ESAIM: PS 24 (2020) 435–453
https://doi.org/10.1051/ps/2020017

ESAIM: Probability and Statistics
www.esaim-ps.org

RAKING-RATIO EMPIRICAL PROCESS WITH AUXILIARY INFORMATION LEARNING

Mickael Albertus*

Abstract. The raking-ratio method is a statistical and computational method which adjusts the empirical measure to match the true probability of sets of a finite partition. The asymptotic behavior of the raking-ratio empirical process indexed by a class of functions is studied when the auxiliary information is given by estimates. These estimates are supposed to result from the learning of the probability of sets of partitions from another sample larger than the sample of the statistician, as in the case of two-stage sampling surveys. Under some metric entropy hypothesis and conditions on the size of the information source sample, the strong approximation of this process and in particular the weak convergence are established. Under these conditions, the asymptotic behavior of the new process is the same as the classical raking-ratio empirical process. Some possible statistical applications of these results are also given, like the strengthening of the Z-test and the chi-square goodness of fit test.

Mathematics Subject Classification. 62G09, 62G20, 60F17, 60F05.

Received May 28, 2019. Accepted April 28, 2020.

1. Introduction

Description. The raking-ratio method is a statistical and computational method aiming to incorporate auxiliary information given by the knowledge of probability of a set of several partitions. The algorithm modifies a sample frequency table in such a way that the marginal totals satisfy the known auxiliary information. At each turn, the method performs a simple cross-multiplication and assigns new weights to individuals belonging to the same set of a partition in order to satisfy the known constraints: it is the "ratio" step of this method. After each modification, the previous constraints are no longer fulfilled in general. Nevertheless, under the conditions that all initial frequencies are strictly positive, if the ratio step are cycling through a finite number of partitions, the method converges to a frequency table satisfying the expected values – see [12]. It is the "raking" step of the algorithm. The goal of these operations is therefore to improve the quality of estimators or the power of statistical tests based on the exploitation of the sample frequency table by lowering the quadratic risk when the sample size is large enough. For a numerical example of the raking-ratio method, see Appendix A.1 of [1]. For an example of a simple statistic using the new weights from the raking-ratio method see Appendix A. The following paragraph summarizes the known results for this method.

Keywords and phrases: Uniform central limit theorems, nonparametric statistics, empirical processes, raking ratio process, auxiliary information, learning.

Institut de Mathématiques de Toulouse, Université Paul Sabatier UMR5219, 31400 Toulouse, France.

* Corresponding author: mickael.albertus@math.univ-toulouse.fr

Literature. The raking-ratio method was suggested by Deming and Stephan and called in a first time "iterative proportions" – see Section 5 of [8]. This algorithm has been initially proposed to adjust the frequency table in the aim to converge it towards the least squares solution. Stephan [13] then showed that this last statement was wrong and proposed a modification to correct it. Ireland and Kullback [9] proved that the raking-ratio method converges to the unique projection of the empirical measure with Kullback-Leibler divergence on the set of discrete probability measures verifying all knowing constraints. In some specific cases, estimates for the variance of cell probabilities in the case of a two-way contingency table were established: Brackstone and Rao [6] for $N \leq 4$, Konijn [10] or Choudhry and Lee [7], Bankier [2] for $N = 2$ and Binder and Théberge [4] for any N. Results of these papers suggest the decrease of variance for the raked estimators of the cells of the table and for a finite number of iterations by providing a complex approximation of the variance of these estimators. Albertus and Berthet [1] defined the empirical measure and process associated to the raking-ratio method and have proved the asymptotic bias cancellation, the asymptotic reduction of variance and so the diminution of the quadratic risk for these process. To prove it, they showed that the raking-ratio empirical process indexed by a class of functions satisfying some metric entropy conditions converges weakly to a specific centered Gaussian process with a lower variance than the usual Brownian bridge. Under general and natural conditions that are recalled below, they proved that the variance decreases by raking among the same cycle of partitions. This paper is a plugin work of that of these authors.

Auxiliary information learning. The main motivation of this paper is when the statistician does not have the true probability of sets of a given partition but has a source of information which gives him an estimation of this probability more precisely than if he used his own sample. This source can be of different types: preliminary survey of a large sample of individuals, database processing, purchase of additional data at a lower cost, the knowledge of an expert. The model should suppose that only the estimate of the auxiliary information is transmitted by the source. It is not then necessary for the statistician to know the entire sample of the auxiliary information source: this assumption is essential if the sample size of this source is much too large for the sample to be transferred to the statistician or else for security or privacy concerns. This hypothesis ensures a fast speed of data acquisition and allows a plurality of sources of information and a diversity of partitions. It is a common situation in statistics since today's technologies like streaming data allow the collection and the transmission of such information in real time. The statistician can use this learned information as auxiliary information which is an estimate of the true one. The raking-ratio method makes it possible to combine shared information of several sources. The main statistical question of this article is whether the statistician can still apply the raking-ratio method by using the estimate of inclusion probabilities rather than the true ones as auxiliary information. Let show that the answer to this question is positive provided that the minimum size of the samples of the different sources of auxiliary information is controlled.

Organization. This paper is organized as follow. Main notation and results are respectively grouped at Sections 2.1 and 2.2. Some statistical applications are given at Section 2.3. One end up by exposing all the proofs at Section 3. Appendix A contains a numerical example of the calculation of a raked mean on a generated sample. At Appendix B the calculation of the asymptotic variance of the raked Gaussian process in a simple case is done.

2. Results of the paper

2.1. Main notation

Framework. Let $X_1, \ldots, X_n, X$ be i.i.d. random variables defined on the same probability space $(\Omega, \mathcal{T}, \mathbb{P})$ with same unknown law $P = \mathbb{P}^{X_1}$ on some measurable space $(\mathcal{X}, \mathcal{A})$. The measurable space $(\mathcal{X}, \mathcal{A})$ is endowed with P.

Class of functions. Let $\mathcal{M}$ denote the set of real valued measurable functions on $(\mathcal{X}, \mathcal{A})$. Let consider a class of functions $\mathcal{F} \subset \mathcal{M}$ such that $\sup_{f \in \mathcal{F}} |f| \leq M_{\mathcal{F}} < +\infty$ for some $M_{\mathcal{F}} > 0$ and satisfying the pointwise measurability condition, that is there exists a countable subset $\mathcal{F}_* \subset \mathcal{F}$ such that for all $f \in \mathcal{F}$ there exists a

sequence $\{f_m\} \subset \mathcal{F}_*$ with f as simple limit, that is $\lim_{m\to+\infty} f_m(x) = f(x)$ for all $x \in \mathcal{X}$. This condition is often used to ensure the P-measurability of $\mathcal{F}$ – see example 2.3.4 of [15]. For a probability measure Q on $(\mathcal{X}, \mathcal{A})$ and $f, g \in \mathcal{M}$ let $d_Q^2(f,g) = \int_{\mathcal{X}} (f-g)^2 dQ$. Let $N(\mathcal{F}, \varepsilon, d_Q)$ be the minimum number of balls with d_Q-radius ε necessary to cover $\mathcal{F}$ and $N_{[\,]}(\mathcal{F}, \varepsilon, d_Q)$ be the least number of ε-brackets necessary to cover $\mathcal{F}$, that is elements of the form $[g_-, g_+] = \{f \in \mathcal{F} : g_- \leq f \leq g_+\}$ with $d_P(g_-, g_+) < \varepsilon$. Let also assume that $\mathcal{F}$ satisfies one of the two metric entropy conditions (VC) or (BR) discussed below.

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

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

If all elements $f\mathbb{1}_{A_j^{(N)}}$ for every $N > 0, 1 \leq j \leq m_N$ and $f \in \mathcal{F}$ are added to $\mathcal{F}$, this class still satisfies the same entropy condition but with a new constant c_0 or b_0. Let denote $\ell^\infty(\mathcal{F})$ the set of real-valued functions bounded on $\mathcal{F}$ endowed with the supremum norm $||\cdot||_{\mathcal{F}}$. In this paper the following notations are used: for all $f \in \mathcal{F}, A \in \mathcal{A}$ let denote $P(f) = \mathbb{E}[f(X)]$, $P(A) = P(\mathbb{1}_A)$, $\mathbb{E}[f|A] = P(f\mathbb{1}_A)/P(A)$, $\sigma_f^2 = \mathrm{Var}(f(X))$ and $\sigma_{\mathcal{F}}^2 = \sup_{f\in\mathcal{F}} \sigma_f^2$.

