



HAL
open science

Processus empirique avec informations auxiliaires

Mickael Albertus

► **To cite this version:**

Mickael Albertus. Processus empirique avec informations auxiliaires. Statistiques [math.ST]. Université Paul Sabatier - Toulouse III, 2019. Français. NNT : 2019TOU30095 . tel-02501305

HAL Id: tel-02501305

<https://theses.hal.science/tel-02501305>

Submitted on 6 Mar 2020

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

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



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *17 septembre 2019* par :

Mickael Albertus

Processus empirique avec informations auxiliaires

JURY

JEAN-MARC AZAIS
PHILIPPE BERTHET
GILLES BLANCHARD
JEAN-CLAUDE FORT
IRÈNE GIJBELS
AGNÈS LAGNOUX

Université Toulouse 3
Université Toulouse 3
Université de Potsdam
Université Paris 5
KU Leuven
Université Toulouse 2

Membre du Jury
Directeur de thèse
Rapporteur
Membre du Jury
Rapporteur
Membre du Jury

École doctorale et spécialité :

MITT : Domaine Mathématiques : Mathématiques appliquées

Unité de Recherche :

Institut de Mathématiques de Toulouse (UMR 5219)

Directeur de Thèse :

Philippe Berthet

Rapporteurs :

Gilles Blanchard et Irène Gijbels

Remerciements

Je tiens tout d'abord à remercier Philippe Berthet, mon directeur de thèse qui m'a en tout premier lieu donné l'envie de continuer mes études de mathématiques à la période de ma vie où je me demandais vers quelle voie m'orienter. Plus que cela, il m'a fait aimer le domaine des probabilités et de la statistique alors que j'avais ceux-là en horreur à cette époque. Son dynamisme et ses idées à foison m'ont poussé à me dépasser au niveau de cette thèse. Je le remercie dans un second temps de m'avoir fait confiance pour accomplir ces trois années de thèse ainsi que de m'avoir accordé du temps malgré ses fonctions. Je n'ai malheureusement pas eu le temps de traiter toutes les idées évoquées lors de nos entretiens mais chacune d'entre elles était une source d'inspiration et je l'en remercie pour cela.

Je suis reconnaissant envers Gilles Blanchard et Irène Gijbels d'avoir accepté de relire mon travail et pris le temps de me dire comment améliorer mon manuscrit. Un merci également aux membres du jury, Agnès Lagnoux, Jean-Marc Azais et Jean-Claude Fort de me faire l'honneur de leur présence.

Je tiens également à remercier certains membre de l'IMT, pas seulement membre de l'ESP, qui m'ont communiqué leur engouement des mathématiques : Laure Coutin, Clément Pellegrini, Claire Dartyge, Katy Paroux, Franck Barthe, Stéphane Lamy, Patrick Cattiaux, Joseph Najnudel, Joan Millès,... Aux permanents de l'IMT qui animent les lieux : Guillaume Cebon, Laurent Miclo, Réda Chaïbi et Sebastien Gadat. Dans la continuité des remerciements des membres de l'université Paul Sabatier, je tenais à remercier Sylvie Crabos, secrétaire souriante et dévouée aux biens des élèves, ainsi que toutes les personnes qui m'ont aidé au niveau administratif, à savoir Marie-Line Domenjole, Agnès Requis et Martine Labruyère.

Je remercie tous mes collègues bien entendu. À commencer par les doctorants avec qui j'ai partagé mon bureau, mon espace de travail ! Dans l'ordre chronologique, je devrais commencer par citer Laure (basketteuse et amatrice de vin blanc comme moi !), Hugo (aveyronnais

