_{\boldsymbol{e}}\right)\right]$$
$$+ \mathbb{C}ov\left(\mathbb{E}\left[\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\middle| \{U_{\boldsymbol{e}'}\}_{\boldsymbol{e}'\in\mathcal{E}_1}\right]\middle| U_{\boldsymbol{e}}\right], \mathbb{E}\left[\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{2}-\boldsymbol{e}}} f(Y_{\ell,\mathbf{2}-\boldsymbol{e}})\middle| \left\{U_{(\mathbf{2}-\boldsymbol{e})\odot\boldsymbol{e}'}\right\}_{\boldsymbol{e}'\in\mathcal{E}_1}\right]\middle| U_{\boldsymbol{e}}\right]\right).$$

The unique common element between $\{U_{\boldsymbol{e}'}\}_{\boldsymbol{e}'\in\mathcal{E}_1}$ and $\left\{U_{(\mathbf{2}-\boldsymbol{e})\odot\boldsymbol{e}'}\right\}_{\boldsymbol{e}'\in\mathcal{E}_1}$ is $U_{\boldsymbol{e}}$. So the independence of the U ensures that $\{U_{\boldsymbol{e}'}\}_{\boldsymbol{e}'\in\mathcal{E}_1} \perp\!\!\!\perp \left\{U_{(\mathbf{2}-\boldsymbol{e})\odot\boldsymbol{e}'}\right\}_{\boldsymbol{e}'\in\mathcal{E}_1} |U_{\boldsymbol{e}}$ and next:

$$\mathbb{C}ov\left(\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\middle| \{U_{\boldsymbol{e}'}\}_{\boldsymbol{e}'\in\mathcal{E}_1}\right], \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{2}-\boldsymbol{e}}} f(Y_{\ell,\mathbf{2}-\boldsymbol{e}})\middle| \left\{U_{(\mathbf{2}-\boldsymbol{e})\odot\boldsymbol{e}'}\right\}_{\boldsymbol{e}'\in\mathcal{E}_1}\right]\middle| U_{\boldsymbol{e}}\right) = 0.$$

Because for any $\boldsymbol{e} \in \mathcal{E}_1$, $(N_{\mathbf{1}}, (Y_{\ell,\mathbf{1}})_{l\geq 1})|U_{\boldsymbol{e}} \overset{d}{=} (N_{\mathbf{2}-\boldsymbol{e}}, (Y_{\ell,\mathbf{2}-\boldsymbol{e}})_{l\geq 1})|U_{\boldsymbol{e}}$, and $(\mathbf{2}-\boldsymbol{e})\odot \boldsymbol{e} = \boldsymbol{e}$ the law of iterated expectation ensures:

$$\mathbb{E}\left[\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{2}-\boldsymbol{e}}} f(Y_{\ell,\mathbf{2}-\boldsymbol{e}})\middle| \left\{U_{(\mathbf{2}-\boldsymbol{e})\odot\boldsymbol{e}'}\right\}_{\boldsymbol{e}'\in\mathcal{E}_1}\right]\middle| U_{\boldsymbol{e}}\right] = \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{2}-\boldsymbol{e}}} f(Y_{\ell,\mathbf{2}-\boldsymbol{e}})\middle| U_{\boldsymbol{e}}\right]$$
$$= \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\middle| U_{\boldsymbol{e}}\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\middle| \{U_{\boldsymbol{e}'}\}_{\boldsymbol{e}'\in\mathcal{E}_1}\right]\middle| U_{\boldsymbol{e}}\right].$$

It follows that for any $\boldsymbol{e} \in \mathcal{E}_1$: $\eta_{\boldsymbol{e}} = \mathbb{V}\left(\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\middle| U_{\boldsymbol{e}}\right]\right)$ and

$$A = \sum_{\boldsymbol{e}\in\mathcal{E}_1} \frac{1}{\prod_{i:e_i\neq 0} C_i} \mathbb{V}\left(\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\middle| U_{\boldsymbol{e}}\right]\right) + O\left(\underline{C}^{-2}\right)$$

(b) $B = O(\underline{C}^{-2})$

There are $\Pi_C \times \prod_{i=1}^k (C_i - 1)$ pairs $(\boldsymbol{j}, \boldsymbol{j}')$ without any common component. For these pairs, the independence of the U ensures $\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \perp\!\!\!\perp \sum_{\ell=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell,\boldsymbol{j}'}) \Big| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}, U_{\boldsymbol{j}'\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}$ which yields:

$$\mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}), \sum_{\ell=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell,\boldsymbol{j}'}) \Bigg| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}, U_{\boldsymbol{j}'\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}\right) = 0.$$

There are $\Pi_C \times \sum_{i=1}^{k} \prod_{i'\neq i}(C_i - 1) = O(\Pi_C^2 \times \underline{C}^{-1})$ pairs $(\boldsymbol{j}, \boldsymbol{j}')$ sharing only one component. For these pairs we also have $\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \perp\!\!\!\perp \sum_{\ell=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell,\boldsymbol{j}'}) \Big| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}, U_{\boldsymbol{j}'\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}$, and

$$\mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}), \sum_{\ell=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell,\boldsymbol{j}'}) \Bigg| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}, U_{\boldsymbol{j}'\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}\right) = 0.$$

For $r > 1$ there are $O(\Pi_C^2 \times \underline{C}^{-r})$ pairs $(\boldsymbol{j}, \boldsymbol{j}')$ sharing r components. And for such pairs, $\mathbb{E}\left[\mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}), \sum_{\ell=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell,\boldsymbol{j}'}) \Big| \left\{U_{\boldsymbol{j}\odot\boldsymbol{e}}, U_{\boldsymbol{j}'\odot\boldsymbol{e}}\right\}_{\boldsymbol{e}\in\mathcal{E}_1}\right)\right]$ does not depends on the particular choice of $(\boldsymbol{j}, \boldsymbol{j}')$.
Then $B = \frac{1}{\Pi_C^2}\left(\sum_{r=2}^{k} O(\Pi_C^2 \times \underline{C}^{-r})\right) = O(\underline{C}^{-2})$.

Remember that $\mathbb{V}(\mathbb{P}_C f) = A + B$, so we have:

$$\begin{aligned}\mathbb{V}(\mathbb{P}_C f) &= \sum_{\boldsymbol{e}\in\mathcal{E}_1} \frac{1}{\prod_{i:e_i\neq 0} C_i} \mathbb{V}\left(\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}) \Big| U_{\boldsymbol{e}}\right]\right) + O(\underline{C}^{-2}) \\ &= \mathbb{V}(H_1 f) + O(\underline{C}^{-2}).\end{aligned}$$

As a result, when $\underline{C} \to \infty$:

$$\mathbb{V}\left(\mathbb{G}_C f\right) = \underline{C}\mathbb{V}(\mathbb{P}_C f) = \mathbb{V}(\sqrt{\underline{C}} H_1 f) + O(\underline{C}^{-1}),$$

3. **Conclusion on asymptotic normality of $\mathbb{G}_C f$**
Note that $\lim_{\underline{C}\uparrow\infty} \mathbb{V}(\sqrt{\underline{C}} H_1 f) = \sum_{i=1}^{k} \lambda_i \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}), \sum_{\ell=1}^{N_{\boldsymbol{2}_i}} f(Y_{\ell,\boldsymbol{2}_i})\right) < \infty$.
If $\sum_{i=1}^{k} \lambda_i \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}), \sum_{\ell=1}^{N_{\boldsymbol{2}_i}} f(Y_{\ell,\boldsymbol{2}_i})\right) = 0$ then $\lim_{\underline{C}\uparrow\infty} \mathbb{V}(\mathbb{G}_C(f)) = 0$. Because $\mathbb{G}_C(f)$ is centered, this means that $\mathbb{G}_C(f)$ converges in L^2 hence in distribution to $0 \overset{d}{=} \mathcal{N}(0,0)$. Otherwise, $\frac{\mathbb{V}(\mathbb{G}_C f)}{\mathbb{V}(\sqrt{\underline{C}} H_1 f)}$. Theorem 11.2 in van der Vaart (2000) applies and allows us to state that

$$\frac{\mathbb{G}_C f}{\mathbb{V}(\mathbb{G}_C f)^{1/2}} - \frac{\sqrt{\underline{C}} H f}{\mathbb{V}(\sqrt{\underline{C}} H f)^{1/2}} = o_p(1).$$

It follows from the Slutsky Lemma and asymptotic normality of $\frac{\sqrt{\underline{C}} H f}{\mathbb{V}(\sqrt{\underline{C}} H f)^{1/2}}$:

$$\frac{\mathbb{G}_C f}{\mathbb{V}(\mathbb{G}_C f)^{1/2}} \xrightarrow{d} \mathcal{N}\left(0, 1\right).$$

Because $\mathbb{V}(\mathbb{G}_C f) = \mathbb{V}(\sqrt{\underline{C}}Hf) + o(1) = \sum_{i=1}^{k} \lambda_i \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}), \sum_{\ell=1}^{N_{\mathbf{2}_i}} f(Y_{\ell,\mathbf{2}_i})\right) + o(1)$, the Slutsky Lemma ensures:

$$\mathbb{G}_C f \xrightarrow{d} \mathcal{N}\left(0, \sum_{i=1}^{k} \lambda_i \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}), \sum_{\ell=1}^{N_{\mathbf{2}_i}} f(Y_{\ell,\mathbf{2}_i})\right)\right).$$

B.3.2 Asymptotic equicontinuity

The asymptotic continuity of $\mathbb{G}_C$ can be stated as

$$\lim_{\delta\to 0} \limsup_{\underline{C}\to+\infty} \mathbb{P}\left(\sup_{f\in\mathcal{F}_\delta} |\mathbb{G}_C f| > \epsilon\right) = 0. \tag{16}$$

Note that for every $\epsilon > 0$, Markov's inequality ensures

$$\mathbb{P}\left(\sup_{\mathcal{F}_\delta} |\mathbb{G}_C f| > \epsilon\right) \leq \frac{\sqrt{\underline{C}}}{\epsilon} \mathbb{E}\left[\sup_{\mathcal{F}_\delta} |\mathbb{P}_C f - \mathbb{P} f|\right],$$

so that it is sufficient to control the expectation term.

Step 1: Construction of an upper bound for the supremum of the empirical process with the entropy integral

Due to Lemma B.1, we have $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}} \stackrel{a.s.}{=} \left(\tau\left((U_{\boldsymbol{j}\odot\boldsymbol{e}})_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\right)\right)_{\boldsymbol{j}\geq\mathbf{1}}$ for $(U_{\boldsymbol{c}})_{\boldsymbol{c}>\mathbf{0}}$ independent family of uniform-$(0,1)$.

Lemma C.3 applied to $Z_{\boldsymbol{j}} = \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)$, $\mathcal{G} = \pi \circ \mathcal{F}_\delta = \{g\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right) = \sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}); f \in \mathcal{F}_\delta\}$ and $\Phi = Id$ ensures:

$$\mathbb{E}\left[\sup_{\mathcal{F}_\delta} |\mathbb{P}_C f - \mathbb{P} f|\right] \leq 2 \sum_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}} \mathbb{E}\left[\sup_{\mathcal{F}_\delta}\left|\frac{1}{\Pi_C} \sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}} \varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} f\left(Y_{\ell,\boldsymbol{j}}\right)\right|\right]$$

The term $2\sum_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}} \mathbb{E}\left[\sup_{\mathcal{F}_\delta}\left|\frac{1}{\Pi_C} \sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}} \varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} f\left(Y_{\ell,\boldsymbol{j}}\right)\right|\right]$ can be controlled thanks to Lemma C.8, which ensure the existence of $K(F, F_{N_{\mathbf{1}},\vec{Y}_{\mathbf{1}}})$ depending only of F and the distribution of $(N_{\mathbf{1}}, \vec{Y}_{\mathbf{1}})$ such that

$$\mathbb{E}\left[\sup_{\mathcal{F}_\delta} |\mathbb{P}_C f - \mathbb{P} f|\right] \leq \frac{K(F, F_{N_{\mathbf{1}},\vec{Y}_{\mathbf{1}}})}{\sqrt{\underline{C}}}\left\{\sqrt{\mathbb{E}\left[\sigma_C^2\right]} + J_{2,\mathcal{F}}\left(\frac{1}{4}\sqrt{\frac{\mathbb{E}\left[\sigma_C^2\right]}{\mathbb{E}\left[A_F\right]}}\right)\right\},$$

where $A_F = N_{\mathbf{1}} \sum_{\ell=1}^{N_{\mathbf{1}}} F\left(Y_{\ell,\mathbf{1}}\right)^2$ and $\sigma_C^2 = \sup_{\mathcal{F}_\delta} \frac{1}{\Pi_C} \sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}} \left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right)^2$ under Assumptions 1, 2, and 3, or

$$\mathbb{E}\left[\sup_{\mathcal{F}_\delta} |\mathbb{P}_C f - \mathbb{P} f|\right] =\leq \frac{K(F, F_{N_{\mathbf{1}},\vec{Y}_{\mathbf{1}}})}{\sqrt{\underline{C}}}\left\{\sqrt{\mathbb{E}\left[\sigma_C^2\right]} + J_{\infty,\beta,\mathcal{F}}\left(\frac{1}{4}\sqrt{\frac{\mathbb{E}\left[\sigma_C^2\right]}{||F||_{\infty,\beta}^2\,\mathbb{E}\left[A_\beta\right]}}\right)\right\},$$

where $A_\beta := \left(\sum_{\ell=1}^{N_{\mathbf{1}}}(1+|Y_{\ell,\mathbf{1}}|^2)^{-\beta/2}\right)^2$, under Assumptions 1, 2 and 3'.

Step 2: Show $\lim_{\delta\to 0}\limsup_{\underline{C}\to+\infty}\mathbb{E}\left[\sigma_C^2\right]=0$ Because

$$\begin{aligned}
\sigma_C^2 &= \sup_{\mathcal{F}_\delta}\frac{1}{\Pi_C}\sum_{\mathbf{1}\le\boldsymbol{j}\le\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\\
&\le \sup_{\mathcal{F}_\delta}\left\{\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\le\boldsymbol{j}\le\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2-\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\right]\right|+\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\right]\right\}\\
&\le \sup_{\mathcal{F}_\delta}\left\{\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\le\boldsymbol{j}\le\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2-\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\right]\right|\right\}+\sup_{\mathcal{F}_\delta}\left\{\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\right]\right\}\\
&\le \sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\le\boldsymbol{j}\le\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2-\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\right]\right|+\delta^2,
\end{aligned}$$

what needs to be shown is

$$\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\le\boldsymbol{j}\le\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2-\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\right]\right|\xrightarrow[\underline{C}\to\infty]{L_1}0.$$

Lemma C.3 applied to $Z_{\boldsymbol{j}}=\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)$, $\mathcal{G}=\pi\circ\mathcal{F}_\infty=\left\{g\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)=\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2;f\in\mathcal{F}_\infty\right\}$ and $\Phi=Id$ ensures

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\le\boldsymbol{j}\le\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2-\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\right]\right|\right]\\
&\le 2\sum_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\le\boldsymbol{j}\le\boldsymbol{C}}\varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f\left(Y_{\ell,\boldsymbol{j}}\right)\right)^2\right|\right].
\end{aligned}\tag{17}$$

Let $\widetilde{F}\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)=\sum_{\ell=1}^{N_{\boldsymbol{j}}}F(Y_{\ell,\boldsymbol{j}})$. For every $f\in\mathcal{F}_\infty$ we have $|f|\le 2F$, $\mathbb{E}(\widetilde{F}^2)\le\mathbb{E}(A_F^2)<\infty$ under Assumption 3 and $\mathbb{E}(\widetilde{F}^2)\le||F||_{\infty,\beta}^2\mathbb{E}(A_\beta)<\infty$ under Assumption 3. Next, we split the expectations in the upper bound in two, depending on whether $4\widetilde{F}^2\le M$ or not, for some arbitrary M.

For every $\boldsymbol{e}$ such that $\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}$, the triangle inequality ensures

$$\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\le\boldsymbol{j}\le\boldsymbol{C}}\varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f\left(Y_{\ell,\boldsymbol{j}}\right)\right)^2\mathbb{1}\left\{4\widetilde{F}^2>M\right\}\right|\right]\le 4\mathbb{E}\left[\widetilde{F}\left(N_{\mathbf{1}},\vec{Y}_{\mathbf{1}}\right)^2\mathbb{1}\left\{4\widetilde{F}^2>M\right\}\right],$$

and next,

$$2\sum_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\le\boldsymbol{j}\le\boldsymbol{C}}\varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f\left(Y_{\ell,\boldsymbol{j}}\right)\right)^2\mathbb{1}\left\{4\widetilde{F}^2>M\right\}\right|\right]\le 8(2^k-1)\mathbb{E}\left[\widetilde{F}\left(N_{\mathbf{1}},\vec{Y}_{\mathbf{1}}\right)^2\mathbb{1}\left\{4\widetilde{F}^2>M\right\}\right]$$

which vanishes when $M \to +\infty$ by the Dominated Convergence Theorem (under Assumption 3 or 3').

Lemma C.9 ensures that there exists u, a non increasing function from $]0,+\infty[$ to $[0,+\infty[$ and $K'(F, F_{N_{\mathbf{1}},\vec{Y}_{\mathbf{1}}}) > 0$, such that for every $M > 0$ and every $\eta > 0$,

$$2\sum_{\mathbf{0}\prec \boldsymbol{e}\preceq \mathbf{1}} \mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}} \varepsilon_{\boldsymbol{j}\odot \boldsymbol{e}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f\left(Y_{\ell,\boldsymbol{j}}\right)\right)^2\right| \mathbb{1}\{4\widetilde{F}^2 \leq M\}\right] = K'(F, F_{N_{\mathbf{1}},\vec{Y}_{\mathbf{1}}})\left(\frac{\sqrt{M}u(\eta)}{\sqrt{\underline{C}}} + \eta\right).$$

It follows that for every $M > 0$ and every $\eta > 0$

$$\begin{aligned}&\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right)^2 - \mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right)^2\right]\right|\right]\\ &= K''(F, F_{N_{\mathbf{1}},\vec{Y}_{\mathbf{1}}})\left(\mathbb{E}\left[\widetilde{F}\left(N_{\mathbf{1}},\vec{Y}_{\mathbf{1}}\right)^2 \mathbb{1}\left\{4\widetilde{F}^2 > M\right\}\right] + \frac{Mu(\eta)}{\sqrt{\underline{C}}} + \eta\right),\end{aligned} \tag{18}$$

for some K'' depending only on the envelope F and on the distribution of $(N_{\mathbf{1}},\vec{Y}_{\mathbf{1}})$.

In the Equation (18), we can fix M and η such that $\mathbb{E}\left[\widetilde{F}\left(N_{\mathbf{1}},\vec{Y}_{\mathbf{1}}\right)^2 \mathbb{1}\left\{4\widetilde{F}^2 > M\right\}\right] + \eta$ is arbitrary small, next for $\underline{C}$ sufficiently large, $\mathbb{E}\left[\widetilde{F}\left(N_{\mathbf{1}},\vec{Y}_{\mathbf{1}}\right)^2 \mathbb{1}\left\{4\widetilde{F}^2 > M\right\}\right] + \frac{Mu(\eta)}{\sqrt{\underline{C}}} + \eta$ is arbitrary small. It follows that

$$\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right)^2 - \mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right)^2\right]\right| \xrightarrow[\underline{C}\to\infty]{L_1} 0.$$

This is enough to conclude that $\lim_{\delta\to 0}\limsup_{\underline{C}\to+\infty}\mathbb{E}\left[\sigma_C^2\right] = 0$.

Step 3: Conclusion on asymptotic equicontinuity

Under Assumptions 1, 2 and 3 (respectivelly 3), we have shown in the previous step that $\mathbb{E}\left[\sigma_C^2\right] = \delta^2 + o(1)$ and

$$\mathbb{E}\left[\sup_{\mathcal{F}_\delta}|\mathbb{P}_C f - \mathbb{P}f|\right] \leq \frac{K(F, F_{N_{\mathbf{1}},\vec{Y}_{\mathbf{1}}})}{\sqrt{\underline{C}}}\left(\sqrt{\mathbb{E}(\sigma_C^2)} + J_*\left(\frac{1}{2}\sqrt{\frac{\mathbb{E}\left[\sigma_C^2\right]}{Deno}}\right)\right),$$

with $* = 2, \mathcal{F}$ and $Deno = \mathbb{E}(A_F)$ (respectivelly $* = \infty, \beta, \mathcal{F}$ and $Deno = ||F||^2_{\infty,\beta}\mathbb{E}\left[A_\beta\right]$). The continuity of J_* at 0 ensures that $\lim_{\delta\to 0}\limsup_{\underline{C}\to+\infty}\sqrt{\underline{C}}\mathbb{E}\left[\sup_{\mathcal{F}_\delta}|\mathbb{P}_C f - \mathbb{P}f|\right] = 0$. As discussed at the start of the proof, this ensures (16) is satisfied by Markov's inequality.

B.3.3 Total boundedness

Let $||f||_{P,2} = \sqrt{\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right)^2\right]}$.

Under Assumptions 1-2 and 3, the proof of the total boundedness of $\mathcal{F}$ under the $||\cdot||_{P,2}$ semimetric is similar to that in van der Vaart and Wellner (1996), adapted to our framework. We first fix $\varepsilon > 0$ and next we will sgow that $N\left(\varepsilon, \mathcal{F}, ||\cdot||_{P,2}\right) < \infty$.
In the previous step, we have shown that

$$\sup_{\mathcal{F}_\infty} \left| \frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \right)^2 - \mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}) \right)^2 \right] \right| \xrightarrow[\underline{C}\to\infty]{L_1} 0.$$

This convergence holds in probability as well. Since any sequence that converges in probability admits an almost surely converging subsequence, we can pick one of those subsequences for some ω in Ω' the set of probability one where convergence occurs. Any such subsequence is deterministic. This implies that it is possible to construct a sequence of deterministic measures $\frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \delta_{\left(N_{\boldsymbol{j}}(\omega), \vec{Y}_{\boldsymbol{j}}(\omega)\right)}$ such that the previous convergence holds in a non-random sense (it is implicit that $(C_i)_{i=1}^k$ corresponds to a re-indexing since we focus on subsequences). As a result, we can take Π_C large enough so that the previous supremum is bounded by ε^2. This implies that for every $(f, g) \in (\mathcal{F})^2$

$$0 \leq ||f - g||_{P,2}^2 \leq ||\pi \circ f - \pi \circ g||_{\mu_C(\omega),2}^2 + \varepsilon^2.$$

This implies that

$$N\left(\varepsilon, \mathcal{F}, ||\cdot||_{P,2}\right) \leq N\left(\frac{\varepsilon}{\sqrt{2}}, \pi \circ \mathcal{F}, ||\cdot||_{\mu_C(\omega),2}\right).$$

If $\overline{N}_2(\omega) = 0$ then $\mu_C(\omega)$ is the null measure on $\pi \circ \mathcal{F}$, and it follows:

$$N\left(\varepsilon, \mathcal{F}, ||\cdot||_{P,2}\right) = N\left(\frac{\varepsilon}{\sqrt{2}}, \pi \circ \mathcal{F}, ||\cdot||_{\mu_C(\omega),2}\right) = 1.$$

If $\overline{N}_2(\omega) > 0$, Lemmas C.6 $i)$ and C.1 and the uniform entropy condition ensure

$$\begin{aligned}
N\left(\frac{\varepsilon}{\sqrt{2}}, \pi \circ \mathcal{F}, ||\cdot||_{\mu_C(\omega),2}\right) &\leq N\left(\frac{\varepsilon}{\sqrt{2\overline{N}_2(\omega)}}, \mathcal{F}, ||\cdot||_{\mathbb{Q}_C^2(\omega),2}\right) \\
&\leq N\left(\frac{\varepsilon}{\sqrt{2\overline{N}_2(\omega)}\, ||F||_{\mathbb{Q}_C^2(\omega),2}} ||F||_{\mathbb{Q}_C^2(\omega),2}, \mathcal{F}, ||\cdot||_{\mathbb{Q}_C^2(\omega),2}\right) \\
&\leq N\left(\frac{\varepsilon}{\sqrt{2\overline{N}_2(\omega)}\, ||F||_{\mathbb{Q}_C^2(\omega),2}} ||F||_{\mathbb{Q}_C^2(\omega),2}, \mathcal{F}, ||\cdot||_{\mathbb{Q}_C^2(\omega),2}\right)
\end{aligned}$$

with $\overline{N}_2(\omega) = \frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} N_{\boldsymbol{j}}^2(\omega)$ and $||f||_{\mathbb{Q}_C^2(\omega),2}^2 = \frac{1}{\overline{N}_2(\omega)\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} N_{\boldsymbol{j}}(\omega) \sum_{\ell=1}^{N_{\boldsymbol{j}}(\omega)} f(Y_{\ell,\boldsymbol{j}}(\omega))^2$.
The uniform entropy condition ensures that

$$N\left(\eta ||F||_{\mathbb{Q}_C^2(\omega),2}, \mathcal{F}, ||\cdot||_{\mathbb{Q}_C^2(\omega),2}\right) \leq \sup_{Q: Supp(Q) < \infty} N\left(\eta ||F||_{Q,2}, \mathcal{F}, ||\cdot||_{Q,2}\right) < +\infty \text{ for any } \eta > 0.$$

This is in particular true for $\eta = \frac{\varepsilon}{\sqrt{2\overline{N}_2(\omega)}||F||_{\mathbb{Q}^2_C(\omega),2}}$, we conclude that:

$$N\left(\varepsilon, \mathcal{F}, ||\cdot||_{P,2}\right) < +\infty \text{ for any } \varepsilon > 0.$$

So we conclude that $(\mathcal{F}, ||.||_{P,2})$ is totally bounded under Assumption 3.
Under Assumptions 1-2 and 3', total boundedness of $(\mathcal{F}, ||.||_{P,2})$ under $||\cdot||_{P,2}$ is more straightforward. Notice first that for every $(f,g) \in (\mathcal{F})^2$

$$||f||_{P,2} \leq ||f||_{\infty,\beta}\sqrt{\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}\langle Y_{\ell,\mathbf{1}}\rangle^{-\beta}\right)^2\right]}.$$

Then Lemma C.5 *ii*) and *i*) ensure that:

$$\begin{aligned}
N(\varepsilon, \mathcal{F}, ||.||_{P,2}) &\leq N\left(\varepsilon, \mathcal{F}, \sqrt{\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}\langle Y_{\ell,\mathbf{1}}\rangle^{-\beta}\right)^2\right]}||.||_{\infty,\beta}\right)\\
&\leq N\left(\frac{\varepsilon}{\sqrt{\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}\langle Y_{\ell,\mathbf{1}}\rangle^{-\beta}\right)^2\right]}}, \mathcal{F}, ||.||_{\infty,\beta}\right)\\
&\leq N\left(\eta||F||_{\infty,\beta}, \mathcal{F}, ||.||_{\infty,\beta}\right) \text{ for } \eta = \frac{\varepsilon}{\sqrt{\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}\langle Y_{\ell,\mathbf{1}}\rangle^{-\beta}\right)^2\right]}||F||_{\infty,\beta}}
\end{aligned}$$

The integral condition of Assumption 3' ensures that this last quantity is finite. Then $(\mathcal{F}, ||.||_{P,2})$ is totally bounded under Assumption 3'.

B.4 Proof of Theorem 3.2

The proof is divided in several steps that mirror the steps of the proof of Theorem 3.1. We first prove the consistency of $(\mathbb{G}^*_C(f_1), ..., \mathbb{G}^*_C(f_m))$ for any $m \geq 1$ and $(f_1, ..., f_m) \in (\mathcal{F})^m$. In a second step, we prove the asymptotic equicontinuity of the boostrap process. Note that the total boundedness is a property of $\mathcal{F}$ that has already been established in the proof of Theorem 3.1.

B.4.1 Consistency of $(\mathbb{G}^*_C(f_1), ..., \mathbb{G}^*_C(f_m))$

The Cramer-Wold device ensures that we only have to prove the asymptotic normality for a single function f such that $\mathbb{E}\left(\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right)^2\right) < \infty$. The proof of the finite dimensional convergence of the bootstrap process follows the steps of the proof for the initial process: characterization of the Hájek projection, variance comparison and conclusion.