Empirical measures and processes. Let denote the empirical measure $\mathbb{P}_n(\mathcal{F}) = \{\mathbb{P}_n(f) : f \in \mathcal{F}\}$ defined by $\mathbb{P}_n(f) = \frac{1}{n}\sum_{i=1}^n f(X_i)$ and the empirical process $\alpha_n(\mathcal{F}) = \{\alpha_n(f) : f \in \mathcal{F}\}$ defined by $\alpha_n(f) = \sqrt{n}(\mathbb{P}_n(f) - P(f))$. For $N \in \mathbb{N}$, let

$$\mathcal{A}^{(N)} = \{A_1^{(N)}, \ldots, A_{m_N}^{(N)}\} \subset \mathcal{A},$$

be a partition of $\mathcal{X}$ such that

$$P[\mathcal{A}^{(N)}] = (P(A_1^{(N)}), \ldots, P(A_{m_N}^{(N)})) \neq \mathbf{0}.$$

Let denote B_{n,N_0} be the set defined by

$$B_{n,N_0} = \left\{ \min_{0\leq N\leq N_0} \min_{1\leq j\leq m_N} \mathbb{P}_n(A_j^{(N)}) > 0 \right\} \in \mathcal{T}. \tag{2.1}$$

For n large enough and N fixed, according to the law of large number, the event B_{n,N_0} almost surely happens. Let $\mathbb{P}_n^{(N)}(\mathcal{F}) = \{\mathbb{P}_n^{(N)}(f) : f \in \mathcal{F}\}$ be the Nth raking-ratio empirical measure defined recursively on the set $B_{n,N}$ by $\mathbb{P}_n^{(0)} = \mathbb{P}_n$ and for all $f \in \mathcal{F}$,

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

This process is well defined since on B_{n,N_0} one have $\mathbb{P}_n^{(N)}(A) > 0$ for any $N \in \mathbb{N}^*$ and $A \in \mathcal{A}$ such that $\mathbb{P}_n(A) > 0$. The empirical measure $\mathbb{P}_n^{(N)}(\mathcal{F})$ uses the auxiliary information given by $P[\mathcal{A}^{(N)}]$ to modify $\mathbb{P}_n(\mathcal{F})$ in such a way that

$$\mathbb{P}_n^{(N)}[\mathcal{A}^{(N)}] = (\mathbb{P}_n^{(N)}(A_1^{(N)}), \ldots, \mathbb{P}_n(A_{m_N}^{(N)})) = P[\mathcal{A}^{(N)}].$$

Let denote $\alpha_n^{(N)}(\mathcal{F}) = \{\alpha_n^{(N)}(f) : f \in \mathcal{F}\}$ the Nth raking-ratio empirical process defined for all $f \in \mathcal{F}$ by

$$\alpha_n^{(N)}(f) = \sqrt{n}(\mathbb{P}_n^{(N)}(f) - P(f)). \tag{2.2}$$

This process satisfies the following property

$$\alpha_n^{(N)}[\mathcal{A}^{(N)}] = (\alpha_n^{(N)}(A_1^{(N)}), \ldots, \alpha_n^{(N)}(A_{m_N}^{(N)})) = \mathbf{0}.$$

Gaussian processes. Under (VC) or (BR), $\mathcal{F}$ is a Donsker class, that is $\alpha_n(\mathcal{F})$ converges weakly in $\ell^\infty(\mathcal{F})$ to the P-Brownian bridge $\mathbb{G}(\mathcal{F}) = \{\mathbb{G}(f) : f \in \mathcal{F}\}$, the Gaussian process such that $f \mapsto \mathbb{G}(f)$ is linear and for all $f, g \in \mathcal{F}$,

$$\mathbb{E}[\mathbb{G}(f)] = 0, \ \operatorname{Cov}(\mathbb{G}(f), \mathbb{G}(g)) = P(fg) - P(f)P(g).$$

For short, let denote $\mathbb{G}(A) = \mathbb{G}(\mathbb{1}_A)$ for any $A \in \mathcal{A}$. Let $\mathbb{G}^{(N)}(\mathcal{F}) = \{\mathbb{G}^{(N)}(f) : f \in \mathcal{F}\}$ be the Nth raking-ratio P-Brownian bridge, that is a centered Gaussian process defined recursively by $\mathbb{G}^{(0)} = \mathbb{G}$ and for any $N > 0, f \in \mathcal{F}$,

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

Albertus and Berthet established the strong approximation and the weak convergence when n goes to infinity in $\ell^\infty(\mathcal{F})$ of $\alpha_n^{(N)}(\mathcal{F})$ to $\mathbb{G}^{(N)}(\mathcal{F})$ for N fixed – see Proposition 4 and Theorem 2.1 of [1]. For that they used the strong approximation of the empirical process indexed by a function class satisfying (VC) or (BR) – see Theorem 1 and 2 of [3]. They gave the exact value of $\sigma_f^{(N)} = \operatorname{Var}(\mathbb{G}^{(N)}(f))$ and showed in particular for all $f \in \mathcal{F}$ and $N_0 \in \mathbb{N}$ that $\sigma_f^{(N_0)} \leq \sigma_f^{(0)} = \sigma_f$ and $\sigma_f^{(N_1)} \leq \sigma_f^{(N_0)}$ if $N_1 \geq 2N_0$ is such that $\mathcal{A}^{(N_0-k)} = \mathcal{A}^{(N_1-k)}$ for $0 \leq k \leq N_0$ – see Propositions 7–9.

Auxiliary information. For $N > 0$ let $\mathbb{P}'_N[\mathcal{A}^{(N)}] = (\mathbb{P}'_N(A_1^{(N)}), \ldots, \mathbb{P}'_N(A_{m_N}^{(N)}))$ be a random vector with multinomial law, n_N trials and event probabilities $P[\mathcal{A}^{(N)}]$. This random vector corresponds to the estimation of the auxiliary information of the Nth auxiliary information source based on a sample of size $n_N = n_N(n) \gg n$ not necessarily independent of $X_1, \ldots, X_n$. Let study the asymptotic behavior of the raking-ratio empirical process which uses $\mathbb{P}'_N[\mathcal{A}^{(N)}]$ as auxiliary information instead of $P[\mathcal{A}^{(N)}]$. By defining the sequence $\{n_N\}$ it is supposed that this information can be estimated by different sources that would not necessarily have the same sample size but still have a sample size larger than n. Let $\widetilde{\mathbb{P}}_n^{(N)}(\mathcal{F}) = \{\widetilde{\mathbb{P}}_n^{(N)}(f) : f \in \mathcal{F}\}$ be the Nth raking-ratio empirical measure with learned auxiliary information defined recursively by $\widetilde{\mathbb{P}}_n^{(0)} = \mathbb{P}_n$ and for all $N > 0, f \in \mathcal{F}$,

$$\widetilde{\mathbb{P}}_n^{(N)}(f) = \sum_{j=1}^{m_N} \frac{\mathbb{P}'_N(A_j^{(N)})}{\widetilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} \widetilde{\mathbb{P}}_n^{(N-1)}(f\mathbb{1}_{A_j^{(N)}}).$$

This empirical measure satisfies the learned auxiliary information since

$$\begin{aligned}\widetilde{\mathbb{P}}_n^{(N)}[\mathcal{A}^{(N)}] &= (\widetilde{\mathbb{P}}_n^{(N)}(A_1^{(N)}), \ldots, \widetilde{\mathbb{P}}_n^{(N)}(A_{m_N}^{(N)})) \\ &= \mathbb{P}'_N[\mathcal{A}^{(N)}].\end{aligned}$$

Let define $\widetilde{\alpha}_n^{(N)}(\mathcal{F}) = \{\widetilde{\alpha}_n^{(N)}(f) : f \in \mathcal{F}\}$ the Nth raking-ratio empirical with estimated auxiliary information defined for $f \in \mathcal{F}$ by

$$\widetilde{\alpha}_n^{(N)}(f) = \sqrt{n}(\widetilde{\mathbb{P}}_n^{(N)}(f) - P(f)). \tag{2.4}$$

2.2. Main results

Deviation inequality. For $N_0 > 0$, denote $K_{\mathcal{F}} = \max(1, M_{\mathcal{F}})$ and

$$\begin{aligned} p_{(N_0)} &= \min_{1\le N\le N_0} \min_{1\le j\le m_N} P(A_j^{(N)}), \\ m_{(N_0)} &= \sup_{0\le N\le N_0} m_N, \\ n_{(N_0)} &= \min_{1\le N\le N_0} n_N > n. \end{aligned}$$

Remind that empirical measures $\mathbb{P}_n^{(0)}(\mathcal{F}), \dots, \mathbb{P}_n^{(N_0)}(\mathcal{F})$ and $\widetilde{\mathbb{P}}_n^{(0)}(\mathcal{F}), \dots, \widetilde{\mathbb{P}}_n^{(N)}(\mathcal{F})$ are defined on the set B_{n,N_0} defined by (2.1) and satisfying

$$\mathbb{P}(B_{n,N_0}^C) \le \sum_{N=1}^{N_0} m_N(1-p_N)^n \le N_0 m_{(N_0)}(1-p_{(N_0)})^n,$$

where $B_{n,N_0}^C = \Omega \setminus B_{n,N_0}$. The following proposition is a Talagrand type inequality which bounds the probability that $||\widetilde{\alpha}_n^{(N)}||_{\mathcal{F}}$ deviates from a certain value.

Proposition 2.1. *For any $N_0 \in \mathbb{N}$, $n > 0$ and $t > 0$, it holds under the event B_{n,N_0}*

$$\begin{aligned} \mathbb{P}\left(\sup_{0\le N\le N_0} ||\widetilde{\alpha}_n^{(N)}||_{\mathcal{F}} > t\right) \le N_0 \mathbb{P}\left(||\widetilde{\alpha}_n^{(0)}||_{\mathcal{F}} > \frac{t p_{(N_0)}^{N_0}}{4^{N_0} m_{(N_0)}^{N_0} K_{\mathcal{F}}^{N_0} (1+t/\sqrt{n})^{N_0}}\right) \\ + 2N_0^3 m_{(N_0)} \exp\left(-\frac{n_{(N_0)} p_{(N_0)}^2 t^2}{2n m_{(N_0)}^2 K_{\mathcal{F}}^2}\right). \end{aligned} \tag{2.5}$$

Under (VC) and the event B_{n,N_0} there exists $t_0 > 0$ such that for all $t_0 < t < 2M_{\mathcal{F}}\sqrt{n}$,

$$\begin{aligned} \mathbb{P}\left(\sup_{0\le N\le N_0} ||\widetilde{\alpha}_n^{(N)}||_{\mathcal{F}} > t\right) \le D_1 t^{\nu_0} \exp(-D_2 t^2) \\ + 2N_0^3 m_{(N_0)} \exp\left(-\frac{n_{(N_0)} p_{(N_0)}^2 t^2}{2n m_{(N_0)}^2 K_{\mathcal{F}}^2}\right), \end{aligned} \tag{2.6}$$

where $D_1, D_2 > 0$ are defined by (3.7). Under (BR) and the event B_{n,N_0} there exists $t_0, C > 0$ such that for all $t_0 < t < C\sqrt{n}$,

$$\begin{aligned} \mathbb{P}\left(\sup_{0\le N\le N_0} ||\widetilde{\alpha}_n^{(N)}||_{\mathcal{F}} > t\right) \le D_3 \exp(-D_4 t^2) \\ + 2N_0^3 m_{(N_0)} \exp\left(-\frac{n_{(N_0)} p_{(N_0)}^2 t^2}{2n m_{(N_0)}^2 K_{\mathcal{F}}^2}\right), \end{aligned} \tag{2.7}$$

where $D_3, D_4 > 0$ are defined by (3.9).