selon ses dires et souffre-douleur bien que je sois arrivé après lui) et Guillaume (bien qu'il faut des gens comme lui, défenseur hors pair des ourangs-outans et par conséquence *hateur* du Nutella !) qui m'ont fait de suite me sentir chez moi. J'ai apprécié chaque moment avec eux et mon seul regret est de ne pas avoir pu partager ma dernière année avec eux ! Après un déménagement de bureaux, je me suis retrouvé dans le bureau de Maylis, Valentin et An avec qui j'ai passé le plus de temps en tant que co-bureau. Je les remercie tous les trois bien entendu. Maylis avec qui j'ai fait le Master 2 Recherche, avec qui nous nous sommes entraînés et co-aidés durant ces années, avec qui j'ai beaucoup ri notamment sur ses lancers de cornes de brume de petits enfants ! Valentin qui m'a de suite cerné (entre autre mes envies de « bien » manger, oh disons-le franchement, de gras) et qui me doit un bucket ! An pour ces discussions sur le thème gastronomie qui m'ont à chaque fois fait rêver et qui m'impressionne sur sa maîtrise du français (je ne suis pas sûr de pouvoir en faire autant avec le vietnamien même au bout de plusieurs années). Je tiens à remercier tous les autres doctorants qui m'ont fait me sentir à l'aise durant ces trois années de thèse et qui me donnaient la pêche. Fabien pour ses blagues parfois drôles (merci de m'avoir traumatisé des têtards) parfois douteuses avec qui on s'est affronté pendant de nombreux matchs Smash BrosTM. Pierre pour son stoïcisme à toute épreuve (sauf à celle de Keny !) avec qui j'ai pu me dépenser au tennis. Baptiste, président des doctorants, prêt à louer un appart pour qu'on puisse faire des soirées ensemble ! Eva avec qui j'ai partagé ma découverte de l'Amérique latine et une chevauchée de lama de sel ! Le couple Brangelina de l'IMT, William (donne nous des nouvelles et du chocolat suisse !) & Antoine (ex-président des doctorants à qui tu manques), avec qui nous avons fait d'excellents escape-games ! Les petits nouveaux qui vont prendre la relève et avec qui j'ai eu l'occasion de passer un peu de temps : Clément et Laetitia. Et bien entendu tous les autres doctorants avec qui j'ai partagé des bons moments que je n'oublierai pas : Oleksander, Anton, Camille, Jose, Kevin, Trang, Joseph, Sylvère, Jade, Florian, Mael, Claire, Magali, Phong... Je remercie les amis que je me suis fait durant ma scolarité, Marion, Cédric, Raphaël, Laure, Axelle, Anthony, Alix, Théo et tant d'autres ! Je remercie également tous ceux qui m'en voudront de ne pas les avoir cités.

Que seraient ces remerciements si je ne prenais pas le temps de remercier toute ma famille ! En particulier mes parents qui ont su faire preuve de patience et qui ont toujours su m'exprimer leur amour. Mon frère et ma sœur avec qui je partage beaucoup de choses, notamment la curiosité scientifique ! Des remerciements à ma belle-famille, Dominique, Maurice, Mathieu et Anaïs pour leur accueil chaleureux. Ces remerciements vont également à mes meilleurs amis que je considère comme faisant parti de ma famille : Sébastien, mon meilleur ami (my BFF!), Mathilde et Julien qui sont respectivement pour moi comme une seconde sœur et frère, Céline, Vincent et Marie-Charlotte. À toi Pierre, des remerciements sans limite ! Chaque instant à tes côtés est un bonheur, je t'aime +3000 ! ♡

La question principale de mon sujet de thèse est de savoir comment utiliser une information auxiliaire, c'est-à-dire une information que l'on aurait a priori ou qui a été obtenue par une source d'information, afin d'améliorer nos statistiques. Grâce à une information auxiliaire, il paraît naturel que l'on puisse améliorer nos estimations dans le sens où l'on pourrait faire baisser le risque quadratique de nos estimateurs. On pourrait également améliorer certains tests statistiques, dans le sens où on augmente la puissance de ces tests, en exploitant l'information auxiliaire. Les questions sous-jacentes auxquelles j'essaie de répondre dans cette thèse sont donc quelle définition peut-on donner de l'information auxiliaire, comment combiner différentes informations auxiliaires et quelle est la contribution d'une information auxiliaire dans l'amélioration de nos statistiques ?