1. **Hájek Projection** For $\boldsymbol{c} \in \mathcal{I}_r$, let $W_{\boldsymbol{c}}^{\boldsymbol{C}} = \prod_{i:c_i=1} W_{c_i}^i$ (when $\boldsymbol{c} \geq \mathbf{1}$, we have $W_{\boldsymbol{c}}^{\boldsymbol{C}} = \prod_{i=1}^k W_{c_i}^i$). Because the $(W_j^i)_{j\geq 1}$ are mutually independent across i and independent of the array $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}$, the Hájek projection of $\mathbb{G}_C^*(f)$ on

$$\sum_{\boldsymbol{e}\in\mathcal{E}_1} g_{\boldsymbol{e}}\left(\left(W_{\boldsymbol{c}}^{\boldsymbol{C}}\right)_{\boldsymbol{c}:\boldsymbol{c}\wedge\mathbf{1}=\boldsymbol{e}}, \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right)$$

with $g_{\boldsymbol{e}}$ square integrable functions, is:

$$\begin{aligned}\sum_{\boldsymbol{e}\in\mathcal{E}_1}^{k} \mathbb{E}\left(\mathbb{G}_C^*(f)\middle| \left(W_{\boldsymbol{c}}^{\boldsymbol{C}}\right)_{\boldsymbol{c}:\boldsymbol{c}\wedge\mathbf{1}=\boldsymbol{e}}, \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right) &= \sum_{\boldsymbol{e}\in\mathcal{E}_1} \frac{\sqrt{\underline{C}}}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}} \left(W_{\boldsymbol{c}}^{\boldsymbol{C}} - 1\right) a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c}) \\ &= \sum_{\boldsymbol{e}\in\mathcal{E}_1} \frac{\sqrt{\underline{C}}}{\prod_{i:e_i=1} \sqrt{C_i}} \mathbb{H}_{\boldsymbol{e}}(f),\end{aligned}$$

with $a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c}) = \frac{1}{\prod_{s:e_s=0} C_s} \sum_{\mathbf{1}-\boldsymbol{e}\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{e})} \sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}'}} f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}'}) - \frac{1}{\Pi_C} \sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}} \sum_{l\geq 1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})$, and $\mathbb{H}_{\boldsymbol{e}}(f) = \frac{1}{\prod_{i:e_i=1}\sqrt{C_i}} \sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}} \left(W_{\boldsymbol{c}}^{\boldsymbol{C}} - 1\right) a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})$. Because the $(W_j^i)_{j\geq 1}$ are mutually independent across i and independent of the array $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}$, we have:

$$\begin{aligned}&\mathbb{V}\left(\sum_{\boldsymbol{e}\in\mathcal{E}_1} \frac{\sqrt{\underline{C}}}{\prod_{i:e_i=1}\sqrt{C_i}} \mathbb{H}_{\boldsymbol{e}}(f)\middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right) \\ &= \sum_{\boldsymbol{e}\in\mathcal{E}_1} \frac{\underline{C}}{\prod_{i:e_i=1} C_i^2} \sum_{\boldsymbol{e}\leq\boldsymbol{c},\boldsymbol{c}'\leq\boldsymbol{C}\odot\boldsymbol{e}} \mathbb{E}\left(\left(W_{\boldsymbol{c}}^{\boldsymbol{C}} - 1\right)\left(W_{\boldsymbol{c}'}^{\boldsymbol{C}} - 1\right)\right) a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c}) a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c}').\end{aligned}$$

Noting that $\mathbb{E}(W_j^i) = 1$, $\mathbb{E}(W_j^i W_{j'}^i) = \mathbb{1}_{j=j'} + 1 - C_i^{-1}$, and for any $\boldsymbol{e} \in \mathcal{E}_1$ and $\boldsymbol{c}, \boldsymbol{c}'$ such that $\boldsymbol{e} \leq \boldsymbol{c}, \boldsymbol{c}' \leq \boldsymbol{C} \odot \boldsymbol{e}$, we have

$$\mathbb{E}\left(\left(W_{\boldsymbol{c}}^{\boldsymbol{C}} - 1\right)\left(W_{\boldsymbol{c}'}^{\boldsymbol{C}} - 1\right)\right) = \mathbb{E}\left(W_{\boldsymbol{c}}^{\boldsymbol{C}} W_{\boldsymbol{c}'}^{\boldsymbol{C}}\right) - 1 = \mathbb{1}_{\{\boldsymbol{c}=\boldsymbol{c}'\}} - \frac{1}{\prod_{i:e_i=1} C_i},$$

we have:

$$\begin{aligned}&\mathbb{V}\left(\sum_{\boldsymbol{e}\in\mathcal{E}_1} \frac{\sqrt{\underline{C}}}{\prod_{i:e_i=1}\sqrt{C_i}} \mathbb{H}_{\boldsymbol{e}}(f)\middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right) \\ &= \sum_{\boldsymbol{e}\in\mathcal{E}_1} \frac{\underline{C}}{\prod_{i:e_i=1} C_i} \left(\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}^{C_i} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2 - \left[\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}} a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right]^2\right).\end{aligned}$$

Because $\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}} a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c}) = 0$, we have:

$$\mathbb{V}\left(\sum_{\boldsymbol{e}\in\mathcal{E}_1} \frac{\sqrt{\underline{C}}}{\prod_{i:e_i=1}\sqrt{C_i}} \mathbb{H}_{\boldsymbol{e}}(f)\middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right) = \sum_{\boldsymbol{e}\in\mathcal{E}_1} \frac{\underline{C}}{\prod_{i:e_i=1} C_i^2} \sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2.$$

Lemma C.13 ensures that

$$\mathbb{V}\left(\sum_{\boldsymbol{e}\in\mathcal{E}_1}\frac{\sqrt{C}}{\prod_{i:e_i=1}\sqrt{C_i}}\mathbb{H}_{\boldsymbol{e}}(f)\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}\right)\xrightarrow{a.s.}\sum_{i=1}^{k}\lambda_i\mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}}f(Y_{\ell,\boldsymbol{1}}),\sum_{\ell=1}^{N_{\boldsymbol{2}_i}}f(Y_{\ell,\boldsymbol{2}_i})\right)$$

We have to prove the asymptotic normality conditionally on $\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}$, for almost-every $\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}$, ie for $\underline{C}\to\infty$ and any $t_{\boldsymbol{e}}\in\mathbb{R}$:

$$\mathbb{E}\left(exp(\mathrm{i}t_{\boldsymbol{e}}\mathbb{H}_{\boldsymbol{e}}(f))|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}\right)\xrightarrow{a.s.}\exp\left(-\frac{t_{\boldsymbol{e}}}{2}\mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}}f(Y_{\ell,\boldsymbol{1}}),\sum_{\ell=1}^{N_{\boldsymbol{2}-\boldsymbol{e}}}f(Y_{\ell,\boldsymbol{2}-\boldsymbol{e}})\right)\right).$$

To do so, we will apply the Lindeberg-Feller theorem. If $(N_{\boldsymbol{j}}^{\boldsymbol{e}*},(Y_{\ell,\boldsymbol{j}}^{\boldsymbol{e}*})_{\ell\geq1})_{\boldsymbol{C}\geq\boldsymbol{j}\geq\boldsymbol{1}}$ denotes the bootstrap sample obtained by the selection of the cells when sampling clusters of component i correponding to the non null componente of $\boldsymbol{e}$, $\mathbb{H}_{\boldsymbol{e}}(f)$ is also equal to:

$$\begin{aligned}\mathbb{H}_{\boldsymbol{e}}(f)&=\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\frac{1}{\prod_{i:e_i=1}\sqrt{C_i}}\left(a_{\boldsymbol{e}}^{*\boldsymbol{C}}(\boldsymbol{c})-a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)\\&=\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\frac{1}{\prod_{i:e_i=1}\sqrt{C_i}}\left[a_{\boldsymbol{e}}^{*\boldsymbol{C}}(\boldsymbol{c})-\mathbb{E}\left(a_{\boldsymbol{e}}^{*\boldsymbol{C}}(\boldsymbol{c})\mid\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}\right)\right],\end{aligned}$$

with $a_{\boldsymbol{e}}^{*\boldsymbol{C}}(\boldsymbol{c})=\frac{1}{\prod_{s:e_s=0}C_s}\sum_{\boldsymbol{1}-\boldsymbol{e}\leq\boldsymbol{c'}\leq\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})}\sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c'}}^{\boldsymbol{e}*}}f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c'}}^{\boldsymbol{e}*})-\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\sum_{l\geq1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})$.
Because the boostrap sampling in each component are done with replacement and equal probability, for any function g and any $\boldsymbol{e}\in\mathcal{E}_1$, we have:

$$\mathbb{E}\left[g(a_{\boldsymbol{e}}^{*\boldsymbol{C}}(\boldsymbol{c}))\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}\right]=\frac{1}{\prod_{i:e_i=1}C_i}\sum_{\boldsymbol{e}\leq\boldsymbol{j}\leq\boldsymbol{C}\odot\boldsymbol{e}}g(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{j})).$$

It follows that:

$$\begin{aligned}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\mathbb{V}\left(\frac{a_{\boldsymbol{e}}^{*\boldsymbol{C}}(\boldsymbol{c})}{\prod_{i:e_i=1}\sqrt{C_i}}\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}\right)&=\frac{1}{\prod_{i:e_i=1}C_i}\sum_{\boldsymbol{e}\leq\boldsymbol{j}\leq\boldsymbol{C}\odot\boldsymbol{e}}\left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{j})\right)^2\\&\quad-\left(\frac{1}{\prod_{i:e_i=1}C_i}\sum_{\boldsymbol{e}\leq\boldsymbol{j}\leq\boldsymbol{C}\odot\boldsymbol{e}}a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{j})\right)^2\\&=\frac{1}{\prod_{i:e_i=1}C_i}\sum_{\boldsymbol{e}\leq\boldsymbol{j}\leq\boldsymbol{C}\odot\boldsymbol{e}}\left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{j})\right)^2\end{aligned}$$

and for any $\varepsilon>0$:

$$\begin{aligned}&\frac{1}{\prod_{i:e_i=1}C_i}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\mathbb{E}\left(\left(a_{\boldsymbol{e}}^{*\boldsymbol{C}}(\boldsymbol{c})\right)^2\mathbb{1}\left\{\left|a_{\boldsymbol{e}}^{*\boldsymbol{C}}(\boldsymbol{c})\right|>(\Pi_{i:e_i=1}C_i)^{1/2}\varepsilon\right\}\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}\right)\\&=\frac{1}{\prod_{i:e_i=1}C_i}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\frac{1}{\prod_{i:e_i=1}C_i}\sum_{\boldsymbol{e}\leq\boldsymbol{j}\leq\boldsymbol{C}\odot\boldsymbol{e}}\left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{j})\right)^2\mathbb{1}\left\{\left|a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{j})\right|>(\Pi_{i:e_i=1}C_i)^{1/2}\varepsilon\right\}\\&=\frac{1}{\prod_{i:e_i=1}C_i}\sum_{\boldsymbol{e}\leq\boldsymbol{j}\leq\boldsymbol{C}\odot\boldsymbol{e}}\left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{j})\right)^2\mathbb{1}\left\{\left|a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{j})\right|>(\Pi_{i:e_i=1}C_i)^{1/2}\varepsilon\right\}.\end{aligned}$$

Lemma C.13 ensures that:

$$\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e}\le \boldsymbol{j}\le \boldsymbol{C}\odot\boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{j})\right)^2 \xrightarrow{a.s.} \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}), \sum_{\ell=1}^{N_{\boldsymbol{2}-\boldsymbol{e}}} f(Y_{\ell,\boldsymbol{2}-\boldsymbol{e}})\right) < \infty.$$

$$\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e}\le \boldsymbol{j}\le \boldsymbol{C}\odot\boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{j})\right)^2 \mathbb{1}\left\{\left|a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{j})\right| > \left(\Pi_{i:e_i=1} C_i\right)^{1/2}\varepsilon\right\} \xrightarrow{a.s.} 0$$

The Lindeberg-Feller Theorem (see for instance van der Vaart (2000), Section 2.8) ensures that, conditionally on $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}$, $\mathbb{H}_{\boldsymbol{e}}(f)$ converges weakly to

$$\mathcal{N}\left(0, \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}), \sum_{\ell=1}^{N_{\boldsymbol{2}-\boldsymbol{e}}} f(Y_{\ell,\boldsymbol{2}-\boldsymbol{e}})\right)\right),$$

for almost-every $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}$.

Because the $\mathbb{H}_{\boldsymbol{e}}(f)$ are mutually independent conditionally on $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}$, we have the joint asymptotic normality.

Next, the Slutsky's Lemma ensures that conditionally on $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}$ and for almost-surely $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}$, $\sum_{\boldsymbol{e}\in\mathcal{E}_1} \frac{\sqrt{\underline{C}}}{\prod_{i:e_i=1}\sqrt{C_i}} \mathbb{H}_{\boldsymbol{e}}(f)$ converges in distribution to

$$\mathcal{N}\left(0, \sum_{i=1}^{k} \lambda_i \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}), \sum_{\ell=1}^{N_{\boldsymbol{2}_i}} f(Y_{\ell,\boldsymbol{2}_i})\right)\right).$$

2. **Variance comparison**

Let $V = \mathbb{V}\left(\mathbb{G}_C^*(f)\middle|\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right)$, we have:

$$\begin{aligned}
V &= \frac{\underline{C}}{(\Pi_C)^2} \sum_{\boldsymbol{1}\le\boldsymbol{j},\boldsymbol{j}'\le\boldsymbol{C}} \left[\mathbb{E}(W_{\boldsymbol{j}}^{\boldsymbol{C}} W_{\boldsymbol{j}'}^{\boldsymbol{C}}) - 1\right] \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}_C(f)\right) \left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'}) - \mathbb{P}_C(f)\right) \\
&= \frac{\underline{C}}{(\Pi_C)^2} \sum_{\boldsymbol{1}\le\boldsymbol{j},\boldsymbol{j}'\le\boldsymbol{C}} \left[\prod_{i=1}^{k}\left(\mathbb{1}_{j_i=j_i'} + 1 - \frac{1}{C_i}\right) - 1\right] \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}_C(f)\right) \left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'}) - \mathbb{P}_C(f)\right) \\
&= \frac{\underline{C}}{(\Pi_C)^2} \sum_{\boldsymbol{0}\le\boldsymbol{e}\le\boldsymbol{1}} \sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_{\boldsymbol{e}}} \left[\prod_{i:e_i=1}\left(2 - \frac{1}{C_i}\right)\prod_{i:e_i=0}\left(1 - \frac{1}{C_i}\right) - 1\right] \\
&\qquad \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}_C(f)\right) \left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'}) - \mathbb{P}_C(f)\right)
\end{aligned}$$

Let focus on the term corresponding to $\boldsymbol{e}=\mathbf{0}$. Because

$$\sum_{\mathbf{1}\le \boldsymbol{j},\boldsymbol{j}'\le \boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})-\mathbb{P}_C(f)\right)\left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'})-\mathbb{P}_C(f)\right)=0,$$

we have:

$$\begin{aligned}
&\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_{\mathbf{0}}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})-\mathbb{P}_C(f)\right)\left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'})-\mathbb{P}_C(f)\right)\\
&=-\sum_{\mathbf{0}<\boldsymbol{e}\le\mathbf{1}}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_{\boldsymbol{e}}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})-\mathbb{P}_C(f)\right)\left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'})-\mathbb{P}_C(f)\right)
\end{aligned}$$

Then:

$$\begin{aligned}
V=\sum_{\mathbf{0}<\boldsymbol{e}\le\mathbf{1}}&\left[\prod_{i:e_i=0}\left(1-\frac{1}{C_i}\right)\left(\prod_{i:e_i=1}\left(2-\frac{1}{C_i}\right)-\prod_{i:e_i=1}\left(1-\frac{1}{C_i}\right)\right)\right]\\
&\qquad\frac{\underline{C}}{(\Pi_C)^2}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_{\boldsymbol{e}}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})-\mathbb{P}_C(f)\right)\left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'})-\mathbb{P}_C(f)\right)
\end{aligned}$$

We have $\mathcal{B}_{\boldsymbol{e}}\cap\mathcal{B}_{\boldsymbol{e}'}=\mathcal{B}_{\boldsymbol{e}\vee\boldsymbol{e}'}$ and $\mathcal{B}_{\boldsymbol{e}}=\mathcal{A}_{\boldsymbol{e}}\cup(\cup_{\boldsymbol{e}<\boldsymbol{e}'\le\mathbf{1}}\mathcal{B}_{\boldsymbol{e}'})$. Because $\mathcal{A}_{\boldsymbol{e}}\cap\mathcal{B}_{\boldsymbol{e}'}=\emptyset$ for any $\boldsymbol{e}'>\boldsymbol{e}$, we have $\mathbb{1}_{\mathcal{A}_{\boldsymbol{e}}}=\mathbb{1}_{\mathcal{B}_{\boldsymbol{e}}}-\mathbb{1}_{\cup_{\boldsymbol{e}<\boldsymbol{e}'\le\mathbf{1}}\mathcal{B}_{\boldsymbol{e}'}}$.

The inclusion-exclusion principle ensures:

$$\mathbb{1}_{\mathcal{A}_{\boldsymbol{e}}}=\mathbb{1}_{\mathcal{B}_{\boldsymbol{e}}}-\sum_{\boldsymbol{e}<\boldsymbol{e}'\le\mathbf{1}}(-1)^{\sum_{i=1}^k e'_i}\mathbb{1}_{\mathcal{B}_{\boldsymbol{e}'}},$$

then

$$\begin{aligned}
&\frac{\underline{C}}{(\Pi_C)^2}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_{\boldsymbol{e}}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})-\mathbb{P}_C(f)\right)\left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'})-\mathbb{P}_C(f)\right)\\
&=\frac{\underline{C}}{(\Pi_C)^2}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_{\boldsymbol{e}}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})-\mathbb{P}_C(f)\right)\left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'})-\mathbb{P}_C(f)\right)\\
&-\sum_{\boldsymbol{e}<\boldsymbol{e}'\le\mathbf{1}}(-1)^{\sum_{i=1}^k e'_i}\frac{\underline{C}}{(\Pi_C)^2}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_{\boldsymbol{e}'}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})-\mathbb{P}_C(f)\right)\left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'})-\mathbb{P}_C(f)\right).
\end{aligned}$$

Let $r=1,...,k-1$ for $\boldsymbol{e}$ such that $\boldsymbol{e}\in\mathcal{E}_r$, Lemma C.13 ensures:

$$\begin{aligned}
&\frac{\underline{C}}{(\Pi_C)^2}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_{\boldsymbol{e}}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})-\mathbb{P}_C(f)\right)\left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'})-\mathbb{P}_C(f)\right)\\
&=\frac{\underline{C}}{\prod_{i:e_i=1}C_i^2}\sum_{\boldsymbol{e}\le\boldsymbol{c}\le\boldsymbol{C}\odot\boldsymbol{e}}a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})^2=O(\underline{C}^{1-r})\text{ almost-surely,}
\end{aligned}$$

and for $\boldsymbol{e} \in \mathcal{E}_1$, the limit almost-sure is $\lambda_i \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}), \sum_{\ell=1}^{N_{\mathbf{2}_i}} f(Y_{\ell,\mathbf{2}_i})\right)$, for i the non null component of $\boldsymbol{e}$.

Last, Lemma 7.35 in Kallenberg (2005) ensures that $\frac{1}{\Pi_C} \sum_{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \xrightarrow{a.s.} \mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right)$ and $\frac{1}{\Pi_C} \sum_{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C}} \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right)^2 \xrightarrow{a.s.} \mathbb{E}\left(\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right)^2\right)$, so we have:

$$\begin{aligned}
&\frac{\underline{C}}{(\Pi_C)^2} \sum_{(\boldsymbol{j},\boldsymbol{j}') \in \mathcal{B}_{\mathbf{1}}} \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}_C(f)\right) \left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'}) - \mathbb{P}_C(f)\right) \\
&= \frac{\underline{C}}{(\Pi_C)^2} \sum_{\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{C}} \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}_C(f)\right)^2 = O(\underline{C}^{1-k}) \text{ almost-surely,}
\end{aligned}$$

Last, because for any $\boldsymbol{e} \in \mathcal{E}_r$, we have

$$\lim_{\underline{C} \to \infty} \left[\prod_{i:e_i=0} \left(1 - \frac{1}{C_i}\right) \left(\prod_{i:e_i=1} \left(1 - \frac{1}{C_i}\right) - \prod_{i:e_i=1} \left(1 - \frac{1}{C_i}\right)\right)\right] = 2^r - 1 + O(\underline{C}^{-1})$$

It follows that

$$\begin{aligned}
V = &\sum_{\mathbf{0} < \boldsymbol{e} \le \mathbf{1}} \left[\prod_{i:e_i=0} \left(1 - \frac{1}{C_i}\right) \left(\prod_{i:e_i=1} \left(1 - \frac{1}{C_i}\right) - \prod_{i:e_i=1} \left(1 - \frac{1}{C_i}\right)\right)\right] \\
&\qquad \frac{\underline{C}}{(\Pi_C)^2} \sum_{(\boldsymbol{j},\boldsymbol{j}') \in \mathcal{A}_{\boldsymbol{e}}^+} \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}_C(f)\right) \left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'}) - \mathbb{P}_C(f)\right) \\
= &\sum_{\boldsymbol{e} \in \mathcal{E}_1} \frac{\underline{C}}{(\Pi_C)^2} \sum_{(\boldsymbol{j},\boldsymbol{j}') \in \mathcal{A}_{\boldsymbol{e}}^+} \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}_C(f)\right) \left(\sum_{l'=1}^{N_{\boldsymbol{j}'}} f(Y_{\ell',\boldsymbol{j}'}) - \mathbb{P}_C(f)\right) + O(\underline{C}^{-1}) \\
= &\mathbb{V}\left(\sum_{\boldsymbol{e} \in \mathcal{E}_1} \frac{\sqrt{\underline{C}}}{\prod_{i:e_i=1} \sqrt{C_i}} \mathbb{H}_{\boldsymbol{e}}(f) \middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j} \ge \mathbf{1}}\right) + O(\underline{C}^{-1}),
\end{aligned}$$

with the $O(\underline{C}^{-1})$, being uniform on an almost-sure set Ω.

3. **Conclusion on asymptotic normality of $\mathbb{G}_C^*(f)$**

In step one, we have proved that

$$\mathbb{V}\left(\sum_{\boldsymbol{e} \in \mathcal{E}_1} \frac{\sqrt{\underline{C}}}{\prod_{i:e_i=1} \sqrt{C_i}} \mathbb{H}_{\boldsymbol{e}}(f) \middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j} \ge \mathbf{1}}\right) \xrightarrow{a.s.} \sum_{i=1}^{k} \lambda_i \mathbb{C}ov\left(\sum_{i=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}), \sum_{i=1}^{N_{\mathbf{2}_i}} f(Y_{\ell,\mathbf{2}_i})\right) < \infty.$$

In the previous step, we have shown that

$$\mathbb{V}\left(\mathbb{G}_C^*(f) \middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j} \ge \mathbf{1}}\right) - \mathbb{V}\left(\sum_{\boldsymbol{e} \in \mathcal{E}_1} \frac{\sqrt{\underline{C}}}{\prod_{i:e_i=1} \sqrt{C_i}} \mathbb{H}_{\boldsymbol{e}}(f) \middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j} \ge \mathbf{1}}\right) \xrightarrow{a.s.} 0$$

If $\sum_{i=1}^{k}\lambda_i \mathbb{C}ov\left(\sum_{i=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}), \sum_{i=1}^{N_{\mathbf{2}_i}} f(Y_{\ell,\mathbf{2}_i})\right)$, because $\mathbb{E}\left(\mathbb{G}_C^*(f)\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right)=0$ then $\mathbb{G}_C^*(f)$ converges in L^2 to 0 (almost-surely, conditionally on $\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}$) and next,

$$\mathbb{G}_C^*(f) \xrightarrow{d} \mathcal{N}(0,0), \text{ almost-surely.}$$

If $\sum_{i=1}^{k}\lambda_i \mathbb{C}ov\left(\sum_{i=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}), \sum_{i=1}^{N_{\mathbf{2}_i}} f(Y_{\ell,\mathbf{2}_i})\right) > 0$, because $\sum_{\boldsymbol{e}\in\mathcal{E}_1}\frac{\sqrt{C}}{\prod_{i:e_i=1}\sqrt{C_i}}\mathbb{H}_{\boldsymbol{e}}(f)$ is asymptotically normal, with the asymptotic variance being the almost-sure limit of $\mathbb{V}\left(\sum_{\boldsymbol{e}\in\mathcal{E}_1}\frac{\sqrt{\underline{C}}}{\prod_{i:e_i=1}\sqrt{C_i}}\mathbb{H}_{\boldsymbol{e}}(f)\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right)$, Theorem 11.2 in van der Vaart (2000) combined with the Slustky's Lemma ensures that

$$\mathbb{G}_C^*(f) \xrightarrow{d} \mathcal{N}\left(0, \sum_{i=1}^{k}\lambda_i \mathbb{C}ov\left(\sum_{i=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}), \sum_{i=1}^{N_{\mathbf{2}_i}} f(Y_{\ell,\mathbf{2}_i})\right)\right), \text{ almost-surely.}$$

B.4.2 Asymptotic equicontinuity for the bootstrap

What we want to show is $\lim_{\delta\to 0}\limsup_{\underline{C}\to+\infty}\sqrt{\underline{C}}\mathbb{E}\left[\sup_{\mathcal{F}_\delta}|\mathbb{P}_C^* f-\mathbb{P}_C f| \mid \left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]=0$ with probability approaching one.

Step 1: Construction of an upper bound for the supremum of the empirical process with the entropy integral

Conditional on $\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}$, $\left(N^*_{\boldsymbol{j}},\vec{Y}^*_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}$ is a separately exchangeable and dissociated random array. As a result, Lemma B.1 applies and we have $\left(N^*_{\boldsymbol{j}},\vec{Y}^*_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}} \overset{a.s.}{=} \left(\tau\left(\left(U_{\boldsymbol{j}\odot\boldsymbol{e}}\right)_{\mathbf{0}<\boldsymbol{e}\leq\mathbf{1}}\right)\right)_{\boldsymbol{j}\geq\mathbf{1}}$ for $(U_{\boldsymbol{c}})_{\boldsymbol{c}>\mathbf{0}}$ an independent family of uniform-$(0,1)$s and τ that depends on $\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}$.