Proposition 2.1 proves that if $\mathcal{F}$ satisfies (VC) or (BR) then almost surely $||\alpha_n||_{\mathcal{F}} = O(\sqrt{\log(n)})$. If $\mathcal{F}$ satisfies (VC), let define $v_n = n^{-\alpha_0}(\log n)^{\beta_0}$ with $\alpha_0 = 1/(2+5\nu_0) \in (0, 1/2)$ and $\beta_0 = (4+5\nu_0)/(4+10\nu_0)$. If $\mathcal{F}$ satisfies (BR), let define $v_n = (\log n)^{-\gamma_0}$ with $\gamma_0 = (1-r_0)/2r_0$.

Weak convergence. If the sample size of the sources are large enough then the asymptotic behavior for $\widetilde{\alpha}_n^{(N)}(\mathcal{F})$ is the same as $\alpha_n^{[N)}(\mathcal{F})$.

Proposition 2.2. *If* $n\log(n) = o(n_{(N_0)})$ *then the sequence* $(\widetilde{\alpha}_n^{(0)}(\mathcal{F}), \ldots, \widetilde{\alpha}_n^{(N_0)}(\mathcal{F}))$ *converges weakly to* $(\mathbb{G}^{(0)}(\mathcal{F}), \ldots, \mathbb{G}^{(N_0)}(\mathcal{F}))$ *on* $\ell^\infty(\mathcal{F} \to \mathbb{R}^{N_0+1})$

Proposition 2.2 is a direct consequence of the strong approximation of $\widetilde{\alpha}_n^{(N)}(\mathcal{F})$ to $\mathbb{G}^{(N)}(\mathcal{F})$ given by the following theorem – see the beginning of Section 3.5 of [1] for a proof of a similar statement.

Theorem 2.3. *Let* $N_0 \in \mathbb{N}$. *There exists* $d_0, n_0 > 0$, *a sequence* $\{X_n\}$ *of independent random variables with law* P *and a sequence* $\{\mathbb{G}_n\}$ *of versions of* $\mathbb{G}$ *supported on a same probability space such that for all* $n > n_0$,

$$\mathbb{P}\left(\sup_{0\le N\le N_0} ||\widetilde{\alpha}_n^{(N)} - \mathbb{G}_n^{(N)}||_{\mathcal{F}} > d_0\left(v_n + \sqrt{\frac{n\log(n)}{n_{(N_0)}}}\right)\right) < \frac{1}{n^2}, \tag{2.8}$$

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

Under the conditions of Theorem 2.3, by Borel-Cantelli lemma it holds almost surely for large n,

$$\sup_{0\le N\le N_0} ||\widetilde{\alpha}_n^{(N)} - \mathbb{G}_n^{(N)}||_{\mathcal{F}} \le d_2\left(v_n + \sqrt{\frac{n\log(n)}{n_{(N_0)}}}\right). \tag{2.9}$$

Sequence v_n in the previous bound is the deviation from $\alpha_n^{(N)}(\mathcal{F})$ to $\mathbb{G}_n^{(N)}(\mathcal{F})$ while $\sqrt{n\log(n)/n_{(N_0)}}$ represents the deviation from $\widetilde{\alpha}_n^{(N)}(\mathcal{F})$ to $\alpha_n^{(N)}(\mathcal{F})$. The main argument of these results is that the empirical process $\widetilde{\alpha}_n^{(N)}(\mathcal{F})$ is close to $\alpha_n^{(N)}(\mathcal{F})$. Notice that this argument will handle the dependence between $X_1, \ldots, X_n$ and vectors $\mathbb{P}'_N[\mathcal{A}^{(N)}]$.

2.3. Statistical applications

Improvement of a statistical test. Any statistical test using the empirical process can be adapted to use auxiliary information to strengthen this test. It suffices to replace in the expression of the test statistic the process $\alpha_n(\mathcal{F})$ by $\alpha_n^{(N)}(\mathcal{F})$ if the true auxiliary information is at our disposal or by $\widetilde{\alpha}_n^{(N)}(\mathcal{F})$ if only an estimation of this information is available. The two following subsections give an example of application in the case of the Z-test and the chi-squared goodness of fit test. In both case, the statistic of theses tests is transformed and the decision procedure is kept. In the first case, we show that this new statistical test has the same significance level but a higher power. For the second case, we prove that the confidence level decreases and that under (H_1), the new statistic goes to infinity as the same way as the usual one.

***Z*-test.** This test is used to compare the mean of a sample to a given value when the variance of the sample is known. The null hypothesis is $(H_0) : P(f) = P_0(f)$, for some $f \in \mathcal{F}$ and a probability measure $P_0 \in \ell^\infty(\mathcal{F})$. The statistic of the classical Z-test is

$$Z_n = \sqrt{n}\frac{\mathbb{P}_n(f) - P_0(f)}{\sigma_f}.$$

Under (H_0), asymptotically the statistic Z_n follows the standard normal distribution. The null hypothesis is rejected at the α level when $|Z_n| > t_\alpha$, $t_\alpha = \Phi(1-\alpha/2)$ with Φ the probit function. Let define the

following statistics

$$Z_n^{(N)} = \sqrt{n}\frac{\mathbb{P}_n^{(N)}(f) - P_0(f)}{\sigma_f^{(N)}},$$
$$\widetilde{Z}_n^{(N)} = \sqrt{n}\frac{\widetilde{\mathbb{P}}_n^{(N)}(f) - P_0(f)}{\sigma_f^{(N)}},$$

Since the law P is unknown, σ_f and $\sigma_f^{(N)}$ for $N \geq 1$ are usually unknown but a consistent estimation of these standard deviation can be used to calculate $Z_n, Z_n^{(N)}$ or $\widetilde{Z}_n^{(N)}$ – a concrete example of this remark is given at the following paragraph. Doing it does not change the asymptotic behavior of the random variables $Z_n, Z_n^{(N)}$ and $\widetilde{Z}_n^{(N)}$, whether the hypothesis (H_0) is verified or not. The statistical tests based on the reject decision $|Z_n^{(N)}| > t_\alpha$ and $|\widetilde{Z}_n^{(N)}| > t_\alpha$ have the same significance level than the usual test based on the decision $|Z_n| > t_\alpha$ since, under (H_0), $Z_n^{(N)}$ and $\widetilde{Z}_n^{(N)}$ converge weakly to $\mathcal{N}(0,1)$ – see Proposition 6 of [1]. The following proposition shows that the ratio of the beta risk of the usual Z-test and the new statistical test with auxiliary information goes to infinity as $n \to +\infty$.

Proposition 2.4. *Assume that $\sigma_f^{(N)} < \sigma_f$. Under (H_1), for all $\alpha \in (0,1)$ and n large enough one have*

$$\frac{\mathbb{P}(|Z_n| \leq t_\alpha)}{\mathbb{P}(|Z_n^{(N)}| \leq t_\alpha)} \geq \exp(b_n), \tag{2.10}$$

with $b_n \sim n(P(f) - P_0(f))^2 \left(1/\sigma_f^{(N)} - 1/\sigma_f\right)$. If $n\log(n) = o(n_{(N_0)})$ then

$$\frac{\mathbb{P}(|Z_n| \leq t_\alpha)}{\mathbb{P}(|\widetilde{Z}_n^{(N)}| \leq t_\alpha)} \geq \exp(b_n). \tag{2.11}$$

Z-test in a simple case. To calculate $Z_n^{(N)}$ or $\widetilde{Z}_n^{(N)}$ one needs the expression of $\sigma_f^{(N)}$. To illustrate how to get it, let work on a simple case, when the auxiliary information is given by probabilities of two partitions of two sets. More formally for $k \in \mathbb{N}^*$ let define $\mathcal{A}^{(2k-1)} = \mathcal{A} = \{A, A^C\}$ and $\mathcal{A}^{(2k)} = \mathcal{B} = \{B, B^C\}$. By using Proposition 7 of [1] one give simple expressions of $\sigma_f^{(N)}$ for $N = 1, 2$. For the sake of simplification, let denote

$$\begin{aligned} p_A = P(A), \quad p_{\overline{A}} = P(A^C), \quad p_B = P(B), \quad p_{\overline{B}} = P(B^C), \\ p_{AB} = P(A \cap B), \quad \Delta_A = \mathbb{E}[f|A] - \mathbb{E}[f], \quad \Delta_B = \mathbb{E}[f|B] - \mathbb{E}[f], \end{aligned} \tag{2.12}$$

