Electronic Journal of Statistics
Vol. 13 (2019) 120–165
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1526

Auxiliary information: the raking-ratio empirical process

Mickael Albertus and Philippe Berthet

Institut de Mathématiques de Toulouse ; UMR5219
Université de Toulouse ; CNRS
UPS IMT, F-31062 Toulouse Cedex 9, France
e-mail: mickael.albertus@math.univ-toulouse.fr
e-mail: philippe.berthet@math.univ-toulouse.fr

Abstract: We study the empirical measure associated to a sample of size n and modified by N iterations of the raking-ratio method. This empirical measure is adjusted to match the true probability of sets in a finite partition which changes each step. We establish asymptotic properties of the raking-ratio empirical process indexed by functions as $n \to +\infty$, for N fixed. We study nonasymptotic properties by using a Gaussian approximation which yields uniform Berry-Esseen type bounds depending on n, N and provides estimates of the uniform quadratic risk reduction. A closed-form expression of the limiting covariance matrices is derived as $N \to +\infty$. In the two-way contingency table case the limiting process has a simple explicit formula.

MSC 2010 subject classifications: Primary 62G30, 62G20; secondary 60F05, 60F17.
Keywords and phrases: Raking-ratio method, empirical processes, strong approximation, nonparametric statistics, auxiliary information, Sinkhorn algorithm.

Received March 2018.

Contents

1. Introduction

1.1. The raking-ratio method

In survey analysis, statistics, economics and computer sciences the raking-ratio iterative procedure aims to exploit the knowledge of one or several marginals of a discrete multivariate distribution to fit the data after sampling. Despite many papers from the methodological and algorithmic viewpoint, and chapters in classical textbooks for statisticians, economists or engineers, no probabilistic study is available to take into account that the entries of the algorithm are random and the initial discrete measure is empirical. We intend to fill this gap. Let us first describe the algorithm, usually considered with deterministic entries, then recall the few known results and state the open question to be addressed.

The raking-ratio algorithm. A sample is drawn from a population P for which $k \geqslant 2$ marginal finite discrete distributions are explicitly known. Initially, each data point has a weight $1/n$. The *ratio* step of the algorithm consists in computing new weights in such a way that the modified empirical joint distribution has the currently desired marginal. The *raking* step consists in iterating the correction according to another known marginal law, changing again all the weights. The k margin constraints are usually treated in a periodic order, only one being fulfilled at the same time. The raking-ratio method stops after N iterations with the implicit hope that the previous constraints are still almost satisfied. See Section A.1 for an elementary numerical example with $k = 2$ and Section 1.4 for notation and mathematical definition of the algorithm.

The limit. This algorithm was called *iterative proportions* by Deming and Stephan [10] who first introduced it. They showed that the k margins converge to the desired ones as $N \to +\infty$. They even claimed that if the frequencies of a multiway contingency table are raked periodically as $N \to +\infty$ they converge to the frequencies minimizing the chi-square distance to the initial frequencies, under the k margin constraints. Two years later, Stephan [25] observed that it is wrong and modified the algorithm accordingly to achieve the chi-square distance minimization. Lewis [18] and Brown [7] studied the case of Bernoulli marginals from the Shannon entropy minimization viewpoint. When $k = 2$ a two-way contingency table can be viewed as a matrix. Sinkhorn [22, 23] proved that a unique doubly stochastic matrix can be obtained from each positive square matrix by alternately normalizing its rows and its columns, which shows that the algorithm converges in this special case. Finally Ireland and Kullback [16] generalized previous arguments to rigorously justify that the raking-ratio converges to the unique projection of the empirical measure in Shannon-Kullback-Leibler relative entropy on the set of discrete distributions satisfying the k margin constraints. From a numerical viewpoint, the rate of convergence of the algorithm is geometric, see Franklin and Lorentz [15].

Remark A. When minimizing contrasts such as discrimination information, chi-square distance or likelihood, the minimizers are not explicit due to the non-linearity of sums of ratios showing up in derivatives. This is why converging algorithms are used in practice. In the case of the iterative proportions algorithm each step is easily and fastly computed. What has been studied concerns the convergence when the iterations $N \to +\infty$, with n fixed and initial empirical frequencies treated as deterministic entries. When the sample size $n \to +\infty$, these entries are close to P itself, which satisfies the marginal constraints, hence one expects that the number N of iterations necessary to converge is small. We shall study the N_0 first iterations in the statistical setting $n \to +\infty$.

Non explicit bias and variance. The initial values being empirical frequencies the converged solution of the algorithm as $N \to +\infty$ is a joint distributions fulfilling the marginal requirements that still deviates from the true population distribution P, and moreover in a rather complicated way. The modified empirical distribution satisfying only the marginal constraint of the current iteration, there is a permanent bias with respect to other margins, and hence with P. The exact covariance matrix and bias vector of the random weights after N iterations are tedious to compute. For instance, estimates for the variance of cell probabilities in the case of a two-way contingency table are given by Brackstone and Rao [6] for $N \leqslant 4$, Konijn [17] or Choudhry and Lee [8] for $N = 2$. Bankier [2] proposed a recursive linearization technique providing an estimator of the asymptotic variance of weights. In Binder and Théberge [4] the variance of the converged solution requires to calculate weights at each iteration.

Open question. Since exact computations lead to intractable formulas for the bias and variance of frequencies and statistics as simple as means, an important open problem is to identify leading terms when n is large compared to N. We derive comprehensive explicit formulas as $n \to +\infty$ for $N \leqslant N_0$ and N_0 fixed, then for $N \to +\infty$. In order to further analyze the raking ratio method it is moreover desirable to control simultaneously large classes of statistics and hence to work at the empirical measure level rather than with the empirical weights or a single statistic only. This is the main motivation for the forthcoming general study of empirical measures indexed by functions and modified through auxiliary information given by partitions.

1.2. Statistical motivation

Representative sample. In a survey analysis context, the raking-ratio method modifies weights of a contingency table built from a sample of size n in order to fit exactly given marginals. Such a strict margin correction is justified when a few properties of the finite population under study are known, like the size of subpopulations. The modified sample frequencies then reflect the marginal structure of the whole population. If the population is large or infinite the information may come from previous and independent statistical inference, from structural properties of the model or from various experts.

Remark B. Making the sample representative of the population is an ad hoc approach based on common sense. The mathematical impact is twofold. On the one hand all statistics are affected by the new weights in terms of bias, variance and limit law so that practitioners may very well be using estimators, tests or confident bands that have lost their usual properties. On the other hand, replacing marginal frequencies with the true ones may smooth sample fluctuations of statistics correlated to them while leaving the uncorrelated ones rather unaffected. These statements will be quantified precisely at Section 2.2.

Remark C. Fitting after sampling is a natural method that has been re-invented many times in various fields, and was probably used long time ago. Depending on the setting it may be viewed as stratification, calibration, iterating proportional fitting, matrix scaling and could be used to deal with missing data. Many fitting methods may be reduced to a raking-ratio type algorithm. We initially called it auxiliary information of partitions as we re-invented it as a special case of the nonparametric partial information problem stated in Section 1.3.

Remark D. An asymptotic approach is no more relevant in survey analysis when the underlying population is rather small. In the small population case, the way the sample is drawn has a so deep impact that it may even become the main topic. A study of calibration methods for finite population can be found in Deville and Särndal [11, 12]. This is beyond the scope of our work.

Quadratic risk reduction. Modifying marginals frequencies of a sample may induce serious drawbacks. One should ask whether or not the estimation risk can be controlled. Typically, a statistic has more bias when sample weights are changed by using raking, calibration or stratification methods after sampling. In the spirit of Remark B, a variance reduction is expected if the statistic of interest is strongly correlated to the k known discrete marginal variables. Now, evaluating the quadratic risk of a specific statistic requires tedious expansions for the bias, variance and correlations of weights, whence the very small N studied in the literature. Likewise, no global risk reduction property has been established as $n \to +\infty$ and no multivariate or uniform central limit theorem. These results are established at Propositions 4 to 9.

Contributions. In this paper we consider classes of empirical means raked N times, sampled jointly from any population. We derive closed-form expressions of their Gaussian limits and their limiting covariances as $n \to +\infty$ then $N \to +\infty$. We also quantify the uniform risk reduction phenomenon and provide sharp statistical estimation tools such as uniform Berry-Esseen type bounds. In particular, a Donsker invariance principle for the raked empirical process provides joint limiting laws for additive statistics built from empirical means, and this can be extended to non linear estimators by applying the delta method, argmax theorems or plug-in approaches as in the classical setting – see [27, 28].

Organization of the paper. In Section 1.3 we relate the raking-ratio problem to nonparametric auxiliary information. The raking-ratio empirical process $\alpha_n^{(N)}(\mathcal{F})$ is defined in Section 1.4. Usual assumptions on an indexing class $\mathcal{F}$ of functions are given in Section 2.1. In Section 2.2 we state our results for $\alpha_n^{(N)}(\mathcal{F})$ when the number N of iterations is fixed. Our main theorem is a nonasymptotic strong approximation bound which yields the uniform central limit theorem with rate as $n \to +\infty$, as well as an uniform control of the bias and the covariances for fixed n. The approximating Gaussian process is studied in Section 2.3, which establishes the uniform risk reduction phenomenon provided the iterations are stopped properly. In Section 2.4, in the two partitions case we characterize explicitly the limiting process as $N \to +\infty$. All statements are proved in Sections 3 and 4. The Appendix provides a few examples.

1.3. An auxiliary information viewpoint

Let $X_1, ..., X_n$ be independent random variables with unknown law P on some measurable space $(\mathcal{X}, \mathcal{A})$. Assumptions like separability or Haussdorf property are not necessary for this space. Let δ_x denote the Dirac mass at $x \in \mathcal{X}$ and consider the empirical measure $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ on $\mathcal{A}$.

Auxiliary information. Our interest for the raking-ratio method came while investigating how to exploit various kinds of partial information on P to make $\mathbb{P}_n$ closer to P. The auxiliary information paradigm is as follows. Usually what is assumed on P is formulated in terms of technical or regularity requirements. Sometimes it is relevant to assume that P satisfies simple properties that could be tested or estimated separately. Consider the following two extreme situations. First, a parametric model provides a tremendous amount of information by specifying $P = P_\theta$ up to a finite dimensional parameter θ, so that $\mathbb{P}_n$ can be replaced with the most likely $P_{\theta_n(X_1,...,X_n)}$ among the model. Notice that $\mathbb{P}_n$ is used to minimize the empirical likelihood, but the resulting $P_{\theta_n(X_1,...,X_n)}$ is of a very different nature, far from the initial and discrete $\mathbb{P}_n$ thanks to the valuable parametric information. On the opposite, in a purely nonparametric setting the information mainly comes from the sample itself, so that only slight modifications of $\mathbb{P}_n$ are induced by weak hypotheses on P – like support, regularity, symmetry, logconcavity, Bayesian model, semi-parametric model, etc. In between, we would like to formalize a notion of a priori auxiliary information on P based on partial but concrete clues to be combined with the knowledge of $\mathbb{P}_n$. Such clues may come from experts, models, former inference, statistical learning or distributed data. A generic situation one can start with is when the probabilities $P(A_j)$ of a finite number of sets $A_j \in \mathcal{A}$ are known – which in a parametric setting already determines θ then P.

Information from one partition. If the A_j form a finite partition of $\mathcal{X}$ then the auxiliary information coincides with one discrete marginal distribution and a natural nonparametric redesign $\mathbb{P}_n^{(1)}$ of $\mathbb{P}_n$ is the following. Let $A_1^{(1)}, \ldots, A_{m_1}^{(1)} \subset \mathcal{A}$ be a partition of $\mathcal{X}$ such that $P(\mathcal{A}^{(1)}) = (P(A_1^{(1)}), \ldots, P(A_{m_1}^{(1)}))$ is known. According to Proposition 1 below, the random measure

$$\mathbb{P}_n^{(1)} = \frac{1}{n}\sum_{j=1}^{m_1} \frac{P(A_j^{(1)})}{\mathbb{P}_n(A_j^{(1)})} \sum_{X_i \in A_j^{(1)}} \delta_{X_i}, \tag{1.1}$$

satisfies the auxiliary information $\mathbb{P}_n^{(1)}(A_j^{(1)}) = P(A_j^{(1)})$, for $1 \leqslant j \leqslant m_1$, and is the relative entropy projection of $\mathbb{P}_n$ on these m_1 constraints. The random ratios in (1.1) induce a bias between $\mathbb{P}_n^{(1)}$ and P. We prove that the bias of $\alpha_n^{(1)} = \sqrt{n}(\mathbb{P}_n^{(1)} - P)$ vanishes uniformly and that the limiting Gaussian process of $\alpha_n^{(1)}$ has a smaller variance than the P-Brownian bridge.

Extension to N partitions. If some among the sets A_j are overlapping then the information comes from several marginal partitions. It is not obvious how to optimally combine these sources of information since there is no explicit modification of $\mathbb{P}_n$ matching simultaneously several finite discrete marginals. In other words there is no closed form expression of the relative entropy projection of $\mathbb{P}_n$ on several margin constraints. An alternative consists in recursively updating the current modification $\mathbb{P}_n^{(N-1)}$ of $\mathbb{P}_n$ onto $\mathbb{P}_n^{(N)}$ according to the next known marginal $P(\mathcal{A}^{(N)}) = (P(A_1^{(N)}), \ldots, P(A_{m_N}^{(N)}))$ exactly as in (1.1) for $\mathbb{P}_n^{(1)}$ from $\mathbb{P}_n^{(0)} = \mathbb{P}_n$ and $P(\mathcal{A}^{(1)})$. This coincides with the Deming and Stephan's iterative procedure, that is the raking-ratio algorithm, as formalized in Section 1.4.

1.4. Information from N finite partitions

The raking-ratio empirical measure. For all $N \in \mathbb{N}_*$ let $m_N \geqslant 2$ and $\mathcal{A}^{(N)} = \{A_1^{(N)}, \ldots, A_{m_N}^{(N)}\} \subset \mathcal{A}$ be a partition of $\mathcal{X}$ for which we are given the auxiliary information $P(\mathcal{A}^{(N)}) = (P(A_1^{(N)}), \ldots, P(A_{m_N}^{(N)}))$ to be exploited. Assume that

$$p_N = \min_{1 \leqslant j \leqslant m_N} P(A_j^{(N)}) > 0, \quad N \in \mathbb{N}_*, \tag{1.2}$$

and $\mathcal{A}^{(N_1)} \neq \mathcal{A}^{(N_2)}$ if $|N_1 - N_2| = 1$, otherwise $\mathcal{A}^{(N_1)} = \mathcal{A}^{(N_2)}$ is allowed. For $N = 0$ there is no information and $m_0 = 1$, $\mathcal{A}^{(0)} = \{\mathcal{X}\}$, $P(\mathcal{A}^{(0)}) = \{1\}$, $p_0 = 1$. For any measurable real function f write $\mathbb{P}_n^{(0)}(f) = \mathbb{P}_n(f) = n^{-1}\sum_{i=1}^n f(X_i)$, $P(f) = \int_{\mathcal{X}} f dP$ and $\alpha_n^{(0)}(f) = \sqrt{n}(\mathbb{P}_n^{(0)}(f) - P(f))$. In (1.1) $\mathbb{P}_n^{(1)}$ allocates the random weight $P(A_j^{(1)})/n\mathbb{P}_n(A_j^{(1)})$ to each $X_i \in A_j^{(1)}$. Hence

$$\begin{aligned}\mathbb{P}_n^{(1)}(f) &= \sum_{i=1}^{n} \mathbb{P}_n^{(1)}(\{X_i\})f(X_i) = \sum_{j=1}^{m_1} \sum_{X_i \in A_j^{(1)}} \frac{P(A_j^{(1)})}{n\mathbb{P}_n(A_j^{(1)})} f(X_i) \\ &= \sum_{j=1}^{m_1} \frac{P(A_j^{(1)})}{\mathbb{P}_n(A_j^{(1)})} \left(\frac{1}{n} \sum_{X_i \in A_j^{(1)}} f(X_i) \right) = \sum_{j=1}^{m_1} \frac{P(A_j^{(1)})}{\mathbb{P}_n^{(0)}(A_j^{(1)})} \mathbb{P}_n^{(0)}(f 1_{A_j^{(1)}}).\end{aligned}$$

Let define recursively, for $N \in \mathbb{N}_*$, the N-th raking-ratio empirical measure

$$\mathbb{P}_n^{(N)}(f) = \sum_{j=1}^{m_N} \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} \mathbb{P}_n^{(N-1)}(f 1_{A_j^{(N)}}), \tag{1.3}$$

and the N-th raking-ratio empirical process

$$\alpha_n^{(N)}(f) = \sqrt{n}(\mathbb{P}_n^{(N)}(f) - P(f)). \tag{1.4}$$

For $A \in \mathcal{A}$ we also write $\alpha_n^{(N)}(A) = \alpha_n^{(N)}(1_A)$. By (1.3) and (1.4) we have for all $N \in \mathbb{N}_*$

$$\mathbb{P}_n^{(N)}(A_j^{(N)}) = P(A_j^{(N)}), \quad \alpha_n^{(N)}(A_j^{(N)}) = 0, \quad 1 \leqslant j \leqslant m_N, \tag{1.5}$$

as desired. Both weights and support $\{X_1, \ldots, X_n\}$ of the discrete probability measure $\mathbb{P}_n^{(N)}$ are random since (1.3) also reads

$$\mathbb{P}_n^{(N)}(\{X_i\}) = \mathbb{P}_n^{(N-1)}(\{X_i\}) \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})}, \quad \text{for } X_i \in A_j^{(N)}.$$

A few more formulas concerning $\alpha_n^{(N)}$ and $\mathbb{P}_n^{(N)}$ are derived in Section 3.1.

Iterated Kullback projections. The random discrete measures $\mathbb{P}_n^{(1)}, ..., \mathbb{P}_n^{(N)}$ are well defined provided that

$$\min_{1 \leqslant k \leqslant N} \min_{1 \leqslant j \leqslant m_k} \mathbb{P}_n(A_j^{(k)}) > 0, \tag{1.6}$$

which almost surely holds for all n large enough and N fixed, by (1.2) and the law of large numbers. Given two probability measures Q_n and Q supported by $\{X_1, ..., X_n\}$ we define the relative entropy of Q_n and Q – see for instance [9] – to be

$$d_K(Q_n \,||\, Q) = \sum_{i=1}^{n} Q_n(\{X_i\}) \log\left(\frac{Q_n(\{X_i\})}{Q(\{X_i\})} \right).$$

Proposition 1. *If (1.6) holds then*

$$\mathbb{P}_n^{(N)} = \arg\min \Big\{ d_K(\mathbb{P}_n^{(N-1)} \,||\, Q) : Q(\mathcal{A}^{(N)}) = P(\mathcal{A}^{(N)}), \\ \mathit{supp}(Q) = \{X_1, ..., X_n\} \Big\}.$$

As a consequence, the formula (1.3) means that the N-th iteration $\mathbb{P}_n^{(N)}$ is the Shannon-Kullback-Leibler projection of $\mathbb{P}_n^{(N-1)}$ under the constraint $P(\mathcal{A}^{(N)})$. Therefore the raking-ratio method is an iterated maximum likelihood procedure.

A mixture of conditional empirical processes. By introducing, for $A \in \mathcal{A}$ such that $P(A) > 0$ and $\mathbb{P}_n^{(N)}(A) > 0$, the conditional expectations

$$\mathbb{E}_n^{(N)}(f|A) = \frac{\mathbb{P}_n^{(N)}(f 1_A)}{\mathbb{P}_n^{(N)}(A)}, \quad \mathbb{E}(f|A) = \frac{P(f 1_A)}{P(A)}, \tag{1.7}$$

we see that (1.3) further reads

$$\mathbb{P}_n^{(N)}(f) = \sum_{j=1}^{m_N} P(A_j^{(N)}) \mathbb{E}_n^{(N-1)}(f|A_j^{(N)}), \quad P(f) = \sum_{j=1}^{m_N} P(A_j^{(N)}) \mathbb{E}(f|A_j^{(N)}).$$

Therefore (1.4) can also be formulated into

$$\begin{aligned} \alpha_n^{(N)}(f) &= \sum_{j=1}^{m_N} P(A_j^{(N)}) \alpha_{n,j}^{(N-1)}(f), \\ \alpha_{n,j}^{(N-1)}(f) &= \sqrt{n} \left(\mathbb{E}_n^{(N-1)}(f|A_j^{(N)}) - \mathbb{E}(f|A_j^{(N)}) \right). \end{aligned} \tag{1.8}$$

Each $\alpha_{n,j}^{(N-1)}$ is the conditional empirical process of $\mathbb{P}_n^{(N-1)}$ on a set $A_j^{(N)}$ of the new partition $\mathcal{A}^{(N)}$. Their mixture with weights $P(\mathcal{A}^{(N)})$ is $\alpha_n^{(N)}$. In view of (1.7) and (1.8) we have to study the consequences of (1.5) on $\mathbb{P}_n^{(N-1)}(f 1_{A_j^{(N)}})$ and $\mathbb{E}_n^{(N-1)}(f|A_j^{(N)})$ as $n \to +\infty$, for $f \neq 1_{A_j^{(N)}}$.

Bias and variance problem. The processes $\alpha_{n,j}^{(N-1)}$ from (1.8) are not centered due to the factors $1/\mathbb{P}_n^{(N-1)}(A_j^{(N)})$ in (1.3) and (1.7). In general it holds

$$\mathbb{E}\left(\mathbb{E}_n^{(N-1)}(f|A) - \mathbb{E}(f|A)\right) = \mathbb{E}\left(\mathbb{P}_n^{(N-1)}(f 1_A)\left(\frac{1}{\mathbb{P}_n^{(N-1)}(A)} - \frac{1}{P(A)}\right)\right) \neq 0,$$

except for $(A, f) = (A_j^{(N-1)}, 1_{A_j^{(N-1)}})$ hence $\alpha_n^{(N)}$ is no more centered if $N \geqslant 1$. This unavoidable bias is induced by (1.5) to globally compensate for the local cancellation of the variance of $\mathbb{P}_n^{(N)}(\mathcal{A}^{(N)}) = P(\mathcal{A}^{(N)})$. The bias tends to spread through (1.3) since the information $P(\mathcal{A}^{(N)})$ is applied to the biased $\mathbb{P}_n^{(N-1)}$ instead of the unbiased $\mathbb{P}_n$. The variance of $\mathbb{P}_n^{(N)}(f)$ for the step functions $f = 1_A$ being null if $A \in \mathcal{A}^{(N)}$ one expects that $\mathbb{V}(\alpha_n^{(N)}(f)) \leqslant \mathbb{V}(\alpha_n^{(0)}(f))$ for many more functions f. Our results show that, uniformly over a large class of functions, the bias vanishes asymptotically and the variance decreases, as well as the quadratic risk, thus $\mathbb{E}((\mathbb{P}_n^{(N)}(f) - P(f))^2) \leqslant \mathbb{E}((\mathbb{P}_n^{(0)}(f) - P(f))^2)$ for n large.