Lemma C.3 applied to $Z_{\boldsymbol{j}}=\left(N^*_{\boldsymbol{j}},\vec{Y}^*_{\boldsymbol{j}}\right)$, $\mathcal{G}=\pi\circ\mathcal{F}_\delta=\{g\left(N^*_{\boldsymbol{j}},\vec{Y}^*_{\boldsymbol{j}}\right)=\sum_{\ell=1}^{N^*_{\boldsymbol{j}}} f(Y^*_{\ell,\boldsymbol{j}}); f\in\mathcal{F}_\delta\}$ and $\Phi=Id$ ensures:

$$\mathbb{E}\left[\sup_{\mathcal{F}_\delta}|\mathbb{P}_C^* f-\mathbb{P}_C f|\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right] \leq 2\sum_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\mathbb{E}\left[\sup_{\mathcal{F}_\delta}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\sum_{\ell=1}^{N^*_{\boldsymbol{j}}} f\left(Y^*_{\ell,\boldsymbol{j}}\right)\right|\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]$$

By the law of iterated expectations and independence between $(\varepsilon_{\boldsymbol{j}})_{\boldsymbol{j}\geq\mathbf{1}}$ and the rest, we have

$$\begin{aligned}
&2\sum_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\mathbb{E}\left[\sup_{\mathcal{F}_\delta}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\sum_{\ell=1}^{N^*_{\boldsymbol{j}}} f\left(Y^*_{\ell,\boldsymbol{j}}\right)\right|\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]\\
&=2\sum_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\mathbb{E}\left[\mathbb{E}\left[\sup_{\mathcal{F}_\delta}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\sum_{\ell=1}^{N^*_{\boldsymbol{j}}} f\left(Y^*_{\ell,\boldsymbol{j}}\right)\right|\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}},N^*_{\boldsymbol{j}},\vec{Y}^*_{\boldsymbol{j}}\right)_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\right]\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]
\end{aligned}$$

Conditional on $\left(N_{\boldsymbol{j}}^*, \vec{Y}_{\boldsymbol{j}}^*, N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{1}\le\boldsymbol{j}\le\boldsymbol{C}}$, $f \mapsto \frac{1}{\Pi_C}\sum_{\boldsymbol{1}\le\boldsymbol{j}\le\boldsymbol{C}} \varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}} \sum_{\ell=1}^{N_{\boldsymbol{j}}^*} f\left(Y_{\ell,\boldsymbol{j}}^*\right)$ is a sub-Gaussian process. Lemma that can be handled thanks to Lemma C.8 remains valid to control the expectation of this process conditional on $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}$, so that we obtain $P-a.s$

$$\mathbb{E}\left[\sup_{\mathcal{F}_\delta}|\mathbb{P}_C^* f - \mathbb{P}_C f| \,\middle|\, \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right] \le 2\times 2^{k-1}\times\left\{4\sqrt{\frac{2\mathbb{E}\left[\sigma_C^{*2}\middle|\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right]\log 2}{\underline{C}}}\right.$$
$$\left. +32\sqrt{\frac{2\mathbb{E}\left[A_F^*\middle|\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right]}{\underline{C}}} J_{2,\mathcal{F}}\left(\frac{1}{4}\sqrt{\frac{\mathbb{E}\left[\sigma_C^{*2}\middle|\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right]}{\mathbb{E}\left[A_F^*\middle|\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right]}}\right)\right\}, \tag{19}$$

where $A_F^* := N_{\boldsymbol{1}}^*\sum_{\ell=1}^{N_{\boldsymbol{1}}^*} F\left(Y_{\ell,\boldsymbol{1}}^*\right)^2$ and $\sigma_C^{*2} = \sup_{\mathcal{F}_\delta}\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\le\boldsymbol{j}\le\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}^*} f(Y_{\ell,\boldsymbol{j}}^*)\right)^2$ under Assumptions 1, 2, and 3, or

$$\mathbb{E}\left[\sup_{\mathcal{F}_\delta}|\mathbb{P}_C^* f - \mathbb{P}_C f| \,\middle|\, \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right] \le 2\times 2^{k-1}\times\left\{4\sqrt{\frac{2\mathbb{E}\left[\sigma_C^{*2}\middle|\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right]\log 2}{\underline{C}}}\right.$$
$$\left. +32||F||_{\infty,\beta}\times\sqrt{\frac{2\mathbb{E}\left[A_\beta^*\middle|\left(N_{\boldsymbol{j}}^*, \vec{Y}_{\boldsymbol{j}}^*\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right]}{\underline{C}}} J_{\infty,\beta,\mathcal{F}}\left(\frac{1}{4}\sqrt{\frac{\mathbb{E}\left[\sigma_C^{*2}\middle|\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right]}{||F||_{\infty,\beta}\,\mathbb{E}\left[A_\beta^*\middle|\left(N_{\boldsymbol{j}}^*, \vec{Y}_{\boldsymbol{j}}^*\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right]}}\right)\right\}, \tag{20}$$

where $A_\beta^* := \left(\sum_{\ell=1}^{N_{\boldsymbol{1}}^*}\langle Y_{1,\boldsymbol{1}}^*\rangle^{-\beta}\right)^2$, under Assumptions 1, 2 and 3'.
What is more, by definition of the bootstrap scheme and Lemma 7.35 in Kallenberg (2005)

$$\mathbb{E}\left[A_F^*\middle|\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right] \overset{a.s}{=} \frac{1}{\Pi_C}\sum_{\boldsymbol{1}\le\boldsymbol{j}\le\boldsymbol{C}} N_{\boldsymbol{j}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} F(Y_{\ell,\boldsymbol{j}})^2 \xrightarrow[\underline{C}\to+\infty]{a.s} \mathbb{E}\left[N_{\boldsymbol{1}}\sum_{\ell=1}^{N_{\boldsymbol{1}}} F(Y_{\ell,\boldsymbol{1}})^2\right] > 0,$$

and

$$\mathbb{E}\left[A_\beta^*\middle|\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right] \overset{a.s}{=} \frac{1}{\Pi_C}\sum_{\boldsymbol{1}\le\boldsymbol{j}\le\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}\langle Y_{\ell,\boldsymbol{j}}\rangle^{-\beta}\right)^2 \xrightarrow[\underline{C}\to+\infty]{a.s} \mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}}\langle Y_{\ell,\boldsymbol{1}}\rangle^{-\beta}\right)^2\right] > 0.$$

Step 2: Show $\mathbb{E}\left[\sigma_C^{*2}\middle|\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right] = \delta^2 + o_P(1)$ as $\underline{C}\to+\infty$.

Observe that

$$
\begin{aligned}
&\mathbb{E}\left[\sigma_C^{*2}\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]\\
&\leq \mathbb{E}\left[\sup_{\mathcal{F}_\delta}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}^*}f(Y_{\ell,\boldsymbol{j}}^*)\right)^2-\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2\right|\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]\\
&+\sup_{\mathcal{F}_\delta}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2-\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\right]\right|+\sup_{\mathcal{F}_\delta}\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\right]\\
&\leq \mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}^*}f(Y_{\ell,\boldsymbol{j}}^*)\right)^2-\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2\right|\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]\\
&+\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2-\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\right]\right|+\delta^2. \qquad (21)
\end{aligned}
$$

In Section B.3.2, we showed that

$$
\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2-\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\right]\right|\xrightarrow[\underline{C}\to+\infty]{P}0. \qquad (22)
$$

As a result, we only need to control

$$
\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}^*}f(Y_{\ell,\boldsymbol{j}}^*)\right)^2-\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2\right|\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right].
$$

Lemma C.3 applied to $Z_{\boldsymbol{j}}=\left(N_{\boldsymbol{j}}^*,\vec{Y}_{\boldsymbol{j}}^*\right)$, $\mathcal{G}=\pi\circ\mathcal{F}_\infty=\left\{g\left(N_{\boldsymbol{j}}^*,\vec{Y}_{\boldsymbol{j}}^*\right)=\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}^*}f(Y_{\ell,\boldsymbol{j}}^*)\right)^2; f\in\mathcal{F}_\infty\right\}$ and $\Phi=Id$ ensures

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}^*}f(Y_{\ell,\boldsymbol{j}}^*)\right)^2-\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2\right|\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]\\
&\leq 2\sum_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}^*}f\left(Y_{\ell,\boldsymbol{j}}^*\right)\right)^2\right|\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right].
\end{aligned}
$$

Let $\widetilde{F}\left(N_{\boldsymbol{j}}^*,\vec{Y}_{\boldsymbol{j}}^*\right)=\sum_{\ell=1}^{N_{\boldsymbol{j}}^*}F(Y_{\ell,\boldsymbol{j}}^*)$. Note that for every $f\in\mathcal{F}_\infty$ we have $|f|\leq 2F$. Split the expectations in the upper bound in two, depending on whether $4\widetilde{F}^2\leq M$ or not, for some arbitrary M.

For every $\boldsymbol{e}$ such that $\mathbf{0} \prec \boldsymbol{e} \preceq \mathbf{1}$, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N^*_{\boldsymbol{j}}} f\left(Y^*_{\ell,\boldsymbol{j}}\right)\right)^2 \mathbb{1}\left\{4\widetilde{F}\left(N^*_{\boldsymbol{j}},\vec{Y}^*_{\boldsymbol{j}}\right)^2 > M\right\}\right| \left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]\\
&\leq 4\mathbb{E}\left[\widetilde{F}\left(N^*_{\mathbf{1}},\vec{Y}^*_{\mathbf{1}}\right)^2 \mathbb{1}\left\{4\widetilde{F}\left(N^*_{\mathbf{1}},\vec{Y}^*_{\mathbf{1}}\right)^2 > M\right\}\right| \left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]\\
&= \frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\widetilde{F}\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)^2 \mathbb{1}\left\{4\widetilde{F}(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}})^2 > M\right\}.
\end{aligned}$$

And next,

$$\begin{aligned}
&2\sum_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N^*_{\boldsymbol{j}}} f\left(Y^*_{\ell,\boldsymbol{j}}\right)\right)^2 \mathbb{1}\left\{4\widetilde{F}\left(N^*_{\boldsymbol{j}},\vec{Y}^*_{\boldsymbol{j}}\right)^2 > M\right\}\right| \left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]\\
&\leq 8(2^k-1)\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\widetilde{F}\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)^2 \mathbb{1}\left\{4\widetilde{F}(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}})^2 > M\right\}, \qquad (23)
\end{aligned}$$

which vanishes when $\underline{C}\to+\infty$ followed by $M\to+\infty$ thanks to Lemma 7.35 in Kallenberg (2005) and the Dominated Convergence Theorem (under Assumption 3 or 3').
Under Assumptions 1, 2 and 3, Lemma C.9 ensures the existence of $K(F)$ a non negative number depending on the enveloppe F only such that for every $M>0$ and every $\eta>0$,

$$\begin{aligned}
&2\sum_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N^*_{\boldsymbol{j}}} f\left(Y^*_{\ell,\boldsymbol{j}}\right)\right)^2 \mathbb{1}\left\{4\widetilde{F}\left(N^*_{\boldsymbol{j}},\vec{Y}^*_{\boldsymbol{j}}\right)^2 > M\right\}\right| \left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]\\
&\leq K(F)\left(\frac{\sqrt{M}}{\sqrt{\underline{C}}}\left(1+\frac{1}{\eta}\right)+\eta\mathbb{E}\left[N^*_{\mathbf{1}}\sum_{\ell=1}^{N^*_{\mathbf{1}}}F(Y^*_{\ell,\mathbf{1}})^2\right|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]\right)\\
&\leq K(F)\left(\frac{\sqrt{M}}{\sqrt{\underline{C}}}\left(1+\frac{1}{\eta}\right)+\eta\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}N_{\boldsymbol{j}}\sum_{\ell=1}^{N_{\boldsymbol{j}}}F(Y_{\ell,\boldsymbol{j}})^2\right)\\
&\leq K(F)\left(\frac{\sqrt{M}}{\sqrt{\underline{C}}}\left(1+\frac{1}{\eta}\right)+\eta\left(\mathbb{E}\left[N_{\mathbf{1}}\sum_{\ell=1}^{N_{\mathbf{1}}}F(Y_{\ell,\mathbf{1}})^2\right]+o_{a.s}(1)\right)\right) \qquad (24)
\end{aligned}$$

where the last inequality follows from Lemma 7.35 in Kallenberg (2005).
Under Assumptions 1, 2 and 3, Lemmas C.9 and 7.35 in Kallenberg (2005) ensure the existence of $K(F)>0$ such that for every $M>0$ and every $\eta>0$,

$$\begin{aligned}
&2\sum_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\varepsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N^*_{\boldsymbol{j}}} f\left(Y^*_{\ell,\boldsymbol{j}}\right)\right)^2 \mathbb{1}\left\{4\widetilde{F}\left(N^*_{\boldsymbol{j}},\vec{Y}^*_{\boldsymbol{j}}\right)^2 > M\right\}\right| \left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]\\
&\leq K(F)\left(\frac{\sqrt{M}}{\sqrt{\underline{C}}}\left(1+\frac{1}{\eta}\right)+\eta\left(\mathbb{E}\left(\left(\sum_{\ell=1}^{N_{\mathbf{1}}}\langle Y_{\ell,\mathbf{1}}\rangle\right)^2\right)+o_{a.s.}(1)\right)\right). \qquad (25)
\end{aligned}$$

Combine (23) with either (24) or (25), let $\underline{C} \to +\infty$ first and use Kallenberg, and let $M \to +\infty$ and $\delta \to 0$ to get

$$\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\left(\sum_{\ell=1}^{N^*_{\boldsymbol{j}}} f(Y^*_{\ell,\boldsymbol{j}})\right)^2 - \frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right)^2\right| \left| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq \mathbf{1}}\right.\right] \xrightarrow[\underline{C}\to+\infty]{a.s} 0. \tag{26}$$

Combine (21), (22) and (26) to conclude $\mathbb{E}\left[\sigma^{*2}_C\middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq \mathbf{1}}\right] = \delta^2 + o_P(1)$ as $\underline{C} \to +\infty$.

Step 3: Conclusion on asymptotic equicontinuity

Under Assumptions 1, 2 and 3 (respectively 3'), use $\mathbb{E}\left[\sigma^{*2}_C\middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq \mathbf{1}}\right] = \delta^2 + o_P(1)$, $\mathbb{E}\left[A^*_F\middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq \mathbf{1}}\right] = \mathbb{E}\left[N_{\mathbf{1}}\sum_{\ell=1}^{N_{\mathbf{1}}} F(Y_{\ell,\mathbf{1}})^2\right] + o_{a.s}(1)$ (respectively $\mathbb{E}\left[A^*_\beta\middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq \mathbf{1}}\right] = \mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}\langle Y_{\ell,\mathbf{1}}\rangle^{-\beta}\right)^2\right] + o_{a.s}(1)$), continuity of $J_{2,\mathcal{F}}$ (respectively $J_{\infty,\beta,\mathcal{F}}$) at 0, the continuous mapping theorem in probability and Inequality (19) (respectively (20)) to conclude that

$$\lim_{\delta\to 0}\limsup_{\underline{C}\to+\infty}\sqrt{\underline{C}}\,\mathbb{E}\left[\sup_{\mathcal{F}_\delta}|\mathbb{P}^*_C f - \mathbb{P}_C f|\middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq \mathbf{1}}\right] = 0$$

with probability approaching one.

B.5 Proof of Proposition 4.1

Let $\boldsymbol{b}_i = (0, ..., 0, 1, 0, ..., 0)$ with the 1 in position i. Note that $|\mathcal{B}_i| = C_i\prod_{s\neq i} C_s^2$ and:

$$\begin{aligned}\widehat{V}_1 &= \sum_{i=1}^k \frac{\underline{C}}{C_i}\frac{1}{|\mathcal{B}_i|}\left(\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i} S_{\boldsymbol{j}}S_{\boldsymbol{j}'} - 2\widehat{\theta}\left(\prod_{s\neq i} C_s\right)\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}} S_{\boldsymbol{j}} + |\mathcal{B}_i|\widehat{\theta}^2\right)\\ &= \sum_{i=1}^k \frac{\underline{C}}{C_i}\left(\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i} S_{\boldsymbol{j}}S_{\boldsymbol{j}'} - \widehat{\theta}^2\right).\end{aligned}$$

The set $\{S_{\boldsymbol{j}}S_{\boldsymbol{j}'} : (\boldsymbol{j},\boldsymbol{j}') \in \mathcal{B}_i\}$ is equal to $\{S_{\mathbf{c}+\boldsymbol{c}'}S_{\mathbf{c}+\boldsymbol{c}''} : \boldsymbol{b}_i \leq \boldsymbol{c} \leq C_i\boldsymbol{b}_i; \mathbf{1} - \boldsymbol{b}_i \leq \boldsymbol{c}', \boldsymbol{c}'' \leq \boldsymbol{C} \odot (\mathbf{1} - \boldsymbol{b}_i)\}$, so this is a 2k-1 dimensional array indexed by the non null component of $\boldsymbol{c}$, $\boldsymbol{c}'$ and $\boldsymbol{c}''$. This array is jointly exchangeable and dissociated (for a definition of these notions, cf. Kallenberg, 2005). Lemma 7.35 in Kallenberg (2005) ensures that this array is ergodic so $\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i} S_{\boldsymbol{j}}S_{\boldsymbol{j}'}$ converges in L^1 and almost-surely to a constant. Moreover, the first part of the Lemma C.10 applied to $\mathcal{F} = \mathcal{G} = \{Id\}$ ensures that $\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i} S_{\boldsymbol{j}}S_{\boldsymbol{j}'} - \frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_i} S_{\boldsymbol{j}}S_{\boldsymbol{j}'}$ converges in L^1 to 0. Assumption 1 and the representation Lemma B.1 ensure that $\mathbb{E}\left(\frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_i} S_{\boldsymbol{j}}S_{\boldsymbol{j}'}\right) = \mathbb{E}\left(S_{\mathbf{1}}S_{\mathbf{2}_i}\right)$. Then we deduce that $\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i} S_{\boldsymbol{j}}S_{\boldsymbol{j}'} = \mathbb{E}\left(S_{\mathbf{1}}S_{\mathbf{2}_i}\right) + o_p(1)$. The asymptotic normality of $\widehat{\theta}$ ensure that $\widehat{\theta} = \theta_0 + o_p(1)$, then by the continuous mapping Theorem:

$$\widehat{V}_1 = \sum_{i=1}^{k} \lambda_i \left(\mathbb{E}\left(S_{\mathbf{1}} S_{\mathbf{2}_i}\right) - \theta_0^2\right) + o_p(1)$$
$$= \sum_{i=1}^{k} \lambda_i \mathbb{C}ov\left(S_{\mathbf{1}}, S_{\mathbf{2}_i}\right) + o_p(1).$$

The consistency of $\widehat{V}_2 = \sum_{i=1}^k \frac{C}{C_i} \frac{1}{|\mathcal{A}_i|} \sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_i} (S_{\boldsymbol{j}} - \widehat{\theta})(S_{\boldsymbol{j}'} - \widehat{\theta})$ follows from the fact that with probability tending to one, $\widehat{\theta}$ belongs to a compact $\Theta_0 \ni \theta_0$ and the first part of the Lemma C.10 ensuring that

$$\mathbb{E}\left(\sup_{\theta\in\Theta_0}\left|\sum_{i=1}^{k} \frac{C}{C_i}\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i}(S_{\boldsymbol{j}}-\theta)(S_{\boldsymbol{j}'}-\theta) - \sum_{i=1}^{k} \frac{C}{C_i}\frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_i}(S_{\boldsymbol{j}}-\theta)(S_{\boldsymbol{j}'}-\theta)\right|\right) = o(1).$$

Last for the consistency of $\widehat{V}_{cgm}$, note that:

$$\left|\widehat{V}_{cgm} - \widehat{V}_1\right| \leq \frac{\underline{C}}{(\Pi_C)^2} \sum_{\boldsymbol{e}\in\cup_{r=2}^k \mathcal{E}_r} \left|\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_{\boldsymbol{e}}} (S_{\boldsymbol{j}} - \widehat{\theta})(S_{\boldsymbol{j}'} - \widehat{\theta})\right|,$$

the second part of Lemma C.10 ensures that for any $\boldsymbol{e} \in \cup_{r=2}^k \mathcal{E}_r$ and any compact $\Theta_0 \ni \theta_0$ we have:

$$\frac{\underline{C}}{(\Pi_C)^2}\mathbb{E}\left(\sup_{\theta\in\Theta_0}\left|\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_{\boldsymbol{e}}}(S_{\boldsymbol{j}}-\theta)(S_{\boldsymbol{j}'}-\theta)\right|\right) = O(\underline{C}^{-1}).$$

B.6 Proof that moments of quantile IV satisfy Assumption 4.3-4.5

We check Assumption 4.3-4.5 assumnig that the $(Y_{\ell,\mathbf{1}})_{\ell\geq 1}$ are identically distributed and under the following conditions:

Assumption 7

1. *θ_0 belongs to the interior of Θ, a compact subset of $\mathbb{R}^p$.*

2. *The support of X is a compact subset of $\mathbb{R}^p$.*

3. $\mathbb{E}[N_{\mathbf{1}}^2(1 + |Z_{1,\mathbf{1}}|^2)] < +\infty$.

4. *Almost surely, the conditional cdf $F_{W_{1,\mathbf{1}}|X_{1,\mathbf{1}},Z_{1,\mathbf{1}}}(\cdot|X_{1,\mathbf{1}}, Z_{1,\mathbf{1}})$ is continuous everywhere.*

5. *There exists $r > 0$ such that for almost all (x, z), $F_{W_{1,\mathbf{1}}|X_{1,\mathbf{1}},Z_{1,\mathbf{1}}}(\cdot|x, z)$ is differentiable in $\{x'\theta_0 + t, t \in [-r, r]\}$ and*

$$\sup_{(x,z,t)\in\mathcal{S}\times[-r,r]} f_{W_{1,\mathbf{1}}|X_{1,\mathbf{1}},Z_{1,\mathbf{1}}}(x'\theta_0 + t|x, z) < +\infty, \quad (27)$$

where $\mathcal{S}$ denotes the support of $(X_{1,\mathbf{1}}, Z_{1,\mathbf{1}})$.

6. *the rank of* $\mathbb{E}\left[N_{\mathbf{1}}X_{1,\mathbf{1}}Z'_{1,\mathbf{1}}f_{W_{1,\mathbf{1}}|X_{1,\mathbf{1}},Z_{1,\mathbf{1}}}(X'_{\ell,\boldsymbol{j}}\theta_0|X_{1,\mathbf{1}},Z_{1,\mathbf{1}})\right]$ *is equal to* p.

With a slight abuse of notation, we denote by $(a,b$ either the interval $(a,b]$ if $a<b$, or $(b,a]$ if $b<a$. First, we have

$$\begin{aligned}
&\mathbb{E}\left[\left|\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta')-\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right|^2\right]\\
\leq&\mathbb{E}\left[N_{\mathbf{1}}\sum_{\ell=1}^{N_{\mathbf{1}}}|Z_{\ell,\mathbf{1}}|^2\mathbb{1}\{W_{\ell,\boldsymbol{j}}\in(X'_{\ell,\boldsymbol{j}}\theta,X'_{\ell,\boldsymbol{j}}\theta']\}\right]\\
\leq&\mathbb{E}\left[N_{\mathbf{1}}^2|Z_{\ell,\mathbf{1}}|^2\left|F_{W_{1,\mathbf{1}}|X_{1,\mathbf{1}},Z_{1,\mathbf{1}}}(X'_{\ell,\boldsymbol{j}}\theta)-F_{W_{1,\mathbf{1}}|X_{1,\mathbf{1}},Z_{1,\mathbf{1}}}(X'_{\ell,\boldsymbol{j}}\theta')\right|\right],
\end{aligned}$$

where the first inequality follows by the Cauchy-Schwarz inequality and the second uses the fact that the $(Y_{\ell,\mathbf{1}})_{\ell\geq 1}$ are identically distributed and the law of iterated expectation. Then because $\mathbb{E}[N_{\mathbf{1}}^2|Z_{1,\mathbf{1}}|^2]<+\infty$ and $F_{W_{1,\mathbf{1}}|X_{1,\mathbf{1}},Z_{1,\mathbf{1}}}(\cdot|X_{1,\mathbf{1}},Z_{1,\mathbf{1}})$ is continuous everywhere, Assumption 4.3 follows by the dominated convergence theorem.

Turning to 4.4, by the same arguments as to obtain the second inequality above,

$$\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right)=\mathbb{E}\left[N_{\mathbf{1}}Z_{1,\mathbf{1}}\left(\tau-F_{W_{1,\mathbf{1}}|X_{1,\mathbf{1}},Z_{1,\mathbf{1}}}(X'_{\ell,\boldsymbol{j}}\theta|X_{1,\mathbf{1}},Z_{1,\mathbf{1}})\right)\right].$$

By Assumptions 7.2-7.3 and the Cauchy-Schwarz inequality, $\mathbb{E}[N_{\mathbf{1}}|X_{1,\mathbf{1}}Z'_{1,\mathbf{1}}|]<+\infty$. Moreover, still by the Assumptions 7.2 and the Cauchy-Schwarz inequality, there exists a neighborhood $\mathcal{V}$ of θ_0 such that for any $\theta\in\mathcal{V}$, $|x'\theta-x'\theta_0|\leq r$ for all x in the support of $X_{1,\mathbf{1}}$ and where r is defined in Assumption 7.5. Then, by Assumption 7.5, we can apply the dominated convergence theorem to $\theta\mapsto\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right)$ defined on $\mathcal{V}$. This implies that $\theta\mapsto\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right)$ is differentiable at θ_0, with a jacobian matrix J satisfying

$$J=\mathbb{E}\left[N_{\mathbf{1}}X_{1,\mathbf{1}}Z'_{1,\mathbf{1}}f_{W_{1,\mathbf{1}}|X_{1,\mathbf{1}},Z_{1,\mathbf{1}}}(X'_{\ell,\boldsymbol{j}}\theta_0|X_{1,\mathbf{1}},Z_{1,\mathbf{1}})\right].$$

Finally, let us check 4.5. We have to prove that Assumptions 2-3 hold for the class

$$\mathcal{F}_s=\left\{(w,x,z_s)\mapsto z_s\left(\tau-\mathbb{1}\{w-x'\theta\leq 0\}\right),\theta\in\Theta\right\},\ s\in\{1,...,L\}.$$

Reasoning as in the proof of Lemma 8.12 in Kosorok (2006), the class $\{\mathbb{1}\{w-x'\theta\leq 0\},\theta\in\Theta\}$ is pointwise measurable. By Lemma 8.10 in Kosorok (2006), $\mathcal{F}_s$ is pointwise measurable as well.

Finally, we check Assumption 3 for $\mathcal{F}_s$. We have $E[N_{\mathbf{1}}^2]<+\infty$ and by Assumption 7.3, the envelope function $F_s(w,x,z)=2|z_s|$ satisfies $\mathbb{E}\left[N_{\mathbf{1}}\left(\sum_{\ell=1}^{N_{\mathbf{1}}}F(Y_{\ell,\mathbf{1}})\right)^2\right]<+\infty$. Turning to the entropy condition, by Theorem 9.3 in Kosorok (2006), it suffices to prove that $\mathcal{F}_s$ is a

VC class (for a definition of VC classes, see e.g. Kosorok, 2006, Section 9.1.1). The class $\mathcal{G}_s = \{(w,x,z) \mapsto \mathbb{1}\{w - x'\theta \leq 0\}, \theta \in \Theta\}$ is a subset of

$$\left\{(w,x,z) \mapsto \mathbb{1}\{w\eta + x'\theta + z'\gamma \leq \delta\}, (\eta,\theta,\gamma,\delta) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^L \times \mathbb{R}\right\},$$

which is VC by Lemma 9.8 and 9.12 in Kosorok (2006). Hence, $\mathcal{G}_s$ is VC as well. By Lemma 9.9-(v) and (vi), $\mathcal{F}_s$ is also VC. The result follows.

B.7 Proof of Theorem 4.3

The proof is a combination of those of Theorem 1 in Hahn (1996), Theorem 3.3 in Pakes and Pollard (1989) and Theorem 5.21 in van der Vaart (2000).
Because Ξ is a symmetric definite positive matrix, $\widehat{\Xi}$ is also symmetric definite positive with probability tending to 1. We have $\mathbb{1}_{\{det(\widehat{\Xi})>0\}} = 1 + o_p(1)$ and $\mathbb{1}_{\{det(\widehat{\Xi})\leq 0\}} = o_p(1)$.
For any symmetric matrix B, $\rho(B)$ denotes the largest modulus of eigenvalues of B. If B is non negative, $B^{1/2}$ is the symmetric matrix such that $(B^{1/2})^2 = B$. And for any vector b, $|b|$ denotes the Euclidean norm of b.
Hereafter, for any $\theta \in \Theta$, $\mathbb{M}_C(\theta)$ denotes the multidimensional random process from Θ to $\mathbb{R}^L$:

$$\mathbb{G}_C(m(.,\theta)) = \frac{\underline{C}^{1/2}}{\Pi_C} \sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} \left[m(Y_{\ell,\boldsymbol{j}},\theta) - \mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right)\right].$$

Note that Theorem 3.1 implies that $\mathbb{M}_C(\theta) = O_p(1)$ for any θ and even $\sup_{\theta\in\Theta} |\mathbb{M}_C(\theta)| = O_p(1)$.
For any $\theta \in \Theta$, let

$$M(\theta) = \left|\Xi^{1/2}\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right)\right|,$$

let

$$M_C(\theta) = \left|\widehat{\Xi}^{1/2}\left(\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} m(Y_{\ell,\boldsymbol{j}},\theta)\right)\right|,$$

and

$$\overline{M_C}(\theta) = \left|\widehat{\Xi}^{1/2}\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right)\right|.$$

If $det(\widehat{\Xi}) \leq 0$, $M_C(\theta)$ and $\overline{M_C}(\theta)$ could be defined arbitrarily.
Step 1: Consistency

The triangle inequality ensures:

$$
\begin{aligned}
&|M_C(\theta) - M(\theta)| \\
&\leq |M_C(\theta) - M(\theta)| \, \mathbb{1}_{\{det(\widehat{\Xi})>0\}} + |M_C(\theta) - M(\theta)| \, \mathbb{1}_{\{det(\widehat{\Xi})\leq 0\}} \\
&\leq \left(\left|M_C(\theta) - \overline{M_C}(\theta)\right| + \left|\overline{M_C}(\theta) - M(\theta)\right|\right) \mathbb{1}_{\{det(\widehat{\Xi})>0\}} + |M_C(\theta) - M(\theta)| \, \mathbb{1}_{\{det(\widehat{\Xi})\leq 0\}} \\
&\leq \left(\rho(\widehat{\Xi}^{1/2})\underline{C}^{-1/2} \left|\mathbb{M}_C(\theta)\right| + \rho(\widehat{\Xi}^{1/2} - \Xi^{1/2}) \left|\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}}, \theta)\right)\right|\right) \times \mathbb{1}_{\{det(\widehat{\Xi})>0\}} \\
&+ |M_C(\theta) - M(\theta)| \, \mathbb{1}_{\{det(\widehat{\Xi})\leq 0\}}
\end{aligned}
$$

If $\sup_{\theta\in\Theta} |M_C(\theta) - M(\theta)| \, \mathbb{1}_{\{det(\widehat{\Xi})\leq 0\}} > 0$ then $\mathbb{1}_{\{det(\widehat{\Xi})\leq 0\}} = 1$, because $\mathbb{1}_{\{det(\widehat{\Xi})\leq 0\}} = o_p(1)$, we have:

$$\sup_{\theta\in\Theta} |M_C(\theta) - M(\theta)| \, \mathbb{1}_{\{det(\widehat{\Xi})\leq 0\}} = o_p(1).$$

Assumption 4.5 ensures that Theorem 3.1 applies to any m_s for $s = 1, ..., L$ and next:

$$\underline{C}^{-1/2} \sup_{\theta\in\Theta} |\mathbb{M}_C(\theta)| = O_p(\underline{C}^{-1/2}) = o_p(1).$$

It follows that:

$$\sup_{\theta\in\Theta} |M_C(\theta) - M(\theta)| = o_p(1).$$

Because M is continuous (Assumption 4.3) on Θ compact (Assumption 4.1) and attains its minimum only at θ_0 (Assumption 4.2), we have for any $\varepsilon > 0$:

$$\inf_{\theta\in\Theta:|\theta-\theta_0|>\varepsilon} M(\theta) > M(\theta_0).$$

Theorem 5.7 in van der Vaart (2000) applied to $-M_C$ and $-M$ ensures that $\widehat{\theta}$ is well-defined with probability tending to 1 and $\widehat{\theta} = \theta_0 + o_p(1)$.