then,

$$\begin{aligned} (\sigma_f^{(1)})^2 &= \sigma_f^2 - \mathbb{E}[f|\mathcal{A}]^t \cdot \mathrm{Var}(\mathbb{G}[\mathcal{A}]) \cdot \mathbb{E}[f|\mathcal{A}] \\ &= \sigma_f^2 - p_A p_{\overline{A}}(\mathbb{E}[f|A] - \mathbb{E}[f|A^C])^2 = \sigma_f^2 - \frac{p_A}{p_{\overline{A}}}\Delta_A^2, \\ (\sigma_f^{(2)})^2 &= \sigma_f^2 - \mathbb{E}[f|\mathcal{B}]^t \cdot \mathrm{Var}(\mathbb{G}[\mathcal{B}]) \cdot \mathbb{E}[f|\mathcal{B}] = \\ &\quad - \left(\mathbb{E}[f|\mathcal{A}] - \mathbf{P}_{\mathcal{B}|\mathcal{A}} \cdot \mathbb{E}[f|\mathcal{B}]\right)^t \cdot \mathrm{Var}(\mathbb{G}[\mathcal{A}]) \cdot \left(\mathbb{E}[f|\mathcal{A}] - \mathbf{P}_{\mathcal{B}|\mathcal{A}} \cdot \mathbb{E}[f|\mathcal{B}]\right) \\ &= \sigma_f^2 - p_B p_{\overline{B}}(\mathbb{E}[f|B] - \mathbb{E}[f|B^C])^2 \end{aligned}$$

$$- \left(p_A p_{\overline{A}} + \frac{p_B p_{\overline{B}} (p_{AB} - p_A p_B)}{p_A^2 p_{\overline{A}}^2} \right) (\mathbb{E}[f|A] - \mathbb{E}[f|A^C])^2$$
$$= \sigma_f^2 - \frac{p_B}{p_{\overline{B}}} \Delta_B^2 - \left(\frac{p_A}{p_{\overline{A}}} + \frac{p_B p_{\overline{B}} (p_{AB} - p_A p_B)}{p_A^2 p_{\overline{A}}^4} \right) \Delta_A^2,$$

where $\mathbf{P}_{\mathcal{A}|\mathcal{B}}, \mathbf{P}_{\mathcal{B}|\mathcal{A}}$ are stochastic matrices given by (B.1), $\mathbb{E}[f|\mathcal{A}], \mathbb{E}[f|\mathcal{B}]$ are conditional expectation vectors given by (B.2) and $\mathrm{Var}(\mathbb{G}[\mathcal{A}]), \mathrm{Var}(\mathbb{G}[\mathcal{B}])$ are the covariance matrices of $\mathbb{G}[\mathcal{A}] = (\mathbb{G}(A), \mathbb{G}(A^C))$ and $\mathbb{G}[\mathcal{B}] = (\mathbb{G}(B), \mathbb{G}(B^C))$ that is the matrices given by (B.3). These variances can be estimated empirically. Albertus and Berthet proved that the raked Gaussian process $\mathbb{G}^{(N)}$ converges almost surely as $N \to +\infty$ to some centered Gaussian process $\mathbb{G}^{(\infty)}$ with an explicit expression. The stabilization of the raking-ratio method in the case of two marginals when $N \to +\infty$ is fast since the Levy-Prokhorov distance between $\mathbb{G}^{(N)}$ and $\mathbb{G}^{(\infty)}$ is almost surely at most $O(N\lambda^{N/2})$ for some $\lambda \in (0,1)$ – see Proposition 11 of [1]. Let denote $\mathbb{P}_n^{(\infty)}(\mathcal{F})$ the raked empirical measure after stabilization of the raking-ratio algorithm and $(\sigma_f^{(\infty)})^2 = \mathrm{Var}(\mathbb{G}^{(\infty)}(f))$ the asymptotic variance. Let define the following statistic

$$Z_n^{(\infty)} = \sqrt{n} \frac{\mathbb{P}_n^{(\infty)}(f) - P_0(f)}{\sigma_f^{(\infty)}}.$$

According to Proposition 2.4, the statistical test based on the reject decision $|Z_n^{(\infty)}| > t_\alpha$ has the same significance level than the usual Z-test based on $|Z_n| > t_\alpha$ but it is more powerful as n goes to infinity. In the case of two marginals with two partitions, one can give an explicit and simple expression of the asymptotic variance. By using the notations of (2.12) one have

$$(\sigma_f^{(\infty)})^2 = \sigma_f^2 - \frac{p_A p_B \left(p_A \Delta_A^2 + p_B \Delta_B^2 - p_A p_B (\Delta_A - \Delta_B)^2 - 2 p_{AB} \Delta_A \Delta_B \right)}{p_A p_B p_{\overline{A}} p_{\overline{B}} - (p_{AB} - p_A p_B)^2}. \tag{2.13}$$

The calculation of this variance needs the expression of $\mathbb{G}^{(\infty)}$ so it is made at Appendix B. If the values given by (2.12) are unknown, one can use their consistent estimators to estimate the value of $\sigma_f^{(\infty)}$. If $\Delta_A = \Delta_B = 0$ then naturally the auxiliary information is useless since $\sigma_f^{(\infty)} = \sigma_f$, so there is no reduction of the quadratic risk. If A is independent of B then $p_{AB} = p_A p_B$ and

$$(\sigma_f^{(\infty)})^2 = (\sigma_f^{(2)})^2 = \sigma_f^2 - \left(\frac{p_A}{p_{\overline{A}}} \Delta_A^2 + \frac{p_B}{p_{\overline{B}}} \Delta_B^2 \right).$$

Chi-square test. The chi-squared goodness of fit test consists of knowing whether the sample data corresponds to a hypothesized distribution when the interest variable is a categorical variable. Let $\mathcal{B} = \{B_1, \ldots, B_m\}$ be a partition of $\mathcal{X}$. The null hypothesis is

$$(H_0) : P[\mathcal{B}] = P_0[\mathcal{B}], \tag{2.14}$$

where $P[\mathcal{B}] = (P(B_1), \ldots, P(B_m))$ and $P_0[\mathcal{B}] = (P_0(B_1), \ldots, P_0(B_m))$, for some probability measure P_0. The statistic of the classical chi-squared test is

$$T_n = n \sum_{i=1}^{m} \frac{(\mathbb{P}_n(B_i) - P_0(B_i))^2}{P_0(B_i)}.$$

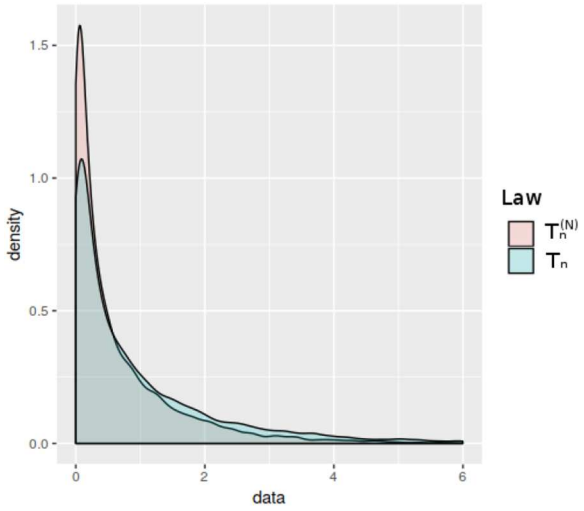

FIGURE 1. Law of T_n and $T_n^{(1)}$.

Under (H_0), asymptotically the statistic T_n follows the χ^2 distribution with $m-1$ degrees of freedom. The null hypothesis is rejected at the level α when $Z_n > t_\alpha^{(m)}$, $t_\alpha^{(m)} = \Phi_m(1-\alpha)$ where Φ_m is the quantile function of $\chi^2(m)$. The question is whether the following statistics

$$T_n^{(N)} = n\sum_{i=1}^{m} \frac{(\mathbb{P}_n^{(N)}(B_i) - P_0(B_i))^2}{P_0(B_i)},$$
$$\widetilde{T}_n^{(N)} = n\sum_{i=1}^{m} \frac{(\widetilde{\mathbb{P}}_n^{(N)}(B_i) - P_0(B_i))^2}{P_0(B_i)},$$

somehow improve the test. The following proposition shows that the tests based on these new statistics have a lower α-risk and are almost surely rejected under the (H_1) hypothesis for a large enough sample size.

Proposition 2.5. *Under (H_0) and for all $\alpha > 0$,*

$$\lim_{n\to+\infty} \mathbb{P}(T_n^{(N)} > t_\alpha^{(m)}) \leq \lim_{n\to+\infty} \mathbb{P}(T_n > t_\alpha^{(m)}) = \alpha, \tag{2.15}$$

and if $n\log(n) = o(n_{(N)})$ then

$$\lim_{n\to+\infty} \mathbb{P}(\widetilde{T}_n^{(N)} > t_\alpha^{(m)}) \leq \alpha. \tag{2.16}$$

Under (H_1) and for all $\alpha > 0$, almost surely there exists $n_0 > 0$ such that for all $n > n_0$,

$$\min(|T_n|, |T_n^{(N)}|, |\widetilde{T}_n^{(N)}|) > t_\alpha^{(m)}. \tag{2.17}$$

Figure 1 is a numerical example of Proposition 2.5 under (H_0). A two-way contingency table with fixed probabilities $P[\mathcal{B}], P[\mathcal{A}]$ is simulated and the chi-square test with the null hypothesis (2.14) is applied. With Monte-Carlo method, the law of T_n and $T_n^{(1)}$ with the auxiliary information given by $P[\mathcal{A}]$ are simulated for $n = 1000$.