2. Main results

2.1. The raking-ratio empirical and Gaussian processes

Let $\mathcal{M}$ denote the set of measurable real valued functions on $(\mathcal{X},\mathcal{A})$. Consider a class $\mathcal{F}\subset\mathcal{M}$ such that $\sup_{f\in\mathcal{F}}|f|\leqslant M<+\infty$ and satisfying the pointwise measurability condition often used to avoid measurability problems. Namely, $\lim_{m\to+\infty} f_m(x)=f(x)$ for all $x\in\mathcal{X}$ and $f\in\mathcal{F}$ where $\{f_m\}\subset\mathcal{F}_*$ depends on f and $\mathcal{F}_*\subset\mathcal{F}$ is countable. With no loss of generality also assume that

$$1_A f\in\mathcal{F},\quad A\in\mathcal{A}_{\cup}^{(N)}=\mathcal{A}^{(1)}\cup...\cup\mathcal{A}^{(N)},\quad f\in\mathcal{F}. \tag{2.1}$$

In addition $\mathcal{F}$ is assumed to have either a small uniform entropy, like Vapnik-Chervonenkis classes – for short, VC-classes – or a small P-bracketing entropy, like many classes of smooth functions. These entropy conditions are defined below and named (VC) and (BR) respectively. For a probability measure Q on $(\mathcal{X},\mathcal{A})$ and $f,g\in\mathcal{M}$ define $d_Q^2(f,g)=\int_{\mathcal{X}}(f-g)^2dQ$. Let $N(\mathcal{F},\varepsilon,d_Q)$ be the minimum number of balls having d_Q-radius ε needed to cover $\mathcal{F}$. Let $N_{[\,]}(\mathcal{F},\varepsilon,d_P)$ be the least number of ε-brackets necessary to cover $\mathcal{F}$, of the form $[g_-,g_+]=\{f:g_-\leqslant f\leqslant g_+\}$ with $d_P(g_-,g_+)<\varepsilon$.

Hypothesis (VC)**.** *For $c_0>0$, $\nu_0>0$ it holds $\sup_Q N(\mathcal{F},\varepsilon,d_Q)\leqslant c_0/\varepsilon^{\nu_0}$ where the supremum is taken over all discrete probability measures Q on $(\mathcal{X},\mathcal{A})$.*

Hypothesis (BR)**.** *For $b_0>0$, $r_0\in(0,1)$ it holds $N_{[\,]}(\mathcal{F},\varepsilon,d_P)\leqslant\exp(b_0^2/\varepsilon^{2r_0})$.*

If one modifies a class $\mathcal{F}$ satisfying (VC) or (BR) by adding the functions necessary to also satisfy the condition (2.1) then (VC) or (BR) still holds with a new constant c_0 or b_0 respectively. Many properties and examples of VC-classes or classes satisfying (BR) can be found in Pollard [19], Van der Vaart and Wellner [27] or Dudley [14]. Uniform boundedness is the less crucial assumption and could be replaced by a moment condition allowing some truncation arguments, however adding technicalities.

Let $\ell^\infty(\mathcal{F})$ denote the set of real-valued functions bounded on $\mathcal{F}$, endowed with the supremum norm $\|\cdot\|_{\mathcal{F}}$. The raking-ratio empirical process $\alpha_n^{(N)}$ defined at (1.4) is now denoted $\alpha_n^{(N)}(\mathcal{F})=\{\alpha_n^{(N)}(f):f\in\mathcal{F}\}$. Under (VC) or (BR) $\mathcal{F}$ is a P-Donsker class – see Sections 2.5.1 and 2.5.2 of [27]. Thus $\alpha_n^{(0)}(\mathcal{F})$ converges weakly in $\ell^\infty(\mathcal{F})$ to the P-Brownian bridge $\mathbb{G}$ indexed by $\mathcal{F}$, that we denote $\mathbb{G}(\mathcal{F})=\{\mathbb{G}(f):f\in\mathcal{F}\}$. Hence $\mathbb{G}(\mathcal{F})$ is a Gaussian process such that $f\mapsto\mathbb{G}(f)$ is linear and, for any $f,g\in\mathcal{F}$,

$$\mathbb{E}\left(\mathbb{G}(f)\right)=0,\quad \mathrm{Cov}(\mathbb{G}(f),\mathbb{G}(g))=P(fg)-P(f)P(g). \tag{2.2}$$

As for $\alpha_n^{(N)}$ we write $\mathbb{G}^{(0)}(\mathcal{F})=\mathbb{G}(\mathcal{F})$ and, for short, $\mathbb{G}(A)=\mathbb{G}(1_A)$ if $A\in\mathcal{A}$. Remind (1.7). Let us introduce a new centered Gaussian process $\mathbb{G}^{(N)}(\mathcal{F})$ indexed by $\mathcal{F}$ that we call the N-th raking-ratio P-Brownian bridge and that is defined recursively, for any $N \in \mathbb{N}_*$ and $f \in \mathcal{F}$, by

$$\mathbb{G}^{(N)}(f) = \mathbb{G}^{(N-1)}(f) - \sum_{j=1}^{m_N} \mathbb{E}(f|A_j^{(N)})\mathbb{G}^{(N-1)}(A_j^{(N)}). \tag{2.3}$$

The distribution of $\mathbb{G}^{(N)}$ is given in Proposition 7. Lastly, the following notation will be useful,

$$\sigma_f^2 = \mathbb{V}(f(X)) = P(f^2) - P(f)^2, \quad \sigma_{\mathcal{F}}^2 = \sup_{f \in \mathcal{F}} \sigma_f^2. \tag{2.4}$$

Notice that $\sigma_f^2 = \mathbb{V}(\alpha_n^{(0)}(f)) = \mathbb{V}(\mathbb{G}^{(0)}(f))$.

2.2. General properties

We now state asymptotic and nonasymptotic properties that always hold after raking N_0 times. The i.i.d. sequence $\{X_n\}$ is defined on a probability space $(\Omega, \mathcal{T}, \mathbb{P})$ so that $\mathbb{P}$ implicitly leads all convergences when $n \to +\infty$ and $(\mathcal{X}, \mathcal{A})$ is endowed with $P = \mathbb{P}^{X_1}$. For all $N \leqslant N_0$ the information $P(\mathcal{A}^{(N)})$ satisfies (1.2). Most of the subsequent constants can be bounded by using only N_0 and

$$p_{(N_0)} = \min_{0 \leqslant N \leqslant N_0} p_N = \min_{0 \leqslant N \leqslant N_0} \min_{1 \leqslant j \leqslant m_N} P(A_j^{(N)}) > 0. \tag{2.5}$$

Write $L(x) = \log(\max(e, x))$ and define $\kappa_{N_0} = \prod_{N=1}^{N_0}(1 + Mm_N), \kappa_0 = 1$.

Proposition 2. *If $\mathcal{F}$ satisfies (VC) or (BR) then for all $N_0 \in \mathbb{N}$ it holds*

$$\limsup_{n \to +\infty} \frac{1}{\sqrt{2L \circ L(n)}} \sup_{0 \leqslant N \leqslant N_0} \left\| \alpha_n^{(N)} \right\|_{\mathcal{F}} \leqslant \kappa_{N_0} \sigma_{\mathcal{F}} \quad a.s.$$

Remark E. The limiting constant $\kappa_{N_0} \leqslant \left(1 + M/p_{(N_0)}\right)^{N_0}$ is large, and possibly largely suboptimal, except for $N_0 = 0$ where $\kappa_0 = 1$ coincides with the classical law of the iterated logarithm – from which the proposition follows.

The next result shows that the nonasymptotic deviation probability for $\|\alpha_n^{(N)}\|_{\mathcal{F}}$ can be controlled by the deviation probability of $\|\alpha_n^{(0)}\|_{\mathcal{F}}$ which in turn can be bounded by using Talagrand [26], van der Vaart and Wellner [27] or more recent bounds from empirical processes theory. However, since the partition changes at each step the constants are penalized by factors similar to κ_{N_0} above, involving

$$P_{N_0} = \prod_{N=1}^{N_0} p_N, \quad M_{N_0} = \prod_{N=1}^{N_0} m_N, \quad S_{N_0} = \sum_{N=1}^{N_0} m_N. \tag{2.6}$$

Proposition 3. *If $\mathcal{F}$ is pointwise measurable, bounded by M then for any $N_0 \in \mathbb{N}$, any $n \in \mathbb{N}_*$ and any $\lambda > 0$ we have*

$$\mathbb{P}\left(\sup_{0 \leqslant N \leqslant N_0} \left\|\alpha_n^{(N)}\right\|_{\mathcal{F}} \geqslant \lambda\right) \leqslant 2^{N_0} N_0 M_{N_0} \mathbb{P}\left(\left\|\alpha_n^{(0)}\right\|_{\mathcal{F}} \geqslant \frac{\lambda P_{N_0}}{(1 + M + \lambda/\sqrt{n})^{N_0}}\right) + S_{N_0}\left(1 - p_{(N_0)}\right)^n.$$

Under (BR) it holds, for $n > n_0$ and $\lambda_0 < \lambda < D_0\sqrt{n}$,

$$\mathbb{P}\left(\sup_{0\leqslant N\leqslant N_0}\left\|\alpha_n^{(N)}\right\|_{\mathcal{F}} \geqslant \lambda\right) \leqslant D_1\exp(-D_2\lambda^2) + S_{N_0}\left(1-p_{(N_0)}\right)^n,$$

where the positive constants D_0, D_1, D_2, n_0, λ_0 are defined at (3.10). Under (VC) it holds, for $n > n_0$ and $\lambda_0 < \lambda < 2M\sqrt{n}$,

$$\mathbb{P}\left(\sup_{0\leqslant N\leqslant N_0}\left\|\alpha_n^{(N)}\right\|_{\mathcal{F}} \geqslant \lambda\right) \leqslant D_3\lambda^{v_0}\exp(-D_4\lambda^2) + S_{N_0}\left(1-p_{(N_0)}\right)^n,$$

where the positive constants D_3, D_4, n_0, λ_0 are defined at (3.11).

Remark F. Clearly, to avoid drawbacks N_0 should be fixed as n increases, and $\mathcal{F}$ limited to the bare necessities for the actual statistical problem. In this case, Proposition 3 shows that $\|\alpha_n^{(N_0)}\|_{\mathcal{F}}$ is of order $C\sqrt{\log n}$ with probability less than $1/n^2$ and $C > 0$. Concentration of measure type probability bounds for $\|\alpha_n^{(N_0)}\|_{\mathcal{F}} - \mathbb{E}(\|\alpha_n^{(N_0)}\|_{\mathcal{F}})$ are more difficult to handle due to the mixture (1.8) of processes $\alpha_{n,j}^{(N-1)}$ involving unbounded random coefficients.

Our main result is that the raking-ratio empirical processes $\alpha_n^{(0)}, ..., \alpha_n^{(N_0)}$ jointly converge weakly at some explicit rate to the raking-ratio P-Brownian bridges $\mathbb{G}^{(0)}, ..., \mathbb{G}^{(N_0)}$ defined at (2.3) and studied in Section 2.3. The $\mathbb{R}^{N_0+1}$-valued version can be stated as follows.

Proposition 4. *If $\mathcal{F}$ satisfies (VC) or (BR) then for all $N_0 \in \mathbb{N}$, as $n \to +\infty$ the sequence $(\alpha_n^{(0)}(\mathcal{F}), ..., \alpha_n^{(N_0)}(\mathcal{F}))$ converges weakly to $(\mathbb{G}^{(0)}(\mathcal{F}), ..., \mathbb{G}^{(N_0)}(\mathcal{F}))$ on $\ell_\infty(\mathcal{F} \to \mathbb{R}^{N_0+1})$.*

By using Berthet and Mason [3] we further obtain the following upper bound for the speed of Gaussian approximation of $\alpha_n^{(N)}$ in $\|\cdot\|_{\mathcal{F}}$ distance. The powers provided at their Propositions 1 and 2 are $\alpha = 1/(2+5\nu_0)$, $\beta = (4+5\nu_0)/(4+10\nu_0)$ and $\gamma = (1-r_0)/2r_0$ – they could be slightly improved.

Theorem 2.1. *Let $\theta_0 > 0$. If $\mathcal{F}$ satisfies (VC) then write $v_n = (\log n)^\beta/n^\alpha$. If $\mathcal{F}$ satisfies (BR) then write $v_n = 1/(\log n)^\gamma$. In both cases, one can define on the same probability space $(\Omega, \mathcal{T}, \mathbb{P})$ a sequence $\{X_n\}$ of independent random variables with law P and a sequence $\{\mathbb{G}_n\}$ of versions of $\mathbb{G}$ satisfying the following property. For any $N_0 \geqslant 0$ there exists $n_0 \in \mathbb{N}$ and $d_0 > 0$ such that we have, for all $n \geqslant n_0$,*

$$\mathbb{P}\left(\sup_{0\leqslant N\leqslant N_0}\left\|\alpha_n^{(N)} - \mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} \geqslant d_0 v_n\right) < \frac{1}{n^{\theta_0}},$$

where $\mathbb{G}_n^{(N)}$ is the version of $\mathbb{G}^{(N)}$ derived from $\mathbb{G}_n^{(0)} = \mathbb{G}_n$ through (2.3).

Remark G. Applied with $\theta_0 > 1$, Theorem 2.1 makes the study of weak convergence of functions of $\alpha_n^{(N)}(\mathcal{F})$ easier by replacing $\alpha_n^{(N)}$ by $\mathbb{G}_n^{(N)}$ through

$$\limsup_{n\to+\infty}\frac{1}{v_n}\sup_{0\leqslant N\leqslant N_0}\left\|\alpha_n^{(N)} - \mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} \leqslant d_0 < +\infty \quad a.s.$$

then exploiting the properties induced by (2.3) as in Section 2.3. For instance the finite dimensional laws of $\mathbb{G}^{(N)}$ are computed explicitly at Proposition 7. For nonasymptotic applications, given a class $\mathcal{F}$ of interest it is possible to compute crude bounds for n_0 and d_0 since most constants are left explicit in our proofs as well as in [3]. Indeed d_0 depends on p_{N_0} from (2.5), on $P_{N_0}, M_{N_0}, S_{N_0}$ from (2.6), on ν_0, c_0, r_0, b_0 from (VC) or (BR), on N_0, M, θ_0 and on some universal constants from the literature.

Clearly, Theorem 2.1 implies that the speed of weak convergence of $\alpha_n^{(N)}$ to $\mathbb{G}^{(N)}$ in Lévy-Prokhorov distance d_{LP} is at least $d_0 v_n$ – see (3.12) and Section 11.3 of [13] for a definition of this metric. More deeply, from Theorem 2.1 we derive the following rates of uniform convergence for the bias and the variance.

Proposition 5. *If $\mathcal{F}$ satisfies (VC) or (BR) then for $N_0 \in \mathbb{N}$ it holds*

$$\limsup_{n\to+\infty} \frac{\sqrt{n}}{v_n} \max_{0\leqslant N\leqslant N_0} \sup_{f\in\mathcal{F}} \left| \mathbb{E}\left(\mathbb{P}_n^{(N)}(f)\right) - P(f)\right| \leqslant d_0,$$

where $v_n \to 0$ and d_0 are the same as in Theorem 2.1, and

$$\begin{aligned}
&\limsup_{n\to+\infty} \frac{n}{v_n} \sup_{f,g\in\mathcal{F}} \left| \mathbb{E}\left((\mathbb{P}_n^{(N)}(f) - P(f))(\mathbb{P}_n^{(N)}(g) - P(g))\right) \right.\\
&\qquad \left. - \frac{1}{n}\mathrm{Cov}\left(\mathbb{G}^{(N)}(f), \mathbb{G}^{(N)}(g)\right) \right| \\
&= \limsup_{n\to+\infty} \frac{n}{v_n} \sup_{f,g\in\mathcal{F}} \left| \mathrm{Cov}\left(\mathbb{P}_n^{(N)}(f), \mathbb{P}_n^{(N)}(g)\right) - \frac{1}{n}\mathrm{Cov}\left(\mathbb{G}^{(N)}(f), \mathbb{G}^{(N)}(g)\right) \right| \\
&\leqslant \sqrt{\frac{8}{\pi}} d_0 \sigma_{\mathcal{F}}.
\end{aligned}$$

By Proposition 5, the bias process $\mathbb{E}(\alpha_n^{(N)}(f)) = \sqrt{n}(\mathbb{E}(\mathbb{P}_n^{(N)}(f)) - Pf)$ vanishes at the uniform rate v_n. The covariance of $\mathbb{G}^{(N)}$ is computed in Section 2.3 and the quadratic risk is estimated at Remark I.

A second consequence of Theorem 2.1 is uniform Berry-Esseen type bounds. Let Φ denote the distribution function of the centered standardized normal law.

Proposition 6. *Assume that $\mathcal{F}$ satisfies (VC) or (BR), fix $N_0 \in \mathbb{N}$ and let $d_0 > 0$, $v_n \to 0$ be defined as in Theorem 2.1. If $\mathcal{F}_0 \subset \mathcal{F}$ is such that*

$$\sigma_0^2 = \inf\left\{\mathbb{V}\left(\mathbb{G}^{(N)}(f)\right) : f \in \mathcal{F}_0, 0 \leqslant N \leqslant N_0\right\} > 0,$$

then for any $d_1 > d_0$ there exists $n_1 \in \mathbb{N}$ such that for all $n \geqslant n_1$,

$$\max_{0\leqslant N\leqslant N_0} \sup_{f\in\mathcal{F}_0} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\sqrt{n}\frac{\mathbb{P}_n^{(N)}(f) - P(f)}{\sqrt{\mathbb{V}\left(\mathbb{G}^{(N)}(f)\right)}} \leqslant x\right) - \Phi(x)\right| \leqslant \frac{d_1}{\sqrt{2\pi}\sigma_0} v_n. \quad (2.7)$$

Let $\mathcal{L}$ be a collection of real valued Lipschitz functions φ defined on $\ell_\infty(\mathcal{F})$ with Lipschitz constant bounded by $C_1 < +\infty$ and such that $\varphi(\mathbb{G}^{(N)})$ has a density *bounded by* $C_2 < +\infty$ *for all* $0 \leqslant N \leqslant N_0$. *Then for all* $\varphi \in \mathcal{L}$, $n \geqslant n_1$,

$$\max_{0\leqslant N\leqslant N_0} \sup_{\varphi\in\mathcal{L}} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\varphi(\alpha_n^{(N)}) \leqslant x\right) - \mathbb{P}\left(\varphi(\mathbb{G}^{(N)}) \leqslant x\right) \right| \leqslant d_1 C_1 C_2 v_n. \tag{2.8}$$

Remark H. The formula (2.7) is a special case of the second one (2.8) and reads

$$\max_{0\leqslant N\leqslant N_0} \sup_{f\in\mathcal{F}_0} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\alpha_n^{(N)}(f) \leqslant x\right) - \mathbb{P}\left(\mathbb{G}^{(N)}(f) \leqslant x\right) \right| \leqslant \frac{d_1}{\sqrt{2\pi}\sigma_0} v_n.$$

The functions $f \in \mathcal{F}$ overdetermined by the knowledge of $P(\mathcal{A}^{(N)})$ have a small $\mathbb{V}(\mathbb{G}^{(N)}(f))$ and are excluded from $\mathcal{F}_0$. Proposition 6 is especially useful under (VC) since v_n is then polynomialy decreasing, thus allowing larger C_1C_2 and $\mathcal{L}$. An example is given in Section A.3. Whenever the class $\mathcal{F}$ is finite, the density of the transform $\varphi(\mathbb{G}(\mathcal{F}))$ of the finite dimensional Gaussian vector $\mathbb{G}(\mathcal{F})$ is easily computed. The conditions for (2.8) of Proposition 6 are fulfilled if, for example, all random variables $\varphi(\mathbb{G}(\mathcal{F}))$ can be controlled by discretizing the small entropy class $\mathcal{F}$, by bounding their densities then by taking limits accordingly.

2.3. Limiting variance and risk reduction

In this section we study the covariance structure of $\mathbb{G}^{(N)}(\mathcal{F})$ from (2.3), for N fixed. The following matrix notation is introduced to shorten formulas. The brackets $[\cdot]$ refer to column vectors built from the partition $\mathcal{A}^{(k)}$ appearing inside. Let V^t denote the transpose of a vector V. For $k \leqslant N$ write

$$\mathbb{E}\left[f|\mathcal{A}^{(k)}\right] = \left(\mathbb{E}(f|A_1^{(k)}), \dots, \mathbb{E}(f|A_{m_k}^{(k)})\right)^t,$$
$$\mathbb{G}\left[\mathcal{A}^{(k)}\right] = \left(\mathbb{G}(A_1^{(k)}), \dots, \mathbb{G}(A_{m_k}^{(k)})\right)^t,$$

and, for $l \leqslant k \leqslant N$ define the stochastic matrix $\mathbf{P}_{\mathcal{A}^{(k)}|\mathcal{A}^{(l)}}$ to be

$$\left(\mathbf{P}_{\mathcal{A}^{(k)}|\mathcal{A}^{(l)}}\right)_{i,j} = P(A_j^{(k)}|A_i^{(l)}) = \frac{P(A_j^{(k)} \cap A_i^{(l)})}{P(A_i^{(l)})}, \quad 1 \leqslant i \leqslant m_l, 1 \leqslant j \leqslant m_k.$$

Write Id_k the identity matrix $k \times k$. Remind that $\mathbb{V}(\mathbb{G}(f)) = P(f^2) - (P(f))^2$, $P(\mathcal{A}^{(k)})^t = P[\mathcal{A}^{(k)}]$ and $P(A_i^{(k)} \cap A_j^{(k)}) = 0$ if $i \neq j$. The covariance matrix of the Gaussian vector $\mathbb{G}[\mathcal{A}^{(k)}]$ is $\mathbb{V}(\mathbb{G}[\mathcal{A}^{(k)}]) = diag(P(\mathcal{A}^{(k)})) - P(\mathcal{A}^{(k)})^t P(\mathcal{A}^{(k)})$. Let $\cdot$ denote a product between a matrix and a vector. Finally define

$$\Phi_k^{(N)}(f) = \mathbb{E}\left[f|\mathcal{A}^{(k)}\right] + \sum_{\substack{1\leqslant L\leqslant N-k \\ k<l_1<l_2<\dots<l_L\leqslant N}} (-1)^L \mathbf{P}_{\mathcal{A}^{(l_1)}|\mathcal{A}^{(i)}} \mathbf{P}_{\mathcal{A}^{(l_2)}|\mathcal{A}^{(l_1)}} \dots \mathbf{P}_{\mathcal{A}^{(l_L)}|\mathcal{A}^{(l_{L-1})}} \cdot \mathbb{E}\left[f|\mathcal{A}^{(l_L)}\right]. \tag{2.9}$$

An explicit expression for $\mathbb{G}^{(N)}$ is given in Lemma 3 and the closed form for the covariance function of $\mathbb{G}^{(N)}(\mathcal{F})$ is as follows.

Proposition 7. *For all $N \in \mathbb{N}$, the process $\mathbb{G}^{(N)}(\mathcal{F})$ is Gaussian, centered and linear with covariance function defined to be, for $(f,g) \in \mathcal{F}^2$,*

$$\begin{aligned}&\operatorname{Cov}\left(\mathbb{G}^{(N)}(f), \mathbb{G}^{(N)}(g)\right)\\&\quad= \operatorname{Cov}\left(\mathbb{G}(f), \mathbb{G}(g)\right) - \sum_{k=1}^{N} \Phi_k^{(N)}(f)^t \cdot \mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right) \cdot \Phi_k^{(N)}(g).\end{aligned}$$

Proposition 7 implies the following variance reduction phenomenon.