Step 2: Asymptotic normality

Theorem 3.1 and the Cramer-Wold device ensure that $\mathbb{M}_C(\theta)$ converges weakly to a centered L-multidimensional gaussian process with covariance kernel:

$$\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}}, \theta_1) \sum_{\ell=1}^{N_{\mathbf{1}}} m'(Y_{\ell,\mathbf{1}}, \theta_2)\right) - \mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}}, \theta_1)\right) \mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m'(Y_{\ell,\mathbf{1}}, \theta_2)\right).$$

Let Q the probability measure defined by $dQ(n, y_1, ..., y_n) = dP_{Y_{1,\mathbf{1}},...,Y_{n,\mathbf{1}}|N_{\mathbf{1}}=n}(y_1, ..., y_n)dP_{N_{\mathbf{1}}}(n)$. Consider $\widetilde{\theta} = \theta_0 + o_p(1)$, Assumptions 4.1,4.3 and the continuous mapping theorem ensures that for any $s = 1, ..., L$:

$$\int \left|\sum_{\ell=1}^{n} m_s(y_l, \widetilde{\theta}) - m_s(y_l, \theta_0)\right|^2 dQ(n, y_1, ..., y_n) = o_p(1).$$

Next by Lemma 19.24 in van der Vaart (2000) we have:

$$\mathbb{M}_C(\theta_0) = \mathbb{M}_C(\widetilde{\theta}) + o_p(1).$$

Assumption 4.4 ensures:

$$\int \sum_{\ell=1}^{n} m(y_l, \widetilde{\theta}) dQ(n, y_1, ..., y_n) = J(\widetilde{\theta} - \theta_0) + o_p((\widetilde{\theta} - \theta_0)),$$

and next:

$$\mathbb{M}_C(\theta_0) = \frac{\underline{C}^{1/2}}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} m(Y_{\ell,\boldsymbol{j}}, \widetilde{\theta}) - \underline{C}^{1/2} J(\widetilde{\theta} - \theta_0) + o_p(\underline{C}^{1/2}(\widetilde{\theta} - \theta_0)) + o_p(1) \tag{28}$$

Let $L_C(\theta) = \Xi^{1/2} J(\theta - \theta_0) + \underline{C}^{-1/2} \widehat{\Xi}^{1/2} \mathbb{M}_C(\theta_0)$. The triangle inequality ensures:

$$\begin{aligned} &\left| L_C(\widetilde{\theta}) - \widehat{\Xi}^{1/2} \frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} m(Y_{\ell,\boldsymbol{j}}, \widetilde{\theta}) \right| \\ &\leq \underline{C}^{-1/2} \rho(\widehat{\Xi}^{1/2}) \left| \mathbb{M}_C(\widetilde{\theta}) - \mathbb{M}_C(\theta_0) \right| + \underline{C}^{-1/2} \rho(\widehat{\Xi}^{1/2} - \Xi^{1/2}) \left| \mathbb{M}_C(\theta_0) \right| + \rho(\Xi^{1/2}) o_p(\underline{C}^{-1/2}) \end{aligned}$$

Equation 28 ensures that $\left| \mathbb{M}_C(\widetilde{\theta}) - \mathbb{M}_C(\theta_0) \right| = o_p(1)$. Moreover $\rho(\widehat{\Xi}^{1/2} - \Xi^{1/2}) \left| \mathbb{M}_C(\theta_0) \right| = o_p(1) O_p(1) = o_p(1)$. We deduce:

$$\left| L_C(\widetilde{\theta}) - \widehat{\Xi}^{1/2} \frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} m(Y_{\ell,\boldsymbol{j}}, \widetilde{\theta}) \right| = o_p(\underline{C}^{-1/2}). \tag{29}$$

Now, we consider $\widetilde{\theta} = \widehat{\theta}$ in Equation 28. We have:

$$\begin{aligned} \left| \frac{\underline{C}^{1/2}}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} m(Y_{\ell,\boldsymbol{j}}, \widehat{\theta}) \right| &\leq \rho(\widehat{\Xi}^{-1/2}) \underline{C}^{1/2} M_C(\widehat{\theta}) \\ &\leq \rho(\widehat{\Xi}^{-1/2}) \underline{C}^{1/2} M_C(\theta_0) \\ &\leq \rho(\Xi^{-1/2}) \rho(\Xi^{1/2}) \left| \mathbb{M}_C(\theta_0) \right| = O_p(1), \end{aligned}$$

and because $\mathbb{M}_C(\theta_0) = O_p(1)$ and J full column rank, we deduce $\widehat{\theta} - \theta_0 = O_p(\underline{C}^{-1/2})$. Equation 29 holds for $\widetilde{\theta} = \widehat{\theta}$ and $\widetilde{\theta} = \theta^* = \theta_0 - \underline{C}^{-1/2} (J' \Xi J)^{-1} J' \Xi^{1/2} \widehat{\Xi}^{1/2} \mathbb{M}_C(\theta_0)$ Because $\widehat{\theta} = \arg\min_\theta M_C(\theta)$ and $\theta^* = \arg\min_\theta |L_C(\theta)|$, we have:

$$\left| L_C(\widehat{\theta}) \right| - o_p(\underline{C}^{1/2}) \leq M_C(\widehat{\theta}) \leq M_C(\theta^*) \leq |L_C(\theta^*)| + o_p(\underline{C}^{-1/2}).$$

It follows that:

$$\left| L_C(\widehat{\theta}) \right| = |L_C(\theta^*)| + o_p(\underline{C}^{-1/2}).$$

Because $L_C(\theta^*) = \underline{C}^{-1/2}\left(I - \Xi^{1/2}J(J'\Xi J)^{-1}J'\Xi^{1/2}\right)\widehat{\Xi}^{1/2}\mathbb{M}_C(\theta_0) = O_p(\underline{C}^{-1/2})$, we have:

$$\left|L_C(\widehat{\theta})\right|^2 = |L_C(\theta^*)|^2 + o_p(\underline{C}^{-1}),$$

and $J'\Xi L_C(\theta^*) = 0$, ensuring that:

$$\left|L_C(\widehat{\theta})\right|^2 = |L_C(\theta^*)|^2 + \left|\Xi^{1/2}J(\widehat{\theta} - \theta^*)\right|^2.$$

Combination of the two previous equations ensures that $\left|\Xi^{1/2}J(\widehat{\theta} - \theta^*)\right| = o_p(\underline{C}^{-1/2})$ and because $\Xi^{1/2}$ is non singular and J is full column rank, we have

$$\widehat{\theta} - \theta^* = o_p(\underline{C}^{-1/2}).$$

It follows that:

$$\begin{aligned}
\underline{C}^{1/2}(\widehat{\theta} - \theta_0) &= \underline{C}^{1/2}(\theta^* - \theta_0) + o_p(1) \\
&= -(J'\Xi J)^{-1}J'\Xi^{1/2}\widehat{\Xi}^{1/2}\mathbb{M}_C(\theta_0) + o_p(1) \\
&= -(J'\Xi J)^{-1}J'\Xi\mathbb{M}_C(\theta_0) + o_p(1)
\end{aligned}$$

We already know that $\mathbb{M}_C(\theta_0) \xrightarrow{d} \mathcal{N}(0, H)$, and next the asymptotic normality of $\widehat{\theta}$ follows from the Slutsky lemma and the continuous mapping theorem.

B.8 Proof of Theorem 4.4

B.8.1 Proof of 1.

The Slutsky lemma and the continuous mapping theorem ensure that $\widehat{V} \xrightarrow{\mathbb{P}} V$ if $\widehat{J} \xrightarrow{\mathbb{P}} J$ and $\widehat{H} \xrightarrow{\mathbb{P}} H$.

We first handle $\widehat{J} = \widehat{J}(\widehat{\theta})$ with $\widehat{J}(\theta) = \frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} d(Y_{\ell,\boldsymbol{j}}, \theta)$

$$\left|\widehat{J} - J\right| \leq \left|\widehat{J} - \mathbb{E}[\widehat{J}(\theta)]_{\theta=\widehat{\theta}}\right| + \left|\mathbb{E}[\widehat{J}(\theta)]_{\theta=\widehat{\theta}} - J\right|.$$

Because $\widehat{\theta} = \theta_0 + o_p(1)$ and Assumption 5.3 holds, the continuous mapping theorem ensures that $\left|\mathbb{E}[\widehat{J}(\theta)]_{\theta=\widehat{\theta}} - J\right| = o_p(1)$.

We also have:

$$\begin{aligned}
&\mathbb{E}\left[\left|\widehat{J} - \mathbb{E}[\widehat{J}(\theta)]_{\theta=\widehat{\theta}}\right|\right] \\
&\leq \sqrt{Lp}\,\mathbb{E}\left[\sum_{s=1}^{L}\sum_{r=1}^{p}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} d_{r,s}(Y_{\ell,\boldsymbol{j}}, \widehat{\theta}) - \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} d_{r,s}(Y_{\ell,\mathbf{1}}, \theta)\right]_{\theta=\widehat{\theta}}\right|\right] \\
&\leq (Lp)^{\frac{3}{2}} \max_{1\leq r\leq p, 1\leq s\leq L} \mathbb{E}\left[\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} d_{r,s}(Y_{\ell,\boldsymbol{j}}, \widehat{\theta}) - \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} d_{r,s}(Y_{\ell,\mathbf{1}}, \theta)\right]_{\theta=\widehat{\theta}}\right|\right] \\
&\leq (Lp)^{\frac{3}{2}} \max_{1\leq r\leq p, 1\leq s\leq L} \mathbb{E}\left[\sup_{\Theta}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} d_{r,s}(Y_{\ell,\boldsymbol{j}}, \theta) - \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} d_{r,s}(Y_{\ell,\mathbf{1}}, \theta)\right]\right|\right].
\end{aligned}$$

Assumptions imposed on the classes $\mathcal{G}_{r,s} = \{y \mapsto d_{r,s}(y,\theta) : \theta \in \Theta\}$ are sufficient to apply Lemma C.12 so that:

$$\max_{r,s} \mathbb{E}\left[\sup_{\Theta}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\le \boldsymbol{j}\le \boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} d_{r,s}(Y_{\ell,\boldsymbol{j}},\theta) - \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} d_{r,s}(Y_{\ell,\mathbf{1}},\theta)\right]\right|\right] = o(1).$$

We can turn to $\widehat{H}$. Let

$$\widehat{H}_i(\theta) = \frac{1}{C_i \prod_{s\neq i} C_s^2}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}} m\left(Y_{\ell,\boldsymbol{j}},\theta\right)\sum_{\ell=1}^{N_{\boldsymbol{j}'}} m'\left(Y_{\ell,\boldsymbol{j}'},\theta\right),$$

where $\mathcal{B}_i := \{(\boldsymbol{j},\boldsymbol{j}') : j_i = j_i'\}$,

$$H_i(\theta) = \mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}} m(Y_{\ell,\mathbf{1}},\theta)\right)\left(\sum_{\ell=1}^{N_{\mathbf{2}_i}} m(Y_{\ell,\mathbf{2}_i},\theta)\right)'\right].$$

We have $H = \sum_i \lambda_i H_i(\theta_0)$ and $\widehat{H} = \sum_i \frac{C}{C_i}\widehat{H}_i(\widehat{\theta})$. As $\frac{C}{C_i} = \lambda_i + o_p(1)$ Slutsky's lemma and the continuous mapping theorem ensure that we only need to show that for every $i = 1, ..., k$, $\widehat{H}_i(\widehat{\theta}) = H_i(\theta_0) + o_p(1)$. Moreover Lemma C.11 ensures that $\mathbb{E}(\sup_{\theta\in\Theta}|\widehat{H}_i(\theta) - H_i(\theta)|) = o(1)$, then by Markov inequality we have $\widehat{H}_i(\widehat{\theta}) = H_i(\widehat{\theta}) + o_p(1)$. The continuity Assumption 5.3 and the continuous mapping theorem ensures that $H_i(\widehat{\theta}) = H_i(\theta_0) + o_p(1)$.

B.8.2 Proof of 2.

We remark that in the proof of Theorem 1 in Hahn (1996), the *iid*-ness assumption on the data only kicks in to ensure weak convergence of the bootstrapped empirical process conditional on the data. Since we show weak convergence of $\mathbb{G}_C^*$ without relying on *iid*-ness in Section 3.2, we can directly follow the proof of Theorem 1 in Hahn (1996) to conclude that conditional on $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\mathbf{1}}$ and with probability approaching one,

$$\sqrt{\underline{C}}\left(\widehat{\theta}^* - \widehat{\theta}\right) \xrightarrow{d} \mathcal{N}(0, V_0),$$

where V_0 is defined in the statement of Theorem 4.3.

C Technical lemmas

C.1 Lemma on Assumption 3

Lemma C.1 *Let $\mathcal{F}$ a non finite class of function from $\mathcal{Y}$ to $\mathbb{R}$.*

i) If Assumption 2 holds for $\mathcal{F}$ it also holds for $\mathcal{F}\cup\{1\}$.

ii) If Assumption, 3 holds for $\mathcal{F}$ it also holds for $\mathcal{F}\cup\{1\}$.

Consequently, under Asumptions 2 and 3 if the class $\mathcal{F}$ is not finite we can assume without loss of generality that F is bounded below by 1.

C.1.1 Proof of Lemma C.1

Because the constante function 1 is measurable Assumption 2 holds.
If $\mathcal{F}' = \mathcal{F} \cup \{1\}$ with $F = \sup_{f\in\mathcal{F}} |f|$ and $F' = \max(F, 1)$ then

$$\int_0^{+\infty} \sup_Q \sqrt{\log N\left(\varepsilon ||F'||_{Q,2}, \mathcal{F}', ||.||_{Q,2}\right)} d\varepsilon = \int_0^{2} \sup_Q \sqrt{\log N\left(\varepsilon ||F'||_{Q,2}, \mathcal{F}', ||.||_{Q,2}\right)} d\varepsilon.$$

We can always complete any covering of $\mathcal{F}$ with a ball centered on 1 and next $N\left(\eta, \mathcal{F}', ||.||_{Q,2}\right) \leq 1 + N\left(\eta, \mathcal{F}, ||.||_{Q,2}\right)$ for any $\eta > 0$. It follows that

$$\begin{aligned} N\left(\varepsilon ||F'||_{Q,2}, \mathcal{F}', ||.||_{Q,2}\right) &\leq 1 + N\left(\varepsilon ||F'||_{Q,2}, \mathcal{F}, ||.||_{Q,2}\right) \\ &\leq 2N\left(\varepsilon ||F'||_{Q,2}, \mathcal{F}, ||.||_{Q,2}\right) \text{ because } N\left(\varepsilon ||F'||_{Q,2}, \mathcal{F}, ||.||_{Q,2}\right) \geq 1 \\ &\leq 2N\left(\varepsilon ||F||_{Q,2}, \mathcal{F}, ||.||_{Q,2}\right) \text{ because } ||F||_{Q,2} \leq ||F'||_{Q,2}. \\ &< \infty. \end{aligned}$$

Then we have:

$$\begin{aligned} \int_0^{+\infty} \sup_Q \sqrt{\log N\left(\varepsilon ||F||_{Q,2}, \mathcal{F}, ||.||_{Q,2}\right)} d\varepsilon &= \int_0^{2} \sup_Q \sqrt{\log N\left(\varepsilon ||F'||_{Q,2}, \mathcal{F}', ||.||_{Q,2}\right)} d\varepsilon \\ &\leq \int_0^{2} \sup_Q \sqrt{\log 2N\left(\varepsilon ||F||_{Q,2}, \mathcal{F}, ||.||_{Q,2}\right)} d\varepsilon \\ &\leq 2\sqrt{log(2)} + \int_0^{2} \sup_Q \sqrt{\log N\left(\varepsilon ||F||_{Q,2}, \mathcal{F}, ||.||_{Q,2}\right)} d\varepsilon \\ &\leq 2\sqrt{log(2)} + \int_0^{+\infty} \sup_Q \sqrt{\log N\left(\varepsilon ||F||_{Q,2}, \mathcal{F}, ||.||_{Q,2}\right)} d\varepsilon \\ &< \infty. \end{aligned}$$

And next, if the integral condition holds for $\mathcal{F}$, this is also the case for $\mathcal{F}'$. The moment condition holds for F' if and only if it holds for F and $\mathbb{E}(N_{\mathbf{1}}^2) < \infty$.

C.2 Lemma on Hajek projections

In the following Lemma, we consider the Hajek projection of statistics of random variables sampled according to the representation Lemma B.1. For any $r = 1, ..., k$, the set $\mathcal{E}_r$ is $\{\boldsymbol{e} \in \{0;1\}^k : \sum_{i=1}^k \boldsymbol{e}_i = r\}$ and the set $\mathcal{I}_r(\boldsymbol{C})$ is $\{\boldsymbol{c} = \boldsymbol{j} \odot \boldsymbol{e} : \boldsymbol{e} \in \mathcal{E}_r, \mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}\}$ (with $\odot$ the Hadamard/componentwise product of $\mathbb{R}^k$).

Lemma C.2
Let $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}$ *a family of random variables such that*

$$\left\{N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right\}_{\boldsymbol{j}\geq\mathbf{1}} = \left\{\tau\left(\left(U_{\boldsymbol{j}\odot\boldsymbol{e}}\right)_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\right)\right\}_{\boldsymbol{j}\geq\mathbf{1}},$$

for some measurable function τ and some $(U_{\boldsymbol{c}})_{\boldsymbol{c}\geq\boldsymbol{0}}$ family of mutually independent uniform random variables on $(0,1)$.
Let f such that $\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}})\right)^2\right] < \infty$ and for any $r = 1, ..., k$. Let $\underline{C} \to \infty$ and for any $\boldsymbol{e} \in \mathcal{E}_r$, $\frac{\underline{C}^r}{\prod_{i:\boldsymbol{e}_i=1} C_i} \to \lambda_{\boldsymbol{e}} \geq 0$.
The Hajek projection $H_r f$ of $\mathbb{P}_C f - \mathbb{P}(f)$ on the set of statistics of the form $\sum_{\boldsymbol{c}\in\mathcal{I}_1(\boldsymbol{C})} g_{\boldsymbol{c}}(U_{\boldsymbol{c}})$ with $g_{\boldsymbol{c}}(U_{\boldsymbol{c}})$ square-integrable functions of the random variables $U_{\boldsymbol{c}}$, is such that:

$$\underline{C}^{r/2} H_r f \xrightarrow{d} \mathcal{N}\left(0, \sum_{\boldsymbol{e}\in\mathcal{E}_r} \lambda_{\boldsymbol{e}} \mathbb{V}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}) \middle| U_{\boldsymbol{e}}\right)\right)\right),$$

$$\mathbb{V}\left(\underline{C}^{r/2} H_r f\right) = \sum_{\boldsymbol{e}\in\mathcal{E}_r} \frac{\underline{C}^r}{\prod_{i:\boldsymbol{e}_i=1} C_i} \mathbb{V}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}) \middle| U_{\boldsymbol{e}}\right)\right),$$

and if $\{N_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{l\geq 1}\}_{\boldsymbol{j}\geq\boldsymbol{1}} = \left\{\tau\left((U_{\boldsymbol{j}\odot\boldsymbol{e}})_{\underline{\boldsymbol{e}}\preceq\boldsymbol{e}\preceq 1}\right)\right\}_{\boldsymbol{j}\geq\boldsymbol{1}}$ for $\underline{\boldsymbol{e}} \in \mathcal{E}_{\underline{r}}$, then for any $\boldsymbol{e} \in \mathcal{E}_{\underline{r}}$:

$$\mathbb{V}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}) \middle| U_{\boldsymbol{e}}\right)\right) = \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}), \sum_{\ell=1}^{N_{\boldsymbol{2}-\boldsymbol{e}}} f(Y_{\ell,\boldsymbol{2}-\boldsymbol{e}})\right).$$

C.2.1 Proof of Lemma C.2

The Hájek projection $H_r f$ is characterized by:

$$\mathbb{E}\left((\mathbb{P}_C f - \mathbb{P}f - H_r f) \times \sum_{\boldsymbol{c}\in\mathcal{I}_1(\boldsymbol{C})} g_{\boldsymbol{c}}(U_{\boldsymbol{c}})\right) = 0 \text{ for any } (g_{\boldsymbol{c}})_{\boldsymbol{c}\in\mathcal{I}_r(\boldsymbol{C})} \in \left(L^2([0;1])\right)^{\#\mathcal{I}_r(\boldsymbol{C})}.$$

As a result, we have:

$$\mathbb{E}\left(\mathbb{P}_C f - \mathbb{P}f | U_{\boldsymbol{c}}\right) = \mathbb{E}(H_r f | U_{\boldsymbol{c}}), \text{ for any } \boldsymbol{c} \in \mathcal{I}_r(\boldsymbol{C})$$

Because the range of H_r is a closed subspace of the space of square integrable random variables, we also have:

$$H_r f = \sum_{\boldsymbol{c}\in\mathcal{I}_1(\boldsymbol{C})} \mathbb{E}(Hf|U_{\boldsymbol{c}}),$$

and next:

$$H_r f = \sum_{\boldsymbol{c}\in\mathcal{I}_1(\boldsymbol{C})} \mathbb{E}\left(\mathbb{P}_C f - \mathbb{P}f | U_{\boldsymbol{c}}\right).$$

Note that for any $\boldsymbol{c} \in \mathcal{I}_r(\boldsymbol{C})$, $\boldsymbol{c} \wedge \boldsymbol{1}$ is the unique element $\boldsymbol{e} \in \mathcal{E}_r$ such that $\boldsymbol{c} = \boldsymbol{j} \odot \boldsymbol{e}$ for some $\boldsymbol{j}$ (note that $\boldsymbol{j}$ is not unique). Moreover, for any $\boldsymbol{c} \in \mathcal{I}_1(\boldsymbol{C})$ the independence the U ensures that $\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \perp\!\!\!\perp U_{\boldsymbol{c}}$ if $\boldsymbol{j} \odot \boldsymbol{e} \neq \boldsymbol{c}$, hence:

$$\begin{aligned} \mathbb{E}\left(\mathbb{P}_C f - \mathbb{P}f | U_{\boldsymbol{c}}\right) &= \tfrac{1}{\Pi_C} \textstyle\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}} \mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}f | U_{\boldsymbol{c}}\right) \\ &= \tfrac{1}{\Pi_C} \textstyle\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}} \mathbb{1}_{\boldsymbol{j}\odot\boldsymbol{e}=\boldsymbol{c}} \mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) - \mathbb{P}f | U_{\boldsymbol{c}}\right) \end{aligned}$$

For any $\boldsymbol{j}$ such that $\boldsymbol{j}\odot\boldsymbol{e}=\boldsymbol{c}$, we have $\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})-\mathbb{P}f|U_{\boldsymbol{c}}\right)=\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}})-\mathbb{P}f|U_{\boldsymbol{c}}\right)$.

$$\begin{aligned}
\mathbb{E}\left(\mathbb{P}_C f-\mathbb{P}f|U_{\boldsymbol{c}}\right) &= \tfrac{1}{\Pi_C}\textstyle\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\mathbb{1}_{\boldsymbol{j}\odot\boldsymbol{e}=\boldsymbol{c}}\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}})-\mathbb{P}f|U_{\boldsymbol{c}}\right)\\
&= \tfrac{\prod_{i:\boldsymbol{c}_i=0}C_i}{\Pi_C}\mathbb{E}\left(\textstyle\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}})-\mathbb{P}f|U_{\boldsymbol{c}}\right)\\
&= \tfrac{1}{\prod_{i:\boldsymbol{c}_i\neq0}C_i}\left(\mathbb{E}\left(\textstyle\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}})|U_{\boldsymbol{c}}\right)-\mathbb{P}(f)\right)
\end{aligned}$$

It follows that:

$$H_r f=\sum_{\boldsymbol{c}\in\mathcal{I}_r(\boldsymbol{C})}\frac{1}{\prod_{i:\boldsymbol{c}_i\neq0}C_i}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}})\middle|U_{\boldsymbol{c}}\right)-\mathbb{P}(f)\right).$$

Note that for any $\boldsymbol{c}\in\mathcal{I}_r(\boldsymbol{C})$ and any $i=1,...,k$, we have $\boldsymbol{c}_i\neq0$ if and only if $(\boldsymbol{c}\wedge\boldsymbol{1})_i=1$. Then a rearrangement of the terms in $H_r f$ ensures that:

$$H_r f=\sum_{\boldsymbol{e}\in\mathcal{E}_r}\frac{1}{\prod_{i:\boldsymbol{e}_i=1}C_i}\sum_{\boldsymbol{c}\in\mathcal{I}_r(\boldsymbol{C}):\boldsymbol{c}\wedge\boldsymbol{1}=\boldsymbol{e}}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}})\middle|U_{\boldsymbol{c}}\right)-\mathbb{P}(f)\right).$$

By independence of the U, the random variables $\left\{\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}})|U_{\boldsymbol{c}}\right)-\mathbb{P}(f)\right\}$ are iid across $\boldsymbol{c}\in\mathcal{I}_r(\boldsymbol{C})$ such that $\boldsymbol{c}\wedge\boldsymbol{1}=\boldsymbol{e}$ and centered with common variance

$$\mathbb{V}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}})\middle|U_{\boldsymbol{c}}\right)\right)=\mathbb{V}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}})\middle|U_{\boldsymbol{c}\wedge\boldsymbol{1}}\right)\right).$$

Indeed, there exists h such that:

$$\begin{aligned}
\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}})|U_{\boldsymbol{c}}\right) &= \mathbb{E}\left(h\left((U_{(\boldsymbol{c}\vee\boldsymbol{1})\odot\boldsymbol{e}'})_{\boldsymbol{e}'\in\cup_{r'=1}^{k}\mathcal{E}_{r'}}\right)|U_{\boldsymbol{c}}\right)\\
&\stackrel{d}{=}\mathbb{E}\left(h\left((U_{\boldsymbol{1}\odot\boldsymbol{e}'})_{\boldsymbol{e}'\in\cup_{r'=1}^{k}\mathcal{E}_{r'}}\right)|U_{\boldsymbol{c}\wedge\boldsymbol{1}}\right)\\
&=\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}})|U_{\boldsymbol{c}\wedge\boldsymbol{1}}\right),
\end{aligned}$$

where the equality in distribution holds because the U are iid and because $(\boldsymbol{c}\wedge\boldsymbol{1})\vee\boldsymbol{1}=\boldsymbol{1}$. Because for a given $\boldsymbol{e}\in\mathcal{E}_r$, we have $\#\{\boldsymbol{c}\in\mathcal{I}_r(\boldsymbol{C}):\boldsymbol{c}\wedge\boldsymbol{1}=\boldsymbol{e}\}=\prod_{i:\boldsymbol{e}_i=1}C_i$, the CLT ensures that

$$\frac{1}{\left(\prod_{i:\boldsymbol{e}_i=1}C_i\right)^{1/2}}\sum_{\boldsymbol{c}\in\mathcal{I}_r(\boldsymbol{C}):\boldsymbol{c}\wedge\boldsymbol{1}=\boldsymbol{e}}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}})\middle|U_{\boldsymbol{c}}\right)-\mathbb{P}(f)\right)\stackrel{d}{\longrightarrow}\mathcal{N}\left(0,\mathbb{V}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}})\middle|U_{\boldsymbol{e}}\right)\right)\right),$$