Costing data. Another possible statistical application is to study how to share resources – economic resource, temporal resource, material resource, ...– to learn auxiliary information from inexpensive data in order to

improve the study of statistics on expensive objects. More formally one have a budget B, for our estimates an individual X_i can be purchased at a fixed price $C > 0$ and for the estimation of auxiliary information $P[\mathcal{A}^{(N)}], N = 1, \dots, N_0$, the information $\mathbb{P}'_N[\mathcal{A}^{(N)}]$ can be bought at a price $c_N n_N$ where c_N is the price for one individual far less than C. The objective is therefore to minimize the bound $v_n + \sqrt{n \log(n)/n_{(N_0)}}$ proposed by Theorem 2.3 by choosing n high-cost individuals and the $n_1, \dots, n_{N_0}$ low-cost individuals while respecting the imposed budget. So the following constraint has to be satisfied

$$Cn + c_1 n_1 + \cdots + c_{N_0} n_{N_0} \leq B. \tag{2.18}$$

To simplify the problem, let suppose that for all $1 \leq N \leq N_0$, $n_N = n_0$ and $c_N = c_0/N_0$ for some $c_0 > 0$. It is the case if one pay the auxiliary information from the same auxiliary information source and if one pay all N_0 information only once time. Inequality (2.18) becomes

$$Cn + c_0 n_0 \leq B. \tag{2.19}$$

There are several ways to answer this problem. If we want only the strong approximation rate of $\alpha_n^{(N)}$ by $\mathbb{G}^{(N)}$ dominates in the uniform error of (2.9), one have to choose n_0 such that $n_0 \geq n \log(n)/v_n^2$. If $n_0 = \lceil n \log(n)/v_n^2 \rceil$ the maximum value of n satisfying (2.19) could be found. Since $v_n > \sqrt{\log(n)/n}$ then

$$n \geq n_{\min} = \left\lfloor \frac{\sqrt{C^2 + 4c_0 B} - C}{2c_0} \right\rfloor. \tag{2.20}$$

If there is no way of finding the optimal n – if the rate v_n is unknown or to avoid additional calculations – values $n = n_{\min}$ and $n_0 = \lfloor (B - Cn)/c_0 \rfloor$ can be taken if one want to use the entire budget or $n_0 = \lceil n \log(n)/v_n^2 \rceil$ otherwise.

3. Proof

For all this section let fix $N_0 > 0$ and let $\Lambda_n, \Lambda'_n > 0$ be the following supremum deviations

$$\begin{aligned}
\Lambda_n &= \max\left(\sup_{0 \leq N \leq N_0} ||\widetilde{\alpha}_n^{(N)}||_{\mathcal{F}}, \sup_{0 \leq N \leq N_0} ||\alpha_n^{(N)}||_{\mathcal{F}} \right), \\
\Lambda'_n &= \sup_{1 \leq N \leq N_0} \sup_{1 \leq j \leq m_N} |\alpha'_N(A_j^{(N)})|,
\end{aligned}$$

where $\alpha'_N(A_j^{(N)}) = \sqrt{n_N}(\mathbb{P}'_N(A_j^{(N)}) - P(A_j^{(N)}))$. Immediately, by Hoeffding inequality one have for all $\lambda > 0$,

$$\mathbb{P}\left(\Lambda'_n > \lambda\right) \leq 2N_0 m_{(N_0)} \exp\left(-2\lambda^2\right). \tag{3.1}$$

Useful decomposition of $\alpha_n^{(N)}(\mathcal{F})$ and $\widetilde{\alpha}_n^{(N)}(\mathcal{F})$ are given since they will be used in the following proofs. By using Definition (2.2) of $\alpha_n^{(N)}(\mathcal{F})$ one have

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

As the same way, by using (2.4) one have

$$\widetilde{\alpha}_n^{(N)}(f) = \sum_{j=1}^{m_N} \frac{\mathbb{P}'_N(A_j^{(N)})}{\widetilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} \widetilde{\alpha}_n^{(N-1)}(f\mathbb{1}_{A_j^{(N)}}) - \frac{P(f\mathbb{1}_{A_j^{(N)}})}{\widetilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} \left(\widetilde{\alpha}_n^{(N-1)}(A_j^{(N)}) - \sqrt{\frac{n}{n_N}} \alpha'_N(A_j^{(N)}) \right). \tag{3.3}$$

3.1. Proof of Proposition 2.1

Let prove (2.5), (2.6) and (2.7) respectively at Step 1, Step 2 and Step 3.

Step 1. Let $0 \leq N \leq N_0$. With (3.3) one can write that

$$\begin{aligned}
&\mathbb{P}(||\widetilde{\alpha}_n^{(N)}||_{\mathcal{F}} > t) \\
&\leq \mathbb{P}\left(\frac{K_{\mathcal{F}} m_{(N)} \left(2||\widetilde{\alpha}_n^{(N-1)}||_{\mathcal{F}} + \sqrt{\frac{n}{n_{(N)}}} \Lambda'_n \right)}{p_{(N)} - ||\widetilde{\alpha}_n^{(N-1)}||_{\mathcal{F}}/\sqrt{n}} > t \right) \\
&\leq \mathbb{P}\left(\Lambda'_n > \sqrt{\frac{n_{(N)}}{n}} \frac{t p_{(N)}}{2 m_{(N)} K_{\mathcal{F}}} \right) \\
&\quad + \mathbb{P}\left(||\widetilde{\alpha}_n^{(N-1)}||_{\mathcal{F}} > \frac{t p_{(N)}}{4 m_{(N)} K_{\mathcal{F}} (1 + t/\sqrt{n})} \right) \\
&\leq \mathbb{P}\left(\Lambda'_n > \sqrt{\frac{n_{(N_0)}}{n}} \frac{t p_{(N_0)}}{2 m_{(N_0)} K_{\mathcal{F}}} \right) \\
&\quad + \mathbb{P}\left(||\widetilde{\alpha}_n^{(N-1)}||_{\mathcal{F}} > \frac{t p_{(N)}}{4 m_{(N)} K_{\mathcal{F}} (1 + t/\sqrt{n})} \right).
\end{aligned} \tag{3.4}$$

By (3.1) and induction on (3.4), one find

$$\mathbb{P}\left(||\widetilde{\alpha}_n^{(N)}||_{\mathcal{F}} > t \right) \leq \mathbb{P}\left(||\widetilde{\alpha}_n^{(0)}||_{\mathcal{F}} > \frac{t p_{(N)}^N}{4^N m_{(N)}^N K_{\mathcal{F}}^N (1 + t/\sqrt{n})^N} \right) + 2N_0^2 m_{(N_0)} \exp\left(-\frac{n_{(N_0)} p_{(N_0)}^2 t^2}{2 n m_{(N_0)}^2 K_{\mathcal{F}}^2} \right).$$

The right-hand side of the last inequality is increasing with N which leads to (2.5). Since

$$\widetilde{\alpha}_n^{(0)}(\mathcal{F}) = \alpha_n(\mathcal{F}) = \alpha_n^{(0)}(\mathcal{F}), \tag{3.5}$$

Talagrand inequality can be applied to control the deviation probability of $||\widetilde{\alpha}_n^{(0)}||_{\mathcal{F}}$ as described in the next two steps.

Step 2. According to Theorem 1.3 of [14] or Theorem 2.14.9 of [15], if $\mathcal{F}$ satisfies (VC) there exists a constant $D = D(c_0) > 0$ such that, for t_0 large enough and $t \geq t_0$,

$$\mathbb{P}\left(||\widetilde{\alpha}_n^{(0)}||_{\mathcal{F}} > t \right) \leq \left(\frac{Dt}{M_{\mathcal{F}} \sqrt{\nu_0}} \right)^{\nu_0} \exp\left(\frac{-2t^2}{M_{\mathcal{F}}^2} \right). \tag{3.6}$$

Inequalities (2.5) and (3.6) imply (2.6) for all $t_0 \leq t \leq 2M_{\mathcal{F}}\sqrt{n}$, where $D_1, D_2 > 0$ are defined by

$$D_1 = N_0 \left(\frac{D p_{(N_0)}^{N_0}}{\nu_0 4^{N_0} m_{(N_0)}^{N_0} K_{\mathcal{F}}^{N_0+1}} \right)^{\nu_0},$$
$$D_2 = \frac{p_{(N_0)}^{2N_0}}{72^{N_0} m_{(N_0)}^{2N_0} K_{\mathcal{F}}^{3N_0+1}}. \tag{3.7}$$

Step 3. According to Corollaries 4.3 of [11] and 2 of [5] – or Theorems 2.14.2 and 2.14.25 of [15] – if $\mathcal{F}$ satisfies (BR), there exists universal constants $D, D' > 0$ such that for all $t_0 < t < t_1$,

$$\mathbb{P}\left(||\widetilde{\alpha}_n^{(0)}||_{\mathcal{F}} > t \right) \leq \exp(-D'' t^2), \tag{3.8}$$

where $t_0 = 2DM_{\mathcal{F}}(1 + b_0/(1 - r_0))$, $t_1 = 2D\sigma_{\mathcal{F}}^2\sqrt{n}/M_{\mathcal{F}}$, $D'' = D'/4D^2\sigma_{\mathcal{F}}^2$. Therefore (2.5) and (3.8) yields (2.7) where $D_3, D_4 > 0$ are defined by

$$D_3 = N_0, \quad D_4 = \frac{D'' p_{(N_0)}^{2N_0}}{8^{N_0} m_{(N_0)}^{2N_0} K_{\mathcal{F}}^{2N_0} (1 + 2D\sigma_{\mathcal{F}}^2/M_{\mathcal{F}})^{2N_0}}. \tag{3.9}$$

3.2. Proof of Theorem 2.3

According to Proposition 2.1, inequality (3.1) and Proposition 3 of [1], there exists $D > 0$ such that

$$\mathbb{P}\left(\{\Lambda_n > D\sqrt{\log(n)}\} \bigcup \{\Lambda_n' > D\sqrt{\log(n)}\} \right) \leq \frac{1}{3n^2}. \tag{3.10}$$