Proposition 8. *For any $\{f_1, ..., f_m\} \subset \mathcal{F}$ and $N \in \mathbb{N}$ the covariance matrices $\Sigma_m^{(N)} = \mathbb{V}((\mathbb{G}^{(N)}(f_1), ..., \mathbb{G}^{(N)}(f_k)))$ are such that $\Sigma_m^{(0)} - \Sigma_m^{(N)}$ is positive definite.*

Remark I. In particular we have $\mathbb{V}(\mathbb{G}^{(N)}(f)) \leqslant \mathbb{V}(\mathbb{G}^{(0)}(f)) = \sigma_f^2$, $f \in \mathcal{F}$. The asymptotic risk reduction after raking is quantified by combining Propositions 5 and 8. Given $\varepsilon_0 > 0$ and $0 < \sigma_0 < \sigma_{\mathcal{F}}$ there exists some $n_0 = n_0(\varepsilon_0, \mathcal{F})$ such that if $n > n_0$ then any $f \in \mathcal{F}$ with initial quadratic risk $\sigma_f^2/n > \sigma_0/n$ has a new risk, after raking N times, equal to

$$\mathbb{E}\left((\mathbb{P}_n^{(N)}(f) - P(f))^2\right) = \frac{\sigma_f^2}{n}(\Delta(f) + e(f)v_n),$$

where $v_n \to 0$ and d_0 are as in Theorem 2.1 and

$$\begin{aligned}\Delta(f) &= \frac{\mathbb{V}(\mathbb{G}^{(N)}(f))}{\sigma_f^2} \in [0,1]\,,\\ \sup_{f\in\mathcal{F},\ \sigma_f \geqslant \sigma_0} |e(f)| &< (1+\varepsilon_0)\sqrt{\frac{8}{\pi}}d_0\frac{\sigma_{\mathcal{F}}}{\sigma_0},\\ \mathbb{V}\left(\mathbb{G}^{(N)}(f)\right) &= \sigma_f^2 - \sum_{k=1}^{N} \Phi_k^{(N)}(f)^t \cdot \mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right) \cdot \Phi_k^{(N)}(f), \qquad (2.10)\end{aligned}$$

so that the risk is reduced whenever $\Delta(f) < 1$ and n is large enough.

When $N_1 > N_0 > 0$ it is not automatically true that the covariance structure of $\alpha_n^{(N_1)}(\mathcal{F})$ decreases compared to that of $\alpha_n^{(N_0)}(\mathcal{F})$. According to the next statement, a simple sufficient condition is to rake two times along the same cycle of partitions.

Proposition 9. *Let $N_0, N_1 \in \mathbb{N}$ be such that $N_1 \geqslant 2N_0$ and*

$$\mathcal{A}^{(N_0-k)} = \mathcal{A}^{(N_1-k)},\ \textit{for } 0 \leqslant k < N_0.$$

Then it holds $\mathbb{V}(\mathbb{G}^{(N_1)}(f)) \leqslant \mathbb{V}(\mathbb{G}^{(N_0)}(f))$ for all $f \in \mathcal{F}$ and $\Sigma_m^{(N_0)} - \Sigma_m^{(N_1)}$ is positive definite for all $\{f_1, ..., f_m\} \subset \mathcal{F}$.

Remark J. In Appendix A.2 a counter-example with $N_1 = N_0 + 1$ shows that the variance does not decrease for all functions at each iteration. This case is excluded from Proposition 9 since $N_1 = N_0 + 1 < 2N_0$ if $N_0 > 1$ and, whenever $N_0 = 1$ and $N_1 = 2$ the requirement $\mathcal{A}^{(N_0)} = \mathcal{A}^{(N_1)}$ is not allowed.

2.4. The case of two marginals

We now consider the original method where k partitions are raked in a periodic order. Let us focus on the case $k=2$ of the two-way contingency table. The Deming and Stephan algorithm coincides with the Sinkhorn-Knopp algorithm for matrix scaling [24]. Denote $\mathcal{A} = \mathcal{A}^{(1)} = \{A_1, ..., A_{m_1}\}$ and $\mathcal{B} = \mathcal{A}^{(2)} = \{B_1, ..., B_{m_2}\}$ the two known margins, thus $\mathcal{A}^{(2m+1)} = \mathcal{A}$ and $\mathcal{A}^{(2m)} = \mathcal{B}$. Likewise for $1 \leqslant i \leqslant m_1$ and $1 \leqslant j \leqslant m_2$ rewrite $(\mathbf{P}_{\mathcal{A}|\mathcal{B}})_{i,j} = P(A_j|B_i)$, $(\mathbf{P}_{\mathcal{B}|\mathcal{A}})_{i,j} = P(B_j|A_i)$ and

$$\mathbb{G}[\mathcal{A}] = (\mathbb{G}(A_1), \ldots, \mathbb{G}(A_{m_1}))^t, \qquad \mathbb{E}[f|\mathcal{A}] = (\mathbb{E}(f|A_1), \ldots, \mathbb{E}(f|A_{m_1}))^t,$$
$$\mathbb{G}[\mathcal{B}] = (\mathbb{G}(B_1), \ldots, \mathbb{G}(B_{m_2}))^t, \qquad \mathbb{E}[f|\mathcal{B}] = (\mathbb{E}(f|B_1), \ldots, \mathbb{E}(f|B_{m_2}))^t.$$

The matrix $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ is $m_1 \times m_1$ and $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ is $m_2 \times m_2$. A sum with a negative upper index is null, a matrix with a negative power is also null, and a square matrix with power zero is the identity matrix. For $N \in \mathbb{N}_*$ define

$$S_{1,\text{even}}^{(N)}(f) = \sum_{k=0}^{N} \left(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}\right)^k \cdot \left(\mathbb{E}[f|\mathcal{A}] - \mathbf{P}_{\mathcal{B}|\mathcal{A}} \cdot \mathbb{E}[f|\mathcal{B}]\right) \text{ is } m_1 \times 1, \tag{2.11}$$

$$S_{2,\text{odd}}^{(N)}(f) = \sum_{k=0}^{N} \left(\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}\right)^k \cdot \left(\mathbb{E}[f|\mathcal{B}] - \mathbf{P}_{\mathcal{A}|\mathcal{B}} \cdot \mathbb{E}[f|\mathcal{A}]\right) \text{ is } m_2 \times 1, \tag{2.12}$$

$$S_{1,\text{odd}}^{(N)}(f) = S_{1,\text{even}}^{(N)} + \left(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}\right)^{N+1} \cdot \mathbb{E}[f|\mathcal{A}] \text{ is } m_1 \times 1, \tag{2.13}$$

$$S_{2,\text{even}}^{(N)}(f) = S_{2,\text{odd}}^{(N)} + \left(\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}\right)^{N+1} \cdot \mathbb{E}[f|\mathcal{B}] \text{ is } m_2 \times 1. \tag{2.14}$$

Proposition 10. *Let $m \in \mathbb{N}$. We have*

$$\mathbb{G}^{(2m)}(f) = \mathbb{G}(f) - S_{1,even}^{(m-1)}(f)^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,even}^{(m-2)}(f)^t \cdot \mathbb{G}[\mathcal{B}], \tag{2.15}$$

$$\mathbb{G}^{(2m+1)}(f) = \mathbb{G}(f) - S_{1,odd}^{(m-1)}(f)^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,odd}^{(m-1)}(f)^t \cdot \mathbb{G}[\mathcal{B}]. \tag{2.16}$$

Remark K. The limiting process $\mathbb{G}^{(N)}$ evaluated at f is then simply $\mathbb{G}(f)$ with a correction depending on the Gaussian vectors $\mathbb{G}[\mathcal{A}]$ and $\mathbb{G}[\mathcal{B}]$ through the two deterministic matrices $\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ and $\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ carrying the information and operating on the conditional expectation vectors $\mathbb{E}[f|\mathcal{A}]$ and $\mathbb{E}[f|\mathcal{B}]$.

The following assumption simplifies the limits and ensures a geometric rate of convergence for matrices $S_{i,\text{even}}^{(N)}$ and $S_{i,\text{odd}}^{(N)}$ as $N \to +\infty$ for $i = 1, 2$.

Hypothesis (ER). *The matrices $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ and $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ are ergodic.*

Remark L. Notice that (ER) holds whenever the matrices have strictly positive coefficients. This is true for $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ if $\sum_{j=1}^{m_2} P(A \cap B_j)P(B_j \cap A') > 0$ for all $A, A' \in \mathcal{A}$ hence if each pair $A, A' \in \mathcal{A}$ is intersected by some $B \in \mathcal{B}$ with positive probability. The latter requirement is for instance met if $\mathcal{X} = \mathbb{R}^d$, P has a positive density and the partitions concern two distinct coordinates.

Proposition 11. *Under (ER) the matrices $S^{(N)}_{l,even}(f)$ and $S^{(N)}_{l,odd}(f)$ for $l = 1, 2$ converge uniformly on $\mathcal{F}$ to $S_{l,even}(f)$ and $S_{l,odd}(f)$ satisfying*

$$S_{1,odd}(f) = S_{1,even}(f) + P_1[f], \quad S_{2,even}(f) = S_{2,odd}(f) + P_2[f],$$

where $P_l[f] = (P(f), \ldots, P(f))^t$ are $m_l \times 1$ vectors. More precisely, given any vector norms $\|\cdot\|_{m_l}$ for $l = 1, 2$, there exists $c_l > 0$ and $0 < \lambda_l < 1$ such that

$$\sup_{f\in\mathcal{F}} \left\| S^{(N)}_{l,even}(f) - S_{l,even}(f) \right\|_{m_l} \leqslant c_l \lambda_l^N,$$

$$\sup_{f\in\mathcal{F}} \left\| S^{(N)}_{l,odd}(f) - S_{l,odd}(f) \right\|_{m_l} \leqslant c_l \lambda_l^N.$$

The main result of this section is the expression of the limiting process for a two partitions raking procedure. Let d_{LP} denote the Lévy-Prokhorov distance. The matrices $S_{1,\text{even}}(f), S_{2,\text{odd}}(f)$ and scalars λ_1, λ_2 are as in Proposition 11.

Theorem 2.2. *Under (ER) the sequence $\left\{\mathbb{G}^{(N)}(\mathcal{F})\right\}$ defined at (2.3) converges almost surely to the centered Gaussian process $\mathbb{G}^{(\infty)}(\mathcal{F})$ defined to be*

$$\mathbb{G}^{(\infty)}(f) = \mathbb{G}(f) - S_{1,even}(f)^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,odd}(f)^t \cdot \mathbb{G}[\mathcal{B}], \quad f \in \mathcal{F}.$$

Moreover we have, for all N large and $c_3 > 0$ depending on $\lambda_1, \lambda_2, P(\mathcal{A}), P(\mathcal{B})$,

$$d_{LP}(\mathbb{G}^{(N)}, \mathbb{G}^{(\infty)}) \leqslant c_3 \sqrt{N} \max(\lambda_1, \lambda_2)^{N/2}.$$

Theorem 2.2 may be viewed as a stochastic counterpart of the deterministic rate obtained by Franklin and Lorentz [15] for the Sinkhorn algorithm. Mixing both approaches could strengthen the following two remarks.

Remark M. The matrices $\mathbf{P}_{\mathcal{A}|\mathcal{B}}$, $\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ and the vectors $\mathbb{E}[f|\mathcal{A}]$, $\mathbb{E}[f|\mathcal{B}]$ are not known without additional information. They can be estimated uniformly over $\mathcal{F}$ as $n \to +\infty$ to evaluate the distribution of $\mathbb{G}^{(N)}$ and $\mathbb{G}^{(\infty)}$, thus giving access to adaptative tests or estimators. Since λ_1, λ_2 and c_3 are related to eigenvalues of $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ and $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ they can be estimated adaptively at rate $1/\sqrt{n}$ in probability. This in turn provides an evaluation of $d_{LP}(\mathbb{G}^{(N)}, \mathbb{G}^{(\infty)})$.

Remark N. In the case of an auxiliary information reduced to $P(A), P(B)$ one should use $\mathcal{A} = \{A, A^c\}$, $\mathcal{B} = \{B, B^c\}$, estimate the missing $P(A \cap B)$ in $\mathbf{P}_{\mathcal{A}|\mathcal{B}}$, $\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ and the conditional expectations on the four sets, then $S_{1,\text{even}}$, $S_{2,\text{odd}}$. If the probabilities of more overlapping sets are known the above characterization of the limiting process $\mathbb{G}^{(\infty)}(\mathcal{F})$ can be generalized to a recursive raking among k partitions $\{A_j, A_j^c\}$ in the same order.

3. Proofs of general results

3.1. Raking formulas

Write

$$B_{n,N_0} = \left\{ \min_{0\leqslant N\leqslant N_0} \min_{1\leqslant j\leqslant m_N} \mathbb{P}_n \left(A_j^{(N)} \right) > 0 \right\}, \tag{3.1}$$

and $B_{n,N_0}^c = \Omega \setminus B_{n,N_0}$. By (1.2), the probability that $\alpha_n^{(N_0)}$ is undefined is

$$P\left(B_{n,N_0}^c\right) \leqslant \sum_{N=1}^{N_0} m_N \left(1-p_N\right)^n \leqslant S_{N_0}\left(1-p_{(N_0)}\right)^n.$$

On B_{n,N_0} we have, by (1.3) and since $\mathcal{A}^{(N)}$ is a partition,

$$\begin{aligned}
&\alpha_n^{(N)}(f)\\
&= \sqrt{n}(\mathbb{P}_n^{(N)}(f) - P(f))\\
&= \sqrt{n}\left(\sum_{j=1}^{m_N} \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})}\mathbb{P}_n^{(N-1)}(f1_{A_j^{(N)}}) - \sum_{j=1}^{m_N} P\left(f1_{A_j^{(N)}}\right)\right)\\
&= \sum_{j=1}^{m_N}\left(\frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})}\alpha_n^{(N-1)}(f1_{A_j^{(N)}}) - \frac{P\left(f1_{A_j^{(N)}}\right)}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})}\alpha_n^{(N-1)}(A_j^{(N)})\right)\\
&= \sum_{j=1}^{m_N} \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})}\alpha_n^{(N-1)}\left((f-\mathbb{E}(f|A_j^{(N)}))1_{A_j^{(N)}}\right). \qquad (3.2)
\end{aligned}$$

In particular by (1.7), (3.2) implies (1.8) since for any $A \in \mathcal{A}$ we have

$$\alpha_n^{(N-1)}(f1_A) - \mathbb{E}(f|A)\alpha_n^{(N-1)}(A) = \mathbb{P}_n^{(N-1)}(f1_A) - \mathbb{E}(f|A)\mathbb{P}_n^{(N-1)}(A).$$

Define $\mathcal{A}_\cap^{(0)} = \{\Omega\}$ and, for $N \geqslant 1$,

$$\mathcal{A}_\cap^{(N)} = \left\{A : A = A_{j_1}^{(1)} \cap A_{j_2}^{(2)} \cap \ldots \cap A_{j_N}^{(N)}, 1 \leqslant j_k \leqslant m_k, 1 \leqslant k \leqslant N\right\}. \qquad (3.3)$$

Let us show that for any $A \in \mathcal{A}_\cap^{(N)}$, $\mathbb{P}_n^{(N)}$ associates to each $X_i \in A$ the weight

$$\omega_n^{(N)}(A) = \frac{1}{n}\prod_{k=1}^{N} \frac{P(A_{j_k}^{(k)})}{\mathbb{P}_n^{(k-1)}(A_{j_k}^{(k)})}. \qquad (3.4)$$

The case $N = 1$ yields (1.1). By induction on (1.3), (3.4) and since $\mathcal{A}_\cap^{(N)}$ is a refined finite partition of $\mathcal{X}$, we get

$$\begin{aligned}
\mathbb{P}_n^{(N+1)}(f) &= \sum_{j=1}^{m_{N+1}} \frac{P(A_j^{(N+1)})}{\mathbb{P}_n^{(N)}(A_j^{(N+1)})}\mathbb{P}_n^{(N)}(f1_{A_j^{(N+1)}})\\
&= \sum_{j=1}^{m_{N+1}} \frac{P(A_j^{(N+1)})}{\mathbb{P}_n^{(N)}(A_j^{(N+1)})} \sum_{A \in \mathcal{A}_\cap^{(N)}} \sum_{i=1}^{n} \omega_n^{(N)}(A) f(X_i) 1_{A \cap A_j^{(N+1)}}(X_i)\\
&= \sum_{i=1}^{n} f(X_i) \sum_{A \in \mathcal{A}_\cap^{(N)}} \sum_{j=1}^{m_{N+1}} 1_{A \cap A_j^{(N+1)}}(X_i)\omega_n^{(N+1)}(A \cap A_j^{(N+1)})\\
&= \sum_{i=1}^{n} f(X_i) \sum_{A \in \mathcal{A}_\cap^{(N+1)}} 1_A(X_i)\omega_n^{(N+1)}(A).
\end{aligned}$$

3.2. Proof of Proposition 1

The partition $\mathcal{A}_{\cap}^{(N)}$ is defined at (3.3). By using (3.4) it holds, for $N \geqslant 1$,

$$\begin{aligned}
& d_K\left(\mathbb{P}_n^{(N-1)} \mid\mid \mathbb{P}_n^{(N)}\right) \\
& = \sum_{i=1}^{n} \mathbb{P}_n^{(N-1)}(\{X_i\}) \log\left(\frac{\mathbb{P}_n^{(N-1)}(\{X_i\})}{\mathbb{P}_n^{(N)}(\{X_i\})}\right) \\
& = \sum_{i=1}^{n} \mathbb{P}_n^{(N-1)}(\{X_i\}) \sum_{A \in \mathcal{A}_{\cap}^{(N-1)}} \sum_{j=1}^{m_N} 1_{A \cap A_j^{(N)}}(X_i) \log\left(\frac{\omega_n^{(N-1)}(A)}{\omega_n^{(N)}(A \cap A_j^{(N)})}\right) \\
& = \sum_{j=1}^{m_N} \log\left(\frac{\mathbb{P}_n^{(N-1)}(A_j^{(N)})}{P(A_j^{(N)})}\right) \sum_{i=1}^{n} \mathbb{P}_n^{(N-1)}(\{X_i\}) \sum_{A \in \mathcal{A}_{\cap}^{(N-1)}} 1_{A \cap A_j^{(N)}}(X_i) \\
& = \sum_{j=1}^{m_N} \mathbb{P}_n^{(N-1)}(A_j^{(N)}) \log\left(\frac{\mathbb{P}_n^{(N-1)}(A_j^{(N)})}{P(A_j^{(N)})}\right),
\end{aligned} \tag{3.5}$$

since $\mathcal{A}_{\cap}^{(N-1)}$ is a partition of $\mathcal{X}$. Hence the contrast between $\mathbb{P}_n^{(N-1)}$ and $\mathbb{P}_n^{(N)}$ viewed as discrete distributions on $\{X_1, ..., X_n\}$ or on $\mathcal{A}^{(N)}$ are the same. Now, by convexity of $-\log(x)$ it follows, for any probability distribution Q supported by $\{X_1, ..., X_n\}$,

$$\begin{aligned}
& d_K\left(\mathbb{P}_n^{(N-1)} \mid\mid Q\right) \\
& = -\sum_{i=1}^{n} \log\left(\frac{Q(\{X_i\})}{\mathbb{P}_n^{(N-1)}(\{X_i\})}\right) \sum_{j=1}^{m_N} \mathbb{P}_n^{(N-1)}(\{X_i\}) 1_{A_j^{(N)}}(X_i) \\
& = -\sum_{j=1}^{m_N} \mathbb{P}_n^{(N-1)}(A_j^{(N)}) \sum_{i=1}^{n} \frac{\mathbb{P}_n^{(N-1)}(\{X_i\})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} 1_{A_j^{(N)}}(X_i) \log\left(\frac{Q(\{X_i\})}{\mathbb{P}_n^{(N-1)}(\{X_i\})}\right) \\
& \geqslant -\sum_{j=1}^{m_N} \mathbb{P}_n^{(N-1)}(A_j^{(N)}) \log\left(\sum_{i=1}^{n} \frac{Q(\{X_i\})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} 1_{A_j^{(N)}}(X_i)\right) \\
& = \sum_{j=1}^{m_N} \mathbb{P}_n^{(N-1)}(A_j^{(N)}) \log\left(\frac{\mathbb{P}_n^{(N-1)}(A_j^{(N)})}{P(A_j^{(N)})}\right) = d_K\left(\mathbb{P}_n^{(N-1)} \mid\mid \mathbb{P}_n^{(N)}\right),
\end{aligned}$$

where the final identification relies on (3.5) and $\mathbb{P}_n^{(N)} = P$ on $\mathcal{A}^{(N)}$.

3.3. Proof of Proposition 2

The classical law of the iterated logarithm holds for the empirical process $\alpha_n^{(0)}$ indexed by $\mathcal{F}$ under (VC) and (BR). See Alexander [1], in particular Theorem 2.12 for (BR) and Theorem 2.13 based on Theorem 2.8 that uses in its proof the consequence of Lemma 2.7, which is indeed (VC). Namely, for any $\varepsilon > 0$, with probability one there exists $n(\omega)$ such that, for all $n > n(\omega)$,

$$u_n||\mathbb{P}_n^{(0)} - P||_{\mathcal{F}} < 1 + \varepsilon, \quad u_n = \sqrt{\frac{n}{2\sigma_{\mathcal{F}}^2 L \circ L(n)}}, \tag{3.6}$$

where by (2.4), $\sigma_{\mathcal{F}}^2 = \sup_{\mathcal{F}} \mathbb{V}(f) \leqslant M^2$. Let $1 \leqslant N \leqslant N_0$. By (1.3) and (1.7) it holds

$$\begin{aligned}\mathbb{P}_n^{(N)}(f) - \mathbb{P}_n^{(N-1)}(f) &= \sum_{j=1}^{m_N} \mathbb{P}_n^{(N-1)}(f 1_{A_j^{(N)}}) \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} - \sum_{j=1}^{m_N} \mathbb{P}_n^{(N-1)}(f 1_{A_j^{(N)}}) \\ &= \sum_{j=1}^{m_N} \mathbb{E}_n^{(N-1)}(f \mid A_j^{(N)}) \left(P(A_j^{(N)}) - \mathbb{P}_n^{(N-1)}(A_j^{(N)}) \right).\end{aligned}$$

Since $\mathbb{P}_n^{(N)}$ is a probability measure we have $\|\mathbb{P}_n^{(N)}(f 1_A)\|_{\mathcal{F}} \leqslant M\mathbb{P}_n^{(N)}(A)$ hence $\|\mathbb{E}_n^{(N-1)}(f \mid A_j^{(N)})\|_{\mathcal{F}} \leqslant M$ and $|\mathbb{P}_n^{(N)}(f) - \mathbb{P}_n^{(N-1)}(f)| \leqslant M m_N \|\mathbb{P}_n^{(N-1)} - P\|_{\mathcal{F}}$. Also observe that (1.2) combined with the fact that $\mathcal{A}^{(N)}$ is a partition implies $m_N \leqslant 1/p_N$ and $p_N \geqslant p_{(N)} \geqslant p_{(N_0)}$. Therefore

$$\begin{aligned}u_n \left\| \mathbb{P}_n^{(N)} - P \right\|_{\mathcal{F}} &\leqslant u_n \left\| \mathbb{P}_n^{(N-1)} - P \right\|_{\mathcal{F}} + u_n \left\| \mathbb{P}_n^{(N)} - \mathbb{P}_n^{(N-1)} \right\|_{\mathcal{F}} \\ &\leqslant u_n \left(1 + M m_N\right) \left\| \mathbb{P}_n^{(N-1)} - P \right\|_{\mathcal{F}} \\ &\leqslant u_n \kappa_N \left\| \mathbb{P}_n^{(0)} - P \right\|_{\mathcal{F}},\end{aligned}$$

where $\kappa_0 = 1$ and, for $N > 0$,

$$\kappa_N = \prod_{k=1}^{N} (1 + M m_k) \leqslant \prod_{k=1}^{N} \left(1 + \frac{M}{p_k}\right) \leqslant \left(1 + \frac{M}{p_{(N_0)}}\right)^{N_0},$$

which by (3.6) remains true for $N = N_0 = 0$. This proves that, given $N_0 \in \mathbb{N}$ and for all $\varepsilon > 0$,

$$\limsup_{n \to +\infty} \sqrt{\frac{n}{2L \circ L(n)}} \sup_{0 \leqslant N \leqslant N_0} \left\| \mathbb{P}_n^{(N)} - P \right\|_{\mathcal{F}} \leqslant (1 + \varepsilon) \kappa_N \sigma_{\mathcal{F}} \quad a.s.$$

and Proposition 2 follows.