Moreover, since the families $(U_{\boldsymbol{c}})_{\boldsymbol{c}\in\mathcal{I}_r(\boldsymbol{C}):\boldsymbol{c}\wedge\boldsymbol{1}}$ are mutually independent across $\boldsymbol{c}$, we have:

$$\begin{aligned}
\underline{C}^{r/2}H_r f &= \sum_{e\in\mathcal{E}_r}\left(\frac{\underline{C}^r}{\prod_{i:\boldsymbol{e}_i=1}C_i}\right)^{1/2}\frac{1}{\left(\prod_{i:\boldsymbol{e}_i=1}C_i\right)^{1/2}}\sum_{\boldsymbol{c}\in\mathcal{I}_r(\boldsymbol{C}):\boldsymbol{c}\wedge\boldsymbol{1}=\boldsymbol{e}}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{c}\vee\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{c}\vee\boldsymbol{1}})\middle|U_{\boldsymbol{c}}\right)-\mathbb{P}(f)\right)\\
&\stackrel{d}{\longrightarrow}\mathcal{N}\left(0,\sum_{\boldsymbol{e}\in\mathcal{E}_r}\lambda_{\boldsymbol{e}}\mathbb{V}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}})\middle|U_{\boldsymbol{e}}\right)\right)\right).
\end{aligned}$$

And we also have:

$$\mathbb{V}\left(\underline{C}^{r/2}H_r f\right)=\sum_{\boldsymbol{e}\in\mathcal{E}_r}\frac{\underline{C}^r}{\prod_{i:\boldsymbol{e}_i=1}C_i}\mathbb{V}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\middle|U_{\boldsymbol{e}}\right)\right).$$

Last, to achieve the proof we also have to show that for $\boldsymbol{e}\in\mathcal{E}_{\underline{r}}$

$$\mathbb{V}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\middle|U_{\boldsymbol{e}}\right)\right)=\mathbb{C}ov\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}}),\sum_{\ell=1}^{N_{\mathbf{2}-\boldsymbol{e}}}f(Y_{\ell,\mathbf{2}-\boldsymbol{e}})\right).$$

Because $\{N_{\boldsymbol{j}},(Y_{\ell,\boldsymbol{j}})_{l\geq 1}\}_{\boldsymbol{j}\geq\mathbf{1}}=\left\{\tau\left((U_{\boldsymbol{j}\odot\boldsymbol{e}})_{\boldsymbol{e}\in\cup_{r=\underline{r}}^k\mathcal{E}_r}\right)\right\}_{\boldsymbol{j}\geq\mathbf{1}}$ with iid U, we have $\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\middle|U_{\boldsymbol{e}}\right)=\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\middle|U_{\boldsymbol{e}}\right)$ for any $\boldsymbol{j}$ such that $\boldsymbol{j}\odot\boldsymbol{e}=\mathbf{1}\odot\boldsymbol{e}=\boldsymbol{e}$. Because $(\mathbf{2}-\boldsymbol{e})\odot\boldsymbol{e}=\boldsymbol{e}$, we have: $\mathbb{V}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\middle|U_{\boldsymbol{e}}\right)\right)=\mathbb{C}ov\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\middle|U_{\boldsymbol{e}}\right),\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{2}-\boldsymbol{e}}}f(Y_{\ell,\mathbf{2}-\boldsymbol{e}})\middle|U_{\boldsymbol{e}}\right)\right)$.
For any $\boldsymbol{e}\in\mathcal{E}_{\underline{r}}$, we have $\mathbf{2}-\boldsymbol{e}\neq\mathbf{1}$, so the independence of the U ensures that $(U_{\mathbf{1}\odot\boldsymbol{e}'})_{\boldsymbol{e}'\in\cup_{r=\underline{r}}^k\mathcal{E}_r\backslash\boldsymbol{e}}\perp\perp(U_{(\mathbf{2}-\boldsymbol{e})\odot\boldsymbol{e}'})_{\boldsymbol{e}'\in\cup_{r=\underline{r}}^k\mathcal{E}_r\backslash\boldsymbol{e}}|U_{\boldsymbol{e}}$ and next $\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\perp\perp\sum_{\ell=1}^{N_{\mathbf{2}-\boldsymbol{e}}}f(Y_{\ell,\mathbf{2}-\boldsymbol{e}})|U_{\boldsymbol{e}}$.
Hence, for $\boldsymbol{e}\in\mathcal{E}_{\underline{r}}$:

$$\mathbb{E}\left(\mathbb{C}ov\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f_1(Y_{\ell,\mathbf{1}}),\sum_{\ell=1}^{N_{\mathbf{2}-\boldsymbol{e}}}f_2(Y_{\ell,\mathbf{2}-\boldsymbol{e}})\middle|U_{\boldsymbol{e}}\right)\right)=0.$$

And by the law of total covariance, we have:

$$\mathbb{V}\left(\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\middle|U_{\boldsymbol{e}}\right)\right)=\mathbb{C}ov\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}}),\sum_{\ell=1}^{N_{\mathbf{2}-\boldsymbol{e}}}f(Y_{\ell,\mathbf{2}-\boldsymbol{e}})\right)$$

C.3 Symmetrization

We prove a symmetrization Lemma

Lemma C.3
Let $(Z_{\boldsymbol{j}})_{\boldsymbol{j}\geq\mathbf{1}}$ a family of random variables with values in Polish space in by $\boldsymbol{j}\in(\mathbb{N}^*)^k$, such that

$$(Z_{\boldsymbol{j}})_{\boldsymbol{j}\geq\mathbf{1}}\overset{a.s.}{=}\left(\tau\left((U_{\boldsymbol{j}\odot\boldsymbol{e}})_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\right)\right)_{\boldsymbol{j}\geq\mathbf{1}}$$

for $(U_{\mathbf{c}})_{\mathbf{c}>\mathbf{0}}$ independent family of uniform-$(0,1)$ and some measurable function τ. Let $\mathcal{G}$ a pointwise measurable class of integrable functions of $Z_{\mathbf{1}}$, and Φ a non decreasing convex function from $\mathbb{R}^+$ to $\mathbb{R}$, we have:

$$\mathbb{E}\left[\Phi\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}g\left(Z_{\boldsymbol{j}}\right)-\mathbb{E}\left[g\left(Z_{\mathbf{1}}\right)\right]\right|\right)\right]$$
$$\leq\frac{1}{2^k-1}\sum_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\mathbb{E}\left[\Phi\left(2(2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}g\left(Z_{\boldsymbol{j}}\right)\right|\right)\right],$$

with $(\epsilon_{\mathbf{c}})_{\mathbf{c}\geq\mathbf{1}}$ a familly of mutually independent Rademacher variables, independent of $(Z_{\boldsymbol{j}})_{\boldsymbol{j}\geq\mathbf{1}}$.

C.3.1 Proof of Lemma C.3

Let $(U_{\boldsymbol{c}}^{(1)})_{\boldsymbol{c}>\boldsymbol{0}}$ independent family of uniform-$(0,1)$, independent of $(U_{\boldsymbol{c}})_{\boldsymbol{c}>\boldsymbol{0}}$, and let $\left(Z_{\boldsymbol{j}}^{(1)}\right)_{\boldsymbol{j}\geq\boldsymbol{1}} = \left(\tau\left(\left(U_{\boldsymbol{j}\odot\boldsymbol{e}'}^{(1)}\right)_{\boldsymbol{0}\prec\boldsymbol{e}'\preceq\boldsymbol{1}}\right)\right)_{\boldsymbol{j}\geq\boldsymbol{1}}$.

We have $\mathbb{E}\left[g\left(Z_{\boldsymbol{1}}\right)\right] = \frac{1}{\Pi_C}\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\mathbb{E}\left[g\left(Z_{\boldsymbol{j}}^{(1)}\right)\right]$, then the Jensen inequality applied for $|.|$, $\sup_{g\in\mathcal{G}}$ and Φ ensures:

$$\mathbb{E}\left[\Phi\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}g\left(Z_{\boldsymbol{j}}\right)-\mathbb{E}\left[g\left(Z_{\boldsymbol{1}}\right)\right]\right|\right)\right]\leq\mathbb{E}\left[\Phi\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}g\left(Z_{\boldsymbol{j}}\right)-g\left(Z_{\boldsymbol{j}}^{(1)}\right)\right|\right)\right].$$

For $\boldsymbol{e}$ such that $\boldsymbol{0}\prec\boldsymbol{e}\preceq\boldsymbol{1}$, let $(Z_{\boldsymbol{j}}(\boldsymbol{e}))_{\boldsymbol{j}\geq\boldsymbol{1}} = \left(\tau\left(\left(U_{\boldsymbol{j}\odot\boldsymbol{e}'}\right)_{0\prec\boldsymbol{e}'\preceq\boldsymbol{e}},\left(U_{\boldsymbol{j}\odot\boldsymbol{e}'}^{(1)}\right)_{\boldsymbol{e}\prec\boldsymbol{e}'\preceq\boldsymbol{1}}\right)\right)_{\boldsymbol{j}\geq\boldsymbol{1}}$ and $\left(Z_{\boldsymbol{j}}^{(1)}(\boldsymbol{e})\right)_{\boldsymbol{j}\geq\boldsymbol{1}} = \left(\tau\left(\left(U_{\boldsymbol{j}\odot\boldsymbol{e}'}\right)_{0\prec\boldsymbol{e}'\prec\boldsymbol{e}},\left(U_{\boldsymbol{j}\odot\boldsymbol{e}'}^{(1)}\right)_{\boldsymbol{e}\preceq\boldsymbol{e}'\preceq\boldsymbol{1}}\right)\right)_{\boldsymbol{j}\geq\boldsymbol{1}}$.

We have $(Z_{\boldsymbol{j}}(\boldsymbol{1}))_{\boldsymbol{j}\geq\boldsymbol{1}} = (Z_{\boldsymbol{j}})_{\boldsymbol{j}\geq\boldsymbol{1}}$. Moreover, if $s(\boldsymbol{e})$ is the successor of $\boldsymbol{e}$ for the total order $\prec$, we have $\left(Z_{\boldsymbol{j}}^{(1)}(s(\boldsymbol{0}))\right)_{\boldsymbol{j}\geq\boldsymbol{1}} = \left(Z_{\boldsymbol{j}}^{(1)}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}$ and for $\boldsymbol{0}\prec\boldsymbol{e}\prec\boldsymbol{1}$ we have $\left(Z_{\boldsymbol{j}}^{(1)}(s(\boldsymbol{e}))\right)_{\boldsymbol{j}\geq\boldsymbol{1}} = (Z_{\boldsymbol{j}}(\boldsymbol{e}))_{\boldsymbol{j}\geq\boldsymbol{1}}$. It follows that:

$$\begin{aligned}
&\mathbb{E}\left[\Phi\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(g(Z_{\boldsymbol{j}})-g(Z_{\boldsymbol{j}}^{(1)})\right)\right|\right)\right]\\
&=\mathbb{E}\left[\Phi\left(\sup_{g\in\mathcal{G}}\left|\sum_{\boldsymbol{0}\prec\boldsymbol{e}\preceq\boldsymbol{1}}\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(g(Z_{\boldsymbol{j}}(\boldsymbol{e}))-g(Z_{\boldsymbol{j}}^{(1)}(\boldsymbol{e}))\right)\right|\right)\right]\\
&\leq\mathbb{E}\left[\Phi\left(\sum_{\boldsymbol{0}\prec\boldsymbol{e}\preceq\boldsymbol{1}}\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(g(Z_{\boldsymbol{j}}(\boldsymbol{e}))-g(Z_{\boldsymbol{j}}^{(1)}(\boldsymbol{e}))\right)\right|\right)\right]\\
&\leq\frac{1}{2^k-1}\sum_{\boldsymbol{0}\prec\boldsymbol{e}\preceq\boldsymbol{1}}\mathbb{E}\left[\Phi\left((2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(g(Z_{\boldsymbol{j}}(\boldsymbol{e}))-g(Z_{\boldsymbol{j}}^{(1)}(\boldsymbol{e}))\right)\right|\right)\right].
\end{aligned}$$

For any $\boldsymbol{e}$ such that $\boldsymbol{0}\prec\boldsymbol{e}\preceq\boldsymbol{1}$, we have $\boldsymbol{j}=\boldsymbol{j}\odot\boldsymbol{e}+\boldsymbol{j}\odot(\boldsymbol{1}-\boldsymbol{e})$, and observe that for any function q on $(\mathbb{N}^*)^k$ we have:

$$\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}q(\boldsymbol{j})=\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\;\sum_{\boldsymbol{1}-\boldsymbol{e}\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})}q(\boldsymbol{c}+\boldsymbol{c}').$$

Then for any $\boldsymbol{e}$ such that $\boldsymbol{0}\prec\boldsymbol{e}\preceq\boldsymbol{1}$:

$$\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(g(Z_{\boldsymbol{j}}(\boldsymbol{e}))-g(Z_{\boldsymbol{j}}^{(1)}(\boldsymbol{e}))\right)=\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\;\sum_{(\boldsymbol{1}-\boldsymbol{e})\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})}\left(g(Z_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e}))-g(Z_{\boldsymbol{c}+\boldsymbol{c}'}^{(1)}(\boldsymbol{e}))\right).$$

Because $(\boldsymbol{c}+\boldsymbol{c}')\odot\boldsymbol{e}=\boldsymbol{c}\odot\boldsymbol{e}=\boldsymbol{c}$ for any $(\boldsymbol{c},\boldsymbol{c}')$ such that $\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}$ and $(\boldsymbol{1}-\boldsymbol{e})\leq\boldsymbol{c}'\leq$

$\boldsymbol{C} \odot (\mathbf{1}-\boldsymbol{e})$, we have:

$$g(Z_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})) = g \circ \tau\left(\left(U_{(\boldsymbol{c}+\boldsymbol{c}')\odot\boldsymbol{e}'}\right)_{\boldsymbol{e}'\prec\boldsymbol{e}}, U_{\boldsymbol{c}}, \left(U^{(1)}_{(\boldsymbol{c}+\boldsymbol{c}')\odot\boldsymbol{e}'}\right)_{\boldsymbol{e}\prec\boldsymbol{e}'}\right),$$
$$g(Z^{(1)}_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})) = g \circ \tau\left(\left(U_{(\boldsymbol{c}+\boldsymbol{c}')\odot\boldsymbol{e}'}\right)_{\boldsymbol{e}'\prec\boldsymbol{e}}, U^{(1)}_{\boldsymbol{c}}, \left(U^{(1)}_{(\boldsymbol{c}+\boldsymbol{c}')\odot\boldsymbol{e}'}\right)_{\boldsymbol{e}\prec\boldsymbol{e}'}\right).$$

Let $R_{\boldsymbol{e}} = \left(\left(U_{\boldsymbol{j}\odot\boldsymbol{e}'}\right)_{\boldsymbol{e}'\prec\boldsymbol{e}}, \left(U^{(1)}_{\boldsymbol{j}\odot\boldsymbol{e}'}\right)_{\boldsymbol{e}\prec\boldsymbol{e}'}\right)_{\boldsymbol{j}\geq\mathbf{1}}$. For any $\boldsymbol{c}$ such that $\boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{C} \odot \boldsymbol{e}$, we have:

$$\mathbb{E}\left(\sum_{(\mathbf{1}-\boldsymbol{e})\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{e})} g(Z_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})) - g(Z^{(1)}_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e}))\middle| R_{\boldsymbol{e}}\right) = 0,$$

and for any $\boldsymbol{c}^1, \boldsymbol{c}^2$ such that $\boldsymbol{e} \leq \boldsymbol{c}^1, \boldsymbol{c}^2 \leq \boldsymbol{C} \odot \boldsymbol{e}$, we have:

$$\left(\sum_{(\mathbf{1}-\boldsymbol{e})\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{e})} g(Z_{\boldsymbol{c}^1+\boldsymbol{c}'}(\boldsymbol{e})) - g(Z^{(1)}_{\boldsymbol{c}^1+\boldsymbol{c}'}(\boldsymbol{e}))\right) \perp\!\!\!\perp \left(\sum_{(\mathbf{1}-\boldsymbol{e})\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{e})} g(Z_{\boldsymbol{c}^2+\boldsymbol{c}'}(\boldsymbol{e})) - g(Z^{(1)}_{\boldsymbol{c}^2+\boldsymbol{c}'}(\boldsymbol{e}))\right)\middle| R_{\boldsymbol{e}}.$$

For any $\boldsymbol{e}$, $\#\{\boldsymbol{c} : \boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{C} \odot \boldsymbol{e}\} = \prod_{i:e_i=1} C_i$. By symmetry, for any $\epsilon_{\boldsymbol{c}} \in \{-1;1\}^{\prod_{i:e_i=1} C_i}$

$$\left(\sum_{(\mathbf{1}-\boldsymbol{e})\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{e})} g(Z_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})) - g(Z^{(1)}_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e}))\right)_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}} \Bigg| R_{\boldsymbol{e}}$$
$$\stackrel{d}{=} \left(\epsilon_{\boldsymbol{c}} \sum_{(\mathbf{1}-\boldsymbol{e})\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{e})} g(Z_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})) - g(Z^{(1)}_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e}))\right)_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}} \Bigg| R_{\boldsymbol{e}}$$

Then considering $(\epsilon_{\boldsymbol{j}})_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}$ some independent Rademacher random variables, independent of the U and the $U^{(1)}$, we have:

$$\mathbb{E}\left[\Phi\left((2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(g(Z_{\boldsymbol{j}}(\boldsymbol{e})) - g(Z^{(1)}_{\boldsymbol{j}}(\boldsymbol{e}))\right)\right|\right)\right]$$
$$= \mathbb{E}\left[\Phi\left((2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\ \sum_{\mathbf{1}-\boldsymbol{e}\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{e})}\left(g(Z_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})) - g(Z^{(1)}_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e}))\right)\right|\right)\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[\Phi\left((2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\ \sum_{\mathbf{1}-\boldsymbol{e}\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{e})}\left(g(Z_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})) - g(Z^{(1)}_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e}))\right)\right|\right)\middle|(\epsilon_{\boldsymbol{j}})_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}, R_{\boldsymbol{e}}\right]\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[\Phi\left((2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\epsilon_{\boldsymbol{c}}\sum_{\mathbf{1}-\boldsymbol{e}\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{e})}\left(g(Z_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})) - g(Z^{(1)}_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e}))\right)\right|\right)\middle|(\epsilon_{\boldsymbol{j}})_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}, R_{\boldsymbol{e}}\right]\right]$$
$$= \mathbb{E}\left[\Phi\left((2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(g(Z_{\boldsymbol{j}}(\boldsymbol{e})) - g(Z^{(1)}_{\boldsymbol{j}}(\boldsymbol{e}))\right)\right|\right)\right]$$

The triangle inequality and the convexity of Φ ensures:

$$\Phi\left((2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(g(Z_{\boldsymbol{j}}(\boldsymbol{e}))-g(Z_{\boldsymbol{j}}^{(1)}(\boldsymbol{e}))\right)\right|\right)$$

$$\leq\Phi\left((2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(g(Z_{\boldsymbol{j}}(\boldsymbol{e}))\right)\right|+(2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(g(Z_{\boldsymbol{j}}^{(1)}(\boldsymbol{e}))\right)\right|\right)$$

$$\leq\frac{1}{2}\Phi\left(2(2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(g(Z_{\boldsymbol{j}}(\boldsymbol{e}))\right)\right|\right)+\frac{1}{2}\Phi\left(2(2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(g(Z_{\boldsymbol{j}}^{(1)}(\boldsymbol{e}))\right)\right|\right)$$

Because $\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(g(Z_{\boldsymbol{j}}(\boldsymbol{e}))\right)\stackrel{d}{=}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(g(Z_{\boldsymbol{j}}^{(1)}(\boldsymbol{e}))\right)\stackrel{d}{=}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(g(Z_{\boldsymbol{j}})\right)$, we get:

$$\mathbb{E}\left[\Phi\left((2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(g(Z_{\boldsymbol{j}}(\boldsymbol{e}))-g(Z_{\boldsymbol{j}}^{(1)}(\boldsymbol{e}))\right)\right|\right)\right]$$

$$\leq\mathbb{E}\left[\Phi\left(2(2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}g(Z_{\boldsymbol{j}}(\boldsymbol{e}))\right|\right)\right]$$

$$=\mathbb{E}\left[\Phi\left(2(2^k-1)\sup_{g\in\mathcal{G}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}g(Z_{\boldsymbol{j}})\right|\right)\right]$$

C.4 Lemmas for uniform CLT and inference based on asymptotic normality

We state the lemmas we use in our proof, before showing them in turn.

Lemma C.4 *Let $\{a_j\}_{j=1}^m$ be a sequence of n-dimensional Euclidean vectors and $\{\nu_i\}_{i=1}^n$ an i.i.d sequence of Rademacher random variables. For every $m\geq 1$*

$$\mathbb{E}\left[\max_{j\in\{1,\ldots,m\}}\left|\sum_{i=1}^n\nu_i a_{ji}\right|\right]\leq\sqrt{2\log 2m}\max_{j\in\{1,\ldots,m\}}|a_j|.$$

Lemma C.5 *Let $\varepsilon>0$,*

(i) if $||\cdot||$ is a pseudo-norm on $\mathcal{G}$ and $\lambda>0$ then :

$$N(\varepsilon,\mathcal{G},\lambda||.||)=N(\varepsilon/\lambda,\mathcal{G},||.||).$$

(ii) if $||.||_a$ and $||.||_b$ are two pseudo-norms on a class $\mathcal{G}$ of functions such that $||g||_a\leq||g||_b$ for any $g\in\mathcal{G}$, then:

$$N(\varepsilon,\mathcal{G},||.||_a)\leq N(\varepsilon,\mathcal{G},||.||_b).$$

(iii) if $||\cdot||$ is a pseudo-norm on $\mathcal{G}'\supset\mathcal{G}$ then :

$$N(\varepsilon,\mathcal{G},||.||)\leq N(\varepsilon/2,\mathcal{G}',||.||).$$

(iv) if $||\cdot||$ is a pseudo-norm on $\mathcal{G}$ and $\mathcal{G}_\infty$ then :

$$N(\varepsilon, \mathcal{G}_\infty, ||.||) \leq N^2(\varepsilon/2, \mathcal{G}, ||.||).$$

Lemma C.6 *For any $\varepsilon > 0$, $\delta \in]0, +\infty]$ and $r \geq 1$*

(i) if $\overline{N}_r > 0$ then $N\left(\varepsilon, \pi \circ \mathcal{F}, ||\cdot||_{\mu_C, r}\right) \leq N\left(\frac{\varepsilon}{\overline{N}_r^{\frac{1}{r}}}, \mathcal{F}, ||\cdot||_{\mathbb{Q}_C^r, r}\right)$

(ii) if $\overline{N}_r > 0$ then $N\left(\varepsilon, \pi \circ \mathcal{F}_\delta, ||\cdot||_{\mu_C, r}\right) \leq N^2\left(\frac{\varepsilon}{4\overline{N}_r^{\frac{1}{r}}}, \mathcal{F}, ||\cdot||_{\mathbb{Q}_C^r, r}\right)$

(iii) if $A_r > 0$ then $N\left(\varepsilon, \pi \circ \mathcal{F}, ||\cdot||_{\mu_C, r}\right) \leq N\left(\frac{\varepsilon}{A_r^{\frac{1}{r}}}, \mathcal{F}, ||\cdot||_{\infty, \beta}\right)$.

(iv) if $A_r > 0$ then $N\left(\varepsilon, \pi \circ \mathcal{F}_\delta, ||\cdot||_{\mu_C, r}\right) \leq N^2\left(\frac{\varepsilon}{4A_r^{\frac{1}{r}}}, \mathcal{F}, ||\cdot||_{\infty, \beta}\right)$.

Lemma C.7 *For every $\varepsilon > 0$*

$$N\left(2\varepsilon||\pi^2 \circ F||_{\mu_C, 1}, \pi^2 \circ \mathcal{F}_\infty, ||\cdot||_{\mu_C, 1}\right) \leq N^2\left(\varepsilon||\pi \circ F||_{\mu_C, 2}, \pi \circ \mathcal{F}, ||\cdot||_{\mu_C, 2}\right).$$

Lemma C.8 *Let $\boldsymbol{e} \in \cup_{i=1}^k \mathcal{E}_i$. Under Assumptions 1, 2 and 3*

$$\mathbb{E}\left[\sup_{\mathcal{F}_\delta}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}} \epsilon_{\boldsymbol{j}\odot\boldsymbol{e}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right|\right] \leq 4\sqrt{\frac{2\mathbb{E}\left[\sigma_C^2\right]\log 2}{\prod_{s:\boldsymbol{e}_s=1} C_s}} + 32\sqrt{\frac{\mathbb{E}\left[A_F\right]}{\prod_{s:\boldsymbol{e}_s=1} C_s}} J_{2,\mathcal{F}}\left(\frac{1}{4}\sqrt{\frac{\mathbb{E}\left[\sigma_C^2\right]}{\mathbb{E}\left[A_F\right]}}\right),$$

where $A_F := N_{\mathbf{1}} \sum_{\ell=1}^{N_{\mathbf{1}}} F\left(Y_{\ell,\mathbf{1}}\right)^2$.
Under Assumptions 1, 2 and 3'

$$\mathbb{E}\left[\sup_{\mathcal{F}_\delta}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}} \epsilon_{\boldsymbol{j}\odot\boldsymbol{e}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right|\right]$$
$$\leq 4\sqrt{\frac{2\mathbb{E}\left[\sigma_C^2\right]\log 2}{\prod_{s:\boldsymbol{e}_s=1} C_s}} + 32||F||_{\infty,\beta} \times \sqrt{\frac{\mathbb{E}\left[A_\beta^2(\mathbf{1})\right]}{\prod_{s:\boldsymbol{e}_s=1} C_s}} J_{\infty,\beta,\mathcal{F}}\left(\frac{1}{4\,||F||_{\infty,\beta}}\sqrt{\frac{\mathbb{E}\left[\sigma_C^2\right]}{\mathbb{E}\left[A_\beta^2(\mathbf{1})\right]}}\right),$$

where $A_\beta(\mathbf{1}) := \sum_{\ell=1}^{N_{\mathbf{1}}} \langle Y_{1,\mathbf{1}}\rangle^{-\beta}$.

Lemma C.9 *Let $M > 0$, $\eta > 0$ and $\boldsymbol{e} \in \cup_{i=1}^k \mathcal{E}_i$ and $(\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}})_{\boldsymbol{j}\geq\mathbf{1}}$ a family of independent Rademacher variables independent from $(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}})_{\boldsymbol{j}\geq\mathbf{1}}$. Under Assumptions 1, 2 and 3*

$$\mathbb{E}\left(\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}} \epsilon_{\boldsymbol{j}\odot\boldsymbol{e}} \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right)^2 \mathbb{1}\left\{\sum_{\ell=1}^{N_{\boldsymbol{j}}} F(Y_{\ell,\boldsymbol{j}}) \leq M\right\}\right|\right)$$
$$\leq \frac{M}{\sqrt{\prod_{s:\boldsymbol{e}_s=1} C_s}} \times \left\{2\sqrt{2\log 2} + \frac{4}{\eta} J_{2,F}(\infty)\right\} + 2\eta\mathbb{E}\left(A_F\right)$$

while under Assumptions 1, 2 and 3'

$$\mathbb{E}\left(\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2\mathbb{1}\left\{\sum_{\ell=1}^{N_{\boldsymbol{j}}}F(Y_{\ell,\boldsymbol{j}})\leq M\right\}\right|\right)$$
$$\leq\frac{M}{\sqrt{\prod_{s:\boldsymbol{e}_s=1}C_s}}\times\left\{2\sqrt{2\log 2}+\frac{4}{\eta}J_{\infty,\beta,\mathcal{F}}(\infty)\right\}+4\eta\,||F||_{\infty,\beta}^2\,\mathbb{E}\left[A_\beta^2(\boldsymbol{1})\right].$$

Lemma C.10 *Let $\mathcal{F}$ and $\mathcal{G}$ be two classes of functions pointwise measurable with respective enveloppe F and G.*
Under Assumptions 1, if $\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}}F(Y_{\ell,\boldsymbol{1}})\right)^2\right]\vee\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}}G(Y_{\ell,\boldsymbol{1}})\right)^2\right]<\infty$ then for any $i\in\{1,...,k\}$:

$$\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j'})\in\mathcal{B}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j'}}}g(Y_{\ell,\boldsymbol{j'}})-\frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j'})\in\mathcal{A}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j'}}}g(Y_{\ell,\boldsymbol{j'}})\right|\right]=o(1)$$

and for any $\boldsymbol{e}\in\mathcal{E}_r$ with $r\geq 1$:

$$(\Pi_C)^{-2}\,\mathbb{E}\left(\sup_{\mathcal{F}\times\mathcal{G}}\left|\sum_{(\boldsymbol{j},\boldsymbol{j'})\in\mathcal{B}_{\boldsymbol{e}}}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell'=1}^{N_{\boldsymbol{j'}}}g(Y_{\ell',\boldsymbol{j'}})\right|\right)=O(\underline{C}^{-r})$$

Lemma C.11 *Let $\mathcal{F}$ and $\mathcal{G}$ be two classes of functions that satisfy Assumptions 1, 2 and either Assumptions 3 or 3'. Then for every $i\in\{1,...,k\}$*

$$\lim_{\underline{C}\to+\infty}\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j'})\in\mathcal{B}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j'}}}g(Y_{\ell,\boldsymbol{j'}})-\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{1}}}f(Y_{\ell,\boldsymbol{1}})\sum_{\ell=1}^{N_{\boldsymbol{2}_i}}g(Y_{\ell,\boldsymbol{2}_i})\right]\right|\right]=0,$$

and

$$\lim_{\underline{C}\to+\infty}\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j'})\in\mathcal{A}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j'}}}g(Y_{\ell,\boldsymbol{j'}})-\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{1}}}f(Y_{\ell,\boldsymbol{1}})\sum_{\ell=1}^{N_{\boldsymbol{2}_i}}g(Y_{\ell,\boldsymbol{2}_i})\right]\right|\right]=0.$$

Lemma C.12 *Let $\mathcal{F}$ a class of functions with envelope F such that $\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{1}}}F(Y_{\ell,\boldsymbol{1}})\right]<\infty$, $\sup_Q\log N(\eta||.||_{Q,1},\mathcal{F},||.||_{Q,1})<\infty$ for any $\eta>0$ and such that $\mathcal{F}$ fulfills Assumption 2. Then under Assumption 1, we have:*

$$\mathbb{E}\left[\sup_{\mathcal{F}}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})-\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{1}}}f(Y_{\ell,\boldsymbol{1}})\right]\right|\right]=o(1).$$

C.4.1 Proof of Lemma C.4

A square integrable random variable u is subgaussian with parameter $\sigma > 0$ if for any $\lambda \in \mathbb{R}$, $\mathbb{E}\left(e^{\lambda u}\right) \leq e^{\lambda^2\sigma^2/2}$. A Rademacher variable is subgaussian of parameter 1. By independence of the ν_i, it follows that $\sum_{i=1}^n \nu_i a_{ji}$ is subgaussian of parameter $|a_j|$ for any j. Lemma 2.3.4 in Giné and Nickl (2015) ensures the result.