According to Theorem 2.1 of [1], one can define on the same probability space a sequence $\{X_n\}$ of independent random variable with law P and a sequence $\{\mathbb{G}_n\}$ of versions of $\mathbb{G}$ satisfying the following property. There exists $n_1, d_1 > 0$ such that for all $n > n_1$,

$$\mathbb{P}\left(\sup_{0 \leq N \leq N_0} ||\alpha_n^{(N)} - \mathbb{G}_n^{(N)}||_{\mathcal{F}} > d_1 v_n \right) \leq \frac{1}{3n^2},$$

where $\mathbb{G}_n^{(N)}$ is the version of $\mathbb{G}^{(N)}$ derived from $\mathbb{G}_n^{(0)} = \mathbb{G}_n$ through (2.3). To show (2.8) it remains to prove, by (3.5), that for all n large enough and some $d_0 > 0$,

$$\mathbb{P}\left(\sup_{0 \leq N \leq N_0} ||\widetilde{\alpha}_n^{(N)} - \alpha_n^{(N)}||_{\mathcal{F}} > d_0 \sqrt{\frac{n \log(n)}{n_{(N_0)}}} \right) \leq \frac{2}{3n^2}.$$

Let $1 \le N \le N_0$. Decompositions of $\alpha_n^{(N)}$ and $\widetilde{\alpha}_n^{(N)}$ respectively given by (3.2) and (3.3) imply that

$$\begin{aligned}
&\widetilde{\alpha}_n^{(N)}(f) - \alpha_n^{(N)}(f) \\
&= \sum_{j=1}^{m_N} \frac{\mathbb{P}'_N(A_j^{(N)})}{\widetilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} (\widetilde{\alpha}_n^{(N-1)}(f\mathbb{1}_{A_j^{(N)}}) - \alpha_n^{(N-1)}(f\mathbb{1}_{A_j^{(N)}})) \\
&\quad + \alpha_n^{(N-1)}(f\mathbb{1}_{A_j^{(N)}}) \left(\frac{\mathbb{P}'_N(A_j^{(N)})}{\widetilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} - \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} \right) \\
&\quad - P(f\mathbb{1}_{A_j^{(N)}}) \left(\frac{\widetilde{\alpha}_n^{(N-1)}(A_j^{(N)})}{\widetilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} - \frac{\alpha_n^{(N-1)}(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} \right) \\
&\quad + \sqrt{\frac{n}{n_N}} \frac{P(f\mathbb{1}_{A_j^{(N)}})}{\widetilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} \alpha'_N(A_j^{(N)}).
\end{aligned} \tag{3.11}$$

By (3.5) for $N = 1$ it holds in particular

$$\begin{aligned}
\widetilde{\alpha}_n^{(1)}(f) - \alpha_n^{(1)}(f) &= \sum_{j=1}^{m_1} \alpha_n(f\mathbb{1}_{A_j^{(1)}}) \left(\frac{\mathbb{P}'_{n_1}(A_j^{(1)}) - P(A_j^{(1)})}{\mathbb{P}_n(A_j^{(1)})} \right) \\
&\quad + \sqrt{\frac{n}{n_1}} \frac{P(f\mathbb{1}_{A_j^{(1)}})}{\mathbb{P}_n(A_j^{(1)})} \alpha'_{n_1}(A_j^{(1)}),
\end{aligned}$$

which is uniformly and roughly bounded by

$$||\widetilde{\alpha}_n^{(1)} - \alpha_n^{(1)}||_{\mathcal{F}} \le \frac{m_{(N)} K_{\mathcal{F}} \Lambda'_n}{p_{(N)} - \Lambda_n/\sqrt{n}} \sqrt{\frac{n}{n_{(N)}}} (1 + \Lambda_n/\sqrt{n}). \tag{3.12}$$

Let $C_{n,N} = 4m_{(N)} K_{\mathcal{F}}/(p_{(N)} - \Lambda_n/\sqrt{n})^2$. Equality (3.11) implies also

$$\begin{aligned}
&||\widetilde{\alpha}_n^{(N)}(f) - \alpha_n^{(N)}(f)||_{\mathcal{F}} \\
&\le C_{n,N} \left(||\widetilde{\alpha}_n^{(N-1)} - \alpha_n^{(N-1)}||_{\mathcal{F}} + \frac{\Lambda_n^2}{\sqrt{n}} + \frac{\Lambda'_n(\Lambda_n + \sqrt{n})}{\sqrt{n_{(N)}}} \right).
\end{aligned}$$

By induction of the last inequality and noticing that for all $n > 0$, $m_{(N)} K_{\mathcal{F}}/(p_{(N)} - \Lambda_n/\sqrt{n})^2 \ge 1$, one have

$$\begin{aligned}
||\widetilde{\alpha}_n^{(N)}(f) - \alpha_n^{(N)}(f)||_{\mathcal{F}} &\le C_{n,N}^{N-1} ||\widetilde{\alpha}_n^{(1)} - \alpha_n^{(1)}||_{\mathcal{F}} \\
&\quad + (N-1) C_{n,N}^{N-1} \left(\frac{\Lambda_n^2}{\sqrt{n}} + \frac{\Lambda'_n(\Lambda_n + \sqrt{n})}{\sqrt{n_{(N)}}} \right),
\end{aligned}$$

then inequality (3.12) immediately implies that

$$||\widetilde{\alpha}_n^{(N)}(f) - \alpha_n^{(N)}(f)||_{\mathcal{F}} \le N C_{n,N}^N \left(\frac{\Lambda_n^2}{\sqrt{n}} + \frac{\Lambda'_n(\Lambda_n + \sqrt{n})}{\sqrt{n_{(N)}}} \right).$$

Since the right-hand side of the last inequality is increasing with N then for all $t > 0$,

$$\begin{aligned}&\mathbb{P}\left(\sup_{1\leq N\leq N_0}||\widetilde{\alpha}_n^{(N)}(f)-\alpha_n^{(N)}(f)||_{\mathcal{F}}>t\right)\\&\leq\mathbb{P}\left(\frac{C_0}{(p_{(N_0)}-\Lambda_n/\sqrt{n})^{2N_0}}\left(\frac{\Lambda_n^2}{\sqrt{n}}+\frac{\Lambda_n'(\Lambda_n+\sqrt{n})}{\sqrt{n_{(N_0)}}}\right)>t\right),\end{aligned}\tag{3.13}$$

with $C_0 = N_0^2(4m_{(N_0)}K_{\mathcal{F}}N_0)^{N_0} > 0$. There exists $n_2 > 0$ such that for all $n > n_2$ it holds $D\sqrt{\log(n)/n} \leq p_{(N_0)}/2 \leq 1/2$. For $n > n_2$, according to (3.10) and (3.13),

$$\begin{aligned}&\mathbb{P}\left(\sup_{1\leq N\leq N_0}||\widetilde{\alpha}_n^{(N)}(f)-\alpha_n^{(N)}(f)||_{\mathcal{F}}>t\right)\\&\leq\mathbb{P}\left(\Lambda_n>D\sqrt{\log(n)}\right)\\&\quad+\mathbb{P}\left(\Lambda_n'>\frac{1}{2}\sqrt{\frac{n_{(N_0)}}{n}}\left(\frac{tp_{(N_0)}^{2N_0}}{4^{N_0}C_0}-\frac{D^2\log(n)}{\sqrt{n}}\right)\right)\\&\leq\frac{1}{3n^2}+\mathbb{P}\left(\Lambda_n'>\frac{1}{2}\sqrt{\frac{n_{(N_0)}}{n}}\left(\frac{tp_{(N_0)}^{2N_0}}{4^{N_0}C_0}-\frac{D^2\log(n)}{\sqrt{n}}\right)\right).\end{aligned}$$

By using (3.10) again, the last inequality implies

$$\mathbb{P}\left(\sup_{1\leq N\leq N_0}||\widetilde{\alpha}_n^{(N)}(f)-\alpha_n^{(N)}(f)||_{\mathcal{F}}>t_n\right)\leq\frac{2}{3n^2},$$

for all $n > n_2$ and

$$t_n=\frac{4^{N_0+1}C_0D}{p_{(N_0)}^{2N_0}}\left(\sqrt{\frac{n\log(n)}{n_{(N_0)}}}+\frac{D\log(n)}{\sqrt{n}}\right).$$

By definition of v_n, there exists $d_2 > \max(d_1, 4^{N_0+1}C_0D/p_{(N_0)}^{2N_0})$ and $n_3 > 0$ such that for all $n > n_3$,

$$d_2\left(v_n+\sqrt{\frac{n\log(n)}{n_{(N_0)}}}\right)>d_1v_n+t_n.$$

Then (2.8) is proved for $d_0 = d_2$ and $n_0 = \max(n_0, n_1, n_3)$.

3.3. Proof of Proposition 2.4

Let assume that $n\log(n) = o(n_{(N_0)})$. According to Theorem 2.1 of [1] and Theorem 2.3, i.i.d random variables $X_1, \ldots, X_n$ with law P and a sequence $z_n \sim \mathcal{N}(0, 1)$ can be defined such that for $n > n_1$ for some $n_1 > 0$, $\mathbb{P}(\mathcal{Z}_n) \leq 1/n^2$ with

$$\begin{aligned}\mathcal{Z}_n^{(N)}=\ &\{|\alpha_n(f)/\sigma_f-z_n|>u_n\}\bigcup\left\{|\alpha_n^{(N)}(f)/\sigma_f^{(N)}-z_n|>u_n\right\}\\&\bigcup\left\{|\widetilde{\alpha}_n^{(N)}(f)/\sigma_f^{(N)}-z_n|>u_n\right\},\end{aligned}$$

where u_n is a sequence with null limit. The strong approximation implies that

$$\begin{aligned}
\lim_{n\to+\infty} \frac{\mathbb{P}(|Z_n| \le t_\alpha)}{\mathbb{P}(|z_n + M_n/\sigma_f| \le t_\alpha)} &= 1,\\
\lim_{n\to+\infty} \frac{\mathbb{P}(|Z_n^{(N)}| \le t_\alpha)}{\mathbb{P}(|z_n + M_n/\sigma_f^{(N)}| \le t_\alpha)} &= 1,\\
\lim_{n\to+\infty} \frac{\mathbb{P}(|\widetilde{Z}_n^{(N)}| \le t_\alpha)}{\mathbb{P}(|z_n + M_n/\sigma_f^{(N)}| \le t_\alpha)} &= 1,
\end{aligned} \tag{3.14}$$