Ces questions sont d'autant plus pertinentes que nous avançons dans un monde connecté qui a la possibilité de stocker une très grande masse d'informations, peut diffuser de manière quasiment instantanée une partie de cette information et recherche constamment des manières d'exploiter cette abondance d'information, des manières d'affiner les prédictions ou les tests statistiques effectués à partir d'un échantillon donné. Bien que cette problématique soit d'actualité, force est de constater que peu de résultats dans la littérature aborde ce problème d'un point de vue empirique.

Premier chapitre

Dans ce premier chapitre nous regroupons des définitions et résultats qui seront utilisés dans les chapitres suivants. Nous donnons la définition de convergence en loi ainsi que les problèmes de mesurabilité qui peuvent découler. Nous en profitons pour rappeler le théorème de représentation de Skorokhod qui justifie l'outil approximation forte. Nous continuons ce chapitre en donnant la définition, les propriétés et résultats principaux de la mesure et du processus empirique. Nous donnons la définition de ces objets quand ils sont indexés par une classe de fonctions mesurables à valeurs réelles. Nous donnons les résultats liés à l'approximation forte du processus empirique et notamment l'approximation forte de Berthet-Mason, qui concerne l'approximation forte du processus empirique indexé par des classes de fonctions mesurables à valeurs réelles vérifiant certaines conditions de mesurabilité et d'entropie. Il s'agit de l'outil de démonstration le plus important de cette thèse et on fera souvent mention de ce résultat.

Deuxième chapitre

Le deuxième chapitre regroupe les premières traces de recherche de cette thèse. Nous nous sommes focalisés sur l'étude du processus empirique quand on disposait de la probabilité d'ensembles d'une partition donnée comme information auxiliaire. Il s'agit d'un cas particulier du chapitre suivant mais nous pouvons établir des résultats plus puissants.

Nous définissons la mesure et le processus empirique avec ce type d'information auxiliaire. Ces derniers semblent difficiles à étudier pour une taille d'échantillon finie puisqu'elle substitue des poids déterministes de chaque individu de l'échantillon par des poids aléatoires dépendant de l'ensemble de la partition connue auquel appartient l'individu. Les résultats de la statistique théorique empirique nous permettent néanmoins de répondre à ce problème. Nous étudions les propriétés asymptotiques et non-asymptotiques de ces objets. Au niveau des propriétés asymptotiques nous établissons en particulier des résultats d'approximation forte et des résultats de convergence en loi. Pour les propriétés non-asymptotiques, nous donnons une loi du logarithme itéré, une borne pour majorer le biais introduit par la modification du processus empirique pour prendre en compte l'information auxiliaire et nous montrons que cette nouvelle mesure empirique n'est autre que la projection de la mesure empirique initiale sous l'ensemble des mesures de probabilité discrètes vérifiant l'information auxiliaire au sens de la distance de Kullback-Leibler. Nous finissons ce chapitre par un exemple concret, l'amélioration du test d'adéquation de Kolmogorov-Smirnov. Nous apportons une information auxiliaire précise et nous prouvons l'augmentation de la puissance de ce test après avoir utilisé l'information auxiliaire.

Troisième chapitre

Le troisième chapitre concerne la méthode du Raking-Ratio, une méthode statistique permettant d'utiliser l'information auxiliaire donnée par la connaissance de la probabilité d'ensembles de plusieurs partitions, pour améliorer nos estimations par exemple. C'est l'un des chapitres les plus importants de cette thèse puisqu'on aborde la méthode du Raking-Ratio du point de vue de la théorie du processus empirique, ce qui permet d'établir des résultats puissants d'approximation forte et de convergence faible. Le fait qu'on généralise nos travaux à des processus empiriques indexés par des classes de fonction nous permet de rester général au niveau des résultats établis.