C.4.2 Proof of Lemma C.5

(i): A ball of radius ε/λ for $||.||$ is also a ball of radius ε for $\lambda||.||$.
(ii): Take a collection of closed balls $(B_n^b)_{n=1,...,N}$ for $||.||_b$, with radius ε and centers in $\mathcal{G}$, covering $\mathcal{G}$. The balls $B_n^a, n = 1, ..., N$ with the same centers and same radius for the norm $||.||_a$ are such that $B_n^b \subset B_n^a$. We conclude that $N(\varepsilon, \mathcal{G}, ||.||_a) \leq N(\varepsilon, \mathcal{G}, ||.||_b)$.
(iii): Take a finite collection of closed balls $(B'_n)_{n=1,...,N}$ for $||.||_b$, with radius $\varepsilon/2$ and centers in $\mathcal{G}'$, covering $\mathcal{G}'$. For all balls B'_n with non-empty intersection with $\mathcal{G}$, select an element of $\mathcal{G} \cap B'_n$ as a center of B_n, a new ball of radius ε. Then $B'_n \subset B_n$, and since $\mathcal{G} \subset \mathcal{G}'$, the collection of such balls B_n covers $\mathcal{G}$ and have centers in $\mathcal{G}$.
(iv): Take a finite collection of closed balls B_n, $n = 1, ..., N$ for $||.||$, with radius $\varepsilon/2$ and center c_n in $\mathcal{G}$, covering $\mathcal{G}$. Then the collection of $\Delta_{n,n'} = \{f - g, (f, g \in B_n^2\}$, for $n, n' = 1, ..., N$, covers $\mathcal{G}_\infty$. Moreover any $\Delta_{n,n'}$ is included in a ball of $\mathcal{G}_\infty$ centered at $c_n - c_{n'} \in \mathcal{G}_\infty$ of radius ε.

C.4.3 Proof of Lemma C.6

(i): For any $f \in \mathcal{F}$:

$$||\pi \circ f||_{\mu_C,r}^r = \frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq C} \left| \sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \right|^r \leq \frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq C} N_{\boldsymbol{j}}^{r-1} \sum_{\ell=1}^{N_{\boldsymbol{j}}} |f(Y_{\ell,\boldsymbol{j}})|^r \leq \overline{N}_r \, ||f||_{\mathbb{Q}_C^r,r}^r .$$

It follows that $||\pi \circ .||_{\mu_C,r} \leq \overline{N}_r^{1/r} ||.||_{\mathbb{Q}_C^r,r}$. Lemma C.5 $i)$ and $ii)$ ensures that

$$N\left(\varepsilon, \mathcal{F}, ||\pi \circ .||_{\mu_C,r}\right) \leq N\left(\frac{\varepsilon}{\overline{N}_r^{\frac{1}{r}}}, \mathcal{F}, ||\cdot||_{\mathbb{Q}_C^r,r}\right).$$

Because $N(\varepsilon, \mathcal{F}, ||\pi \circ .||_{\mu_C,r}) = N(\varepsilon, \pi \circ \mathcal{F}, ||.||_{\mu_C,r})$, we have:

$$N\left(\varepsilon, \pi \circ \mathcal{F}, ||\cdot||_{\mu_C,r}\right) \leq N\left(\frac{\varepsilon}{\overline{N}_r^{\frac{1}{r}}}, \mathcal{F}, ||\cdot||_{\mathbb{Q}_C^r,r}\right).$$

(ii): Because $\pi \circ \mathcal{F}_\delta = [\pi \circ \mathcal{F}]_\delta \subset [\pi \circ \mathcal{F}]_\infty = \pi \circ \mathcal{F}_\infty$, Lemma.C.5 *iii*), *iv*) and Lemma C.6 *i*) ensure that:

$$\begin{aligned} N\left(\varepsilon, \pi \circ \mathcal{F}_\delta, ||.||_{\mu_C,r}\right) &\le N\left(\frac{\varepsilon}{2}, \pi \circ \mathcal{F}_\infty, ||.||_{\mu_C,r}\right) \\ &\le N^2\left(\frac{\varepsilon}{4}, \pi \circ \mathcal{F}, ||.||_{\mu_C,r}\right) \\ &\le N^2\left(\frac{\varepsilon}{4\overline{N}_r^{\frac{1}{r}}}, \mathcal{F}, ||\cdot||_{\mathbb{Q}_C^r,r}\right). \end{aligned}$$

The proofs of *iii*) and *iv*) follow the same line as the proof of *i*) and *ii*) after having noted that $||\pi \circ .||_{\mu_C,r} \le A_r^{1/r}||.||_{\infty,\beta}$.

C.4.4 Proof of Lemma C.7

On $\mathcal{F}$, we have $||\pi \circ \cdot||^2_{\mu_C,2} = \left|\left|\pi^2 \circ \cdot\right|\right|_{\mu_C,1}$ and $\frac{||\pi\circ\cdot||_{\mu_C,2}}{||\pi\circ F||_{\mu_C,2}} \le 1$ which implies $\frac{\left|\left|\pi^2\circ\cdot\right|\right|_{\mu_C,1}}{||\pi^2\circ F||_{\mu_C,1}} = \frac{\left|\left|\pi^2\circ\cdot\right|\right|^2_{\mu_C,2}}{||\pi\circ F||^2_{\mu_C,2}} \le \frac{||\pi\circ\cdot||_{\mu_C,2}}{||\pi\circ F||_{\mu_C,2}}$. Then applying successively Lemma C.5 *iv*), *i*), *ii*) and *i*) again, we have:

$$\begin{aligned} N\left(2\varepsilon||\pi^2 \circ F||_{\mu_C,1}, \pi^2 \circ \mathcal{F}_\infty, ||\cdot||_{\mu_C,1}\right) &= N\left(2\varepsilon||\pi^2 \circ F||_{\mu_C,1}, \mathcal{F}_\infty, \left|\left|\pi^2 \circ \cdot\right|\right|_{\mu_C,1}\right) \\ &\le N^2\left(\varepsilon||\pi^2 \circ F||_{\mu_C,1}, \mathcal{F}, \left|\left|\pi^2 \circ \cdot\right|\right|_{\mu_C,1}\right) \\ &= N^2\left(\varepsilon, \mathcal{F}, \frac{\left|\left|\pi^2 \circ \cdot\right|\right|_{\mu_C,1}}{||\pi^2 \circ F||_{\mu_C,1}}\right) \\ &\le N^2\left(\varepsilon, \mathcal{F}, \frac{||\pi \circ \cdot||_{\mu_C,2}}{||\pi \circ F||_{\mu_C,2}}\right) \\ &= N^2\left(\varepsilon||\pi \circ F||_{\mu_C,2}, \pi \circ \mathcal{F}, ||\cdot||_{\mu_C,2}\right). \end{aligned}$$

C.4.5 Proof of Lemma C.8

A weighted sum of Rademacher variable is subgaussian with respect to the Euclidean norm $|.|$ of the vectors of weights. As a process indexed by the vector of weights. this is a subgaussian process for the Euclidean norm of the weights. Then, conditional on the original data, we can apply Theorem 2.3.6 in Giné and Nickl (2015). It follows that we have for any $\boldsymbol{e} \in \cup_{i=1}^k \mathcal{E}_i$:

$$\begin{aligned} &\mathbb{E}\left[\sup_{\mathcal{F}_\delta}\left|\frac{1}{\prod_{s:\boldsymbol{e}_s=1} C_s}\sum_{\boldsymbol{e}\le\boldsymbol{c}\le\boldsymbol{C}\odot\boldsymbol{e}} \epsilon_{\boldsymbol{c}}\frac{1}{\prod_{s:\boldsymbol{e}_s=0} C_s}\sum_{(\boldsymbol{1}-\boldsymbol{e})\le\boldsymbol{c}'\le\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})}\sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}'}} f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}'})\right| \mid \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right] \\ &\le \frac{4\sqrt{2}}{\sqrt{\prod_{s:\boldsymbol{e}_s=1} C_s}}\int_0^{\sigma_C^{\boldsymbol{e}}}\sqrt{\log 2N\left(\varepsilon, \pi \circ \mathcal{F}_\delta, ||\cdot||_{\mu_C^{\boldsymbol{e}},2}\right)}d\varepsilon \end{aligned}$$

The Jensen's inequality ensures $||.||_{\mu_C^{\boldsymbol{e}},2} \leq ||.||_{\mu_C,2}$ and $\sigma_C^{\boldsymbol{e}} \leq \sigma_C$. The linearity of the integration and Lemma C.5 *ii*) ensures:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\mathcal{F}_\delta}\left|\frac{1}{\prod_{s:\boldsymbol{e}_s=1} C_s}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\epsilon_{\boldsymbol{c}}\frac{1}{\prod_{s:\boldsymbol{e}_s=0} C_s}\sum_{(\boldsymbol{1}-\boldsymbol{e})\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})}\sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}'}} f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}'})\right| \mid \left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}\right]\\
&\leq \frac{4\sqrt{2}}{\sqrt{\prod_{s:\boldsymbol{e}_s=1} C_s}}\int_0^{\sigma_C}\sqrt{\log 2N\left(\varepsilon,\pi\circ\mathcal{F}_\delta,||\cdot||_{\mu_C,2}\right)}d\varepsilon\\
&= \frac{4\sqrt{2\log(2)}\sigma_C}{\sqrt{\prod_{s:\boldsymbol{e}_s=1} C_s}} + \frac{4\sqrt{2}}{\sqrt{\prod_{s:\boldsymbol{e}_s=1} C_s}}\int_0^{\sigma_C}\sqrt{\log N\left(\varepsilon,\pi\circ\mathcal{F}_\delta,||\cdot||_{\mu_C,2}\right)}d\varepsilon. \qquad (30)
\end{aligned}$$

Consider cases where $\overline{N}_2 > 0$. Lemma C.6 *ii*) ensures:

$$\int_0^{\sigma_C}\sqrt{\log N\left(\varepsilon,\pi\circ\mathcal{F}_\delta,||\cdot||_{\mu_C,2}\right)}d\varepsilon \leq \sqrt{2}\int_0^{\sigma_C}\sqrt{\log N\left(\frac{\varepsilon}{4\sqrt{\overline{N}_2}},\mathcal{F},||\cdot||_{\mathbb{Q}_C^2,2}\right)}d\varepsilon$$

Lemma C.1 ensures that we can assume without loss of generality that $||F||_{\mathbb{Q}_C^2,2} \geq 1$, and next we apply the change of variable $\varepsilon' = \frac{\varepsilon}{4\sqrt{\overline{N}_2}||F||_{\mathbb{Q}_C^2,2}}$ to get

$$\int_0^{\sigma_C}\sqrt{\log N\left(\varepsilon,\pi\circ\mathcal{F}_\delta,||\cdot||_{\mu_C,2}\right)}d\varepsilon \leq 4\sqrt{2}\sqrt{\overline{N}_2}\,||F||_{\mathbb{Q}_C^2,2}\,J_{2,\mathcal{F}}\left(\frac{\sigma_C}{4\sqrt{\overline{N}_2}\,||F||_{\mathbb{Q}_C^2,2}}\right)$$

Because $u\in]0,+\infty[\mapsto J_{2,\mathcal{F}}(u/4)$ is an increansing and concave function, $(x,y)\in]0,+\infty[^2\mapsto \sqrt{y}J_{2,\mathcal{F}}\left(\frac{\sqrt{x}}{4\sqrt{y}}\right)$ is concave and it follows from the Jensen's inequality that:

$$\begin{aligned}
\mathbb{E}\left(\sqrt{\overline{N}_2}\,||F||_{\mathbb{Q}_C^2,2}\,J_{2,\mathcal{F}}\left(\frac{\sigma_C}{4\sqrt{\overline{N}_2}\,||F||_{\mathbb{Q}_C^2,2}}\right)\middle|\overline{N}_2>0\right) &\leq \sqrt{\mathbb{E}\left(\overline{N}_2\,||F||_{\mathbb{Q}_C^2,2}^2\,|\overline{N}_2>0\right)}\\
&\quad J_{2,\mathcal{F}}\left(\frac{1}{4}\sqrt{\frac{\mathbb{E}(\sigma_C^2|\overline{N}_2>0)}{\mathbb{E}(\overline{N}_2\,||F||_{\mathbb{Q}_C^2,2}^2\,|\overline{N}_2>0)}}\right).
\end{aligned}$$

Because all the random variables in the expectations of the previous inequality are null when $\overline{N}_2 = 0$, we have:

$$\begin{aligned}
\mathbb{E}\left(\sqrt{\overline{N}_2}\,||F||_{\mathbb{Q}_C^2,2}\,J_{2,\mathcal{F}}\left(\frac{\sigma_C}{4\sqrt{\overline{N}_2}\,||F||_{\mathbb{Q}_C^2,2}}\right)\right) &\leq \sqrt{\mathbb{P}(\overline{N}_2>0)}\sqrt{\mathbb{E}\left(\overline{N}_2\,||F||_{\mathbb{Q}_C^2,2}^2\right)}\\
&\quad J_{2,\mathcal{F}}\left(\frac{1}{4}\sqrt{\frac{\mathbb{E}(\sigma_C^2)}{\mathbb{E}(\overline{N}_2\,||F||_{\mathbb{Q}_C^2,2}^2)}}\right)\\
&\leq \sqrt{\mathbb{E}\left(A_F\right)}J_{2,\mathcal{F}}\left(\frac{1}{4}\sqrt{\frac{\mathbb{E}(\sigma_C^2)}{\mathbb{E}(A_F)}}\right),
\end{aligned}$$

because $\mathbb{E}(\overline{N}_2 \,||F||^2_{\mathbb{Q}^2_C,2}) = \mathbb{E}(A_F)$. This implies that:

$$\mathbb{E}\left[\sup_{\mathcal{F}_\delta}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\le\boldsymbol{j}\le\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right|\right] \le 4\sqrt{\frac{2\mathbb{E}\left[\sigma_C^2\right]\log 2}{\prod_{s:\boldsymbol{e}_s=1} C_s}} + 32\sqrt{\frac{\mathbb{E}\left[A_F\right]}{\prod_{s:\boldsymbol{e}_s=1} C_s}} J_{2,\mathcal{F}}\left(\frac{1}{4}\sqrt{\frac{\mathbb{E}\left[\sigma_C^2\right]}{\mathbb{E}\left[A_F\right]}}\right),$$

Let $A_2 = \frac{1}{\Pi_C}\sum_{\boldsymbol{1}\le\boldsymbol{j}\le\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}\langle Y_{\ell,\boldsymbol{j}}\rangle^{-\beta}\right)^2 = \frac{1}{\Pi_C}A^2_\beta(\boldsymbol{j})$. Under Assumption 3', we deduce from Lemma C.6 *iv*) that

$$\begin{aligned}\int_0^{\sigma_C}\sqrt{\log N\left(\varepsilon,\pi\circ\mathcal{F}_\delta,||\cdot||_{\mu_C,2}\right)}d\varepsilon &\le \sqrt{2}\int_0^{\sigma_C}\sqrt{\log N\left(\frac{\varepsilon}{4\sqrt{A_2}},\mathcal{F},||\cdot||_{\infty,\beta}\right)}d\varepsilon\\ &\le 4\sqrt{2}||F||_{\infty,\beta}\sqrt{A_2}J_{\infty,\beta,\mathcal{F}}\left(\frac{1}{4||F||_{\infty,\beta}}\sqrt{\frac{\sigma_C^2}{A_2}}\right).\end{aligned}$$

Integration of the Inequality (30) over $\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}$, combined with Jensen's inequality ensures:

$$\begin{aligned}&\mathbb{E}\left[\sup_{\mathcal{F}_\delta}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\le\boldsymbol{j}\le\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right|\right]\\ &\le 4\sqrt{\frac{2\mathbb{E}\left[\sigma_C^2\right]\log 2}{\prod_{s:\boldsymbol{e}_s=1} C_s}} + 32||F||_{\infty,\beta}\times\sqrt{\frac{\mathbb{E}\left[A^2_\beta(\boldsymbol{1})\right]}{\prod_{s:\boldsymbol{e}_s=1} C_s}} J_{\infty,\beta,\mathcal{F}}\left(\frac{1}{4\,||F||_{\infty,\beta}}\sqrt{\frac{\mathbb{E}\left[\sigma_C^2\right]}{\mathbb{E}\left[A^2_\beta(\boldsymbol{1})\right]}}\right),\end{aligned}$$

C.4.6 Proof of Lemma C.9

Let $\mathbb{1}_M(\boldsymbol{j}) := \mathbb{1}\left\{\sum_{\ell=1}^{N_{\boldsymbol{j}}} F(Y_{\ell,\boldsymbol{j}}) \le M\right\}$ and $\left(\pi^2\circ\mathcal{F}_\infty\right)\mathbb{1}_M := \left\{\left(\pi^2\circ f\right)\mathbb{1}\left\{\pi\circ F\le M\right\} : f\in\mathcal{F}_\infty\right\}$. Select $m = N\left(\eta_1,\left(\pi^2\circ\mathcal{F}_\infty\right)\mathbb{1}_M,||\cdot||_{\mu^{\boldsymbol{e}}_C,1}\right)$ balls to cover the class $\left(\pi^2\circ\mathcal{F}_\infty\right)\mathbb{1}_M$. In each ball $B_{\mathring{m}}$ ($\mathring{m}= 1,...,m$) of the covering, select its center $f^*_{\mathring{m}} \in \left(\pi^2\circ\mathcal{F}_\infty\right)\mathbb{1}_M$. The triangle inequality ensures that:

$$\begin{aligned}&\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{e}\le\boldsymbol{c}\le\boldsymbol{C}\odot\boldsymbol{e}}\epsilon_{\boldsymbol{c}}\sum_{\boldsymbol{1}-\boldsymbol{e}\le\boldsymbol{c}'\le\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})}\left(\sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}'}} f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}'})\right)^2\mathbb{1}_M(\boldsymbol{c}+\boldsymbol{c}')\right|\;\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right]\\ &\le\mathbb{E}\left[\sup_{\mathring{m}=1,\dots,m}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{e}\le\boldsymbol{c}\le\boldsymbol{C}\odot\boldsymbol{e}}\epsilon_{\boldsymbol{c}}\sum_{\boldsymbol{1}-\boldsymbol{e}\le\boldsymbol{c}'\le\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})} f^*_{\mathring{m}}(N_{\boldsymbol{c}+\boldsymbol{c}'},\vec{Y}_{\boldsymbol{c}+\boldsymbol{c}'})\right|\;\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\ge\boldsymbol{1}}\right]+\eta_1\end{aligned}$$

Because the Euclidean norm of $\left(\frac{1}{\prod_{s:\boldsymbol{e}_s=0} C_s}\sum_{\boldsymbol{1}-\boldsymbol{e}\le\boldsymbol{c}'\le\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})} f^*_{\mathring{m}}(N_{\boldsymbol{c}+\boldsymbol{c}'},\vec{Y}_{\boldsymbol{c}+\boldsymbol{c}'})\right)_{\boldsymbol{e}\preceq\boldsymbol{c}\preceq\boldsymbol{C}\odot\boldsymbol{e}}$ is bounded by $2\sqrt{\prod_{s:\boldsymbol{e}_s=1} C_s} \times M$ for any m, Lemma C.4 ensures that we have for any $\eta_1 > 0$

$$\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\odot\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\epsilon_{\boldsymbol{c}}\sum_{\boldsymbol{1}-\boldsymbol{e}\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})}\left(\sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}'}} f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}'})\right)^2 \mathbb{1}_M(\boldsymbol{c}+\boldsymbol{c}')\right| \,\middle|\, \left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}\right]$$
$$\leq 2\sqrt{2\log 2N\left(\eta_1,(\pi^2\circ\mathcal{F}_\infty)\,\mathbb{1}_M,||\cdot||_{\mu_C^{\boldsymbol{e}},1}\right)}\frac{M}{\sqrt{\prod_{s:\boldsymbol{e}_s=1}C_s}}+\eta_1$$

Because $||.||_{\mu_C^{\boldsymbol{e}},1} \leq ||.||_{\mu_C,1}$, Lemma C.5 *ii)* ensures:

$$\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right)^2 \mathbb{1}_M(\boldsymbol{j})\right| \,\middle|\, \left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}\right]$$
$$= \mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\odot\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\epsilon_{\boldsymbol{c}}\sum_{\boldsymbol{1}-\boldsymbol{e}\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})}\left(\sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}'}} f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}'})\right)^2 \mathbb{1}_M(\boldsymbol{c}+\boldsymbol{c}')\right| \,\middle|\, \left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}\right]$$
$$\leq 2\sqrt{2\log 2N\left(\eta_1,\pi^2\circ\mathcal{F}_\infty,||\cdot||_{\mu_C,1}\right)}\frac{M}{\sqrt{\prod_{s:\boldsymbol{e}_s=1}C_s}}+\eta_1. \tag{31}$$

Note that if $N_{\boldsymbol{j}} = 0$ for all $\boldsymbol{j}$, then the random measure μ_C is null and the previous inequality also hold for $\eta_1 = 0$.

Fix $\eta > 0$.

Let us first focus on Assumption 3.

Apply the previous inequality to the non negative random variable

$$\eta_1 = 2\eta||F||_{\mathbb{Q}_C^2,2}\sqrt{\overline{N}_2}||\pi\circ F||_{\mu_C,2} = 2\eta||F||_{\mathbb{Q}_C^2,2}\sqrt{\overline{N}_2}\frac{||\pi^2\circ F||_{\mu_C,1}}{||\pi\circ F||_{\mu_C,2}}.$$

Note that Lemma C.1 ensures that $\eta_1 = 0$ if and only if $\overline{N}_2 = 0$. When $\overline{N}_2 > 0$ use Lemma C.7 to deduce that:

$$\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right)^2 \mathbb{1}_M(\boldsymbol{j})\right| \,\middle|\, \left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}\right]$$
$$\leq 2\sqrt{2\log 2N\left(2\eta||F||_{\mathbb{Q}_C^2,2}\sqrt{\overline{N}_2}\frac{||\pi^2\circ F||_{\mu_C,1}}{||\pi\circ F||_{\mu_C,2}},\pi^2\circ\mathcal{F}_\infty,||\cdot||_{\mu_C,1}\right)}\frac{M}{\sqrt{\prod_{s:\boldsymbol{e}_s=1}C_s}}$$
$$+ 2\eta||F||_{\mathbb{Q}_C^2,2}\sqrt{\overline{N}_2}||\pi\circ F||_{\mu_C,2}$$
$$\leq \frac{M}{\sqrt{\prod_{s:\boldsymbol{e}_s=1}C_s}}\times\left\{2\sqrt{2\log 2}+4\sqrt{\log N\left(\eta||F||_{\mathbb{Q}_C^2,2}\sqrt{\overline{N}_2},\pi\circ\mathcal{F},||\cdot||_{\mu_C,2}\right)}\right\}$$
$$+ 2\eta||F||_{\mathbb{Q}_C^2,2}\sqrt{\overline{N}_2}||\pi\circ F||_{\mu_C,2}.$$

Using Lemma C.6 i) for the first term and Cauchy-Schwarz inequality for the second term, we deduce:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2\mathbb{1}_M(\boldsymbol{j})\right|\;\middle|\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right]\\
&\leq \frac{M}{\sqrt{\prod_{s:\boldsymbol{e}_s=1}C_s}}\times\left\{2\sqrt{2\log 2}+4\sqrt{\log N\left(\eta||F||_{\mathbb{Q}_C^2,2},\mathcal{F},||\cdot||_{\mathbb{Q}_C^2,2}\right)}\right\}\\
&+2\eta||F||^2_{\mathbb{Q}_C^2,2}\overline{N}_2
\end{aligned}$$

Because $\eta\mapsto\sqrt{\log N\left(\eta||F||_{\mathbb{Q}_C^2,2},\pi\circ\mathcal{F},||\cdot||_{\mathbb{Q}_C^2,2}\right)}$ is decreasing, we have:

$$\begin{aligned}
\sqrt{\log N\left(\eta||F||_{\mathbb{Q}_C^2,2},\pi\circ\mathcal{F},||\cdot||_{\mathbb{Q}_C^2,2}\right)} &\leq \frac{1}{\eta}\int_0^\eta\sqrt{\log N\left(u||F||_{\mathbb{Q}_C^2,2},\pi\circ\mathcal{F},||\cdot||_{\mathbb{Q}_C^2,2}\right)}du\\
&\leq\frac{1}{\eta}J_{2,\mathcal{F}}(\eta)\\
&\leq\frac{1}{\eta}J_{2,\mathcal{F}}(\infty).
\end{aligned}$$

Note that $J_{2,\mathcal{F}}(\infty)<\infty$ by Assumption 3. When $\overline{N}_2=0$ then $N_{\boldsymbol{j}}=0$ for any $\boldsymbol{j}$ and next the measure $\mathbb{Q}_C^2$ is null (by convention) and $\sqrt{\log N\left(\eta||F||_{\mathbb{Q}_C^2,2},\pi\circ\mathcal{F},||\cdot||_{\mathbb{Q}_C^2,2}\right)}=0\leq\frac{1}{\eta}J_{2,\mathcal{F}}(\infty)$. Last, by integration with respect to $\left(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}$, we get:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}}\epsilon_{\boldsymbol{j}\odot\boldsymbol{e}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\right)^2\mathbb{1}_M(\boldsymbol{j})\right|\right]\\
&\leq\frac{M}{\sqrt{\prod_{s:\boldsymbol{e}_s=1}C_s}}\times\left\{2\sqrt{2\log 2}+\frac{4}{\eta}J_{2,F}(\infty)\right\}+2\eta\mathbb{E}\left(||F||^2_{\mathbb{Q}_C^2,2}\overline{N}_2\right)\\
&=\frac{M}{\sqrt{\prod_{s:\boldsymbol{e}_s=1}C_s}}\times\left\{2\sqrt{2\log 2}+\frac{4}{\eta}J_{2,F}(\infty)\right\}+2\eta\mathbb{E}\left(N_{\mathbf{1}}\sum_{\ell=1}^{N_{\mathbf{1}}}F^2(Y_{\ell,\mathbf{1}})\right)
\end{aligned}$$

This concludes the proof of the first part of the Lemma.
Under Assumption 3', for any pair $(f,g)\in(\mathcal{F})^2$, the triangle and Jensen inequalities give us

$$\begin{aligned}
\left|\left|\pi^2\circ(f-g)\right|\right|_{\mu_C,1}&\leq\frac{2}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}F(Y_{\ell,\boldsymbol{j}})\right)\left|\sum_{\ell=1}^{N_{\boldsymbol{j}}}[f(Y_{\ell,\boldsymbol{j}})-g(Y_{\ell,\boldsymbol{j}})]\right|\\
&\leq 2\,||F||_{\infty,\beta}\,A_2\times\left\{||f-g||_{\infty,\beta}\right\},
\end{aligned}\tag{32}$$

where $A_2=\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}}\langle Y_{\ell,\boldsymbol{j}}\rangle^{-\beta}\right)^2$.