with $M_n = \sqrt{n}(P(f) - P_0(f))$. If f_{μ,σ^2} is the density function of $\mathcal{N}(\mu, \sigma^2)$ then

$$\begin{aligned}
\mathbb{P}(|z_n + M_n/\sigma_f| \le t_\alpha) &\ge 2t_\alpha \inf_{[-t_\alpha, t_\alpha]} f_{M_n,1}\\
&\ge \frac{2t_\alpha}{\sqrt{2\pi}} \exp\left(-(M_n/\sigma_f + t_\alpha)^2\right),\\
\mathbb{P}(|z_n + M_n/\sigma_f^{(N)}| \le t_\alpha) &\le 2t_\alpha \sup_{[-t_\alpha, t_\alpha]} f_{M_n,1}\\
&\le \frac{2t_\alpha}{\sqrt{2\pi}} \exp\left(-(M_n/\sigma_f^{(N)} - t_\alpha)^2\right).
\end{aligned}$$

which implies

$$\frac{\mathbb{P}(|z_n + M_n/\sigma_f| \le t_\alpha)}{\mathbb{P}(|z_n + M_n/\sigma_f^{(N)}| \le t_\alpha)} \ge \exp\left(M_n^2\left(\frac{1}{\sigma_f^{(N)}} - \frac{1}{\sigma_f}\right) - 2t_\alpha|M_n|\left(\frac{1}{\sigma_f^{(N)}} + \frac{1}{\sigma_f}\right)\right)$$

For n large enough,

$$\frac{\mathbb{P}(|z_n + M_n/\sigma_f| \le t_\alpha)}{\mathbb{P}(|z_n + M_n/\sigma_f^{(N)}| \le t_\alpha)} \ge \exp(b_n), \tag{3.15}$$

where b_n is a sequence such that $b_n \sim M_n^2\left(1/\sigma_f^{(N)} - 1/\sigma_f\right)$. Then (3.14) and (3.15) imply (2.10) and (2.11).

3.4. Proof of Proposition 2.5

Denote $X \cdot Y$ the product scalar of X and Y and $\mathcal{C} \in \mathbb{R}^m$ the random vector defined by

$$\mathcal{C} = (C_1, \ldots, C_m) = (\mathbb{1}_{B_1}/\sqrt{P(B_1)}, \ldots, \mathbb{1}_{B_m}/\sqrt{P(B_m)}).$$

The case (H_0) is dealt at Step 1 and the case (H_1) at Step 2.

Step 1. Under (H_0), $T_n = \alpha_n[\mathcal{C}] \cdot \alpha_n[\mathcal{C}]^T$, $T_n^{(N)} = \alpha_n^{(N)}[\mathcal{C}] \cdot \alpha_n^{(N)}[\mathcal{C}]^T$ and $\widetilde{T}_n^{(N)} = \widetilde{\alpha}_n^{(N)}[\mathcal{C}] \cdot \widetilde{\alpha}_n^{(N)}[\mathcal{C}]^T$. The statistic $\alpha_n[\mathcal{C}]$ converges weakly to a multi-normal random variable $Y \sim \mathcal{N}(\mathbf{0}, \Sigma)$ while $\alpha_n^{(N)}[\mathcal{C}], \widetilde{\alpha}_n^{(N)}$ converge weakly to $Y^{(N)} \sim \mathcal{N}(\mathbf{0}, \Sigma^{(N)})$ according to Theorem 2.1 of [1] and Theorem 2.3. By Proposition 7 of [1], $\Sigma - \Sigma^{(N)}$ is

positive definite which implies for all $\alpha > 0$,

$$\mathbb{P}(Y \cdot Y^T \geq t_\alpha) \geq \mathbb{P}(Y^{(N)} \cdot (Y^{(N)})^T \geq t_\alpha),$$

and consequently (2.15), (2.16) by definition of weak convergence.

Step 2. Under (H_1), there exists $i \in \{1, \ldots, m\}$ such that $P_0(B_i) \neq P(B_i)$ which implies

$$\begin{aligned}\min(|T_n|, |T_n^{(N)}|, \widetilde{T}_n^{(N)}|) > &-\Lambda_n^2 - 2\sqrt{n}\Lambda_n |P_0(C_i) - P(C_i)| \\ &+ n(P_0(C_i) - P(C_i))^2.\end{aligned}$$

By Borel-Cantelli and (3.10) with probability one there exists $n_1 > 0$ such that for all $n > n_1$, $\Lambda_n < D\sqrt{\log(n)}$. For $n > n_1$, one have

$$\begin{aligned}t_n &< \min(|T_n|, |T_n^{(N)}|, \widetilde{T}_n^{(N)}|), \\ t_n &= -D^2 \log(n) - 2D\sqrt{n \log(n)} |P_0(C_i) - P(C_i)| \\ &\quad + n(P_0(C_i) - P(C_i))^2.\end{aligned}$$

Since $\lim_{n \to +\infty} t_n = +\infty$, for all $\alpha \in (0, 1)$ there exists $n_2 > 0$ such that $t_n > t_\alpha$ for all $n > n_2$. Inequality (2.17) is satisfied for $n_0 = \max(n_1, n_2)$.

Appendix A. Numerical example of a raked mean

The usual way to calculate the mean of $X_1, \ldots, X_n$ is to sum the data X_i multiplied by the weights $w_i = 1/n$. If one have the auxiliary information $P[\mathcal{A}^{(N)}] = (P(A_1^{(N)}), \ldots, P(A_{m_N}^{(N)}))$ for $1 \leq N \leq N_0$ one want to change iteratively the initial weights w_i in new weights $w_i^{(N)}$ such that $\sum_{i=1}^n w_i^{(N)} = 1$ and

$$\sum_{i=1}^n w_i^{(N)} \mathbb{1}_{A_j}^{(N)}(X_i) = P(A_j^{(N)}),$$

for any $1 \leq N \leq N_0$ and $1 \leq j \leq m_N$. Recall that it does not imply that $\sum_{i=1}^n w_i^{(N_1)} \mathbb{1}_{A_j}^{(N_2)}(X_i) = P(A_j^{(N_2)})$ with $N_1 \neq N_2$ and $1 \leq j \leq N_2$. For this example one takes $N_0 = 2, \mathcal{A}^{(1)} = \mathcal{A} = \{A_1, A_2, A_3\}, \mathcal{A}^{(2)} = \mathcal{B} = \{B_1, B_2\}$ and one generates normal random values X_i with fixed variances $\sigma^2 = 0.1$ and such that the probabilities and conditional expectations are given by the Tables A.1 and A.2. In particular,

$$\begin{aligned}P[\mathcal{A}] &= (P(A_1), P(A_2), P(A_3)) = (0.45, 0.35, 0.2), \\ P[\mathcal{B}] &= (P(B_1), P(B_2)) = (0.55, 0.45), \\ P(X) &= \mathbb{E}[X] = 0.225, \\ \mathbb{E}[X|\mathcal{A}] &= (\mathbb{E}[X|A_1], \mathbb{E}[X|A_2], \mathbb{E}[X|A_3]) \simeq (0.611, -0.286, 0.25), \\ \mathbb{E}[X|\mathcal{B}] &= (\mathbb{E}[X|B_1], \mathbb{E}[X|B_2]) \simeq (0.227, 0.222).\end{aligned}$$

By generating $n = 10$ values, the following data are obtained from Table A.3. In this case, the usual mean is the sum of all X_i over 10 that is the weight $1/n = 0.1$ is assigned at each X_i and $\mathbb{P}_n(X) \simeq 0.055$. After one step, the weights $0.15, 0.07, 0.1$ are assigned at individuals belonging respectively to A_1, A_2, A_3. The raked mean for $N = 1$ is

$$\mathbb{P}_n^{(1)}(X) = 0.15 \times \frac{P(A_1)}{\mathbb{P}_n(A_1)} + 0.07 \times \frac{P(A_2)}{\mathbb{P}_n(A_2)} + 0.1 \times \frac{P(A_3)}{\mathbb{P}_n(A_3)} \simeq 0.2.$$

TABLE A.1. Probabilities of sets.

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

TABLE A.2. Conditional expectations of the generated random variables.

$\mathbb{E}[X\|A_i \cap B_j]$	A_1	A_2	A_3
B_1	0.75	-0.5	1
B_2	0.5	0.25	-0.5

TABLE A.3. Generated random variables.

X_i	$\mathcal{A}$	$\mathcal{B}$
0.953	1	1
0.975	1	1
0.058	1	1
-0.766	2	1
-0.644	2	1
-0.819	2	1
0.028	2	2
0.627	2	2
1.04	3	1
-0.904	3	2

TABLE A.4. Final raked weights.

$w_i^{(\infty)}$	A_1	A_2	A_3
B_1	0.15	0.024	0.029
B_2	X	0.139	0.17

When the algorithm is stabilized in this case the final weights are given by Table A.4. Notice that the cross means that no random variable belonging to $A_1 \cap B_2$ was generated due to a low value of n. This kind of situation can sometimes prevent the convergence of the method. The final raked mean is $\mathbb{P}_n^{(\infty)}(X) \simeq 0.212$ which is closer of $P(X)$ than the usual mean $\mathbb{P}_n(X)$.