3.4. Proof of Proposition 3

Step 1. We work on B_{n,N_0} from (3.1), which means that all the probabilities considered below concern events that we implicitly assume to be intersected with B_{n,N_0}. By (1.7) and (1.8) we have, for $N \geqslant 1$,

$$\alpha_{n,j}^{(N-1)}(f) = \frac{1}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} \alpha_n^{(N-1)} \left((f - \mathbb{E}(f | A_j^{(N)})) 1_{A_j^{(N)}} \right),$$

with $\left|\mathbb{E}(f|A_j^{(N)})\right| \leqslant M$, and

$$\mathbb{P}\left(\left\|\alpha_n^{(N)}\right\|_{\mathcal{F}} \geqslant \lambda\right) \leqslant \mathbb{P}\left(\sum_{j=1}^{m_N} P(A_j^{(N)})\left\|\alpha_{n,j}^{(N-1)}\right\|_{\mathcal{F}} \geqslant \sum_{j=1}^{m_N} P(A_j^{(N)})\lambda\right)$$
$$\leqslant \sum_{j=1}^{m_N} \mathbb{P}\left(\left\|\alpha_{n,j}^{(N-1)}\right\|_{\mathcal{F}} \geqslant \lambda\right).$$

Each term in the latter sum satisfies, for any positive numbers $K \leqslant P(A_j^{(N)})$ and $K' \leqslant P(A_j^{(N)}) - K$,

$$\begin{aligned}
&\mathbb{P}\left(\left\|\alpha_{n,j}^{(N-1)}\right\|_{\mathcal{F}} \geqslant \lambda\right) &(3.7)\\
&= \mathbb{P}\left(\frac{1}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})}\left\|\alpha_n^{(N-1)}(f) - \mathbb{E}(f|A_j^{(N)})\alpha_n^{(N-1)}(A_j^{(N)})\right\|_{\mathcal{F}} \geqslant \lambda\right)\\
&\leqslant \mathbb{P}\left((1+M)\left\|\alpha_n^{(N-1)}\right\|_{\mathcal{F}} \geqslant K\lambda\right) + \mathbb{P}\left(\mathbb{P}_n^{(N-1)}(A_j^{(N)}) \leqslant K\right)\\
&\leqslant \mathbb{P}\left(\left\|\alpha_n^{(N-1)}\right\|_{\mathcal{F}} \geqslant \frac{\lambda K}{1+M}\right) + \mathbb{P}\left(\alpha_n^{(N-1)}(A_j^{(N)}) \leqslant -K'\sqrt{n}\right)\\
&\leqslant 2\mathbb{P}\left(\left\|\alpha_n^{(N-1)}\right\|_{\mathcal{F}} \geqslant \frac{\lambda K}{1+M}\right), &(3.8)
\end{aligned}$$

where the last bound holds provided that $K'\sqrt{n} \geqslant \lambda K/(1+M)$. Define

$$\beta = \frac{1}{1+\lambda/(1+M)\sqrt{n}} \in (0,1), \quad K = \beta p_N, \quad K' = p_N(1-\beta),$$

where p_N is defined in (1.2). Since $p_N \leqslant 1/2$ for any $N \geqslant 1$ it holds $K' > 0$ and $K'\sqrt{n} = \lambda K/(1+M)$. We have shown that for any $N \geqslant 1$,

$$\mathbb{P}\left(\left\|\alpha_n^{(N)}\right\|_{\mathcal{F}} \geqslant \lambda\right) \leqslant 2m_N\mathbb{P}\left(\left\|\alpha_n^{(N-1)}\right\|_{\mathcal{F}} \geqslant \frac{\lambda\beta p_N}{1+M}\right).$$

Applying (3.8) again with λ turned into the smaller $\lambda\beta p_N/(1+M)$ then iterating backward from N_0 we get, for $P_{N_0} = \prod_{N=1}^{N_0} p_N$ and $M_{N_0} = \prod_{N=1}^{N_0} m_N \leqslant 1/P_{N_0}$,

$$\mathbb{P}\left(\left\|\alpha_n^{(N_0)}\right\|_{\mathcal{F}} \geqslant \lambda\right) \leqslant 2^{N_0}M_{N_0}\mathbb{P}\left(\left\|\alpha_n^{(0)}\right\|_{\mathcal{F}} \geqslant \frac{\lambda P_{N_0}}{(1+M+\lambda/\sqrt{n})^{N_0}}\right). \tag{3.9}$$

The latter upper bound being increasing with N_0 we conclude that

$$\mathbb{P}\left(\sup_{0\leqslant N\leqslant N_0}\left\|\alpha_n^{(N)}\right\|_{\mathcal{F}} \geqslant \lambda\right) \leqslant \sum_{N=1}^{N_0}\mathbb{P}\left(\left\|\alpha_n^{(N)}\right\|_{\mathcal{F}} \geqslant \lambda\right) \leqslant N_0\mathbb{P}\left(\left\|\alpha_n^{(N_0)}\right\|_{\mathcal{F}} \geqslant \lambda\right).$$

Step 2. By Theorem 2.14.25 of van der Vaart and Wellner [27] or Corollary 2 of [5], for $n \geqslant 1$, $t > 0$, we have for some universal constants $D_1' > 0$, $D_2' > 0$,

$$\mathbb{P}\left(\left\|\alpha_n^{(0)}\right\|_{\mathcal{F}} > D_1'\left(\mu_n + t\right)\right) \leqslant \exp\left(-D_2'\min\left(\frac{t^2}{\sigma_{\mathcal{F}}^2}, \frac{t\sqrt{n}}{M}\right)\right),$$

where, by the last maximal inequality in Theorem 2.14.2 of [27] applied to $\mathcal{F}$ with envelop function constant to M, it holds

$$\mu_n = \mathbb{E}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n f(X_i)\right\|_{\mathcal{F}} \leqslant M\int_0^1 \sqrt{1+\log N_{[\ ]}(\mathcal{F}, M\varepsilon, d_P)}d\varepsilon.$$

Under (BR), we have $\mu_n < C$ with $C = M(1 + b_0/(1-r_0))$. For $\lambda_0 = 2D_1'C$ we get, for any $n > 0$ and $\lambda_0 < \lambda < 2D_1'\sigma_{\mathcal{F}}^2\sqrt{n}/M$,

$$\mathbb{P}\left(\left\|\alpha_n^{(0)}\right\|_{\mathcal{F}} > \lambda\right) \leqslant \mathbb{P}\left(\left\|\alpha_n^{(0)}\right\|_{\mathcal{F}} > D_1'\left(\mu_n + \frac{\lambda}{2D_1'}\right)\right) \leqslant \exp\left(-D_2''\lambda^2\right),$$

where $D_2'' = D_2'/4(D_1')^2\sigma_{\mathcal{F}}^2$. Therefore, according to (3.9), taking

$$D_0 = \frac{2D_1'\sigma_{\mathcal{F}}^2}{M}, \quad D_1 = N_0 2^{N_0} M_{N_0}, \quad D_2 = \frac{D_2'' P_{N_0}^2}{(1+M+D_0)^{2N_0}}, \tag{3.10}$$

yields $\mathbb{P}(\sup_{0\leqslant N\leqslant N_0}\|\alpha_n^{(N_0)}\|_{\mathcal{F}} \geqslant \lambda) \leqslant D_1\exp\left(-D_2\lambda^2\right)$ for $\lambda_0 < \lambda < D_0\sqrt{n}$.

Step 3. By Theorem 2.14.9 in [27], under (VC) there exists a constant $D(c_0)$ such that, for t_0 large enough and all $t \geqslant t_0$,

$$\mathbb{P}\left(\left\|\alpha_n^{(0)}\right\|_{\mathcal{F}} \geqslant t\right) \leqslant \left(\frac{D(c_0)t}{M\sqrt{v_0}}\right)^{v_0} \exp\left(-\frac{2t^2}{M^2}\right).$$

Denote $\lambda_{1,n}$ and $\lambda_{2,n}$ the two solutions of $\lambda P_{N_0} = t_0(1+M+\lambda/\sqrt{n})^{N_0}$. Notice that, for n large, $\lambda_{1,n}$ is close to $t_0(1+M)^{N_0}/P_{N_0}$ and $\lambda_{2,n} = O(n^{N_0/2(N_0-1)})$. Combined with (3.9) it ensues that for some n_0, λ_0 it holds, for all $n > n_0$ and $\lambda_0 < \lambda < 2M\sqrt{n}$, $\mathbb{P}(\sup_{0\leqslant N\leqslant N_0}\|\alpha_n^{(N)}\|_{\mathcal{F}} \geqslant \lambda) \leqslant D_3\lambda^{v_0}\exp(-D_4\lambda^2)$ where

$$D_3 = \frac{N_0 2^{N_0} M_{N_0}}{(1+M)^{v_0 N_0}}\left(\frac{D(c_0)P_{N_0}}{M\sqrt{v_0}}\right)^{v_0}, \quad D_4 = \frac{2P_{N_0}^2}{M^2(3M+1)^{2N_0}}. \tag{3.11}$$

Finally, at each step, add $S_{N_0}\left(1-p_{(N_0)}\right)^n$ to take B_{n,N_0}^c from (3.1) into account.

3.5. Proof of Proposition 4 and Theorem 2.1

Theorem 2.1 implies Proposition 4 since the weak convergence on $(\ell^\infty(\mathcal{F}), \|\cdot\|_{\mathcal{F}})$ is metrized by the Lévy-Prokhorov distance between $\alpha_n^{(N)}$ and $\mathbb{G}_n^{(N)}$ which is

$$\inf\left\{\varepsilon > 0 : \mathbb{P}^{\alpha_n^{(N)}}(A) \leqslant \mathbb{P}^{\mathbb{G}_n^{(N)}}(A^\varepsilon) + \varepsilon, \mathbb{P}^{\mathbb{G}_n^{(N)}}(A) \leqslant \mathbb{P}^{\alpha_n^{(N)}}(A^\varepsilon) + \varepsilon\right\} \leqslant d_0 v_n. \tag{3.12}$$

To see this, recall that we have $v_n > 1/n^{\theta_0}$ for $\theta_0 > 1/2$ and $v_n \to 0$ in Theorem 2.1, remind (1.6) and (3.1) then observe that

$$\begin{aligned}\mathbb{P}\left(\alpha_n^{(N)} \in A\right) &\leqslant \mathbb{P}\left(\left\{\alpha_n^{(N)} \in A\right\} \cap \left\{\left\|\alpha_n^{(N)} - \mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} \leqslant d_0 v_n\right\} \cap B_{n,N_0}\right) \\ &\quad + \mathbb{P}\left(\left\{\left\|\alpha_n^{(N)} - \mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} > d_0 v_n\right\} \cap B_{n,N_0}\right) + \mathbb{P}\left(B_{n,N_0}^c\right) \\ &\leqslant \mathbb{P}(\mathbb{G}_n^{(N)} \in A^{d_0 v_n}) + \frac{1}{n^{\theta_0}} + S_{N_0}\left(1-p_{(N_0)}\right)^n,\end{aligned}$$

which obviously remains true by exchanging $\alpha_n^{(N)}$ and $\mathbb{G}_n^{(N)}$. Since v_n is the slowest sequence as $n \to +\infty$, if n_0 satisfies $v_{n_0} > 1/n_0^{\theta_0} + S_{N_0}\left(1 - p_{(N_0)}\right)^{n_0}$ then $v_n > 1/n^{\theta_0} + S_{N_0}\left(1 - p_{(N_0)}\right)^{n}$ for all $n > n_0$. Whence (3.12).

We next establish Theorem 2.1. Fix $N_0 \in \mathbb{N}$.

Step 1. Let introduce the transforms, for $f \in \mathcal{F}$, $N \geqslant 1$ and $1 \leqslant j \leqslant m_N$,

$$\phi_{(j,N)}f = \left(f - \mathbb{E}\left(f \mid A_j^{(N)}\right)\right) 1_{A_j^{(N)}},$$
$$\phi_{(N)}f = \sum_{j=1}^{m_N} \phi_{(j,N)}f = f - \sum_{j=1}^{m_N} \mathbb{E}(f \mid A_j^{(N)}) 1_{A_j^{(N)}}.$$

It holds $P(\phi_{(N)}f) = P(\phi_{(j,N)}f) = 0$ and, since $\mathcal{A}^{(N)}$ is a partition of $\mathcal{X}$,

$$(\phi_{(j,N)}f)(\phi_{(j',N)}g) = 0, \quad 1 \leqslant j \neq j' \leqslant m_N. \tag{3.13}$$

Moreover, the $L_2(P)$ property of conditional expectations yields, with the notation (2.4),

$$\sigma^2_{\phi_{(j,N)}f} = P(f^2_{(j,N)}) \leqslant \sigma^2_{\phi_{(N)}f} = P(f^2_{(N)}) = \sum_{j=1}^{m_N} \sigma^2_{\phi_{(j,N)}f} \leqslant \sigma^2_f. \tag{3.14}$$

Next consider the class of backward iterated transforms

$$\mathcal{F}_{(N)} = \phi_{(1)} \circ ... \circ \phi_{(N)}(\mathcal{F}),$$
$$\mathcal{H}_{(N)} = \bigcup_{1 \leqslant k \leqslant N} \bigcup_{1 \leqslant j \leqslant m_k} \phi_{(j,k)} \circ \phi_{(k+1)} \circ ... \circ \phi_{(N)}(\mathcal{F}),$$

where $\phi_{(k+1)} \circ ... \circ \phi_{(N)} = id$ if $k = N \geqslant 1$ and $\mathcal{F}_{(0)} = \mathcal{H}_{(0)} = \mathcal{F}$. Also write $\mathcal{F}_0 = \bigcup_{0 \leqslant N \leqslant N_0} \mathcal{F}_{(N)}$ and $\mathcal{H}_0 = \bigcup_{0 \leqslant N \leqslant N_0} \mathcal{H}_{(N)}$. By iterating (3.14) it comes $\sigma^2_{\mathcal{H}_0} \leqslant \sigma^2_{\mathcal{F}_0} \leqslant \sigma^2_{\mathcal{F}}$. We first show that properties of $\mathcal{F}$ transfer to $\mathcal{F}_{(N)}$, $\mathcal{H}_{(N)}$ for $0 \leqslant N \leqslant N_0$ and thus to $\mathcal{F}_0$ and $\mathcal{H}_0$. Remind (1.2) and the constants defined at (2.6).

Lemma 1. *If $\mathcal{F}$ is pointwise measurable and bounded by M then $\mathcal{F}_{(N)}$ and $\mathcal{H}_{(N)}$ (resp. $\mathcal{F}_0$ and $\mathcal{H}_0$) are pointwise measurable and bounded by $(2M)^N/P_N$ (resp. $(2M)^{N_0}/P_{N_0}$). If (VC) (resp. (BR)) holds then $\mathcal{F}_0$ and $\mathcal{H}_0$ also satisfy (VC) (resp. (BR)) with the same power ν_0 (resp. r_0) as $\mathcal{F}$.*

Proof. If $\mathcal{F}$ is uniformly bounded by M then for $N \leqslant N_0$ we have

$$\sup_{\mathcal{F}} \sup_{\mathcal{X}} \left|\phi_{(N)}f\right| = \sup_{\mathcal{F}} \max_{1 \leqslant j \leqslant m_N} \sup_{\mathcal{X}} \left|\phi_{(j,N)}f\right| \leqslant M\left(1 + \frac{1}{p_N}\right) \leqslant \frac{2M}{p_N},$$

thus, by backward induction from N_0 to 1, $\mathcal{F}_{(N_0)}$ and $\mathcal{H}_{(N_0)}$ are uniformly bounded by $(2M)^{N_0}/P_{N_0}$. It readily follows that $\mathcal{F}_0$ and $\mathcal{H}_0$ are bounded by $(2M)^{N_0}/P_{N_0}$. Assume that $f_k \in \mathcal{F}_*$ converges pointwise on $\mathcal{X}$ to $f \in \mathcal{F}$. From

$$\lim_{k \to +\infty} 1_{A_j^{(N)}}(X) f_k(X) = 1_{A_j^{(N)}}(X) f(X) \quad \text{and} \quad P(1_{A_j^{(N)}}|f_k|) \leqslant P\left(|f_k|\right) \leqslant M,$$

we deduce by dominated convergence that $\lim_{k\to+\infty}\mathbb{E}(f_k \mid A_j^{(N)}) = \mathbb{E}(f \mid A_j^{(N)})$. Thus $\phi_{(j,N)}f_k$ converges pointwise to $\phi_{(j,N)}f$ and $\phi_{(N)}f_k = \sum_{j=1}^{m_N}\phi_{(j,N)}f_k$ to $\phi_{(N)}f = \sum_{j=1}^{m_N}\phi_{(j,N)}f$. By iterating this reasoning backward from N to 1 we obtain that $\mathcal{F}_{(N)}$ and $\mathcal{H}_{(N)}$ are pointwise measurable, by using the countable classes $\phi_{(1)}\circ\ldots\circ\phi_{(N)}(\mathcal{F}_*)$ and $\bigcup_{1\leqslant k\leqslant N}\bigcup_{1\leqslant j\leqslant m_k}\phi_{(j,k)}\circ\phi_{(k+1)}\circ\ldots\circ\phi_{(N)}(\mathcal{F}_*)$ respectively. Assume next that $\mathcal{F}$ satisfies (VC). By (3.13) we have

$$\begin{aligned}
d_Q^2(\phi_{(N)}f,\phi_{(N)}g) &= \int_{\mathcal{X}}\left(\textstyle\sum_{j=1}^{m_N}(\phi_{(j,N)}f-\phi_{(j,N)}g)\right)^2 dQ\\
&= \sum_{j=1}^{m_N} d_Q^2(\phi_{(j,N)}f,\phi_{(j,N)}g)\\
&= \sum_{j=1}^{m_N}\int_{A_j^{(N)}}\left(f-g-\mathbb{E}(f-g \mid A_j^{(N)})\right)^2 dQ\\
&\leqslant \sum_{j=1}^{m_N}\int_{A_j^{(N)}}\left(f-g-(Qf-Qg)\right)^2 dQ\\
&= d_Q^2(f,g)-(Qf-Qg)^2,
\end{aligned}$$

thus $d_Q(f,g)<\varepsilon$ implies $d_Q^2(\phi_{(N)}f,\phi_{(N)}g)\leqslant d_Q^2(f,g)<\varepsilon^2$. If $\mathcal{F}$ can be covered by $N(\mathcal{F},\varepsilon,d_Q)$ balls of d_Q-radius ε centered at some g then $\phi_{(N)}(\mathcal{F})$ can be covered by the same number of balls, centered at the corresponding $\phi_{(N)}g$ and hence the same number of centers $\phi_{(1)}\circ\ldots\circ\phi_{(N)}g$ suffices to cover $\mathcal{F}_{(N)}$. All the $\phi_{(j,k)}\circ\phi_{(k+1)}\circ\ldots\circ\phi_{(N)}g$ are needed to cover $\mathcal{H}_{(N)}$, that is $S_N N(\mathcal{F},\varepsilon,d_Q)$. This shows that $\mathcal{F}_0$ (resp. $\mathcal{H}_0$) obeys (VC) with the same power ν_0 and a constant $c_0(N_0+1)$ (resp. $c_0\sum_{N=0}^{N_0}S_N$). Assume now that $\mathcal{F}$ satisfies (BR). If $g_-\leqslant f\leqslant g_+$ then we have

$$\begin{aligned}
h^-_{(j,N)} &= 1_{A_j^{(N)}}g_- - 1_{A_j^{(N)}}\mathbb{E}\left(g_+ \mid A_j^{(N)}\right)\\
&\leqslant \phi_{(j,N)}f \leqslant 1_{A_j^{(N)}}g_+ - 1_{A_j^{(N)}}\mathbb{E}\left(g_- \mid A_j^{(N)}\right) = h^+_{(j,N)},
\end{aligned}$$

and the $L_2(P)$-size of the new bracket $[h^-_{(j,N)},h^+_{(j,N)}]$ is

$$\begin{aligned}
d_P^2(h^-_{(j,N)},h^+_{(j,N)}) &= \int_{A_j^{(N)}}\left(g_+-g_-+\mathbb{E}(g_+-g_- \mid A_j^{(N)})\right)^2 dP\\
&= P(1_{A_j^{(N)}}(g_+-g_-)^2)+P(A_j^{(N)})\mathbb{E}(g_+-g_- \mid A_j^{(N)})^2\\
&\quad + 2\mathbb{E}(g_+-g_- \mid A_j^{(N)})P(1_{A_j^{(N)}}(g_+-g_-)).
\end{aligned}$$

If $d_P(g_+,g_-)<\varepsilon$ the Hölder inequality yields $P(1_{A_j^{(N)}}(g_+-g_-))\leqslant\varepsilon\sqrt{P(A_j^{(N)})}$ and $\mathbb{E}(g_+-g_- \mid A_j^{(N)})\leqslant\varepsilon/\sqrt{P(A_j^{(N)})}$ hence

$$d_P^2(h^-_{(j,N)},h^+_{(j,N)})\leqslant P(1_{A_j^{(N)}}(g_+-g_-)^2)+3\varepsilon^2,$$

so that $\phi_{(N)}f = \sum_{j=1}^{m_N} \phi_{(j,N)}f \in [h^-_{(N)}, h^+_{(N)}]$ where $h^\pm_{(N)} = \sum_{j=1}^{m_N} h^\pm_{(j,N)}$ satisfies

$$d_P^2(h^-_{(N)}, h^+_{(N)}) = \sum_{j=1}^{m_N} d_P^2(h^-_{(j,N)}, h^+_{(j,N)}) \leqslant d_P^2(g_-, g_+) + 3m_N\varepsilon^2 \leqslant 4m_N\varepsilon^2.$$

It ensues $N_{[\,]}(\phi_{(N)}(\mathcal{F}), \varepsilon, d_P) \leqslant N_{[\,]}(\mathcal{F}, \varepsilon/2\sqrt{m_N}, d_P)$ and $N_{[\,]}(\mathcal{F}_{(N)}, \varepsilon, d_P) \leqslant N_{[\,]}(\mathcal{F}, \varepsilon/2^N\sqrt{M_N}, d_P)$. To cover $\phi_{(j,k)} \circ \phi_{(k+1)} \circ ... \circ \phi_{(N)}(\mathcal{F})$ one needs at most $m_k N_{[\,]}(\mathcal{F}, \varepsilon/2^{N-k}\sqrt{m_{k+1}...m_N}, d_P)$ brackets. We have proved that

$$\begin{aligned}
N_{[\,]}(\mathcal{F}_0, \varepsilon, d_P) &\leqslant (N_0+1)N_{[\,]}(\mathcal{F}, \varepsilon/2^{N_0}\sqrt{M_{N_0}}, d_P),\\
N_{[\,]}(\mathcal{H}_0, \varepsilon, d_P) &\leqslant S_{N_0}N_{[\,]}(\mathcal{F}, \varepsilon/2^{N_0}\sqrt{M_{N_0}}, d_P).
\end{aligned}$$

Therefore $\mathcal{F}_0$, $\mathcal{H}_0$ satisfy (BR) with power r_0 and constant $2^{r_0N_0}M_{N_0}^{r_0}b_0$. □

Step 2. By (3.2) we have

$$\alpha_n^{(N)}(f) = \sum_{j=1}^{m_N} \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})}\alpha_n^{(N-1)}(\phi_{(j,N)}f) = \alpha_n^{(N-1)}(\phi_{(N)}f) + \Gamma_n^{(N)}(f), \tag{3.15}$$

$$\Gamma_n^{(N)}(f) = \sum_{j=1}^{m_N} q_n(j,N)\alpha_n^{(N-1)}\left(\phi_{(j,N)}f\right), \quad q_n(j,N) = \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} - 1.$$

Under the convention that $\phi_{(N+1)} \circ \phi_{(N)} = id$, iterating (3.15) leads to

$$\begin{aligned}
\alpha_n^{(N)}(f) &= \alpha_n^{(0)}(\phi_{(1)} \circ ... \circ \phi_{(N)}f) + F_n^{(N)}(f),\\
F_n^{(N)}(f) &= \sum_{k=1}^{N} \Gamma_n^{(k)}(\phi_{(k+1)} \circ ... \circ \phi_{(N)}f).
\end{aligned}$$

Clearly the terms $\Gamma_n^{(k)}$ carry out some bias and variance distortion. However the following lemma states that $\alpha_n^{(0)}(\mathcal{F}_{(N)})$ is the main contribution to $\alpha_n^{(N)}(\mathcal{F})$ and $F_n^{(N)}(\mathcal{F})$ is an error process.