From Equation (32), Lemma C.5 *ii*), *i*) and *iv*) we deduce that for any $\eta_1 > 0$:

$$\begin{aligned} N\left(\eta_1, \pi^2 \circ \mathcal{F}_\infty, ||\cdot||_{\mu_C,1}\right) &\leq N\left(\eta_1, \mathcal{F}_\infty, 2\,||F||_{\infty,\beta}\, A_2\, ||\cdot||_{\infty,\beta}\right) \\ &\leq N\left(\frac{\eta_1}{2\,||F||_{\infty,\beta}\, A_2}, \mathcal{F}_\infty, ||\cdot||_{\infty,\beta}\right) \\ &\leq N^2\left(\frac{\eta_1}{4\,||F||_{\infty,\beta}\, A_2}, \mathcal{F}, ||\cdot||_{\infty,\beta}\right). \end{aligned}$$

Now consider $\eta_1 = 4\eta\,||F||^2_{\infty,\beta}\, A_2$ in Equation (31), to deduce that

$$\begin{aligned} &\mathbb{E}\left[\sup_{\mathcal{F}_\infty}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{C}} \epsilon_{\boldsymbol{j}\odot \boldsymbol{e}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right)^2 \mathbb{1}_M(\boldsymbol{j})\right| \,\middle|\, \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq \mathbf{1}}\right] \\ &\leq 2\sqrt{2\log 2N\left(4\eta\,||F||^2_{\infty,\beta}\, A_2, \pi^2\circ\mathcal{F}_\infty, ||\cdot||_{\mu_C,1}\right)}\frac{M}{\sqrt{\prod_{s:\boldsymbol{e}_s=1} C_s}} + 4\eta\,||F||^2_{\infty,\beta}\, A_2 \\ &\leq \frac{M}{\sqrt{\prod_{s:\boldsymbol{e}_s=1} C_s}} \times \left\{2\sqrt{2\log 2} + 4\sqrt{\log N\left(\eta\,||F||_{\infty,\beta}, \mathcal{F}, ||\cdot||_{\infty,\beta}\right)}\right\} + 4\eta\,||F||^2_{\infty,\beta}\, A_2. \end{aligned}$$

Because $\eta \mapsto \sqrt{\log N\left(\eta\,||F||_{\infty,\beta}, \mathcal{F}, ||\cdot||_{\infty,\beta}\right)}$ is decreasing, we have:

$$\sqrt{\log N\left(\eta\,||F||_{\infty,\beta}, \mathcal{F}, ||\cdot||_{\infty,\beta}\right)} \leq \frac{1}{\eta} J_{\infty,\beta,\mathcal{F}}(\eta) \leq \frac{1}{\eta} J_{\infty,\beta,\mathcal{F}}(\infty)$$

Integration over $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq \mathbf{1}}$ ensures the result.

C.4.7 Proof of Lemma C.10

Let F and G the respective envelope of $\mathcal{F}$ and $\mathcal{G}$.
Recall that

$$\mathcal{A}_i := \left\{(\boldsymbol{j}, \boldsymbol{j}') : \mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}, \mathbf{1} \leq \boldsymbol{j}' \leq \boldsymbol{C}, j_i = j'_i, j_s \neq j'_s \forall s \neq i\right\}$$

and

$$\mathcal{B}_i := \left\{(\boldsymbol{j}, \boldsymbol{j}') : \mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}, \mathbf{1} \leq \boldsymbol{j}' \leq \boldsymbol{C}, j_i = j'_i\right\}.$$

Because $\mathcal{A}_i \subset \mathcal{B}_i$, we have:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j}'}}g(Y_{\ell,\boldsymbol{j}'})-\frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j}'}}g(Y_{\ell,\boldsymbol{j}'})\right|\right]\\
&\leq\left|\frac{1}{|\mathcal{B}_i|}-\frac{1}{|\mathcal{A}_i|}\right|\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j}'}}g(Y_{\ell,\boldsymbol{j}'})\right|\right]\\
&+\frac{1}{|\mathcal{B}_i|}\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i\setminus\mathcal{A}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j}'}}g(Y_{\ell,\boldsymbol{j}'})\right|\right]\\
&\leq\left|\frac{1}{|\mathcal{B}_i|}-\frac{1}{|\mathcal{A}_i|}\right|\times|\mathcal{A}_i|\times\sqrt{\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}F(Y_{\ell,\mathbf{1}})\right)^2\right]}\sqrt{\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}G(Y_{\ell,\mathbf{1}})\right)^2\right]}\\
&+\frac{1}{|\mathcal{B}_i|}\times|\mathcal{B}_i\setminus\mathcal{A}_i|\times\sqrt{\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}F(Y_{\ell,\mathbf{1}})\right)^2\right]}\sqrt{\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}G(Y_{\ell,\mathbf{1}})\right)^2\right]}
\end{aligned}$$

We have $|\mathcal{A}_i| = C_i\prod_{s\neq i}C_s(C_s-1)$, $|\mathcal{B}_i| = C_i\prod_{s\neq i}C_s^2$ and $|\mathcal{B}_i\setminus\mathcal{A}_i| = |\mathcal{B}_i|-|\mathcal{A}_i|$, this implies that $\lim_{\underline{C}\to\infty}\frac{|\mathcal{A}_i|}{|\mathcal{B}_i|}=1$ and next:

$$\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j}'}}g(Y_{\ell,\boldsymbol{j}'})-\frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j}'}}g(Y_{\ell,\boldsymbol{j}'})\right|\right]=o(1)$$

For the second part of the Lemma, note that the envelope conditions and the Cauchy-Schwarz inequality ensure:

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{\mathcal{F}\times\mathcal{G}}\left|\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_{\boldsymbol{e}}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j}'}} g(Y_{\ell,\boldsymbol{j}'})\right|\right)\\
&\leq \mathbb{E}\left(\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_{\boldsymbol{e}}}\sum_{\ell=1}^{N_{\boldsymbol{j}}} F(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j}'}} G(Y_{\ell,\boldsymbol{j}'})\right)\\
&\leq \mathbb{E}\left(\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\ \sum_{\boldsymbol{1}-\boldsymbol{e}\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})}\ \sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}'}} F(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}'})\sum_{\boldsymbol{1}-\boldsymbol{e}\leq\boldsymbol{c}''\leq\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})}\ \sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}''}} G(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}''})\right)\\
&\leq \mathbb{E}\left(\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\left(\sum_{\boldsymbol{1}-\boldsymbol{e}\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})}\ \sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}'}} F(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}'})\right)^2\right)^{1/2}\\
&\quad\times \mathbb{E}\left(\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{C}\odot\boldsymbol{e}}\left(\sum_{\boldsymbol{1}-\boldsymbol{e}\leq\boldsymbol{c}''\leq\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{e})}\ \sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}''}} G(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}''})\right)^2\right)^{1/2}\\
&\leq \left(\prod_{s:e_i=0} C_s\right)\times\mathbb{E}\left(\sum_{\boldsymbol{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} F(Y_{\ell,\boldsymbol{j}})\right)^2\right)^{1/2}\times\mathbb{E}\left(\sum_{\boldsymbol{e}\leq\boldsymbol{j}\leq\boldsymbol{C}}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} G(Y_{\ell,\boldsymbol{j}})\right)^2\right)^{1/2}\\
&\leq \left(\prod_{s:e_i=0} C_s\right)\times\Pi_C\times\mathbb{E}\left(\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} F(Y_{\ell,\boldsymbol{1}})\right)^2\right)^{1/2}\times\mathbb{E}\left(\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} G(Y_{\ell,\boldsymbol{1}})\right)^2\right)^{1/2}
\end{aligned}
$$

C.4.8 Proof of Lemma C.11

We first prove the result for $\mathcal{A}_i$ with $i = 1, ..., k$. The result for $\mathcal{B}_i$ will follow from Lemma C.10.

The representation Lemma B.1 ensures that: $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}} \overset{a.s}{=} \left(\tau\left(\left(U_{\boldsymbol{j}\odot\boldsymbol{e}}\right)_{\boldsymbol{0}<\boldsymbol{e}\leq\boldsymbol{1}}\right)\right)_{\boldsymbol{j}\geq\boldsymbol{1}}$, for some mutually independent uniform random variables on $(U_{\boldsymbol{c}})_{\boldsymbol{c}>\boldsymbol{0}}$. If $(U^{(1)}_{\boldsymbol{c}})_{\boldsymbol{c}>\boldsymbol{0}}$ is an independent copy of $(U_{\boldsymbol{c}})_{\boldsymbol{c}>\boldsymbol{0}}$, then $\left(N^{(1)}_{\boldsymbol{j}}, \vec{Y}^{(1)}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}} \overset{a.s}{=} \left(\tau\left(\left(U^{(1)}_{\boldsymbol{j}\odot\boldsymbol{e}}\right)_{\boldsymbol{0}<\boldsymbol{e}\leq\boldsymbol{1}}\right)\right)_{\boldsymbol{j}\geq\boldsymbol{1}}$ is an independent copy of $\left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\boldsymbol{1}}$. Because the array is separately exchangeable, we have

$$
\mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}})\sum_{\ell=1}^{N_{\boldsymbol{2}_i}} g(Y_{\ell,\boldsymbol{2}_i})\right] = \mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j}'}} g(Y_{\ell,\boldsymbol{j}'})\right]
$$

for any $(\boldsymbol{j}, \boldsymbol{j}') \in \mathcal{A}_i$ and next:

$$\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}) \sum_{\ell=1}^{N_{\mathbf{2}_i}} g(Y_{\ell,\mathbf{2}_i})\right] = \mathbb{E}\left[\frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \sum_{\ell=1}^{N_{\boldsymbol{j}'}} g(Y_{\ell,\boldsymbol{j}'})\right]$$
$$= \mathbb{E}\left[\frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_i}\sum_{\ell=1}^{N^{(1)}_{\boldsymbol{j}}} f^{(1)}(Y_{\ell,\boldsymbol{j}}) \sum_{\ell=1}^{N^{(1)}_{\boldsymbol{j}'}} g(Y^{(1)}_{\ell,\boldsymbol{j}'}) \middle| \left(N_{\boldsymbol{j}}, \vec{Y}_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}\right].$$

The monotonicity of the expectation and the law of iterated expectation ensure:

$$\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \sum_{\ell=1}^{N_{\boldsymbol{j}'}} g(Y_{\ell,\boldsymbol{j}'}) - \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\boldsymbol{j}}) \sum_{\ell=1}^{N_{\mathbf{2}_i}} g(Y_{\ell,\boldsymbol{j}'})\right]\right|\right]$$
$$\leq \mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_i}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \sum_{\ell=1}^{N_{\boldsymbol{j}'}} g(Y_{\ell,\boldsymbol{j}'}) - \sum_{\ell=1}^{N^{(1)}_{\boldsymbol{j}}} f(Y^{(1)}_{\ell,\boldsymbol{j}}) \sum_{\ell=1}^{N^{(1)}_{\boldsymbol{j}'}} g(Y^{(1)}_{\ell,\boldsymbol{j}'})\right)\right|\right] \tag{33}$$

Moreover, Lemma C.10 combined with triangular inequality ensures that:

$$\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \sum_{\ell=1}^{N_{\boldsymbol{j}'}} g(Y_{\ell,\boldsymbol{j}'}) - \mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\boldsymbol{j}}) \sum_{\ell=1}^{N_{\mathbf{2}_i}} g(Y_{\ell,\boldsymbol{j}'})\right]\right|\right]$$
$$\leq \mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \sum_{\ell=1}^{N_{\boldsymbol{j}'}} g(Y_{\ell,\boldsymbol{j}'}) - \sum_{\ell=1}^{N^{(1)}_{\boldsymbol{j}}} f(Y^{(1)}_{\ell,\boldsymbol{j}}) \sum_{\ell=1}^{N^{(1)}_{\boldsymbol{j}'}} g(Y^{(1)}_{\ell,\boldsymbol{j}'})\right)\right|\right] + o(1) \tag{34}$$

Let $S^f_{\boldsymbol{j}} = \sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})$, $\widetilde{S}^f_{\boldsymbol{j}}(\mathbf{0}) = \sum_{\ell=1}^{N^{(1)}_{\boldsymbol{j}}} f(Y^{(1)}_{\ell,\boldsymbol{j}})$, and for any $\boldsymbol{e}$ such that $\mathbf{0} \leq \boldsymbol{e} \leq \mathbf{1}$ let $S^f_{\boldsymbol{j}}(\boldsymbol{e}) = \sum_{\ell=1}^{N_{\boldsymbol{j}}(\boldsymbol{e})} f(Y_{\ell,\boldsymbol{j}}(\boldsymbol{e}))$, $\widetilde{S}^f_{\boldsymbol{j}}(\boldsymbol{e}) = \sum_{\ell=1}^{\widetilde{N}_{\boldsymbol{j}}(\boldsymbol{e})} f(\widetilde{Y}_{\ell,\boldsymbol{j}}(\boldsymbol{e}))$ for

$$\left(N_{\boldsymbol{j}}(\boldsymbol{e}), \vec{Y}_{\boldsymbol{j}}(\boldsymbol{e})\right)_{\boldsymbol{j}\geq\mathbf{1}} \overset{a.s}{=} \left(\tau\left(\left(U_{\boldsymbol{j}\odot\boldsymbol{e}'}\right)_{\mathbf{0}\prec\boldsymbol{e}'\preceq\boldsymbol{e}}, \left(U^{(1)}_{\boldsymbol{j}\odot\boldsymbol{e}'}\right)_{\boldsymbol{e}\prec\boldsymbol{e}'\preceq\mathbf{1}}\right)\right)_{\boldsymbol{j}\geq\mathbf{1}}$$
$$\left(\widetilde{N}_{\boldsymbol{j}}(\boldsymbol{e}), \vec{\widetilde{Y}}_{\boldsymbol{j}}(\boldsymbol{e})\right)_{\boldsymbol{j}\geq\mathbf{1}} \overset{a.s}{=} \left(\tau\left(\left(U_{\boldsymbol{j}\odot\boldsymbol{e}'}\right)_{\mathbf{0}\prec\boldsymbol{e}'\prec\boldsymbol{e}}, \left(U^{(1)}_{\boldsymbol{j}\odot\boldsymbol{e}'}\right)_{\boldsymbol{e}\preceq\boldsymbol{e}'\preceq\mathbf{1}}\right)\right)_{\boldsymbol{j}\geq\mathbf{1}}.$$

We have $S^f_{\boldsymbol{j}}(\mathbf{1})S^g_{\boldsymbol{j}}(\mathbf{1}) = S^f_{\boldsymbol{j}}S^g_{\boldsymbol{j}}$ and $\widetilde{S}^f_{\boldsymbol{j}}(\mathbf{0})\widetilde{S}^g_{\boldsymbol{j}}(\mathbf{0}) = \widetilde{S}^f_{\boldsymbol{j}}\widetilde{S}^g_{\boldsymbol{j}}$. The triangle inequality ensures:

$$\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \sum_{\ell=1}^{N_{\boldsymbol{j}'}} g(Y_{\ell,\boldsymbol{j}'}) - \sum_{\ell=1}^{N^{(1)}_{\boldsymbol{j}}} f(Y^{(1)}_{\ell,\boldsymbol{j}}) \sum_{\ell=1}^{N^{(1)}_{\boldsymbol{j}'}} g(Y^{(1)}_{\ell,\boldsymbol{j}'})\right)\right|\right]$$
$$\leq \sum_{\boldsymbol{b}_i\leq\boldsymbol{e}\leq\mathbf{1}}\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{C_i}\sum_{\boldsymbol{b}_i\leq\boldsymbol{c}\leq C_i\boldsymbol{b}_i}\frac{1}{\prod_{s\neq i}C_s^2}\sum_{\mathbf{1}-\boldsymbol{b}_i\leq\boldsymbol{c}',\boldsymbol{c}''\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{b}_i)} S^f_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})S^g_{\boldsymbol{c}+\boldsymbol{c}''}(\boldsymbol{e}) - \widetilde{S}^f_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})\widetilde{S}^g_{\boldsymbol{c}+\boldsymbol{c}''}(\boldsymbol{e})\right|\right]$$
$$+\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{C_i}\sum_{\boldsymbol{b}_i\leq\boldsymbol{c}\leq C_i\boldsymbol{b}_i}\frac{1}{\prod_{s\neq i}C_s^2}\sum_{\mathbf{1}-\boldsymbol{b}_i\leq\boldsymbol{c}',\boldsymbol{c}''\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{b}_i)} S^f_{\boldsymbol{c}+\boldsymbol{c}'}(\mathbf{0})S^g_{\boldsymbol{c}+\boldsymbol{c}''}(\mathbf{0}) - \widetilde{S}^f_{\boldsymbol{c}+\boldsymbol{c}'}(\mathbf{0})\widetilde{S}^g_{\boldsymbol{c}+\boldsymbol{c}''}(\mathbf{0})\right|\right].$$

For any $\boldsymbol{e}, \boldsymbol{c}$ such that $\mathbf{0} \leq \boldsymbol{e} \leq \mathbf{1}$ and $\boldsymbol{b}_i \leq \boldsymbol{c} \leq C_i \boldsymbol{b}_i$, if $R_{\boldsymbol{e}}$ denotes $\left(\left(U_{\boldsymbol{j}\odot\boldsymbol{e}'}\right)_{\mathbf{0}\prec\boldsymbol{e}'\prec\boldsymbol{e}}, \left(U^{(1)}_{\boldsymbol{j}\odot\boldsymbol{e}'}\right)_{\boldsymbol{e}\prec\boldsymbol{e}'\preceq\mathbf{1}}\right)_{\boldsymbol{j}\geq\mathbf{1}}$ then the terms $\left(\sum_{\mathbf{1}-\boldsymbol{b}_i\leq\boldsymbol{c}',\boldsymbol{c}''\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{b}_i)} S^f_{\boldsymbol{c}+\boldsymbol{c}'}(\mathbf{0})S^g_{\boldsymbol{c}+\boldsymbol{c}''}(\mathbf{0}) - \widetilde{S}^f_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})\widetilde{S}^g_{\boldsymbol{c}+\boldsymbol{c}''}(\boldsymbol{e})\right)$ are independent across $\boldsymbol{c}$ conditionally on $R_{\boldsymbol{e}}$ and have a symmetric conditional distribution. It follows that for any Rademacher process $\epsilon_{\boldsymbol{c},\boldsymbol{e}}$ indexed by $(\boldsymbol{c}, \boldsymbol{e})$ and independent form the $U, U^{(1)}$, we have:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_i}\left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j}'}} g(Y_{\ell,\boldsymbol{j}'}) - \sum_{\ell=1}^{N^{(1)}_{\boldsymbol{j}}} f(Y^{(1)}_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N^{(1)}_{\boldsymbol{j}'}} g(Y^{(1)}_{\ell,\boldsymbol{j}'})\right)\right|\right]\\
&\leq \sum_{\boldsymbol{b}_i\leq\boldsymbol{e}\leq\mathbf{1}}\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{C_i}\sum_{\boldsymbol{b}_i\leq\boldsymbol{c}\leq C_i\boldsymbol{b}_i}\frac{\epsilon_{\boldsymbol{c},\boldsymbol{e}}}{\prod_{s\neq i}C_s^2}\sum_{\mathbf{1}-\boldsymbol{b}_i\leq\boldsymbol{c}',\boldsymbol{c}''\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{b}_i)} S^f_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})S^g_{\boldsymbol{c}+\boldsymbol{c}''}(\boldsymbol{e}) - \widetilde{S}^f_{\boldsymbol{c}+\boldsymbol{c}'}(\boldsymbol{e})\widetilde{S}^g_{\boldsymbol{c}+\boldsymbol{c}''}(\boldsymbol{e})\right|\right]\\
&+\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{C_i}\sum_{\boldsymbol{b}_i\leq\boldsymbol{c}\leq C_i\boldsymbol{b}_i}\frac{\epsilon_{\boldsymbol{c},\mathbf{0}}}{\prod_{s\neq i}C_s^2}\sum_{\mathbf{1}-\boldsymbol{b}_i\leq\boldsymbol{c}',\boldsymbol{c}''\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{b}_i)} S^f_{\boldsymbol{c}+\boldsymbol{c}'}(\mathbf{0})S^g_{\boldsymbol{c}+\boldsymbol{c}''}(\mathbf{0}) - \widetilde{S}^f_{\boldsymbol{c}+\boldsymbol{c}'}(\mathbf{0})\widetilde{S}^g_{\boldsymbol{c}+\boldsymbol{c}''}(\mathbf{0})\right|\right].\\
&\leq 2\sum_{\boldsymbol{b}_i\leq\boldsymbol{e}\leq\mathbf{1}}\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{C_i}\sum_{\boldsymbol{b}_i\leq\boldsymbol{c}\leq C_i\boldsymbol{b}_i}\epsilon_{\boldsymbol{c},\boldsymbol{e}}\frac{1}{\prod_{s\neq i}C_s^2}\sum_{\mathbf{1}-\boldsymbol{b}_i\leq\boldsymbol{c}',\boldsymbol{c}''\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{b}_i)} S^f_{\boldsymbol{c}+\boldsymbol{c}'}S^g_{\boldsymbol{c}+\boldsymbol{c}''}\right|\right]\\
&+2\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{C_i}\sum_{\boldsymbol{b}_i\leq\boldsymbol{c}\leq C_i\boldsymbol{b}_i}\epsilon_{\boldsymbol{c},\mathbf{0}}\frac{1}{\prod_{s\neq i}C_s^2}\sum_{\mathbf{1}-\boldsymbol{b}_i\leq\boldsymbol{c}',\boldsymbol{c}''\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{b}_i)} S^f_{\boldsymbol{c}+\boldsymbol{c}'}S^g_{\boldsymbol{c}+\boldsymbol{c}''}\right|\right]. \qquad (35)
\end{aligned}$$

The last inequality follows from the triangular inequality, from the independence of $\epsilon_{.,.}$ and the $S^{\cdot}(.)$ and the fact that $\left(S^f_{\boldsymbol{j}}(\boldsymbol{e})S^g_{\boldsymbol{j}}(\boldsymbol{e})\right)_{\boldsymbol{j}\geq\mathbf{1}} \overset{d}{=} \left(\widetilde{S}^f_{\boldsymbol{j}}(\boldsymbol{e})\widetilde{S}^g_{\boldsymbol{j}}(\boldsymbol{e})\right)_{\boldsymbol{j}\geq\mathbf{1}} \overset{d}{=} \left(S^f_{\boldsymbol{j}}S^g_{\boldsymbol{j}}\right)_{\boldsymbol{j}\geq\mathbf{1}}$.
Let $\mathbb{1}_M(\boldsymbol{c}) := \mathbb{1}\left\{\frac{1}{\prod_{s\neq i}C_s}\sum_{\mathbf{1}-\boldsymbol{b}_i\leq\boldsymbol{c}'\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{b}_i)} S^F_{\boldsymbol{c}+\boldsymbol{c}'} \vee \frac{1}{\prod_{s\neq i}C_s}\sum_{\mathbf{1}-\boldsymbol{b}_i\leq\boldsymbol{c}''\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{b}_i)} S^G_{\boldsymbol{c}+\boldsymbol{c}''} \leq M\right\}$, for any $\boldsymbol{e}$ we have:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{C_i}\sum_{\boldsymbol{b}_i\leq\boldsymbol{c}\leq C_i\boldsymbol{b}_i}\epsilon_{\boldsymbol{c},\boldsymbol{e}}\frac{1}{\prod_{s\neq i}C_s^2}\sum_{\mathbf{1}-\boldsymbol{b}_i\leq\boldsymbol{c}',\boldsymbol{c}''\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{b}_i)} S^f_{\boldsymbol{c}+\boldsymbol{c}'}S^g_{\boldsymbol{c}+\boldsymbol{c}''}\right|\right]\\
&\leq \mathbb{E}\left[\frac{1}{C_i}\sum_{\boldsymbol{b}_i\leq\boldsymbol{c}\leq C_i\boldsymbol{b}_i}\frac{1}{\prod_{s\neq i}C_s^2}\sum_{\mathbf{1}-\boldsymbol{b}_i\leq\boldsymbol{c}',\boldsymbol{c}''\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{b}_i)} S^F_{\boldsymbol{c}+\boldsymbol{c}'}S^G_{\boldsymbol{c}+\boldsymbol{c}''}(1-\mathbb{1}_M(\boldsymbol{c}))\right]\\
&+\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{C_i}\sum_{\boldsymbol{b}_i\leq\boldsymbol{c}\leq C_i\boldsymbol{b}_i}\epsilon_{\boldsymbol{c},\boldsymbol{e}}\frac{1}{\prod_{s\neq i}C_s^2}\sum_{\mathbf{1}-\boldsymbol{b}_i\leq\boldsymbol{c}',\boldsymbol{c}''\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{b}_i)} S^f_{\boldsymbol{c}+\boldsymbol{c}'}S^g_{\boldsymbol{c}+\boldsymbol{c}''}\mathbb{1}_M(\boldsymbol{c})\right|\right]. \qquad (36)
\end{aligned}$$

The first term tends to 0 for large M. Lemma C.4 and the inequality $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$

ensure that for any $M > 0$ and $\eta > 0$:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{C_i}\sum_{\boldsymbol{b}_i\le\boldsymbol{c}\le C_i\boldsymbol{b}_i}\epsilon_{\boldsymbol{c},\boldsymbol{e}}\frac{1}{\prod_{s\neq i}C_s^2}\sum_{\boldsymbol{1}-\boldsymbol{b}_i\le\boldsymbol{c}',\boldsymbol{c}''\le\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{b}_i)}S^f_{\boldsymbol{c}+\boldsymbol{c}'}S^g_{\boldsymbol{c}+\boldsymbol{c}''}\mathbb{1}_M(\boldsymbol{c})\right|\,\middle|(N_{\boldsymbol{j}},\vec{Y}_{\boldsymbol{j}})_{\boldsymbol{j}\ge\boldsymbol{1}}\right]\\
&\le \eta + M^2\sqrt{\frac{2\ln(2)+2\ln N(\eta,\mathcal{H},||.||_q)}{C_i}}\\
&\le \eta + \frac{M^2}{\sqrt{C_i}}\left(\sqrt{2\ln(2)}+\sqrt{2\ln N(\eta,\mathcal{H},||.||_q)}\right), \qquad (37)
\end{aligned}$$

with $\mathcal{H}$ the class of functions on the $(k-1)^2$-dimensional subarrays $A_{\boldsymbol{c}}$ indexed by $\boldsymbol{c}\in\mathbb{N}^+\boldsymbol{b}_i$ defined by

$$\mathcal{H}=\left\{r(A_{\boldsymbol{c}})=\frac{1}{\prod_{s\neq i}C_s^2}\sum_{\boldsymbol{1}-\boldsymbol{b}_i\le\boldsymbol{c}',\boldsymbol{c}''\le\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{b}_i)}S^f_{\boldsymbol{c}+\boldsymbol{c}'}S^g_{\boldsymbol{c}+\boldsymbol{c}''}\mathbb{1}_M(\boldsymbol{c}),(f,g)\in\mathcal{F}\times\mathcal{G}\right\},$$

for

$$A_{\boldsymbol{c}}=\left\{(\vec{N}_{\boldsymbol{c}+\boldsymbol{c}'},\vec{Y}_{\boldsymbol{c}+\boldsymbol{c}'},\vec{N}_{\boldsymbol{c}+\boldsymbol{c}''},\vec{Y}_{\boldsymbol{c}+\boldsymbol{c}''}),\boldsymbol{1}-\boldsymbol{b}_i\le\boldsymbol{c}',\boldsymbol{c}''\le\boldsymbol{C}\odot(1-\boldsymbol{b}_i)\right\},$$

and $||r||_q = \frac{1}{C_i}\sum_{\boldsymbol{b}_i\le\boldsymbol{c}\le C_i\boldsymbol{b}_i}|r(A_{\boldsymbol{c}})|$. The Cauchy-Schwarz inequality applied repeatedly ensures:

$$\begin{aligned}
||r||_q &\le \frac{1}{C_i}\sum_{\boldsymbol{b}_i\le\boldsymbol{c}\le C_i\boldsymbol{b}_i}\frac{1}{\prod_{s\neq i}C_s^2}\left|\sum_{\boldsymbol{1}-\boldsymbol{b}_i\le\boldsymbol{c}'\le\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{b}_i)}S^f_{\boldsymbol{c}+\boldsymbol{c}'}\right|\times\left|\sum_{\boldsymbol{1}-\boldsymbol{b}_i\le\boldsymbol{c}''\le\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{b}_i)}S^g_{\boldsymbol{c}+\boldsymbol{c}''}\right|\\
&\le\left(\frac{1}{C_i}\sum_{\boldsymbol{b}_i\le\boldsymbol{c}\le C_i\boldsymbol{b}_i}\frac{1}{\prod_{s\neq i}C_s^2}\left|\sum_{\boldsymbol{1}-\boldsymbol{b}_i\le\boldsymbol{c}'\le\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{b}_i)}S^f_{\boldsymbol{c}+\boldsymbol{c}'}\right|^2\right)^{1/2}\\
&\times\left(\frac{1}{C_i}\sum_{\boldsymbol{b}_i\le\boldsymbol{c}\le C_i\boldsymbol{b}_i}\frac{1}{\prod_{s\neq i}C_s^2}\left|\sum_{\boldsymbol{1}-\boldsymbol{b}_i\le\boldsymbol{c}'\le\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{b}_i)}S^g_{\boldsymbol{c}+\boldsymbol{c}'}\right|^2\right)^{1/2}\\
&\le\left(\frac{1}{C_i}\sum_{\boldsymbol{b}_i\le\boldsymbol{c}\le C_i\boldsymbol{b}_i}\frac{1}{\prod_{s\neq i}C_s}\sum_{\boldsymbol{1}-\boldsymbol{b}_i\le\boldsymbol{c}'\le\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{b}_i)}\left(S^f_{\boldsymbol{c}+\boldsymbol{c}'}\right)^2\right)^{1/2}\\
&\times\left(\frac{1}{C_i}\sum_{\boldsymbol{b}_i\le\boldsymbol{c}\le C_i\boldsymbol{b}_i}\frac{1}{\prod_{s\neq i}C_s}\sum_{\boldsymbol{1}-\boldsymbol{b}_i\le\boldsymbol{c}'\le\boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{b}_i)}\left(S^g_{\boldsymbol{c}+\boldsymbol{c}'}\right)^2\right)^{1/2}\\
&=||(\pi\circ f)||_{\mu_{\boldsymbol{C}},2}\times||(\pi\circ g)||_{\mu_{\boldsymbol{C}},2}.
\end{aligned}$$