APPENDIX B. CALCULATION OF $\sigma_f^{(\infty)}$

Notations of the Section 4.4 of [1] – concerning the proof of their Proposition 11 in the aim to establish the expression of $\mathbb{G}^{(\infty)}$ – are used and remind notation (2.12). The calculation uses the two following stochastic matrices

$$\mathbf{P}_{\mathcal{A}|\mathcal{B}} = \begin{pmatrix} P(A|B) & P(A^C|B) \\ P(A|B^C) & P(A^C|B^C) \end{pmatrix} = \begin{pmatrix} p_{AB}/p_B & 1 - p_{AB}/p_B \\ (p_A - p_{AB})/p_{\overline{B}} & 1 - (p_A - p_{AB})/p_{\overline{B}} \end{pmatrix},$$

$$\mathbf{P}_{\mathcal{B}|\mathcal{A}} = \begin{pmatrix} P(B|A) & P(B^C|A) \\ P(B|A^C) & P(B^C|A^C) \end{pmatrix} = \begin{pmatrix} p_{AB}/p_A & 1 - p_{AB}/p_A \\ (p_B - p_{AB})/p_{\overline{A}} & 1 - (p_B - p_{AB})/p_{\overline{A}} \end{pmatrix}, \tag{B.1}$$

the two following conditional expectation vectors

$$\mathbb{E}[f|\mathcal{A}] = (\mathbb{E}[f|A], \mathbb{E}[f|A^C]), \quad \mathbb{E}[f|\mathcal{B}] = (\mathbb{E}[f|B], \mathbb{E}[f|B^C]), \tag{B.2}$$

the two following covariance matrices

$$\mathrm{Var}(\mathbb{G}[A]) = p_A p_{\overline{A}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad \mathrm{Var}(\mathbb{G}[B]) = p_B p_{\overline{B}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \tag{B.3}$$

and the two following vectors

$$\begin{aligned} V_1(f) &= \mathbb{E}[f|\mathcal{A}] - \mathbf{P}_{\mathcal{B}|\mathcal{A}} \cdot \mathbb{E}[f|\mathcal{B}] \\ &= \begin{pmatrix} \mathbb{E}[f|A] \\ \mathbb{E}[f|A^C] \end{pmatrix} - \begin{pmatrix} p_{AB}/p_A & 1 - p_{AB}/p_A \\ (p_B - p_{AB})/p_{\overline{A}} & 1 - (p_B - p_{AB})/p_{\overline{A}} \end{pmatrix} \cdot \begin{pmatrix} \mathbb{E}[f|B] \\ \mathbb{E}[f|B^C] \end{pmatrix} \\ &= (\mathbb{E}[f](p_A - p_{AB}) - \mathbb{E}[f|A]p_A p_{\overline{B}} + \mathbb{E}[f|B](p_{AB} - p_A p_B)) \cdot \begin{pmatrix} -1/p_A p_{\overline{B}} \\ 1/p_{\overline{A}} p_{\overline{B}} \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} V_2(f) &= \mathbb{E}[f|\mathcal{B}] - \mathbf{P}_{\mathcal{A}|\mathcal{B}} \cdot \mathbb{E}[f|\mathcal{A}] \\ &= \begin{pmatrix} \mathbb{E}[f|B] \\ \mathbb{E}[f|B^C] \end{pmatrix} - \begin{pmatrix} p_{AB}/p_B & 1 - p_{AB}/p_B \\ (p_A - p_{AB})/p_{\overline{B}} & 1 - (p_A - p_{AB})/p_{\overline{B}} \end{pmatrix} \cdot \begin{pmatrix} \mathbb{E}[f|A] \\ \mathbb{E}[f|A^C] \end{pmatrix} \\ &= (\mathbb{E}[f](p_B - p_{AB}) - \mathbb{E}[f|B]p_{\overline{A}} p_B + \mathbb{E}[f|A](p_{AB} - p_A p_B)) \cdot \begin{pmatrix} -1/p_{\overline{A}} p_B \\ 1/p_{\overline{A}} p_{\overline{B}} \end{pmatrix}. \end{aligned}$$

The eigenvalues of $\mathbf{P}_{\mathcal{A}|\mathcal{B}} \cdot \mathbf{P}_{\mathcal{B}|\mathcal{A}}$ and $\mathbf{P}_{\mathcal{B}|\mathcal{A}} \cdot \mathbf{P}_{\mathcal{A}|\mathcal{B}}$ are 1 and $T_1 = T_2 = (p_{AB} - p_A p_B)^2 / p_A p_{\overline{A}} p_B p_{\overline{B}}$. Their eigenvectors associated to T_1 and T_2 are respectively $(p_{\overline{B}}/p_B, -1)^t$ and $(p_{\overline{A}}/p_A, -1)^t$ which implies

$$U_1 = \begin{pmatrix} 1 & p_{\overline{A}}/p_A \\ 1 & -1 \end{pmatrix}, \quad U_2 = \begin{pmatrix} 1 & p_{\overline{B}}/p_B \\ 1 & -1 \end{pmatrix}.$$

For the case of two marginals, Albertus and Berthet showed that $\mathbb{G}^{(N)}$ converge almost surely to $\mathbb{G}^{(\infty)}(f) = \mathbb{G}(f) - S_{1,even}(f)^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,odd}(f)^t \cdot \mathbb{G}[\mathcal{B}]$ where

$$S_{1,even}(f) = U_1 \begin{pmatrix} 0 & 0 \\ 0 & (1 - T_1)^{-1} \end{pmatrix} \cdot U_1^{-1} \cdot V_1(f) = C_{1,even}(f) \begin{pmatrix} -p_{\overline{A}} p_B \\ p_A p_B \end{pmatrix},$$

$$C_{1,even}(f) = \frac{\mathbb{E}[f|B](p_{AB} - p_A p_B) - \mathbb{E}[f|A]p_A p_{\overline{B}} - \mathbb{E}[f](p_{AB} - p_A)}{p_A p_B p_{\overline{A}} p_{\overline{B}} - (p_{AB} - p_A p_B)^2},$$

$$S_{2,odd}(f) = U_2 \begin{pmatrix} 0 & 0 \\ 0 & (1 - T_2)^{-1} \end{pmatrix} \cdot U_2^{-1} \cdot V_2(f) = C_{2,odd}(f) \begin{pmatrix} -p_A p_{\overline{B}} \\ p_A p_B \end{pmatrix},$$

$$C_{2,odd}(f) = \frac{\mathbb{E}[f|A](p_{AB} - p_A p_B) - \mathbb{E}[f|B]p_{\overline{A}}p_B - \mathbb{E}[f](p_{AB} - p_B)}{p_A p_B p_{\overline{A}} p_{\overline{B}} - (p_{AB} - p_A p_B)^2}.$$

By linearity of $f \mapsto \mathbb{G}(f)$ and the fact that $\mathbb{G}(a) = 0$ for any constant $a \in \mathbb{R}$ one can write

$$\mathbb{G}^{(\infty)}(f) = \mathbb{G}\left(f + p_B C_{1,even}(f)\mathbb{1}_A + p_A C_{2,odd}(f)\mathbb{1}_B\right),$$

which implies that

$$\begin{aligned}
\sigma_f^{(\infty)} &= \mathrm{Var}(\mathbb{G}^{(\infty)}(f)) \\
&= \mathrm{Var}(f) + \mathrm{Var}(p_B C_{1,even}(f)\mathbb{1}_A + p_A C_{2,odd}(f)\mathbb{1}_B) \\
&\quad + 2\mathrm{Cov}(f, p_B C_{1,even}(f)\mathbb{1}_A + p_A C_{2,odd}(f)\mathbb{1}_B) \\
&= \mathrm{Var}(f) + p_A p_{\overline{A}} p_B^2 C_{1,even}^2(f) + p_A^2 p_B p_{\overline{B}} C_{2,odd}^2(f) \\
&\quad + 2p_A p_B C_{1,even}(f) C_{2,odd}(f)(p_{AB} - p_A p_B) \\
&\quad + 2p_A p_B \left(C_{1,even}(f)\Delta_A + C_{2,odd}(f)\Delta_B\right)
\end{aligned}$$

With some calculations the simple expression of $\sigma_f^{(\infty)}$ given by (2.13) is found.

References

[1] M. Albertus and P. Berthet, Auxiliary information: the raking-ratio empirical process. *Electron. J. Stat.* **13** (2019) 120–165.
[2] M.D. Bankier, Estimators based on several stratified samples with applications to multiple frame surveys. *J. Am. Stat. Assoc.* **81** (1986) 1074–1079.
[3] P. Berthet and D.M. Mason, Revisiting two strong approximation results of Dudley and Philipp. *JSTOR* **51** (2006) 155–172.
[4] D.A Binder and A. Théberge, Estimating the variance of raking-ratio estimators. *Canad. J. Statist.* **16** (1988) 47–55.
[5] L. Birgé and P. Massart, Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4** (1998) 329–375.
[6] G.J. Brackstone and J.N.K. Rao, An investigation of raking ratio estimators. *Indian J. Stat.* **41** (1979) 97–114.
[7] G. Choudhry and H. Lee, Variance estimation for the canadian labour force survey. *Survey Methodol.* **13** (1987) 147–161.
[8] W.E. Deming and F.F. Stephan, On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Stat.* **11** (1940) 427–444
[9] C.T. Ireland and S. Kullback, Contingency tables with given marginals. *Biometrika* **55** (1968) 179–188.
[10] H.S Konijn, Biases, variances and covariances of raking ratio estimators for marginal and cell totals and averages of observed characteristics. *Metrika* **28** (1981) 109–121.
[11] D. Pollard, Asymptotics via empirical processes. *Statist. Sci.* **4** (1989) 341–366.
[12] R. Sinkhorn, A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.* **35** (1964) 876–879.
[13] F.F. Stephan, An iterative method of adjusting sample frequency tables when expected marginal totals are known. *Ann. Math. Stat.* **13** (1942) 166–178.
[14] M. Talagrand, Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** (1994) 28–76.
[15] A.W. van der Vaart and J.A. Wellner, *Weak convergence and empirical processes*. Springer Series in Statistics (Springer-Verlag, New York, 1996). With applications to statistics.