Lemma 2. *Consider the sequence v_n defined at Theorem 2.1. If $\mathcal{F}$ satisfies (VC) or (BR) then there exists $C_0 < +\infty$ such that we almost surely have, for all n large enough, $\max_{0\leqslant N\leqslant N_0} \|F_n^{(N)}\|_{\mathcal{F}} \leqslant C_0 L\circ L(n)/\sqrt{n}$. Moreover, for any $\zeta > 0$ and $\theta > 0$ there exists $n_3(\zeta,\theta)$ such that we have, for all $n > n_3(\zeta,\theta)$,*

$$\mathbb{P}\left(\max_{0\leqslant N\leqslant N_0} \|F_n^{(N)}\|_{\mathcal{F}} > \zeta v_n\right) \leqslant \frac{1}{2n^\theta}.$$

Proof. **(i)** Let us apply Proposition 2 to $\mathcal{F}$ and, thanks to Lemma 1, to $\mathcal{H}_0$ and $\mathcal{H}_{(N)}$. So for all $\varepsilon > 0$ we have for all n large enough,

$$\max_{1\leqslant N\leqslant N_0}\max_{1\leqslant j\leqslant m_N}\left|\alpha_n^{(N-1)}(A_j^{(N)})\right| \leqslant \sigma_{\mathcal{F}}\kappa_{N_0}\sqrt{2L\circ L(n)}(1+\varepsilon).$$

The following statements are almost surely true, for all n large enough. On the one hand, for $\varepsilon = \sqrt{2}-1 > 0$,

$$\max_{1\leqslant N\leqslant N_0}\max_{1\leqslant j\leqslant m_N}\left|\alpha_n^{(N-1)}(A_j^{(N)})\right| \leqslant b_n = 2\sigma_{\mathcal{F}}\kappa_{N_0}\sqrt{L\circ L(n)}. \tag{3.16}$$

On the other hand, having $\sigma_{\mathcal{H}_0} \leqslant \sigma_{\mathcal{F}}$ by (3.14),

$$\begin{aligned}&\max_{1\leqslant N\leqslant N_0}\max_{1\leqslant k\leqslant N}\max_{1\leqslant j\leqslant m_k}\left|\alpha_n^{(k-1)}(\phi_{(j,k)}\circ\phi_{(k+1)}\circ\ldots\circ\phi_{(N)}f)\right|\\ &\quad\leqslant \max_{1\leqslant k\leqslant N_0}\left\|\alpha_n^{(k-1)}\right\|_{\mathcal{H}_0}\leqslant b_n.\end{aligned}$$

By (3.16), $q_n(j,N) = 1/\left(1+\alpha_n^{(N-1)}(A_j^{(N)})/P(A_j^{(N)})\sqrt{n}\right)-1$ satisfies

$$\max_{1\leqslant N\leqslant N_0}\max_{1\leqslant j\leqslant m_N}|q_n(j,N)|\,\frac{\sqrt{n}}{b_n}p_{(N_0)}\leqslant 2, \tag{3.17}$$

which implies

$$\begin{aligned}&\left\|F_n^{(N)}\right\|_{\mathcal{F}}\leqslant\sum_{k=1}^{N}\max_{1\leqslant j\leqslant m_k}|q_n(j,k)|\sum_{j=1}^{m_k}\left|\alpha_n^{(k-1)}\left(\phi_{(j,k)}\circ\phi_{(k+1)}\circ\ldots\circ\phi_{(N)}f\right)\right|,\\ &\max_{1\leqslant N\leqslant N_0}\left\|F_n^{(N)}\right\|_{\mathcal{F}}\leqslant\frac{2b_n}{\sqrt{n}p_{(N_0)}}S_{N_0}\max_{1\leqslant k\leqslant N_0}\left\|\alpha_n^{(k-1)}\right\|_{\mathcal{H}_0}\leqslant\frac{2b_n^2S_{N_0}}{\sqrt{n}p_{(N_0)}}.\end{aligned} \tag{3.18}$$

The almost sure result then holds with $C_0 = 8\sigma_{\mathcal{F}}^2\kappa_{N_0}^2S_{N_0}/p_{(N_0)}$.

(ii) We now work on the event B_{n,N_0} of (3.1). There obviously exists n_1 such that if $n > n_1$ then $S_{N_0}(1-p_{(N_0)})^n \leqslant 1/4n^\theta$. We can also find $\kappa > 0$ so small that $n^{2\kappa}/\sqrt{n} = o(v_n)$ as $n\to+\infty$. Therefore, whatever $\zeta > 0$ there exists $n_2(\kappa, S_{N_0}, \zeta, \mathcal{F}, P)$ such that $\zeta v_n > 2S_{N_0}n^{2\kappa}/p_{(N_0)}\sqrt{n}$ for any $n\geqslant n_2$. Choosing $n\geqslant\max(n_1,n_2)$ we deduce as for (3.17) and (3.18) that

$$\begin{aligned}&\mathbb{P}\left(\max_{0\leqslant N\leqslant N_0}\|F_n^{(N)}\|_{\mathcal{F}}>\zeta v_n\right)\\ &\leqslant\mathbb{P}\left(S_{N_0}\max_{1\leqslant N\leqslant N_0}\left(\left\|\alpha_n^{(N-1)}\right\|_{\mathcal{H}_0}\max_{1\leqslant j\leqslant m_N}|q_n(j,N)|\right)>\zeta v_n\right)\\ &\leqslant\mathbb{P}\left(\left(\max_{1\leqslant N\leqslant N_0}\max_{1\leqslant j\leqslant m_N}|q_n(j,N)|\right)>\frac{2n^\kappa}{p_{(N_0)}\sqrt{n}}\right)\\ &\quad+\mathbb{P}\left(\max_{1\leqslant N\leqslant N_0}\left\|\alpha_n^{(N-1)}\right\|_{\mathcal{H}_0}>n^\kappa\right)\\ &\leqslant 2\mathbb{P}\left(\max_{1\leqslant N\leqslant N_0}\left\|\alpha_n^{(N-1)}\right\|_{\mathcal{H}_0}>n^\kappa\right).\end{aligned}$$

By Proposition 3 we see that under (VC) or (BR) the latter probability can be made less than $1/8n^\theta$ for any $n > n_3(\zeta,\theta)$ and $n_3(\zeta,\theta)$ large enough. Clearly $n_3(\zeta,\theta)$ depends on ζ,θ,n_1,n_2 and on the entropy of $\mathcal{H}_0$ thus all constants in Lemma 1 and Proposition 3 are involved. $\square$

Step 3. Fix $\theta > 0$. By Lemma 1 we can apply Propositions 1 and 2 of Berthet and Mason [3] to $\mathcal{F}_0$, which ensures the following Gaussian approximation. For some constant $c_\theta(\mathcal{F}_0, P) > 0$ and $n_\theta(\mathcal{F}_0, P) > 0$ we can build on a probability space $(\Omega, \mathcal{T}, \mathbb{P})$ a version of the sequence $\{X_n\}$ of independent random variables with law P and a sequence $\{\mathbb{G}_n^{(0)}\}$ of coupling versions of $\mathbb{G}^{(0)}$ in such a way that, for all $n \geqslant n_\theta(\mathcal{F}_0, P)$,

$$\mathbb{P}\left(\left\|\alpha_n^{(0)} - \mathbb{G}_n^{(0)}\right\|_{\mathcal{F}_0} > c_\theta(\mathcal{F}_0, P) v_n\right) \leqslant \frac{1}{2n^\theta}. \tag{3.19}$$

Keep in mind that constants n_θ and c_θ only depend on the entropy of $\mathcal{F}_0$ through the constants M, c_0, ν_0, b_0, r_0. By choosing $\theta > 1$, $d_\theta > c_\theta(\mathcal{F}_0, P)$ then applying Borel-Cantelli lemma to (3.19), it almost surely holds, for all n large enough,

$$\left\|\alpha_n^{(0)} - \mathbb{G}_n^{(0)}\right\|_{\mathcal{F}_0} < d_\theta v_n. \tag{3.20}$$

Step 4. Let $\theta_0 > 0$. We work under the event B_{n,N_0} of (3.1) with a probability at least $1 - 1/4n^{\theta_0}$ provided that $n > n_1$. The process $\mathbb{G}^{(0)}$ being linear on $\mathcal{F}$ we see that the recursive definition (2.3) applied to the version $\mathbb{G}_n^{(0)}$ of $\mathbb{G}^{(0)}$ reads $\mathbb{G}_n^{(N)}(f) = \mathbb{G}_n^{(N-1)}(\phi_{(N)} f)$. This combined with (3.15) readily gives

$$\begin{aligned}
&\max_{1 \leqslant N \leqslant N_0} \left\|\alpha_n^{(N)} - \mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} \\
&= \max_{1 \leqslant N \leqslant N_0} \left\|\alpha_n^{(N-1)}(\phi_{(N)} f) - \mathbb{G}_n^{(N-1)}(\phi_{(N)} f) + \Gamma_n^{(N)}(f)\right\|_{\mathcal{F}} \\
&= \max_{1 \leqslant N \leqslant N_0} \left\|\alpha_n^{(0)}(\phi_{(1)} \circ \ldots \circ \phi_{(N)} f) - \mathbb{G}_n^{(0)}(\phi_{(1)} \circ \ldots \circ \phi_{(N)} f) + F_n^{(N)}(f)\right\|_{\mathcal{F}} \\
&\leqslant \left\|\alpha_n^{(0)} - \mathbb{G}_n^{(0)}\right\|_{\mathcal{F}_0} + \max_{0 \leqslant N \leqslant N_0} \|F_n^{(N)}\|_{\mathcal{F}}.
\end{aligned} \tag{3.21}$$

Remind that $v_n > L \circ L(n)/\sqrt{n}$ and Lemma 2 holds. By (3.20) and (3.21) we almost surely have, for all n large enough and $d_0 = 2d_{\theta_0}$,

$$\max_{1 \leqslant N \leqslant N_0} \left\|\alpha_n^{(N)} - \mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} \leqslant d_{\theta_0} v_n + C_0 \frac{L \circ L(n)}{\sqrt{n}} \leqslant d_0 v_n.$$

By Lemmas 1 and 2, (3.19) and (3.21), for $n_0 > \max(n_1, n_3(\zeta, \theta_0), n_{\theta_0}(\mathcal{F}_0, P))$ and $d_0 > c_{\theta_0}(\mathcal{F}_0, P) + \zeta$ we have, for all $n \geqslant n_0$,

$$\begin{aligned}
\mathbb{P}\left(\max_{1 \leqslant N \leqslant N_0} \left\|\alpha_n^{(N)} - \mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} > d_0 v_n\right) &\leqslant \frac{1}{2n^{\theta_0}} + \mathbb{P}\left(\max_{0 \leqslant N \leqslant N_0} \|F_n^{(N)}\|_{\mathcal{F}} > \zeta v_n\right) \\
&\leqslant \frac{1}{n^{\theta_0}}.
\end{aligned}$$

To conclude observe that the parameters $N_0, M, M_{N_0}, S_{N_0}, P_{N_0}, p_{(N_0)}, \theta_0, \nu_0, c_0, r_0, b_0$ have been used at one or several steps to finally define n_0 and d_0.

3.6. Proof of Proposition 5

Theorem 2.1 implies, for $f \in \mathcal{F}$,

$$\mathbb{P}_n^{(N)}(f) - P(f) = \frac{1}{\sqrt{n}} \mathbb{G}_n^{(N)}(f) + \frac{1}{\sqrt{n}} \mathbb{R}_n^{(N)}(f), \tag{3.22}$$

where $\mathbb{G}_n^{(N)}$ is a sequence of versions of the centered Gaussian process $\mathbb{G}^{(N)}$ from (2.3) and the random sequence $r_n^{(N)} = \|\mathbb{R}_n^{(N)}\|_{\mathcal{F}}$ satisfies

$$r_n^{(N)} \leqslant \left\|\mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} + \left\|\alpha_n^{(N)}\right\|_{\mathcal{F}} \leqslant \left\|\mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} + 2M\sqrt{n}, \qquad \lim_{n\to+\infty} \frac{r_n^{(N)}}{v_n} \leqslant d_0 \quad a.s.$$

We have to be a little careful with the expectation, variance and covariance of the coupling error process $\mathbb{R}_n^{(N)}$.

Step 1. Since $\mathbb{G}_n^{(N)}(f)$ is centered the bias is controlled by

$$\sup_{f\in\mathcal{F}} \frac{\sqrt{n}}{v_n} \left|\mathbb{E}\left(\mathbb{P}_n^{(N)}(f)\right) - P(f)\right| = \sup_{f\in\mathcal{F}} \left|\frac{1}{v_n}\mathbb{E}\left(\mathbb{R}_n^{(N)}(f)\right)\right| \leqslant \mathbb{E}\left(\frac{r_n^{(N)}}{v_n}\right). \tag{3.23}$$

Write $a_n = \sqrt{K \log n}$ where $K > 0$ and $\theta_0 > 1$ from Theorem 2.1 can be chosen as large as needed. Then, for $\theta > 1$, $\varepsilon > 0$ and $k \in \mathbb{N}^*$ consider the events

$$\begin{aligned} A_n &= \left\{r_n^{(N)} \leqslant (d_0+\varepsilon)v_n\right\}, \quad B_n = \left\{\left\|\mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} \leqslant a_n\right\}, \\ C_{n,k} &= \left\{\theta^{k-1}a_n < \left\|\mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} \leqslant \theta^k a_n\right\}. \end{aligned}$$

By Theorem 2.1, $\mathbb{P}(A_n^c) < 1/n^{\theta_0}$ and $v_n > a_n/\sqrt{n}$ for all n large enough, hence

$$\begin{aligned} \frac{1}{v_n}\mathbb{E}\left(r_n^{(N)}\right) &= \mathbb{E}\left(\frac{r_n^{(N)}}{v_n} 1_{A_n}\right) + \mathbb{E}\left(\frac{r_n^{(N)}}{v_n} 1_{A_n^c\cap B_n}\right) + \mathbb{E}\left(\frac{r_n^{(N)}}{v_n} 1_{A_n^c\cap B_n^c}\right) \\ &\leqslant d_0 + \varepsilon + \frac{a_n + 2M\sqrt{n}}{v_n}\mathbb{P}(A_n^c) + \mathbb{E}\left(\frac{r_n^{(N)}}{v_n} 1_{B_n^c}\right) \\ &\leqslant d_0 + 2\varepsilon + \sum_{k=1}^{+\infty} \mathbb{E}\left(\frac{r_n^{(N)}}{v_n} 1_{C_{n,k}}\right). \end{aligned}$$

By Propositions 7 and 8, $\mathbb{G}^{(N)}(f)$ is a centered Gaussian process indexed by $\mathcal{F}$ such that, under (VC) or (BR), it holds

$$\begin{aligned} &\mathbb{E}\left(\left\|\mathbb{G}^{(N)}\right\|_{\mathcal{F}}\right) < +\infty, \quad \sup_{f\in\mathcal{F}} \mathbb{V}(\mathbb{G}^{(N)}(f)) \leqslant \sigma_{\mathcal{F}}^2 < +\infty, \\ &\mathbb{E}\left(\left\|\mathbb{G}^{(N)}\right\|_{\mathcal{F}}^2\right) \leqslant C_{\mathcal{F}}^2 = \sigma_{\mathcal{F}}^2 + \mathbb{E}\left(\left\|\mathbb{G}^{(N)}\right\|_{\mathcal{F}}\right)^2 < +\infty. \end{aligned} \tag{3.24}$$

Thus, by Borell's inequality – see Appendix A.2 of [27] – for any version $\mathbb{G}_n^{(N)}$ of $\mathbb{G}^{(N)}$, we have

$$\mathbb{P}\left(\left\|\mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} > \lambda\right) \leqslant 2\exp\left(-\frac{\lambda^2}{8C_{\mathcal{F}}^2}\right). \tag{3.25}$$

Therefore we have, since $\theta > 1$ and $v_n > 4M/\sqrt{n} > 2a_n/n$ for n large enough,

$$\begin{aligned}\mathbb{E}\left(\frac{r_n^{(N)}}{v_n}1_{C_{n,k}}\right) &\leqslant \frac{\theta^k a_n + 2M\sqrt{n}}{v_n}\mathbb{P}\left(C_{n,k}\right) \leqslant \theta^k n\mathbb{P}\left(\left\|\mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} > \theta^{k-1}a_n\right)\\ &\leqslant 2\theta^k n\exp\left(-\frac{(\theta^{k-1}a_n)^2}{8C_{\mathcal{F}}^2}\right),\end{aligned}$$

and the following series is converging to an arbitrarily small sum,

$$\begin{aligned}\sum_{k=1}^{+\infty}\mathbb{E}\left(\frac{r_n^{(N)}}{v_n}1_{C_{n,k}}\right) &\leqslant 2n\exp\left(-\frac{a_n^2}{8C_{\mathcal{F}}^2}\right)\sum_{k=1}^{+\infty}\theta^k\exp\left(-\left(\frac{\theta^{2(k-1)}-1}{8C_{\mathcal{F}}^2}\right)a_n^2\right)\\ &\leqslant n\exp\left(-\frac{K\log n}{8C_{\mathcal{F}}^2}\right)\sum_{k=1}^{+\infty}2e\theta^k\exp\left(-\theta^{2(k-1)}\right) \leqslant \frac{1}{n^\delta},\end{aligned}$$

where $\delta < K/8C_{\mathcal{F}}^2 - 1$. It follows that (3.23) is ultimately bounded by d_0.

Step 2. Starting from (3.22) and the bias and variance decomposition, the quadratic risk is in turn controlled by

$$\begin{aligned}&\mathbb{E}\left((\mathbb{P}_n^{(N)}(f) - P(f))^2\right) - \frac{1}{n}\mathbb{V}\left(\mathbb{G}^{(N)}(f)\right)\\ &= \left|\mathbb{E}\left(\mathbb{P}_n^{(N)}(f)\right) - P(f)\right|^2 + \frac{1}{n}\mathbb{V}\left(\mathbb{R}_n^{(N)}(f)\right) + \frac{2}{n}\mathrm{Cov}\left(\mathbb{G}_n^{(N)}(f), \mathbb{R}_n^{(N)}(f)\right).\end{aligned} \tag{3.26}$$

(i) By Step 1, the first right-hand term is the squared bias, of order $d_0^2 v_n^2/n$. Concerning the second right-hand term in (3.26), we bound $\mathbb{E}(\mathbb{R}_n^{(N)}(f)^2)$. Fix $\varepsilon > 0$ and assume that n is large enough for the following statements. By setting $s_n^{(N)} = (r_n^{(N)})^2$ then using $v_n > a_n/\sqrt{n}$, $a_n = K\sqrt{\log n} < \sqrt{n}$ we get, for $\theta_0 = 2$,

$$\begin{aligned}\frac{1}{v_n^2}\sup_{f\in\mathcal{F}}\mathbb{E}\left(\mathbb{R}_n^{(N)}(f)^2\right) &\leqslant \mathbb{E}\left(\frac{s_n^{(N)}}{v_n^2}1_{A_n}\right) + \mathbb{E}\left(\frac{s_n^{(N)}}{v_n^2}1_{A_n^c\cap B_n}\right) + \mathbb{E}\left(\frac{s_n^{(N)}}{v_n^2}1_{A_n^c\cap B_n^c}\right)\\ &\leqslant (d_0+\varepsilon)^2 + \left(\frac{a_n + 2M\sqrt{n}}{v_n}\right)^2\mathbb{P}(A_n^c) + \mathbb{E}\left(\frac{s_n^{(N)}}{v_n^2}1_{B_n^c}\right)\\ &\leqslant (d_0+2\varepsilon)^2 + \left(\frac{3M\sqrt{n}}{\log n}\right)^2\frac{1}{n^2} + \sum_{k=1}^{+\infty}\mathbb{E}\left(\frac{s_n^{(N)}}{v_n^2}1_{C_{n,k}}\right)\end{aligned}$$

$$
\begin{aligned}
&\leqslant (d_0+3\varepsilon)^2 + \sum_{k=1}^{+\infty} \theta^{2k} n^2 \mathbb{P}\left(\left\|\mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} > \theta^{k-1} a_n\right) \\
&\leqslant (d_0+4\varepsilon)^2,
\end{aligned}
$$

where the series is equal to its first term $n^2 \exp\left(-a_n^2/8C_{\mathcal{F}}^2\right)$ times a convergent series, by using (3.25) as for Step 1 with $K > 16C_{\mathcal{F}}^2$. We have shown that

$$
\limsup_{n\to+\infty} \frac{1}{v_n^2} \sup_{f\in\mathcal{F}} \mathbb{V}\left(\mathbb{R}_n^{(N)}(f)\right) \leqslant \limsup_{n\to+\infty} \frac{1}{v_n^2} \mathbb{E}\left(s_n^{(N)}\right) \leqslant d_0^2. \tag{3.27}
$$

(ii) Concerning the covariance term in (3.26) it holds

$$
\begin{aligned}
\frac{1}{v_n}\left|\mathrm{Cov}\left(\mathbb{G}_n^{(N)}(f), \mathbb{R}_n^{(N)}(f)\right)\right| &= \frac{1}{v_n}\left|\mathbb{E}\left(\mathbb{G}_n^{(N)}(f)\mathbb{R}_n^{(N)}(f)\right)\right| \\
&\leqslant \frac{1}{v_n}\mathbb{E}\left(\left|\mathbb{G}_n^{(N)}(f)\right| r_n^{(N)}\right) \\
&= T_{A_n}(f) + T_{A_n^c\cap B_n}(f) + T_{A_n^c\cap B_n^c}(f),
\end{aligned}
$$

where $T_D(f) = \mathbb{E}(1_D|\mathbb{G}_n^{(N)}(f)|r_n^{(N)}/v_n)$ for $D \in \{A_n, A_n^c \cap B_n, A_n^c \cap B_n^c\}$. We have, by Proposition 7 and 8,

$$
\begin{aligned}
T_{A_n}(f) &\leqslant \mathbb{E}\left(\left|\mathbb{G}_n^{(N)}(f)\right|(d_0+\varepsilon)1_{A_n}\right) \\
&= (d_0+\varepsilon)\sqrt{\mathbb{V}\left(\mathbb{G}_n^{(N)}(f)\right)}\,\mathbb{E}\left(|\mathcal{N}(0,1)|\right) \leqslant \sqrt{\frac{2}{\pi}}(d_0+\varepsilon)\sigma_f.
\end{aligned}
$$