It follows that for any $\eta > 0$, we have

$$N(\eta,\mathcal{H},||.||_q)\le N(\eta^{1/2},\pi\circ\mathcal{F},||.||_{\mu_{\boldsymbol{C}},2})\times N(\eta^{1/2},\pi\circ\mathcal{G},||.||_{\mu_{\boldsymbol{C}},2}). \qquad (38)$$

We are now lead back to a case study, depending if $\mathcal{F}$ and $\mathcal{G}$ fulfill Assumptions 3 or 3'. We will only treat the case where $\mathcal{F}$ fulfills Assumption 3 and $\mathcal{G}$ fulfills Assumption 3', other cases can be treated similarly up to a simple adaptation. If $\overline{N}_2 = 0$ (respectively $A_2 = 0$) then $N(\eta^{1/2}, \pi \circ \mathcal{F}, ||.||_{\mu_{\boldsymbol{C}},2}) = 1$ (respectively $N(\eta^{1/2}, \pi \circ \mathcal{G}, ||.||_{\mu_{\boldsymbol{C}},2}) = 1$). Otherwise, Lemma C.6 i) combined with Assumption 3 and Lemma C.6 ii) combined with Assumption 3 ensure respectively:

$$
\begin{aligned}
\sqrt{\log N\left(\eta^{1/2}, \pi \circ \mathcal{F}, ||.||_{\mu_{\boldsymbol{C}},2}\right)} &\leq \sqrt{\log N\left(\frac{\eta^{1/2}}{\sqrt{\overline{N}_2}}, \mathcal{F}, ||.||_{\mathbb{Q}^2_{\boldsymbol{C}},2}\right)} \\
&\leq \frac{\overline{N}_2^{1/2}}{\eta^{1/2}} \int_0^{\frac{\eta^{1/2}}{\overline{N}_2^{1/2}}} \sqrt{\log N(u, \mathcal{F}, ||.||_{\mathbb{Q}^2_{\boldsymbol{C}},2})} du \\
&\leq \frac{\overline{N}_2^{1/2} ||F||_{\mathbb{Q}^2_{\boldsymbol{C}},2}}{\eta^{1/2}} J_{2,\mathcal{F}}(\infty), \qquad (39)
\end{aligned}
$$

$$
\begin{aligned}
\sqrt{\log N\left(\eta^{1/2}, \pi \circ \mathcal{G}, ||.||_{\mu_{\boldsymbol{C}},2}\right)} &\leq \sqrt{\log N\left(\frac{\eta^{1/2}}{A_2^{1/2}}, \mathcal{G}, ||.||_{\mu_{\boldsymbol{C}},2}\right)} \\
&\leq \frac{A_2^{1/2} ||G||_{\infty,\beta}}{\eta^{1/2}} J_{\infty,\beta}(\infty). \qquad (40)
\end{aligned}
$$

The combination of Inequalities (36), (37) , (38), (39) (40) and the inequality $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ ensure that for any $\boldsymbol{e}$ we have:

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}} \left| \frac{1}{C_i} \sum_{\boldsymbol{b}_i \leq \boldsymbol{c} \leq C_i \boldsymbol{b}_i} \epsilon_{\boldsymbol{c},\boldsymbol{e}} \frac{1}{\prod_{s\neq i} C_s^2} \sum_{\boldsymbol{1}-\boldsymbol{b}_i \leq \boldsymbol{c}', \boldsymbol{c}'' \leq \boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{b}_i)} S^f_{\boldsymbol{c}+\boldsymbol{c}'} S^g_{\boldsymbol{c}+\boldsymbol{c}''} \right|\right] \\
&\leq \mathbb{E}\left[\frac{1}{C_i} \sum_{\boldsymbol{b}_i \leq \boldsymbol{c} \leq C_i \boldsymbol{b}_i} \frac{1}{\prod_{s\neq i} C_s^2} \sum_{\boldsymbol{1}-\boldsymbol{b}_i \leq \boldsymbol{c}', \boldsymbol{c}'' \leq \boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{b}_i)} S^F_{\boldsymbol{c}+\boldsymbol{c}'} S^G_{\boldsymbol{c}+\boldsymbol{c}''} (1 - \mathbb{1}_M(\boldsymbol{c}))\right] + \eta \\
&+ \frac{M^2}{\sqrt{C_i}} \sqrt{2} \left(\sqrt{\log(2)} + \frac{\mathbb{E}\left(\overline{N}_2^{1/2} ||F||_{\mathbb{Q}^2_{\boldsymbol{C}},2}\right)}{\eta^{1/2}} J_{2,\mathcal{F}}(\infty) + \frac{\mathbb{E}(A_2^{1/2}) ||G||_{\infty,\beta}}{\eta^{1/2}} J_{\infty,\beta}(\infty)\right). \qquad (41)
\end{aligned}
$$

Fix M sufficiently large and η sufficiently small to ensure that

$$
\mathbb{E}\left[\frac{1}{C_i} \sum_{\boldsymbol{b}_i \leq \boldsymbol{c} \leq C_i \boldsymbol{b}_i} \frac{1}{\prod_{s\neq i} C_s^2} \sum_{\boldsymbol{1}-\boldsymbol{b}_i \leq \boldsymbol{c}', \boldsymbol{c}'' \leq \boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{b}_i)} S^F_{\boldsymbol{c}+\boldsymbol{c}'} S^G_{\boldsymbol{c}+\boldsymbol{c}''} (1 - \mathbb{1}_M(\boldsymbol{c}))\right] + \eta
$$

is arbitrarily small. Jensen's inequality ensures that $\mathbb{E}\left(\overline{N}_2^{1/2} ||F||_{\mathbb{Q}^2_{\boldsymbol{C}},2}\right) \leq \mathbb{E}\left(\overline{N}_2 ||F||^2_{\mathbb{Q}^2_{\boldsymbol{C}},2}\right)^{1/2} = \mathbb{E}\left(N_{\boldsymbol{1}} \sum_{\ell=1}^{N_{\boldsymbol{1}}} F(Y_{\ell,\boldsymbol{1}})^2\right)^{1/2}$ and $\mathbb{E}\left(A_2^{1/2}\right) \leq \mathbb{E}\left(A_2\right)^{1/2} = \mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} \left(1 + |Y_{\ell,\boldsymbol{1}}|^2\right)^{-\beta/2}\right)^2\right]^{1/2}$ which are finite by assumption. Then we deduce that when $\underline{C} \to \infty$, we have for any $\boldsymbol{e}$:

$$
\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}} \left| \frac{1}{C_i} \sum_{\boldsymbol{b}_i \leq \boldsymbol{c} \leq C_i \boldsymbol{b}_i} \epsilon_{\boldsymbol{c},\boldsymbol{e}} \frac{1}{\prod_{s\neq i} C_s^2} \sum_{\boldsymbol{1}-\boldsymbol{b}_i \leq \boldsymbol{c}', \boldsymbol{c}'' \leq \boldsymbol{C}\odot(\boldsymbol{1}-\boldsymbol{b}_i)} S^f_{\boldsymbol{c}+\boldsymbol{c}'} S^g_{\boldsymbol{c}+\boldsymbol{c}''} \right|\right] = o(1), \qquad (42)
$$

it follows from (34) and (35) that

$$\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{A}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j'})\in\mathcal{A}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j'}}}g(Y_{\ell,\boldsymbol{j'}})-\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\mathbf{2}_i}}g(Y_{\ell,\boldsymbol{j'}})\right]\right|\right]=o(1).$$

Lemma C.10 combined with triangle inequality ensures that:

$$\mathbb{E}\left[\sup_{\mathcal{F}\times\mathcal{G}}\left|\frac{1}{|\mathcal{B}_i|}\sum_{(\boldsymbol{j},\boldsymbol{j'})\in\mathcal{B}_i}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\boldsymbol{j'}}}g(Y_{\ell,\boldsymbol{j'}})-\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\boldsymbol{j}})\sum_{\ell=1}^{N_{\mathbf{2}_i}}g(Y_{\ell,\boldsymbol{j'}})\right]\right|\right]=o(1).$$

C.5 Proof of Lemma C.12

Lemmas B.1, C.3, C.4 together ensure that that for any $M > 0$ and any random $\eta_1 > 0$:

$$\begin{aligned}&\mathbb{E}\left[\sup_{\mathcal{F}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})-\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right]\right|\right]\\ &\leq 2(2^k-1)\mathbb{E}\left[\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}}F(Y_{\ell,\boldsymbol{j}})(1-\mathbb{1}_M(\boldsymbol{j}))\right]\\ &+2(2^k-1)\mathbb{E}\left[\left(\eta_1+\frac{M\sqrt{2}}{\sqrt{\Pi_C}}\left(\sqrt{\log 2}+\sqrt{\log N(\eta_1,\pi\circ\mathcal{F},||.||_{\mu_C,1})}\right)\right)\right],\end{aligned}$$

with $\mathbb{1}_M(\boldsymbol{j})=\mathbb{1}\{\sum_{\ell=1}^{N_{\boldsymbol{j}}}F(Y_{\ell,\boldsymbol{j}})\leq M\}$.

Consider $\eta_1=\eta\overline{N}_1||F||_{Q^1_C,1}=\eta\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}^{N_{\boldsymbol{j}}}F(Y_{\ell,\boldsymbol{j}})$ and use Lemma C.6 i) to deduce that $\sqrt{\log N(\eta_1,\pi\circ\mathcal{F},||.||_{\mu_C,1})}\leq\sqrt{\sup_Q\log N(\eta||F||_Q,\mathcal{F},||.||_Q)}<\infty$. Moreover we have $\mathbb{E}(\eta_1)=\eta\mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}}F(Y_{\ell,\mathbf{1}})\right)<\infty$.

We have :$\mathbb{E}\left[\frac{1}{\Pi_C}\sum_{1\leq\boldsymbol{j}\leq\boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}}F(Y_{\ell,\boldsymbol{j}})(1-\mathbb{1}_M(\boldsymbol{j}))\right]=\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}}F(Y_{\ell,\mathbf{1}})(1-\mathbb{1}_M(\mathbf{1}))\right]$ which converges to 0 when M tends to ∞. So fixing M sufficiently large and η sufficiently small and after that considering that $\underline{C}$ tends to ∞ ensures that:
$\mathbb{E}\left[\sup_{\mathcal{F}}\left|\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\sum_{\ell=1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}})-\mathbb{E}\left[\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right]\right|\right]=o(1)$

C.6 Lemma for the bootstrap

Lemma C.13 *Let* $\left\{N_{\boldsymbol{j}},(Y_{\ell,\boldsymbol{j}})_{\ell\geq 1}\right\}_{\boldsymbol{j}\geq\mathbf{1}}$ *such that there exists a measurable function* τ *such that*

$$\left\{N_{\boldsymbol{j}},(Y_{\ell,\boldsymbol{j}})_{\ell\geq 1}\right\}_{\boldsymbol{j}\geq\mathbf{1}}\overset{a.s.}{=}\left\{\tau\left(\left(U_{\boldsymbol{j}\odot\boldsymbol{e}}\right)_{\mathbf{0}\prec\boldsymbol{e}\preceq\mathbf{1}}\right)\right\}_{\boldsymbol{j}\geq\mathbf{1}},$$

where $(U_{\boldsymbol{j}})_{\boldsymbol{j}>\mathbf{0}}$ *is a family of mutually independent Uniform-*$(0,1)$ *random variables. Let* f *such that* $\mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}}f(Y_{\ell,\mathbf{1}})\right)^2\right]<\infty$. *For any* $\boldsymbol{e}$ *such that* $\mathbf{0}<\boldsymbol{e}<\mathbf{1}$ *and any* $\boldsymbol{c}$ *such that* $\boldsymbol{c}\wedge 1=\boldsymbol{e}$ *let*

$$a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})=\frac{1}{\prod_{s:e_s=0}C_s}\sum_{\mathbf{1}-\boldsymbol{e}\leq\boldsymbol{c'}\leq\boldsymbol{C}\odot(\mathbf{1}-\boldsymbol{e})}\sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c'}}}f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c'}})-\frac{1}{\Pi_C}\sum_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}\sum_{\ell\geq 1}^{N_{\boldsymbol{j}}}f(Y_{\ell,\boldsymbol{j}}),$$

we have:

$$\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2 \xrightarrow{a.s.} L_{\boldsymbol{e}} \leq \mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right)^2\right],$$

$$\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2 \mathbb{1}\left\{|a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})| \geq (\Pi_{i:e_i=1} C_i)^{1/2} \varepsilon\right\} \xrightarrow{a.s.} 0, \textit{ for any } \varepsilon > 0,$$

and for $\boldsymbol{e} \in \mathcal{E}_1$*:*

$$\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2 \xrightarrow{a.s.} \mathbb{C}ov\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}), \sum_{\ell=1}^{N_{\mathbf{2}_i}} f(Y_{\ell,\mathbf{2}_i})\right),$$

Proof:

We have:

$$\begin{aligned}\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2 &= \frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{C} \odot \boldsymbol{e}} \left(\frac{1}{\prod_{s:e_s=0} C_s} \sum_{\mathbf{1}-\boldsymbol{e} \leq \boldsymbol{c}' \leq \boldsymbol{C} \odot (\mathbf{1}-\boldsymbol{e})} \sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}'}} f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}'})\right)^2 \\ &\quad - \left(\frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \sum_{l \geq 1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right)^2 \\ &\leq \frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right)^2\end{aligned}$$

Lemma 7.35 in Kallenberg (2005) ensures:

$$\frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \left(\sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}})\right)^2 \xrightarrow{a.s.} \mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right)^2\right],$$

then

$$\limsup_{\underline{C} \to \infty} \frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2 \leq \mathbb{E}\left[\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right)^2\right], \text{ almost-surely.}$$

Let

$$b_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c}) = a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c}) + \mathbb{P}_C(f) = \frac{1}{\prod_{s:e_s=0} C_s} \sum_{\mathbf{1}-\boldsymbol{e} \leq \boldsymbol{c}' \leq \boldsymbol{C} \odot (\mathbf{1}-\boldsymbol{e})} \sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}'}} f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}'}).$$

Lemma 7.35 in Kallenberg (2005) ensures:

$$\frac{1}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}} \sum_{\ell=1}^{N_{\boldsymbol{j}}} f(Y_{\ell,\boldsymbol{j}}) \xrightarrow{a.s.} \mathbb{E}\left(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}})\right),$$

so the almost-sure limit $L_{\boldsymbol{e}}$ exists if and only if

$$A = \frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \le \boldsymbol{c} \le \boldsymbol{C} \odot \boldsymbol{e}} \left(b_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2,$$

admits an almost-sure limit. We can rewrite A as:

$$\frac{1}{\prod_{i:e_i=1} C_i} \frac{1}{\prod_{s:e_s=0} C_s^2} \sum_{\boldsymbol{e} \le \boldsymbol{c} \le \boldsymbol{C} \odot \boldsymbol{e}} \sum_{\boldsymbol{1}-\boldsymbol{e} \le \boldsymbol{c}' \le \boldsymbol{C} \odot (\boldsymbol{1}-\boldsymbol{e})} \sum_{\boldsymbol{1}-\boldsymbol{e} \le \boldsymbol{c}'' \le \boldsymbol{C} \odot (\boldsymbol{1}-\boldsymbol{e})} \sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}'}} f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}'}) \sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}''}} f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}''}).$$

A k dimensional jointly exchangeable array, is an array such that Condition 1 in Assumption 1 holds for $\pi_1 = \pi_2 = ... = \pi_k$.

Note that

$$\left\{ \sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}'}} f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}'}) \times \sum_{\ell=1}^{N_{\boldsymbol{c}+\boldsymbol{c}''}} f(Y_{\ell,\boldsymbol{c}+\boldsymbol{c}''}) \right\}_{\boldsymbol{c},\boldsymbol{c}',\boldsymbol{c}'':\boldsymbol{c}\wedge\boldsymbol{1}=\boldsymbol{e},\boldsymbol{c}'\wedge\boldsymbol{1}=\boldsymbol{c}''\wedge\boldsymbol{1}=\boldsymbol{1}-\boldsymbol{e}}$$

is a jointly exchangeable array indexed by non null components of $\boldsymbol{c}, \boldsymbol{c}', \boldsymbol{c}''$ of dimension $l = 2\sum_{i=1}^k (1-e_i) + \sum_{i=1}^k e_i = 2k - \sum_{i=1}^k e_i$. Moreover this array is dissociated (cf. Kallenberg (2005), page 339 or Aldous (1985), page 125). Lemma 7.35 in Kallenberg (2005) ensures that this array is ergodic, then an almost-sure limit of A exists.

We have:

$$\begin{aligned}
&\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \le \boldsymbol{c} \le \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2 \mathbb{1}\left\{ |a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})| \ge \left(\prod_{i:e_i=1} \sqrt{C_i} \right) \varepsilon \right\} \\
&\le \frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \le \boldsymbol{c} \le \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2 \mathbb{1}\left\{ |b_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})| \ge \left(\prod_{i:e_i=1} \sqrt{C_i} \right) \varepsilon/2 \right\} \\
&+ \mathbb{1}\left\{ |\mathbb{P}_C(f)| \ge \left(\prod_{i:e_i=1} \sqrt{C_i} \right) \varepsilon/2 \right\} \frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \le \boldsymbol{c} \le \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2.
\end{aligned}$$

Because $\mathbb{P}_C(f) \xrightarrow{a.s.} \mathbb{E}\left[\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}})\right]$ and $\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \le \boldsymbol{c} \le \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2 \xrightarrow{a.s.} L_{\boldsymbol{e}}$, it is sufficient to show that:

$$\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \le \boldsymbol{c} \le \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})\right)^2 \mathbb{1}\left\{ |b_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})| \ge \left(\prod_{i:e_i=1} \sqrt{C_i} \right) \varepsilon/2 \right\} \xrightarrow{a.s.} 0.$$

We have:

$$\frac{1}{C_{i_{\boldsymbol{e}}}} \sum_{\boldsymbol{e} \le \boldsymbol{c} \le \boldsymbol{C} \odot \boldsymbol{e}} |b_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})|^2 \xrightarrow{a.s.} L_{\boldsymbol{e}} - \left[\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}) \right) \right]^2.$$

Let S^+ denotes the set of strictly increasing function from $\mathbb{N}$ to $\mathbb{N}$. There exists $\Omega' \subset \Omega$ of measure one such that $\forall \omega \in \Omega'$ and $\forall \boldsymbol{s} \in (S^+)^k$ when $\underline{C} \to \infty$:

$$\begin{aligned}
&\frac{1}{\prod_{i:e_i=1} \boldsymbol{s}_i(\underline{C})} \sum_{\boldsymbol{e} \le \boldsymbol{c} \le \boldsymbol{s}(\underline{C}) \odot \boldsymbol{e}} |b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|^2 \to L_{\boldsymbol{e}} - \left[\mathbb{E}\left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}) \right) \right]^2 \\
&\frac{1}{\prod_{i:e_i=1} \boldsymbol{s}_i(\underline{C})} \sum_{\boldsymbol{e} \le \boldsymbol{c} \le \boldsymbol{s}(\underline{C}) \odot \boldsymbol{e}} |a_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|^2 \to L_{\boldsymbol{e}}
\end{aligned}$$

For $\omega \in \Omega'$, $\boldsymbol{s} \in (S^+)^k$, for $\underline{C} > \underline{C}^*$, we have:

$$\begin{aligned}\sum_{(\boldsymbol{s}(\underline{C})-\mathbf{1})\odot\boldsymbol{e}<\boldsymbol{c}\leq\boldsymbol{s}(\underline{C})\odot\boldsymbol{e}} \frac{|b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|^2}{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})} &= \frac{1}{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{s}(\underline{C})\odot\boldsymbol{e}} |b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|^2 \\ &\quad - \frac{\prod_{i:e_i=1}(\boldsymbol{s}_i(\underline{C})-1)}{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})}\frac{1}{\prod_{i:e_i=1}(\boldsymbol{s}_i(\underline{C}^*)-1)}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq(\boldsymbol{s}(\underline{C}^*)-\mathbf{1})\odot\boldsymbol{e}} |b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|^2 \\ &= \frac{1}{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{s}(\underline{C})\odot\boldsymbol{e}} |b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|^2 \\ &\quad - \frac{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})}{\prod_{i:e_i=1}(\boldsymbol{s}_i(\underline{C})-1)^2}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq(\boldsymbol{s}(\underline{C})-\mathbf{1})\odot\boldsymbol{e}} |b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})-\mathbf{1}}(\boldsymbol{c})(\omega)|^2\end{aligned}$$

.

Because $\frac{1}{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{s}(\underline{C})\odot\boldsymbol{e}} |b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|^2$ and $\frac{1}{\prod_{i:e_i=1}(\boldsymbol{s}_i(\underline{C})-1)}\sum_{\boldsymbol{e}\leq\boldsymbol{c}\leq(\boldsymbol{s}(\underline{C})-\mathbf{1})\odot\boldsymbol{e}} |b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})-\mathbf{1}}(\boldsymbol{c})(\omega)|^2$ have the same limit $L_{\boldsymbol{e}} - \mathbb{E}(\sum_{\ell=1}^{N_{\mathbf{1}}} f(Y_{\ell,\mathbf{1}}))^2$ when $\underline{C}$ tends to ∞, we deduce that:

$$\sum_{(\boldsymbol{s}(\underline{C})-\mathbf{1})\odot\boldsymbol{e}<\boldsymbol{c}\leq\boldsymbol{s}(\underline{C})\odot\boldsymbol{e}} \frac{|b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|^2}{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})} \to 0.$$

In particular, when $\underline{C}$ tend to ∞:

$$\max_{(\boldsymbol{s}(\underline{C})-\mathbf{1})\odot\boldsymbol{e}<\boldsymbol{c}\leq\boldsymbol{s}(\underline{C})\odot\boldsymbol{e}} \frac{|b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|^2}{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})} \to 0.$$

So when $\underline{C}^*$ tends to ∞:

$$\max_{\underline{C}\geq\underline{C}^*}\max_{(\boldsymbol{s}(\underline{C})-\mathbf{1})\odot\boldsymbol{e}<\boldsymbol{c}\leq\boldsymbol{s}(\underline{C})\odot\boldsymbol{e}} \frac{|b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|^2}{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})} = \max_{(\boldsymbol{s}(\underline{C}^*)-\mathbf{1})\odot\boldsymbol{e}<\boldsymbol{c};\boldsymbol{c}\wedge\mathbf{1}=\boldsymbol{e}} \frac{|b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|^2}{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})} \to 0.$$

So for any $\eta > 0$ there exists a sufficiently large $\underline{C}^*$ such that for $\underline{C} > \underline{C}^*$, we have $\max_{(\boldsymbol{s}(\underline{C}^*)-\mathbf{1})\odot\boldsymbol{e}<\boldsymbol{c}\leq\boldsymbol{s}(\underline{C})\odot\boldsymbol{e}} \frac{|b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|^2}{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})} < \eta$. Let $m(\underline{C}) = \max_{\boldsymbol{e}\leq\boldsymbol{c}\leq\boldsymbol{s}(\underline{C})\odot\boldsymbol{e}} |b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)|$. We have:

$$\begin{aligned}m(\underline{C}) &= m(\underline{C}^*) \vee \max_{\boldsymbol{s}(\underline{C}^*)\odot\boldsymbol{e}<\boldsymbol{c}\leq\boldsymbol{s}(\underline{C})\odot\boldsymbol{e}} |b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})(\omega)| \\ &\leq m(\underline{C}^*) \vee \sqrt{\eta\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})}.\end{aligned}$$

For $\underline{C}$ sufficiently large, we have $\frac{m(\underline{C}^*)}{\sqrt{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})}} < \sqrt{\eta}$ and next,

$$\frac{m(\underline{C})}{\sqrt{\prod_{i:e_i=1}\boldsymbol{s}_i(\underline{C})}} < \sqrt{\eta}.$$

Because η is arbitrary, we can choose it such that $\sqrt{\eta} < \varepsilon/2$, and next:

$$\frac{1}{\prod_{i:e_i=1} \boldsymbol{s}_i(\underline{C})} \sum_{\boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{s}(\underline{C}) \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c}) \right)^2 (\omega) \mathbb{1} \left\{ |b_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c})|(\omega) \geq \sqrt{\prod_{i:e_i=1} \boldsymbol{s}_i(\underline{C})} \varepsilon/2 \right\}$$
$$\leq \mathbb{1} \left\{ m(\underline{C}) \geq \sqrt{\prod_{i:e_i=1} \boldsymbol{s}_i(\underline{C})} \varepsilon/2 \right\} \frac{1}{\prod_{i:e_i=1} \boldsymbol{s}_i(\underline{C})} \sum_{\boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{s}(\underline{C}) \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{s}(\underline{C})}(\boldsymbol{c}) \right)^2 (\omega)$$

and the right hand side tends to 0 when $\underline{C} \to \infty$ because $\mathbb{1} \left\{ m(\underline{C}) \geq \sqrt{\prod_{i:e_i=1} \boldsymbol{s}_i(\underline{C})} \varepsilon/2 \right\} = 0$ for $\underline{C}$ sufficiently large. This is true for any $\omega \in \Omega'$ and any $\boldsymbol{s} \in (S^+)^k$. This means that

$$\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c}) \right)^2 \mathbb{1} \left\{ |b_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c})| \geq \left(\prod_{i:e_i=1} \sqrt{C_i} \right) \varepsilon/2 \right\} \xrightarrow{a.s.} 0.$$

Last, if $\boldsymbol{e} \in \mathcal{E}_1$, Lemma C.10 (with functional classes reduced to a single element) ensures that, $\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c}) \right)^2$ converges in probability to $\mathbb{C}ov \left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}), \sum_{\ell=1}^{N_{\boldsymbol{2}_i}} f(Y_{\ell,\boldsymbol{2}_i}) \right)$. Because the convergence almost-sure implies the convergence in probability, we have:

$$\frac{1}{\prod_{i:e_i=1} C_i} \sum_{\boldsymbol{e} \leq \boldsymbol{c} \leq \boldsymbol{C} \odot \boldsymbol{e}} \left(a_{\boldsymbol{e}}^{\boldsymbol{C}}(\boldsymbol{c}) \right)^2 \xrightarrow{a.s.} \mathbb{C}ov \left(\sum_{\ell=1}^{N_{\boldsymbol{1}}} f(Y_{\ell,\boldsymbol{1}}), \sum_{\ell=1}^{N_{\boldsymbol{2}_i}} f(Y_{\ell,\boldsymbol{2}_i}) \right).$$