By using again $\mathbb{P}(A_n^c) < 1/n^2$ we see that

$$
T_{A_n^c\cap B_n}(f) \leqslant \mathbb{E}\left(a_n\left(\frac{2M\sqrt{n}+a_n}{v_n}\right)1_{A_n^c\cap B_n}\right) \leqslant a_n\left(\frac{3M\sqrt{n}}{v_n}\right)\frac{1}{n^2} \leqslant \varepsilon.
$$

Lastly, for $g_n^{(N)} = \left\|\mathbb{G}_n^{(N)}\right\|_{\mathcal{F}}$, K large and all n large enough it holds, by (3.24) and (3.25),

$$
\begin{aligned}
T_{A_n^c\cap B_n^c}(f) &\leqslant \mathbb{E}\left(g_n^{(N)}\left(\frac{2M\sqrt{n}+g_n^{(N)}}{v_n}\right)1_{B_n^c}\right) \\
&= \sum_{k=1}^{+\infty}\mathbb{E}\left(g_n^{(N)}\left(\frac{2M\sqrt{n}+g_n^{(N)}}{v_n}\right)1_{C_{n,k}}\right) \\
&\leqslant \sum_{k=1}^{+\infty}\theta^{2k}a_n^2 n\mathbb{P}\left(\left\|\mathbb{G}_n^{(N)}\right\|_{\mathcal{F}} > \theta^{k-1}a_n\right) \leqslant \varepsilon.
\end{aligned}
$$

The above upper bounds imply, by (2.4), (3.26) and (3.27),

$$
\limsup_{n\to+\infty}\frac{n}{v_n}\sup_{f\in\mathcal{F}}\left|\mathbb{E}\left((\mathbb{P}_n^{(N)}(f)-P(f))^2\right) - \frac{1}{n}\mathbb{V}\left(\mathbb{G}^{(N)}(f)\right)\right| \leqslant \sqrt{\frac{8}{\pi}}d_0\sigma_{\mathcal{F}}.
$$

Step 3. Let extend Step 2 to the covariance. By Step 1 we have, for all n large,

$$\begin{aligned}
&\left|\operatorname{Cov}\left(\mathbb{P}_n^{(N)}(f),\mathbb{P}_n^{(N)}(g)\right)-\mathbb{E}\left(\left(\mathbb{P}_n^{(N)}(f)-P(f)\right)\left(\mathbb{P}_n^{(N)}(g)-P(g)\right)\right)\right|\\
&=\left|\left(\mathbb{E}(\mathbb{P}_n^{(N)}(f))-P(f)\right)\left(\mathbb{E}(\mathbb{P}_n^{(N)}(g))-P(g)\right)\right|<d_0^2\frac{v_n^2}{n}.
\end{aligned}$$

Now, by the upper bounds computed at (i) and (ii) of Step 2,

$$\begin{aligned}
&\left|\mathbb{E}\left((\mathbb{P}_n^{(N)}(f)-P(f))(\mathbb{P}_n^{(N)}(g)-P(g))\right)-\frac{1}{n}\operatorname{Cov}\left(\mathbb{G}^{(N)}(f),\mathbb{G}^{(N)}(g)\right)\right|\\
&\leqslant\frac{1}{n}\mathbb{E}\left(\left|\mathbb{G}_n^{(N)}(f)\mathbb{R}_n^{(N)}(g)\right|\right)\\
&\quad+\frac{1}{n}\mathbb{E}\left(\left|\mathbb{G}_n^{(N)}(g)\mathbb{R}_n^{(N)}(f)\right|\right)+\frac{1}{n}\mathbb{E}\left(\left|\mathbb{R}_n^{(N)}(f)\mathbb{R}_n^{(N)}(g)\right|\right)\\
&\leqslant\frac{2}{n}\sup_{f\in\mathcal{F}}\mathbb{E}\left(\left|\mathbb{G}_n^{(N)}(f)\right|r_n^{(N)}\right)+\frac{1}{n}\mathbb{E}\left(s_n^{(N)}\right)\\
&\leqslant\frac{2}{n}(d_0+\varepsilon)\sigma_{\mathcal{F}}\sqrt{\frac{2}{\pi}}v_n+\frac{1}{n}(d_0+\varepsilon)^2v_n^2.
\end{aligned}$$

3.7. Proof of Proposition 6

Fix $N_0\in\mathbb{N}$. Let apply Theorem 2.1 with $\theta_0=2$, from which we also use n_0, d_0 and v_n. We have, for all $0\leqslant N\leqslant N_0$, $\varphi\in\mathcal{L}$, $x\in\mathbb{R}$ and $n\geqslant n_0$,

$$\begin{aligned}
\mathbb{P}\left(\varphi(\alpha_n^{(N)})\leqslant x\right)&\leqslant\frac{1}{n^2}+\mathbb{P}\left(\left\{\varphi(\alpha_n^{(N)})\leqslant x\right\}\cap\left\{\left\|\alpha_n^{(N)}-\mathbb{G}_n^{(N)}\right\|_{\mathcal{F}}<d_0v_n\right\}\right)\\
&\leqslant\frac{1}{n^2}+\mathbb{P}\left(\varphi(\mathbb{G}_n^{(N)})\leqslant x+d_0C_1v_n\right)\\
&\leqslant\frac{1}{n^2}+\mathbb{P}\left(\varphi(\mathbb{G}_n^{(N)})\leqslant x\right)+d_0C_1C_2v_n,
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}\left(\varphi(\mathbb{G}_n^{(N)})\leqslant x-d_0C_1v_n\right)\\
&\leqslant\frac{1}{n^2}+\mathbb{P}\left(\left\{\varphi(\mathbb{G}_n^{(N)})\leqslant x-d_0C_1v_n\right\}\cap\left\{\left\|\alpha_n^{(N)}-\mathbb{G}_n^{(N)}\right\|_{\mathcal{F}}<d_0v_n\right\}\right)\\
&\leqslant\frac{1}{n^2}+\mathbb{P}\left(\varphi(\alpha_n^{(N)})\leqslant\varphi(\mathbb{G}_n^{(N)})+d_0C_1v_n\leqslant x\right),
\end{aligned}$$

so that

$$\mathbb{P}\left(\varphi(\alpha_n^{(N)})\leqslant x\right)\geqslant\mathbb{P}\left(\varphi(\mathbb{G}_n^{(N)})\leqslant x\right)-d_0C_1C_2v_n-\frac{1}{n^2}.$$

This establishes the second statement of Proposition 6 provided $d_1>d_0$ and $n\geqslant n_1\geqslant n_0$ where n_1 is large enough to have $(d_1-d_0C_1C_2)v_n>n^{-2}$. The first statement coincides with the special case $\mathcal{L}=\{\varphi_f:f\in\mathcal{F}_0\}$ where $\varphi_f(g)=g(f)$

are pointwise projectors and we then have a Lipshitz constant $C_1 = 1$ whereas $\varphi_f(\mathbb{G}_n^{(N)}) = \mathbb{G}_n^{(N)}(f)$ has a Gaussian density bounded by

$$\frac{1}{\sqrt{2\pi\mathbb{V}(\mathbb{G}_n^{(N)}(f))}} \leqslant C_2 = \frac{1}{\sqrt{2\pi}\sigma_0} < +\infty.$$

4. Proofs concerning the limiting process

4.1. Proof of Proposition 7

Step 1. Let us first relate $\mathbb{G}^{(N)}(\mathcal{F})$ from (2.3) to $\mathbb{G}(\mathcal{F}) = \mathbb{G}^{(0)}(\mathcal{F})$ from (2.2) by means of the vectors $\Phi_k^{(N)}(f)$ introduced at (2.9) before Proposition 7.

Lemma 3. *For all $N \in \mathbb{N}_*$ and $f \in \mathcal{F}$ it holds*

$$\mathbb{G}^{(N)}(f) = \mathbb{G}(f) - \sum_{k=1}^{N} \Phi_k^{(N)}(f)^t \cdot \mathbb{G}\left[\mathcal{A}^{(k)}\right].$$

Proof. The formula is true for $N = 0$. Assume that it is the case for $N \geqslant 0$. Recall that sets $A \in \mathcal{A}_{\cup}^{(N_0)}$ from (1.9) are identified to $f = 1_A$. By (2.3),

$$\begin{aligned}
\mathbb{G}^{(N+1)}(f) &= \mathbb{G}^{(N)}(f) - \mathbb{E}\left[f \mid \mathcal{A}^{(N+1)}\right]^t \cdot \mathbb{G}^{(N)}\left[\mathcal{A}^{(N+1)}\right] \\
&= \mathbb{G}(f) - \sum_{k=1}^{N} \Phi_k^{(N)}(f)^t \cdot \mathbb{G}\left[\mathcal{A}^{(k)}\right] \\
&\quad - \sum_{j=1}^{m_{N+1}} \mathbb{E}(f \mid A_j^{(N+1)}) \left(\mathbb{G}(A_j^{(N+1)}) - \sum_{k=1}^{N} \Phi_k^{(N)}(A_j^{(N+1)})^t \cdot \mathbb{G}\left[\mathcal{A}^{(k)}\right]\right) \\
&= \mathbb{G}(f) - \mathbb{E}\left[f \mid \mathcal{A}^{(N+1)}\right]^t \cdot \mathbb{G}\left[\mathcal{A}^{(N+1)}\right] \\
&\quad - \sum_{k=1}^{N} \left[\Phi_k^{(N)}(f) - \Phi_k^{(N)}[\mathcal{A}^{(N+1)}] \cdot \mathbb{E}[f \mid \mathcal{A}^{(N+1)}]\right]^t \cdot \mathbb{G}\left[\mathcal{A}^{(k)}\right],
\end{aligned}$$

where the $m_k \times m_{N+1}$ matrix $\Phi_k^{(N)}[\mathcal{A}^{(N+1)}] = (\Phi_k^{(N)}(A_1^{(N+1)}), \ldots, \Phi_k^{(N)}(A_{m_{N+1}}^{(N+1)}))$ satisfies

$$\sum_{j=1}^{m_{N+1}} \mathbb{E}(f \mid A_j^{(N+1)}) \Phi_k^{(N)}(A_j^{(N+1)})^t = \left[\Phi_k^{(N)}[\mathcal{A}^{(N+1)}] \cdot \mathbb{E}[f \mid \mathcal{A}^{(N+1)}]\right]^t.$$

Now observe that $\Phi_{N+1}^{(N+1)}(f) = \mathbb{E}[f \mid \mathcal{A}^{(N+1)}]$ by the definition of $\Phi_k^{(N+1)}$ given by (2.9) when $k = N+1$. It remains to show that

$$\Phi_k^{(N)}(f) - \Phi_k^{(N)}[\mathcal{A}^{(N+1)}] \cdot \mathbb{E}[f \mid \mathcal{A}^{(N+1)}] = \Phi_k^{(N+1)}(f). \tag{4.1}$$

For $1 \leqslant k \leqslant N$ and $1 \leqslant j \leqslant m_{N+1}$ we have

$$\Phi_k^{(N)}(A_j^{(N+1)}) = \mathbb{E}\left[A_j^{(N+1)} \mid \mathcal{A}^{(k)}\right] + \sum_{\substack{1\leqslant L\leqslant N-k \\ k<l_1<\ldots<l_L\leqslant N}} (-1)^L \mathbf{P}_{\mathcal{A}^{(l_1)}|\mathcal{A}^{(k)}} \mathbf{P}_{\mathcal{A}^{(l_2)}|\mathcal{A}^{(l_1)}} \ldots \mathbf{P}_{\mathcal{A}^{(l_L)}|\mathcal{A}^{(l_{L-1})}} \cdot \mathbb{E}\left[A_j^{(N+1)}|\mathcal{A}^{(l_L)}\right],$$

where, for $l = k, k+1, ..., N$ the vector

$$\mathbb{E}\left[A_j^{(N+1)} \mid \mathcal{A}^{(l)}\right] = \left(\frac{P(A_j^{(N+1)} \cap A_1^{(l)})}{P(A_1^{(l)})}, ..., \frac{P(A_j^{(N+1)} \cap A_{m_l}^{(l)})}{P(A_{m_l}^{(l)})}\right)^t,$$

is also the j-th column of $\mathbf{P}_{\mathcal{A}^{(N+1)}|\mathcal{A}^{(l)}}$. Therefore, by turning L into $L' = L+1$,

$$\begin{aligned}
&- \Phi_k^{(N)}\left[\mathcal{A}^{(N+1)}\right] \cdot \mathbb{E}\left[f \mid \mathcal{A}^{(N+1)}\right] \\
&= - \sum_{j=1}^{m_{N+1}} \mathbb{E}(f \mid A_j^{(N+1)}) \Phi_k^{(N)}(A_j^{(N+1)}) \\
&= - \sum_{j=1}^{m_{N+1}} \mathbb{E}(f \mid A_j^{(N+1)}) \mathbb{E}\left[A_j^{(N+1)} \mid \mathcal{A}^{(k)}\right] \\
&+ \sum_{j=1}^{m_{N+1}} \mathbb{E}(f \mid A_j^{(N+1)}) \sum_{\substack{1\leqslant L\leqslant N-k \\ k<l_1<\ldots<l_L\leqslant N}} (-1)^{L+1} \mathbf{P}_{\mathcal{A}^{(l_1)}|\mathcal{A}^{(k)}} \ldots \mathbf{P}_{\mathcal{A}^{(l_L)}|\mathcal{A}^{(l_{L-1})}} \\
&\quad \cdot \mathbb{E}\left[A_j^{(N+1)}|\mathcal{A}^{(l_L)}\right] \\
&= (-1)^1 \mathbf{P}_{\mathcal{A}^{(N+1)}|\mathcal{A}^{(k)}} \cdot \mathbb{E}\left[f|\mathcal{A}^{(N+1)}\right] \\
&+ \sum_{\substack{1\leqslant L'\leqslant N+1-k \\ k<l_1<\ldots<l_{L'}=N+1}} (-1)^{L'} \mathbf{P}_{\mathcal{A}^{(l_1)}|\mathcal{A}^{(k)}} \ldots \mathbf{P}_{\mathcal{A}^{(l_{L'-1})}|\mathcal{A}^{(l_{L'-2})}} \mathbf{P}_{\mathcal{A}^{(N+1)}|\mathcal{A}^{(l_{L'-1})}} \\
&\quad \cdot \mathbb{E}\left[f|\mathcal{A}^{(N+1)}\right],
\end{aligned}$$

where all terms are different from those in

$$\begin{aligned}
\Phi_k^{(N)}(f) = \mathbb{E}\left[f \mid \mathcal{A}^{(k)}\right] + &\sum_{\substack{1\leqslant L'\leqslant N+1-k \\ k<l_1<\ldots<l_{L'}<N+1}} (-1)^L \mathbf{P}_{\mathcal{A}^{(l_1)}|\mathcal{A}^{(k)}} \ldots \mathbf{P}_{\mathcal{A}^{(l_{L'})}|\mathcal{A}^{(l_{L'-1})}} \\
&\cdot \mathbb{E}\left[f|\mathcal{A}^{(l_{L'})}\right].
\end{aligned}$$

Having collected all terms of $\Phi_k^{(N)}(f)$ in (2.9), this establishes (4.1). The proof is completed by induction. $\square$

The functions $\Phi_k^{(N)}$ and the process $\mathbb{G}$ are linear, hence Lemma 3 implies that $\mathbb{G}^{(N)}$ is a linear process. Moreover $\mathbb{G}(f)$ and $\mathbb{G}[\mathcal{A}^{(k)}]$ being centered Gaussian, Lemma 3 proves that $\mathbb{G}^{(N)}(f)$ is a centered Gaussian random variable.

Step 2. To compute the covariance of $\mathbb{G}^{(N)}(\mathcal{F})$ we need the following properties. Recall that $\mathbf{P}_{\mathcal{A}^{(k)}|\mathcal{A}^{(k)}} = \mathrm{Id}_{m_k}$ is the identity matrix of $\mathbb{R}^{m_k}$.

Lemma 4. *For $1 \leqslant k, l \leqslant N$ and $f \in \mathcal{F}$ we have*

$$\mathrm{Cov}\left(\mathbb{G}[\mathcal{A}^{(k)}], \mathbb{G}(f)\right) = \mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right) \cdot \mathbb{E}[f \mid \mathcal{A}^{(k)}], \tag{4.2}$$

$$\mathrm{Cov}\left(\mathbb{G}[\mathcal{A}^{(k)}], \mathbb{G}[\mathcal{A}^{(l)}]\right) = \mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right) \mathbf{P}_{\mathcal{A}^{(l)}|\mathcal{A}^{(k)}}, \tag{4.3}$$

$$\Phi_k^{(N)}(f) = \mathbb{E}\left[f \mid \mathcal{A}^{(k)}\right] - \sum_{k<l\leqslant N} \mathbf{P}_{\mathcal{A}^{(l)}|\mathcal{A}^{(k)}} \cdot \Phi_l^{(N)}(f). \tag{4.4}$$

Proof. The j-th coordinate of the vector $\mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right) \cdot \mathbb{E}[f \mid \mathcal{A}^{(k)}]$ is

$$\begin{aligned}
&P(A_j^{(k)})(1 - P(A_j^{(k)}))\mathbb{E}(f \mid A_j^{(k)}) - \sum_{j \neq i \leqslant m_k} P(A_i^{(k)})P(A_j^{(k)})\mathbb{E}(f \mid A_i^{(k)}) \\
&= \mathbb{E}(1_{A_j^{(k)}} f) - P(A_j^{(k)}) \sum_{1 \leqslant i \leqslant m_k} \mathbb{E}(1_{A_i^{(k)}} f) \\
&= \mathrm{Cov}\left(\mathbb{G}(A_j^{(k)}), \mathbb{G}(f)\right),
\end{aligned}$$

which proves (4.2). Likewise the (i,j)-th coordinate of the matrix $\mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right) \cdot \mathbf{P}_{\mathcal{A}^{(l)}|\mathcal{A}^{(k)}}$ is

$$\begin{aligned}
&P(A_i^{(k)})(1 - P(A_i^{(k)}))P(A_j^{(l)} \mid A_i^{(k)}) - \sum_{j \neq m \leqslant m_k} P(A_i^{(k)})P(A_m^{(k)})P(A_j^{(l)} \mid A_m^{(k)}) \\
&= P(A_j^{(l)} \cap A_i^{(k)}) - P(A_i^{(k)}) \sum_{1 \leqslant m \leqslant m_k} P(A_j^{(l)} \cap A_m^{(k)}) \\
&= \mathrm{Cov}\left(\mathbb{G}(A_i^{(k)}), \mathbb{G}(A_j^{(l)})\right),
\end{aligned}$$

which proves (4.3). By the definition (2.9) of the vectors $\Phi_l^{(N)}(f)$ we get

$$\begin{aligned}
&\sum_{k<l\leqslant N} \mathbf{P}_{\mathcal{A}^{(l)}|\mathcal{A}^{(k)}} \cdot \Phi_l^{(N)}(f) \\
&= \sum_{k<l\leqslant N} \mathbf{P}_{\mathcal{A}^{(l)}|\mathcal{A}^{(k)}} \cdot \mathbb{E}\left[f \mid \mathcal{A}^{(l)}\right] \\
&+ \sum_{\substack{k<l\leqslant N, 1\leqslant L\leqslant N-l \\ l<l_1<\ldots<l_L\leqslant N}} (-1)^L \mathbf{P}_{\mathcal{A}^{(l)}|\mathcal{A}^{(k)}} \mathbf{P}_{\mathcal{A}^{(l_1)}|\mathcal{A}^{(l)}} \ldots \mathbf{P}_{\mathcal{A}^{(l_L)}|\mathcal{A}^{(l_{L-1})}} \cdot \mathbb{E}\left[f \mid \mathcal{A}^{(l_L)}\right]
\end{aligned}$$

$$= \sum_{\substack{1\leqslant L\leqslant N-k \\ k<l_1<\ldots<l_L\leqslant N}} (-1)^{L+1}\mathbf{P}_{\mathcal{A}^{(l_1)}|\mathcal{A}^{(k)}}\ldots\mathbf{P}_{\mathcal{A}^{(l_L)}|\mathcal{A}^{(l_{L-1})}}\cdot\mathbb{E}\left[f\mid\mathcal{A}^{(l_L)}\right]$$
$$= \mathbb{E}\left[f\mid\mathcal{A}^{(k)}\right]-\Phi_k^{(N)}(f),$$

which yields (4.4). $\square$

Step 3. Let us first compute the variance of $\mathbb{G}^{(N)}(f)$. By Lemma 3 we have

$$\mathbb{V}\left(\mathbb{G}^{(N)}(f)\right)-\mathbb{V}\left(\mathbb{G}(f)\right)$$
$$=\sum_{k=1}^{N}\Phi_k^{(N)}(f)^t\cdot\mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right)\cdot\Phi_k^{(N)}(f)-2\sum_{k=1}^{N}\Phi_k^{(N)}(f)^t\cdot\mathrm{Cov}\left(\mathbb{G}[\mathcal{A}^{(k)}],\mathbb{G}(f)\right)$$
$$+2\sum_{1\leqslant k<l\leqslant N}\Phi_k^{(N)}(f)^t\cdot\mathrm{Cov}\left(\mathbb{G}[\mathcal{A}^{(k)}],\mathbb{G}[\mathcal{A}^{(l)}]\right)\cdot\Phi_l^{(N)}(f),$$

hence Lemma 4 gives, through (4.2) and (4.3),

$$\mathbb{V}\left(\mathbb{G}^{(N)}(f)\right)-\mathbb{V}\left(\mathbb{G}(f)\right)=\sum_{k=1}^{N}\Phi_k^{(N)}(f)^t\cdot\mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right)\cdot\Psi_k^{(N)}(f),$$

where, by (4.4),

$$\Psi_k^{(N)}(f)=\Phi_k^{(N)}(f)-2\mathbb{E}\left[f\mid\mathcal{A}^{(k)}\right]+2\sum_{k<l\leqslant N}\mathbf{P}_{\mathcal{A}^{(l)}|\mathcal{A}^{(k)}}\cdot\Phi_l^{(N)}(f)=-\Phi_k^{(N)}(f). \tag{4.5}$$

The formula (2.10) is proved. It extends to the covariance since, by Lemma 3,

$$\mathrm{Cov}(\mathbb{G}^{(N)}(f),\mathbb{G}^{(N)}(g))-\mathrm{Cov}(\mathbb{G}(f),\mathbb{G}(g))$$
$$=\frac{1}{2}\left(\Upsilon_N(f,g)-2\Gamma_N(f,g)\right)+\frac{1}{2}\left(\Upsilon_N(g,f)-2\Gamma_N(g,f)\right),$$

where, by (4.2) and (4.3) again,

$$\Upsilon_N(f,g)=\sum_{1\leqslant k\leqslant l\leqslant N}\Phi_k^{(N)}(f)^t\cdot\mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right)\mathbf{P}_{\mathcal{A}^{(l)}|\mathcal{A}^{(k)}}\cdot\Phi_l^{(N)}(g),$$
$$\Gamma_N(f,g)=\sum_{l=1}^{N}\Phi_k^{(N)}(f)^t\cdot\mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right)\cdot\mathbb{E}\left[g\mid\mathcal{A}^{(k)}\right].$$

By replacing $\Psi_k^{(N)}(g)$ with $-\Phi_k^{(N)}(g)$ according to (4.5), we obtain

$$\frac{1}{2}\left(\Upsilon_N(f,g)-2\Gamma_N(f,g)\right)=-\frac{1}{2}\sum_{k=1}^{N}\Phi_k^{(N)}(f)^t\cdot\mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right)\cdot\Phi_k^{(N)}(g),$$

which is symmetric in f and g. The covariance formula of Proposition 7 is proved.

4.2. Proof of Propositions 8 and 9

Since $\mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right)$ is semi-definite positive, for all $1 \leqslant k \leqslant N$ and $f \in \mathcal{F}$ we have

$$\Phi_k^{(N)}(f)^t \cdot \mathbb{V}(\mathbb{G}[\mathcal{A}^{(k)}]) \cdot \Phi_k^{(N)}(f) \geqslant 0,$$

and the variance part (2.10) of Proposition 8 follows from Proposition 7. For any $m \in \mathbb{N}_*$, $(f_1, ..., f_m) \in \mathcal{F}^m$ and $u \in \mathbb{R}^m$, it further holds,by Proposition 7 again,

$$\begin{aligned}
&u^t \left(\Sigma_m^{(0)} - \Sigma_m^{(N)}\right) u \\
&= \sum_{1\leqslant i,j\leqslant m} u_i u_j \left(\mathrm{Cov}(\mathbb{G}(f_i), \mathbb{G}(f_j)) - \mathrm{Cov}(\mathbb{G}^{(N)}(f_i), \mathbb{G}^{(N)}(f_j))\right) \\
&= \sum_{k=1}^{N} \sum_{1\leqslant i,j\leqslant m} \left(u_i \Phi_k^{(N)}(f_i)\right)^t \cdot \mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right) \cdot \left(u_j \Phi_k^{(N)}(f_j)\right) \\
&= \sum_{k=1}^{N} \left(\sum_{1\leqslant i\leqslant m} u_i \Phi_k^{(N)}(f_i)\right)^t \cdot \mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right) \cdot \left(\sum_{1\leqslant j\leqslant m} u_j \Phi_k^{(N)}(f_j)\right) \geqslant 0.
\end{aligned}$$

Under the wrapping hypothesis of Proposition 9 we have

$$\Phi_{N_0-k}^{(N_0)} = \Phi_{N_1-k}^{(N_1)}, \quad 0 \leqslant k < N_0,$$

since the corresponding (2.9) only involves $\mathcal{A}^{(N_0-k)} = \mathcal{A}^{(N_1-k)}$ for $0 \leqslant k < N_0$. Assuming moreover $N_1 \geqslant 2N_0$ we get, by Proposition 7,

$$\begin{aligned}
&\mathbb{V}\left(\mathbb{G}^{(N_0)}(f)\right) - \mathbb{V}\left(\mathbb{G}^{(N_1)}(f)\right) \\
&= \sum_{k=1}^{N_1} \Phi_k^{(N_1)}(f)^t \cdot \mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right) \cdot \Phi_k^{(N_1)}(f) - \sum_{k=1}^{N_0} \Phi_k^{(N_0)}(f)^t \cdot \mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right) \cdot \Phi_k^{(N_0)}(f) \\
&= \sum_{k=N_0+1}^{N_1-N_0} \Phi_k^{(N_1)}(f)^t \cdot \mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right) \cdot \Phi_k^{(N_1)}(f) \geqslant 0,
\end{aligned}$$

and, for $m \in \mathbb{N}_*$, $(f_1, ..., f_m) \in \mathcal{F}^m$ and $u \in \mathbb{R}^m$,

$$\begin{aligned}
&u^t \cdot \left(\Sigma_m^{(N_0)} - \Sigma_m^{(N_1)}\right) \cdot u \\
&= u^t \cdot \left((\Sigma_m^{(0)} - \Sigma_m^{(N_1)}) - (\Sigma_m^{(0)} - \Sigma_m^{(N_0)})\right) \cdot u \\
&= \sum_{k=N_0+1}^{N_1-N_0} \left(\sum_{1\leqslant i\leqslant m} u_i \Phi_k^{(N)}(f_i)\right)^t \cdot \mathbb{V}\left(\mathbb{G}[\mathcal{A}^{(k)}]\right) \cdot \left(\sum_{1\leqslant j\leqslant m} u_j \Phi_k^{(N)}(f_j)\right) \geqslant 0.
\end{aligned}$$

4.3. Proof of Proposition 10

We show the result by double induction. For $m=0$ we have $\mathbb{G}^{(0)}(f)=\mathbb{G}(f)$ and, by (2.3), $\mathbb{G}^{(1)}(f)=\mathbb{G}(f)-\mathbb{E}[f|\mathcal{A}]^t\cdot\mathbb{G}[\mathcal{A}]$. Assume that (2.15) and (2.16) are true for $m\in\mathbb{N}$. For $m+1$ we have, by the raking ratio transform (2.3),

$$\mathbb{G}^{(2m+2)}(f)=\mathbb{G}^{(2m+1)}(f)-\mathbb{E}[f|\mathcal{B}]^t\cdot\mathbb{G}^{(2m+1)}[\mathcal{B}], \tag{4.6}$$

$$\mathbb{G}^{(2m+3)}(f)=\mathbb{G}^{(2m+2)}(f)-\mathbb{E}[f|\mathcal{A}]^t\cdot\mathbb{G}^{(2m+2)}[\mathcal{A}]. \tag{4.7}$$

For $1\leqslant j\leqslant m_2$ and $f=1_{B_j}$ we get, by (2.11), (2.12), (2.13), (2.14) and (2.16),

$$\begin{aligned}
&\mathbb{G}^{(2m+1)}(B_j)\\
&=\mathbb{G}(B_j)-S_{1,\text{odd}}^{(m-1)}(1_{B_j})^t\cdot\mathbb{G}[\mathcal{A}]-S_{2,\text{odd}}^{(m-1)}(1_{B_j})^t\cdot\mathbb{G}[\mathcal{B}]\\
&=\mathbb{G}(B_j)-\left(\sum_{k=0}^{m-1}\left(\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}\right)^k\cdot\left(\mathbb{E}[1_{B_j}|\mathcal{B}]-\mathbf{P}_{\mathcal{A}|\mathcal{B}}\cdot\mathbb{E}[1_{B_j}|\mathcal{A}]\right)\right)^t\cdot\mathbb{G}[\mathcal{B}]\\
&-\left(\sum_{k=0}^{m-1}\left(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}\right)^k\cdot\left(\mathbb{E}[1_{B_j}|\mathcal{A}]-\mathbf{P}_{\mathcal{B}|\mathcal{A}}\cdot\mathbb{E}[1_{B_j}|\mathcal{B}]\right)\right.\\
&\quad\left.+\left(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}\right)^m\cdot\mathbb{E}[1_{B_j}|\mathcal{A}]\right)^t\cdot\mathbb{G}[\mathcal{A}],
\end{aligned}$$

where $\mathbb{E}[1_{B_j}|\mathcal{A}]$ is the j-th column of $\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ and $\mathbb{E}[1_{B_j}|\mathcal{B}]$ is the j-th unit vector of $\mathbb{R}^{m_2}$. Therefore

$$\begin{aligned}
&\mathbb{G}^{(2m+1)}[\mathcal{B}]\\
&=\mathbb{G}[\mathcal{B}]-\left(\sum_{k=0}^{m-1}(\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k(\mathrm{Id}_{m_2}-\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})\right)^t\cdot\mathbb{G}[\mathcal{B}]\\
&\quad-\left(\sum_{k=0}^{m-1}(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k(\mathbf{P}_{\mathcal{B}|\mathcal{A}}-\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathrm{Id}_{m_2})+(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^m\mathbf{P}_{\mathcal{B}|\mathcal{A}}\right)^t\cdot\mathbb{G}[\mathcal{A}]\\
&=\left((\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^m\right)^t\cdot\mathbb{G}[\mathcal{B}]-\left((\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^m\mathbf{P}_{\mathcal{B}|\mathcal{A}}\right)^t\cdot\mathbb{G}[\mathcal{A}],
\end{aligned}$$

Finally (2.15) and (4.6) then again (2.11), (2.12), (2.13) and (2.14) together imply

$$\begin{aligned}
&\mathbb{G}^{(2m+2)}(f)\\
&=\mathbb{G}(f)-S_{1,\text{odd}}^{(m-1)}(f)^t\cdot\mathbb{G}[\mathcal{A}]-S_{2,\text{odd}}^{(m-1)}(f)^t\cdot\mathbb{G}[\mathcal{B}]-\mathbb{E}[f|\mathcal{B}]^t\cdot\mathbb{G}^{(2m+1)}[\mathcal{B}]\\
&=\mathbb{G}(f)-\left(S_{1,\text{odd}}^{(m-1)}(f)-(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^m\mathbf{P}_{\mathcal{B}|\mathcal{A}}\cdot\mathbb{E}[f|\mathcal{B}]\right)^t\cdot\mathbb{G}[\mathcal{A}]\\
&\quad-\left(S_{2,\text{odd}}^{(m-1)}(f)+(\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^m\cdot\mathbb{E}[f|\mathcal{B}]\right)^t\cdot\mathbb{G}[\mathcal{B}]\\
&=\mathbb{G}(f)-S_{1,\text{even}}^{(m)}(f)^t\cdot\mathbb{G}[\mathcal{A}]-S_{2,\text{even}}^{(m-1)}(f)^t\cdot\mathbb{G}[\mathcal{B}],
\end{aligned}$$

and (2.15) is valid for $m+1$. If $1 \leqslant i \leqslant m_1$ then $\mathbb{E}[1_{A_i}|\mathcal{B}]$ is the i-th column of $\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ and $\mathbb{E}[1_{A_i}|\mathcal{A}]$ is the i-th unit vector of $\mathbb{R}^{m_1}$ thus (2.15) for $m+1$ and $f = 1_{B_i}$ in turn entails

$$\begin{aligned}
&\mathbb{G}^{(2m+2)}[\mathcal{A}] \\
&= \mathbb{G}[\mathcal{A}] - \left(\sum_{k=0}^{m} \left(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}\right)^k \left(\mathrm{Id}_{m_1} - \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}\right)\right)^t \cdot \mathbb{G}[\mathcal{A}] \\
&\quad - \left(\sum_{k=0}^{m-1} \left(\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}\right)^k \left(\mathbf{P}_{\mathcal{A}|\mathcal{B}} - \mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathrm{Id}_{m_1}\right) + \left(\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}\right)^m \mathbf{P}_{\mathcal{A}|\mathcal{B}}\right)^t \cdot \mathbb{G}[\mathcal{B}] \\
&= \left((\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^{m+1}\right)^t \cdot \mathbb{G}[\mathcal{A}] - \left((\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^m \mathbf{P}_{\mathcal{A}|\mathcal{B}}\right)^t \cdot \mathbb{G}[\mathcal{B}],
\end{aligned}$$

and also, thanks to (2.16) and (4.7),

$$\begin{aligned}
&\mathbb{G}^{(2m+3)}(f) \\
&= \mathbb{G}(f) - \left(S_{1,\mathrm{odd}}^{(m)}(f) - (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^{m+1} \cdot \mathbb{E}[f|\mathcal{A}]\right)^t \cdot \mathbb{G}[\mathcal{A}] \\
&\quad - \left(S_{2,\mathrm{odd}}^{(m-1)}(f) + \left(\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}\right)^m \cdot \mathbb{E}[f|\mathcal{B}]\right)^t \cdot \mathbb{G}[\mathcal{B}] - \mathbb{E}[f|\mathcal{A}]^t \cdot \mathbb{G}^{(2m+2)}[\mathcal{A}] \\
&= \mathbb{G}(f) - S_{1,\mathrm{odd}}^{(m)}(f)^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,\mathrm{odd}}^{(m)}(f)^t \cdot \mathbb{G}[\mathcal{B}],
\end{aligned}$$

which is (2.16) for $m+1$.

4.4. Proof of Proposition 11

Step 1. For $m \geqslant 1$ let $\mathbf{0}_{m,m}$ be the $m \times m$ null matrix. Also recall the vectors $P(\mathcal{A}) = (P(A_1), \ldots, P(A_{m_1}))$ and $P(\mathcal{B}) = (P(B_1), \ldots, P(B_{m_2}))$.

Lemma 5. *Assume (ER). For $l = 1, 2$ there exists an invertible $m_l \times m_l$ matrix U_l and an upper triangular $(m_l - 1) \times (m_l - 1)$ matrix T_l such that*

$$\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}} = U_1 \begin{pmatrix} 1 & \mathbf{0}_{1,m_1-1} \\ \mathbf{0}_{m_1-1,1} & T_1 \end{pmatrix} U_1^{-1}, \quad \lim_{k\to+\infty} T_1^k = \mathbf{0}_{m_1-1,m_1-1},$$
$$\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}} = U_2 \begin{pmatrix} 1 & \mathbf{0}_{1,m_2-1} \\ \mathbf{0}_{m_2-1,1} & T_2 \end{pmatrix} U_2^{-1}, \quad \lim_{k\to+\infty} T_2^k = \mathbf{0}_{m_2-1,m_2-1}.$$

Proof. Since $\mathcal{A}$ is a partition, for $1 \leqslant i \leqslant m_2$ the sum of the m_1 terms of row i of $\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ is $\sum_{j=1}^{m_1} P(A_j \mid B_i) = 1$ hence $\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ is stochastic. Likewise $\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ is stochastic and, by stability, so are $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ and $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$. Let the column of 1's associated to their eigenvalue 1 be in first position in their respective matrix U_1, U_2 of eigenvectors. The announced decomposition is always true with some upper triangular matrices T_l having Jordan decomposition $T_l = D_l + N_l$ where $D_l = Q_l \Delta_l Q_l^{-1}$, Δ_l is a diagonal $(m_l - 1) \times (m_l - 1)$ matrix, Q_l is an invertible $(m_l - 1) \times (m_l - 1)$ matrix and N_l is a nilpotent $(m_l - 1) \times (m_l - 1)$ matrix of order $n_l \geqslant 1$ that commute with D_l. Next observe that

$$\begin{aligned}(P(\mathcal{A}) \cdot \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})_k &= \sum_{i=1}^{m_1} P(A_i)(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})_{i,k} \\ &= \sum_{i=1}^{m_1} P(A_i) \sum_{j=1}^{m_2} (\mathbf{P}_{\mathcal{B}|\mathcal{A}})_{i,j}(\mathbf{P}_{\mathcal{A}|\mathcal{B}})_{j,k} \\ &= \sum_{j=1}^{m_2}\sum_{i=1}^{m_1} P(A_i)P(B_j|A_i)P(A_k|B_j) = P(A_k),\end{aligned}$$

which proves that $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ has invariant probability $P(\mathcal{A})$. Similarly, $P(\mathcal{B})$ is invariant for $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$, and the first line of U_1^{-1} and U_2^{-1} is $P[\mathcal{A}]$ and $P[\mathcal{B}]$ respectively. Under (ER) these matrices are ergodic, which ensures that the eigenvalues of Δ_l have moduli strictly less than the dominant 1 since it is the case of eigenvalues of T_l hence D_l. It follows that

$$\lim_{k\to+\infty} \Delta_l^k = \mathbf{0}_{m_l-1,m_l-1}, \quad l = 1, 2.$$

Furthermore, since N_l and D_l commute it holds

$$T_l^k = \sum_{j=0}^{n_l-1} \binom{k}{j} N_l^j D_l^{k-j}, \quad l = 1, 2, \quad k \geqslant n_l.$$

We conclude that $\lim_{k\to+\infty} T_l^k = \mathbf{0}_{m_l-1,m_l-1}$. □

Step 2. Let $V_1(f) = (\mathbb{E}[f|\mathcal{A}] - \mathbf{P}_{\mathcal{B}|\mathcal{A}} \cdot \mathbb{E}[f|\mathcal{B}])$ and $V_2(f) = (\mathbb{E}[f|\mathcal{B}] - \mathbf{P}_{\mathcal{A}|\mathcal{B}} \cdot \mathbb{E}[f|\mathcal{A}])$.

Lemma 6. *Under (ER) we have*

$$\lim_{k\to+\infty} (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k \cdot V_1(f) = \mathbf{0}_{m_1,1}, \ \lim_{k\to+\infty} (\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k \cdot V_2(f) = \mathbf{0}_{m_2,1}.$$

Proof. By Lemma 5 we have

$$\lim_{k\to+\infty} (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k = \begin{pmatrix} P(\mathcal{A}) \\ \vdots \\ P(\mathcal{A}) \end{pmatrix}, \ \lim_{k\to+\infty} (\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k = \begin{pmatrix} P(\mathcal{B}) \\ \vdots \\ P(\mathcal{B}) \end{pmatrix}. \tag{4.8}$$

The scalar product of $P(\mathcal{A})$ by $V_1(f)$ is null since $P(\mathcal{A}) \cdot \mathbb{E}[f|\mathcal{A}] = P(f)$ and

$$\begin{aligned}P(\mathcal{A}) \cdot \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbb{E}[f|\mathcal{B}]) &= \sum_{j=1}^{m_1} P(A_j) \sum_{k=1}^{m_2} P(B_k \mid A_j)\mathbb{E}(f|B_k) \\ &= \sum_{j=1}^{m_1}\sum_{k=1}^{m_2} P(A_j \cap B_k)\mathbb{E}[f|B_k] = P(f).\end{aligned}$$

Likewise we get $P(\mathcal{B}) \cdot V_2(f) = 0$. □

The following convergences hold for any matrix norm. By Lemma 5 we have

$$\sum_{k=0}^{N}(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k = U_1\begin{pmatrix} N+1 & \mathbf{0}_{1,m_1-1} \\ \mathbf{0}_{m_1-1,1} & \sum_{k=0}^{N} T_1^k \end{pmatrix} U_1^{-1},$$
$$\sum_{k=0}^{N}(\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k = U_2\begin{pmatrix} N+1 & \mathbf{0}_{1,m_2-1} \\ \mathbf{0}_{m_2-1,1} & \sum_{k=0}^{N} T_2^k \end{pmatrix} U_2^{-1}.$$

Now, the matrices $\mathrm{Id}_{m_1-1} - T_1$ and $\mathrm{Id}_{m_2-1} - T_2$ are nonsingular since 1 is a dominant eigenvalue of $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ and $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$. Recalling (2.11) and (2.12), by Lemma 5 and 6 it follows that

$$S_{1,\mathrm{even}}^{(N)}(f) = U_1\begin{pmatrix} 0 & \mathbf{0}_{1,m_1-1} \\ \mathbf{0}_{m_1-1,1} & (\mathrm{Id}_{m_1-1} - T_1)^{-1}(\mathrm{Id}_{m_1-1} - T_1^{N+1}) \end{pmatrix} U_1^{-1}\cdot V_1(f),$$
$$S_{2,\mathrm{odd}}^{(N)}(f) = U_2\begin{pmatrix} 0 & \mathbf{0}_{1,m_2-1} \\ \mathbf{0}_{m_2-1,1} & (\mathrm{Id}_{m_2-1} - T_2)^{-1}(\mathrm{Id}_{m_2-1} - T_2^{N+1}) \end{pmatrix} U_2^{-1}\cdot V_2(f),$$

which, by Lemma 5, converge respectively to

$$S_{1,\mathrm{even}}(f) = U_1\begin{pmatrix} 0 & \mathbf{0}_{1,m_1-1} \\ \mathbf{0}_{m_1-1,1} & (\mathrm{Id}_{m_1-1} - T_1)^{-1} \end{pmatrix} U_1^{-1}\cdot V_1(f),$$
$$S_{2,\mathrm{odd}}(f) = U_2\begin{pmatrix} 0 & \mathbf{0}_{1,m_2-1} \\ \mathbf{0}_{m_2-1,1} & (\mathrm{Id}_{m_2-1} - T_2)^{-1} \end{pmatrix} U_2^{-1}\cdot V_2(f).$$

Since we have already seen by using (4.8) and the notations of Proposition 11 that

$$\lim_{k\to+\infty}(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k\cdot\mathbb{E}[f|\mathcal{A}] = P_1[f], \quad \lim_{k\to+\infty}(\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k\cdot\mathbb{E}[f|\mathcal{B}] = P_2[f],$$

we conclude by (2.13) and (2.14) that $S_{1,\mathrm{odd}}^{(N)}(f)$, $S_{2,\mathrm{even}}^{(N)}(f)$ converge to the vectors $S_{1,\mathrm{odd}}(f) = S_{1,\mathrm{even}}(f)+P_1[f]$, $S_{2,\mathrm{even}}(f) = S_{1,\mathrm{odd}}(f)+P_1[f]$ respectively.

Step 3. Given the spectral radius $\rho(T_l) < 1$ of T_l let $\lambda_l = \rho(T_l)+\varepsilon < 1$, $l = 1,2$ for any $\varepsilon > 0$. Then there exists a vector norm $\|\cdot\|_l$ on $\mathbb{C}^{m_l-1}$ such that its induced matrix norm $|||\cdot|||_l$ on matrices $(m_l-1)\times(m_l-1)$ satisfies $|||T_l|||_l \leqslant \lambda_l$. Introduce the vector norm $\|(x_1,...,x_{m_l})^t\|_l' = |x_1|+\|(x_2,...,x_{m_l})^t\|_l$ on $\mathbb{C}^{m_l}$ and the induced operator norm $|||\cdot|||_l'$ for $m_l\times m_l$ matrices. Then we have

$$\left|\left|\left|\begin{pmatrix} 0 & \mathbf{0}_{1,m_l-1} \\ \mathbf{0}_{m_l-1,1} & T \end{pmatrix}\right|\right|\right|_l' = \sup\left\{x\in\mathbb{C}^{m_l} : \frac{0+\|T(x_2,...,x_{m_l})^t\|_l}{|x_1|+\|(x_2,...,x_{m_l})^t\|_l}\right\}$$
$$= |||T|||_l\,,$$

for any $m_l\times m_l$ matrix T. Let $K_l = \left|\left|\left|(\mathrm{Id}_{m_l-1} - T_l)^{-1}\right|\right|\right|_l$, $\widetilde{K}_l = |||U_l|||_l'\,|||U_l^{-1}|||_l'$ and $K_l' > 0$ be such that $\|\cdot\|_l' \leqslant K_l'\,\|\cdot\|_\infty$. By using Lemmas 5 and 6 we get

$$\left|\left|\left|S_{l,even}^{(N)}(f) - S_{l,even}(f)\right|\right|\right|_l'$$

$$
\begin{aligned}
&\leqslant \left|\left|\left| U_l \begin{pmatrix} 0 & \mathbf{0}_{1,m_l-1} \\ \mathbf{0}_{m_l-1,1} & -(\mathrm{Id}_{m_l-1}-T_l)^{-1}T_l^{N+1} \end{pmatrix} U_l^{-1} \right|\right|\right|_l' \|V_l(f)\|_l' \\
&\leqslant \widetilde{K}_l \left|\left|\left|(\mathrm{Id}_{m_l-1}-T_l)^{-1}T_l^{N+1}\right|\right|\right|_l' \|V_l(f)\|_l' \\
&\leqslant \widetilde{K}_l K_l \lambda_l^{N+1} K_l' \|V_l(f)\|_\infty \leqslant c_l \lambda_l^{N+1},
\end{aligned}
$$

where $c_l = \widetilde{K}_l K_l K_l' M$. Similar constants show up for $\left|\left|\left| S_{l,odd}^{(N)}(f) - S_{l,odd}(f) \right|\right|\right|_l'$. The final constants c_1 and c_2 depend on $\lambda_1, \lambda_2, \varepsilon$, both matrices (ER) but also the two implicit constants relating the norms $||\cdot||_{m_1}, ||\cdot||_{m_2}$ of Proposition 11 to the equivalent norms $||\cdot||_1', ||\cdot||_2'$.

4.5. Proof of Theorem 2.2

Write $Z_1 = \max_{1\leqslant j\leqslant m_1} |\mathbb{G}(A_j)|$ and $Z_2 = \max_{1\leqslant j\leqslant m_2} |\mathbb{G}(B_j)|$. According to Proposition 11 the random variables appearing on the right hand side of the following formulae, for $* \in \{\text{even,odd}\}$,

$$
\begin{aligned}
\sup_{f\in\mathcal{F}} \left| (S_{1,*}^{(N)}(f) - S_{1,*}(f))^t \cdot \mathbb{G}[\mathcal{A}] \right| &\leqslant c_1 \sup_{f\in\mathcal{F}} \left\| S_{1,*}^{(N)}(f) - S_{1,*}(f) \right\|_{m_1} Z_1, \\
\sup_{f\in\mathcal{F}} \left| (S_{2,*}^{(N)}(f) - S_{2,*}(f))^t \cdot \mathbb{G}[\mathcal{B}] \right| &\leqslant c_2 \sup_{f\in\mathcal{F}} \left\| S_{2,*}^{(N)}(f) - S_{2,*}(f) \right\|_{m_2} Z_2,
\end{aligned}
$$

almost surely converge to 0 since $\mathbb{P}(||\mathbb{G}||_\mathcal{F} < +\infty) = 1$. Hence the processes $\mathbb{G}^{(2m)}, \mathbb{G}^{(2m+1)}$ converge almost surely in $\ell^\infty(\mathcal{F})$ to $\mathbb{G}_{\text{even}}^{(\infty)}, \mathbb{G}_{\text{odd}}^{(\infty)}$ defined by

$$
\begin{aligned}
\mathbb{G}_{\text{even}}^{(\infty)}(f) &= \mathbb{G}(f) - S_{1,\text{even}}(f)^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,\text{even}}(f)^t \cdot \mathbb{G}[\mathcal{B}] \\
&= \mathbb{G}^{(\infty)}(f) - P_2[f]^t \cdot \mathbb{G}[\mathcal{B}], \\
\mathbb{G}_{\text{odd}}^{(\infty)}(f) &= \mathbb{G}(f) - S_{1,\text{odd}}(f)^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,\text{odd}}(f)^t \cdot \mathbb{G}[\mathcal{B}] \\
&= \mathbb{G}^{(\infty)}(f) - P_1[f]^t \cdot \mathbb{G}[\mathcal{A}],
\end{aligned}
$$

with $\mathbb{G}^{(\infty)}(f) = \mathbb{G}(f) - S_{1,\text{even}}(f)^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,\text{odd}}(f)^t \cdot \mathbb{G}[\mathcal{B}]$ and using (2.13), (2.14). Since $P_1[f]^t \cdot \mathbb{G}[\mathcal{A}] = P(f)\sum_{j=1}^{m_1} \mathbb{G}(A_j) = P(f)\mathbb{G}(1) = 0$ and $P_2[f]^t \cdot \mathbb{G}[\mathcal{B}] = 0$ almost surely, we see that $\mathbb{G}_{\text{even}}^{(\infty)}(\mathcal{F}) = \mathbb{G}_{\text{odd}}^{(\infty)}(\mathcal{F}) = \mathbb{G}^{(\infty)}(\mathcal{F})$. Applying Proposition 11 with the supremum norms $||\cdot||_{m_1}$ and $||\cdot||_{m_2}$ further yields, for any $m \geqslant 0$ and $c_0 = m_1c_1 + m_2c_2$,

$$
\begin{aligned}
\left\| \mathbb{G}^{(2m)} - \mathbb{G}_{\text{even}}^{(\infty)} \right\|_\mathcal{F} &= \left\| (S_{1,\text{even}}^{(m-1)} - S_{1,\text{even}})^t \cdot \mathbb{G}[\mathcal{A}] + (S_{2,\text{even}}^{(m-2)} - S_{2,\text{even}})^t \cdot \mathbb{G}[\mathcal{B}] \right\|_\mathcal{F} \\
&\leqslant c_0 \max(\lambda_1, \lambda_2)^{m-2} Z, \qquad (4.9)
\end{aligned}
$$

where $Z = \max(Z_1, Z_2)$, and

$$
\left\| \mathbb{G}^{(2m+1)} - \mathbb{G}_{\text{odd}}^{(\infty)} \right\|_\mathcal{F} \leqslant c_0 \max(\lambda_1, \lambda_2)^{m-1} Z. \qquad (4.10)
$$

Let $\varepsilon_N = q_N c_0 \max(\lambda_1, \lambda_2)^{N/2}$ and $q_N = F_Z^{-1}(c_0 \max(\lambda_1, \lambda_2)^{N/2})$, which is well defined for N large enough. From (4.9) and (4.10) we deduce that

$$\mathbb{P}\left(\left\|\mathbb{G}^{(N)} - \mathbb{G}^{(\infty)}\right\|_{\mathcal{F}} > \varepsilon_N\right) \leqslant \mathbb{P}(Z > q_N) \leqslant c_0 \max(\lambda_1, \lambda_2)^{N/2},$$

whence an upper bound for the Lévy-Prokhorov distance

$$\begin{aligned} d_{LP}(\mathbb{G}^{(N)}, \mathbb{G}^{(\infty)}) &\leqslant \max\left(\mathbb{P}\left(\left\|\mathbb{G}^{(N)} - \mathbb{G}^{(\infty)}\right\|_{\mathcal{F}} > \varepsilon_N\right), \varepsilon_N\right) \\ &\leqslant c_0 q_N \max(\lambda_1, \lambda_2)^{N/2}. \end{aligned}$$

Let Φ denote the standard Gaussian distribution function, $c_5 = m_1 + m_2$ and $c_4^2 = \max_{D \in \mathcal{A} \cup \mathcal{B}} \{P(D)(1 - P(D))\}$. The union bound

$$\mathbb{P}(Z > \lambda) \leqslant c_5 \left(1 - \Phi\left(\frac{\lambda}{c_4}\right)\right) \leqslant \frac{c_5 c_4}{\sqrt{2\pi}\lambda} \exp\left(-\frac{\lambda^2}{2c_4^2}\right),$$

shows that $q_N = c_6 c_4 \sqrt{N \log(1/c_0 \max(\lambda_1, \lambda_2))}$ for some $c_6 > 0$.

Appendix

A.1. Elementary example

The Raking-Ratio algorithm changes the weights of cells of a contingency table in such a way that given margins are respected, just as if the sample should have respected the expected values of known probabilities. Let us illustrate the method from the following basic two-way contingency table.

$\mathbb{P}_n(1_{A_i^{(1)} \cap A_j^{(2)}})$	$A_1^{(2)}$	$A_2^{(2)}$	$A_3^{(2)}$	Total	Excepted total
$A_1^{(1)}$	0.2	0.25	0.1	0.55	0.52
$A_2^{(1)}$	0.1	0.2	0.15	0.45	0.48
Total	0.3	0.45	0.25	1	
Excepted total	0.31	0.4	0.29		$N = 0$

The margins of this sample differ from the known margins, here called expected total. Firstly the weights of lines are corrected, hence each cell is multiplied by the ratio of the expected total and the actual one, this is step $N = 1$.

$\mathbb{P}_n^{(1)}(1_{A_i^{(1)} \cap A_j^{(2)}})$	$A_1^{(2)}$	$A_2^{(2)}$	$A_3^{(2)}$	Total	Excepted total
$A_1^{(1)}$	0.189	0.236	0.095	0.52	0.52
$A_2^{(1)}$	0.11	0.21	0.16	0.48	0.48
Total	0.299	0.446	0.255	1	
Excepted total	0.31	0.4	0.29		$N = 1$

The totals for each column are similarly corrected at step $N = 2$. Typically the margins of the lines no longer match the expected frequencies. Here they move in the right direction. Some estimators based on $\mathbb{P}_n^{(2)}$ may be improved.

$\mathbb{P}_n^{(2)}(1_{A_i^{(1)}\cap A_j^{(2)}})$	$A_1^{(2)}$	$A_2^{(2)}$	$A_3^{(2)}$	Total	Excepted total
$A_1^{(1)}$	0.196	0.212	0.108	0.516	0.52
$A_2^{(1)}$	0.114	0.188	0.182	0.484	0.48
Total	0.31	0.4	0.29	1	
Excepted total	0.31	0.4	0.29		$N = 2$

The last two operations are repeated until stabilization. The algorithm converges to the Kullback projection of the initial joint law. The rate depends only on the initial table compared to the desired marginals. It takes only 7 iterations in our case to match the expected margins.

$\mathbb{P}_n^{(7)}(1_{A_i^{(1)}\cap A_j^{(2)}})$	$A_1^{(2)}$	$A_2^{(2)}$	$A_3^{(2)}$	Total	Excepted total
$A_1^{(1)}$	0.199	0.212	0.109	0.52	0.52
$A_2^{(1)}$	0.111	0.188	0.181	0.48	0.48
Total	0.31	0.4	0.29	1	
Excepted total	0.31	0.4	0.29		$N = 7$

The final raked frequencies are slightly moved away the initial ones, however this has to be compared with the natural sampling oscillation order $1/\sqrt{n}$ – insidiously n was not mentioned. For small samples such changes are likely to occur that may improve a large class of estimators, and worsen others. Our theoretical results showed that the improvement is uniform over a large class as $n \to +\infty$ and N is fixed.

A.2. Counterexample of Remark J

Let assume that P satisfies the following probability values

$P(A_i \cap B_j)$	A_1	A_2	A_3	$P(B_j)$
B_1	0.2	0.25	0.1	0.55
B_2	0.25	0.1	0.1	0.45
$P(A_i)$	0.45	0.35	0.2	

and that f has the following conditional expectations

$\mathbb{E}(f\|A_i \cap B_j)$	A_1	A_2	A_3	$\mathbb{E}^{(2)}(f) \simeq$
B_1	0.75	-0.5	0.5	0.136
B_2	0.5	0.25	0.5	0.444
$\mathbb{E}^{(1)}(f) \simeq$	0.611	-0.286	0.5	

By supposing also that $\mathbb{V}(f|A_i \cap B_j) = 0.5$ for all $i = 1, 2, 3$ and $j = 1, 2$ we can compute the theoretical limiting variances from Proposition 7. We get $\mathbb{V}(\mathbb{G}^{(0)}(f)) \simeq 0.734$; $\mathbb{V}(\mathbb{G}^{(1)}(f)) \simeq 0.563$; $\mathbb{V}(\mathbb{G}^{(2)}(f)) \simeq 0.569$; $\mathbb{V}(\mathbb{G}^{(3)}(f)) \simeq 0.402$. The fact that $\mathbb{V}(\mathbb{G}^{(2)}(f)) > \mathbb{V}(\mathbb{G}^{(1)}(f))$ shows that the variance doesn't decrease necessarily at each step. As predicted by Propositions 8 and 9 we have $\mathbb{V}(\mathbb{G}^{(N)}(f)) < \mathbb{V}(\mathbb{G}^{(0)}(f))$ for $N = 1, 2, 3$ and $\mathbb{V}(\mathbb{G}^{(3)}(f)) < \mathbb{V}(\mathbb{G}^{(1)}(f))$.

A.3. Raked empirical means over a class

General framework. Many specific settings in statistics may be modeled through $\mathcal{F}$. Typically $\mathcal{X}$ is of very large or infinite dimension and each $f(X)$ is one variable with mean $P(f)$ in the population. To control correlations between such variables one needs to extend $\mathcal{F}$ into $\mathcal{F}_\times = \{fg : f, g \in \mathcal{F}\}$ and consider the covariance process $\alpha_n^{(N)}(\mathcal{F}_\times)$. Random vectors $(Y_1, ..., Y_k) = (f_1(X), ..., f_k(X))$ can in turn be combined into real valued random variables $g_\theta(X) = \varphi_\theta(Y_1, ..., Y_k)$ through parameters θ and functions g_θ that should be included in $\mathcal{F}$ and so on. Consider for instance $g_\theta(X) = \theta_1 Y_1 + ... + \theta_k Y_k + \varepsilon_\sigma(X)$ with a collection of possible residual functions ε_σ turning part of the randomness of X into a noise with variance σ^2. The (VC) or (BR) entropy of $\mathcal{F}$ rules the variety and complexity of models or statistics one can simultaneously deal with. We refer to Pollard [20], Shorack and Wellner [21] and Wellner [29] for classical statistical models where an empirical process indexed by functions is easily identified.

Direct applications. Since the limiting process $\mathbb{G}^{(N)}$ of $\alpha_n^{(N)}$ has less variance than $\mathbb{G}^{(0)}$, Theorem 2.1 can be applied to revisit the limiting behavior of classical estimators or tests by using $\mathbb{P}_n^{(N)}$ instead of $\mathbb{P}_n$ and prove that the induced asymptotic variances or risk decrease. For instance, in the case of goodness of fit tests, the threshold decreases at any given test level while the power increases against any alternative distribution Q that do not satisfy the margin conditions. As a matter of fact, enforcing $\mathbb{P}_n^{(N)}$ to look like P instead of the true Q over all $\mathcal{A}^{(N)}$ makes $\mathbb{P}_n^{(N)}$ go very far from P on sets where Q was already far from P.

Example: two raked distribution functions. Let (X, Y) be a real centered Gaussian vector with covariance matrix $\left(\begin{smallmatrix} 3 & -1 \\ -1 & 1 \end{smallmatrix}\right)$. We consider the raked joint estimation of the two distribution functions F_X, F_Y. An auxiliary information provides their values at points -2 to 2, every 0.5. The class $\mathcal{F}$ we need contains for all $t \in \mathbb{R}$ the functions $f_t^X(x, y) = \mathbf{1}_{]-\infty,t]}(x), f_t^Y(x, y) = \mathbf{1}_{]-\infty,t]}(y)$ thus (VC) holds. For $Z = X, Y$ let $F_{Z,n}^{(N)}(t) = \sum_{Z_i \leqslant t} \mathbb{P}_n^{(N)}(\{Z_i\})$ be the N-th raked empirical distribution function and write $Z_{(1)} \leqslant \cdots \leqslant Z_{(n)}$ the order statistics. To exploit at best the information we use $N = 2m$, $F_{X,n}^{(2m-1)}$ and $F_{Y,n}^{(2m)}$. Consider $d_{Z,n}^{(N)} = \sum_{i=1}^{n-1} (Z_{(i+1)} - Z_{(i)}) |F_{Z,n}^{(N)}(Z_{(i+1)}) - F_Z(Z_{(i+1)})|$ which approximates on $[Z_{(1)}, Z_{(n)}]$ the L^1-distance between $F_{Z,n}^{(N)}$ and F_Z. Denote $\#_{Z,n}^{(N)}$ the random proportion of sample points where $F_{Z,n}^{(N)}$ is closer to F_Z than $F_{Z,n}^{(0)}$. The table below provides Monte-Carlo estimates of $D_{Z,n}^{(N)} = \mathbb{E}(d_{Z,n}^{(N)})$ and $p_{Z,n}^{(N)} = \mathbb{E}(\#_{Z,n}^{(N)})$ from 1000 simulations based on samples of size $n = 200$:

Z	$D_{Z,n}^{(0)}$	$D_{Z,n}^{(10)}$	$D_{Z,n}^{(\infty)}$	$p_{Z,n}^{(10)}$	$p_{Z,n}^{(\infty)}$
X	0.084	0.058	0.065	0.752	0.724
Y	0.085	0.043	0.053	0.731	0.681

This shows some improvement, especially for $N = 10$. For n rather small it seems not always relevant to wait for the stabilization of the algorithm – here denoted $N=\infty$. Our theoretical results provide guaranties only for $N \leqslant N_0$, N_0 fixed and $n \geqslant n_0$ for $n_0>0$. We also observed on graphical representations that the way $F_{Z,n}^{(N)}$ leaves $F_{Z,n}^{(0)}$ to cross F_Z at the known points tends to accentuate the error at a few short intervals where $F_{Z,n}^{(0)}$ is far from F_Z. This is less the case as the auxiliary information partition is refined or the sample size increases.

Example: raked covariance matrices. Given $d \in \mathbb{N}_*$ and $f_1,...,f_d$ let $\mathbb{V}(Y)$ denote the covariance matrix of the random vector $Y=(f_1(X),...,f_d(X))$ which we assume to be centered for simplicity. Instead of the empirical covariance $\mathbb{V}_n^{(0)}(Y)=n^{-1}\sum_{i=1}^n Y_i^t Y_i$ consider its raked version

$$\mathbb{V}_n^{(N)}(Y)=\left(\left(\mathbb{P}_n^{(N)}(f_i f_j)\right)_{i,j}\right).$$

Let $\|\cdot\|$ denote the Froebenius norm and define

$$\varphi_Y(\alpha_n^{(N)})=\sqrt{n}\left\|\mathbb{V}_n^{(N)}(Y)-\mathbb{V}(Y)\right\|.$$

In other words,

$$\varphi_Y^2(\alpha_n^{(N)})=\sum_{i=1}^{d}\sum_{j=1}^{d}\left(\alpha_n^{(N)}(f_i f_j)\right)^2,\quad \varphi_Y^2(\mathbb{G}^{(N)})=\sum_{i=1}^{d}\sum_{j=1}^{d}\left(\mathbb{G}^{(N)}(f_i f_j)\right)^2.$$

In the context of Proposition 6 observe that φ_Y is $(\|\cdot\|_{\mathcal{F}}, \|\cdot\|)$-Lipshitz with parameter $C_1=d$. Clearly $\varphi_Y(\mathbb{G}^{(N)})$ has a bounded density since $\varphi_Y^2(\mathbb{G}^{(N)})$ is a quadratic form with Gaussian components and has a modified $\mathcal{X}^2$ distribution. Choosing a finite collection of such φ_Y ensures that $C_2<+\infty$. More generally by letting $(f_1,...,f_d)$ vary among a small entropy infinite subset $\mathcal{L}_d$ of $\mathcal{F}^d$ and imposing some regularity or localization constraints to the f_i one may have $C_2<+\infty$ while $\{f_i f_j : f_i, f_j \in \mathcal{F}\}$ satisfies (BR). The largest C_2 still works for $\mathcal{L}=\bigcup_{d\leqslant d_0}\mathcal{L}_d$. Therefore Proposition 6 guaranties that

$$\max_{\substack{0\leqslant N\leqslant N_0\\ d\leqslant d_0}}\ \sup_{\substack{(f_1,...,f_d)\in\mathcal{L}_d\\ x>0}}\left|\mathbb{P}\left(\varphi_Y(\alpha_n^{(N)})\leqslant x\right)-\mathbb{P}\left(\varphi_Y(\mathbb{G}^{(N)})\leqslant x\right)\right|\leqslant d_1 C_1 C_2 v_n,$$

where it holds, for all $N\leqslant N_0$, $d_0\leqslant d_1$, $(f_1,...,f_d)\in\mathcal{L}_d$ and $x>0$,

$$\mathbb{P}\left(\varphi_Y(\mathbb{G}^{(N)})\leqslant x\right)\leqslant\mathbb{P}\left(\varphi_Y(\mathbb{G}^{(0)})\leqslant x\right),$$

by the variance reduction property of Proposition 8. Hence we asymptotically have $\mathbb{P}(\varphi_Y(\alpha_n^{(N)})\leqslant x)<\mathbb{P}(\varphi_Y(\alpha_n^{(0)})\leqslant x)-\varepsilon$ uniformly among Y such that $\mathbb{P}\left(\varphi_Y(\mathbb{G}^{(N)})\leqslant x\right)<\mathbb{P}\left(\varphi_Y(\mathbb{G}^{(0)})\leqslant x\right)-2\varepsilon$, for any fixed $\varepsilon>0$.

References

[1] Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *The Annals of Probability* **12** 1041–1067. MR0757769

[2] Bankier, M.D. (1986). Estimators based in several stratified samples with applications to multiple frame surveys. *Journal of the American Statistical Association* **81** 1074–1079.
[3] Berthet, P. and Mason, D. M. (2006). Revisiting two strong approximation results of Dudley and Philipp. *IMS Lecture Notes-Monograph Series High Dimensional Probability* **51** 155–172. MR2387767
[4] Binder, D. A. and Théberge, A (1988). Estimating the variance of raking-ratio estimators. *The Canadian Journal of Statistics* **16** 47–55. MR0997121
[5] Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability* **4** 329–375. MR1653272
[6] Brackstone, G. J. and Rao, J. N. K. (1979). An investigation of raking ratio estimators. *The Indian journal of Statistics* **41** 97–114.
[7] Brown, D. T. (1959). A note on approximations to discrete probability distributions. *Information and control* **2** 386–392. MR0110598
[8] Choudhry, G.H. and Lee, H. (1987). Variance estimation for the Canadian Labour Force Survey. *Survey Methodology* **13** 147–161.
[9] Cover, T. M. and Thomas, J. A. (2012). Elements of information theory. *John Wiley & Sons* MR1122806
[10] Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics* **11** 427–444. MR0003527
[11] Deville, J-C. and Särndal, C-E (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* **87** 376–382. MR1173804
[12] Deville, J-C. and Särndal, C-E (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association* **88** 1013–1020. MR1173804
[13] Dudley, R. M. (1989). Real analysis and probability. *Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA* MR0982264
[14] Dudley, R. M. (2014). Uniform central limit theorems. *Cambridge university press* **142** MR3445285
[15] Franklin, J. and Lorenz, J. (1989). On the scaling of multidimensional matrices. *Linear Algebra and its Applications* **114-115** 717–735. MR0986904
[16] Ireland, C. T. and Kullback, S. (1968). Contingency tables with given marginals. *Biometrika* **55** 179–188. MR0229329
[17] Konijn, H. S. (1981). Biases, variances and covariances of raking ratio estimators for marginal and cell totals and averages of observed characteristics. *Metrika* **28** 109–121. MR0629356
[18] Lewis, P. M. (1959). Approximating probability distributions to reduce storage requirements. *Information and control* **2** 214–225. MR0110597
[19] Pollard, D. (1984). Convergence of stochastic processes. *Springer Science & Business Media* MR0762984

[20] Pollard, D. (1990). Empirical processes: theory and applications. *NSF-CBMS regional conference series in probability and statistics* MR1089429

[21] Shorack, G. R. and Wellner, J. A. (1986). Empirical processes with applications to statistics. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics* MR0838963

[22] Sinkhorn, R. (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices. *The Annals of Mathematical Statistics* **35** 876–879. MR0161868

[23] Sinkhorn, R. (1967). Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly* **74** 402–405. MR0210730

[24] Sinkhorn, R. and Knopp, P. (1967). Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics* **21** 343–348. MR0210731

[25] Stephan, F. F. (1942). An iterative method of adjusting sample frequency tables when expected marginal totals are known. *The Annals of Mathematical Statistics* **13** 166–178. MR0006674

[26] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *The Annals of Probability* **22** 28–76. MR1258865

[27] Van der Vaart, A. W. and Wellner, J. A. (1996). Weak convergence and empirical processes with applications to statistics. *Springer series in statistics* MR1385671

[28] Van der Vaart, A. W. (2000). Asymptotic statistics. *Cambridge university press* MR1652247

[29] Wellner, J. A. (1992). Empirical processes in action: a review. *International Statistical Review/Revue Internationale de Statistique* **60** 247–269. MR0494615