Nous commençons par donner un exemple concret de cette méthode sur un tableau de contingence fictif et nous rappelons la littérature concernant celle-ci. Nous évoquons en particulier les résultats les plus importants qui ont été établis. La plupart d'entre eux sont des résultats limités à un nombre restreint de tour ou proposant des formules difficiles à exploiter. Nous étudions les propriétés asymptotiques d'une collection d'estimateurs après avoir appliqué la méthode du Raking-Ratio. Nous montrons une baisse de la variance uniforme sur une classe de fonctions et nous sommes à même d'établir l'expression formelle des variances de ces estimateurs. Nous nous focalisons sur un cas particulier de cette méthode qui consiste à travailler en cycle sur l'information donnée par la connaissance de deux partitions uniquement. Nous pouvons dans ce cas précis donner une expression formelle de la limite quand la méthode du Raking-Ratio se stabilise. Nous finissons ce chapitre par étudier une particularité de cette méthode, celle où l'information auxiliaire ne serait pas donnée par les probabilités exactes des ensembles des partitions mais par une estimation de ces probabilités. Cette information serait donnée par une source extérieure qui se base sur une estimation à partir d'un échantillon vraisemblablement plus grand que la taille d'échantillon sur lequel nous travaillons. Ce chapitre fait l'objet des articles [2] et [1].

Quatrième chapitre

Dans ce quatrième chapitre, nous donnons une définition de ce qu'est l'information auxiliaire générale. Il y a très peu de littérature et donc peu de résultats autour de cette problématique. Nous allons voir dans ce chapitre que le problème suivant se pose. Donner une définition trop générale de l'information auxiliaire empêche d'établir des résultats puissants autre que des résultats de convergence en loi. Mais donner une définition trop restrictive de l'information auxiliaire ne recouvre pas les cas intéressants sur lesquels on peut tomber dans la pratique et que l'on souhaite pouvoir inclure dans cette thèse. Il faut donc pouvoir définir l'information auxiliaire avec un juste milieu de manière à pouvoir appliquer les outils théoriques que nous avons à disposition.

Nous rappelons dans un premier temps la définition d'un estimateur avec information auxiliaire générale donnée par Tarima et Pavlov ainsi que leurs résultats. Il s'agit du résultat le plus général à notre connaissance qui a été retrouvé dans la littérature statistique concernant l'information auxiliaire. Au prix d'une notation lourde, ils modifient n'importe quelle statistique afin de prendre en compte n'importe quel type d'information auxiliaire, qu'elle soit exacte ou estimée par une source auxiliaire d'information. Sous de nombreuses hypothèses de comportement asymptotiques gaussiens, Tarima et Pavlov ont montré que leur estimateur avec information auxiliaire convergeait en loi vers une variable aléatoire gaussienne dont la variance était inférieure à la statistique initiale.

Nous définissons et étudions deux cas particuliers de ce qui aura été présenté : le processus empirique avec information auxiliaire donnée par des espérances de fonction et le processus empirique avec information auxiliaire donnée par des fonctionnelles de la loi mère supposée inconnue. Dans chacun de ces cas, nous donnons des exemples de statistique avec information auxiliaire rentrant dans le cadre d'étude que nous nous imposons.

Cinquième chapitre

Ce cinquième chapitre n'est pas la suite directe du quatrième chapitre. Il regroupe les résultats d'approximation forte d'un travail effectué avant la thèse. Nous rappelons la littérature et donnons la définition de la mesure et du processus empirique bootstrappés indexés par une classe de fonctions mesurables à valeurs réelles. Il s'agit des objets statistiques concernant la fameuse méthode statistique de ré-échantillonnage introduite par Efron. Si la classe de fonctions vérifie certaines conditions d'entropie, nous établissons l'approximation forte du processus empirique bootstrappé vers un processus gaussien centré ayant la même fonction de covariance que le P -pont brownien. Il s'agit d'une extension des travaux d'Alvarez-Andrade et Bouzebda puisque ces derniers ont étudié l'approximation forte du processus empirique bootstrappé indexé par la classe des fonctions indicatrices $\mathcal{F} = \{\mathbb{1}_{]-\infty, t]} : t \in \mathbf{R}\}$.

Le résultat d'approximation forte permet d'établir des résultats comme la justification de l'utilisation de l'estimateur de la variance d'un estimateur par la méthode du bootstrap, ou encore la justification de l'estimateur de la fonction de répartition d'un estimateur par bootstrap. Le résultat d'approximation forte permet également d'adapter la méthode du bootstrap pour l'utiliser au niveau du processus empirique associé à la méthode du Raking-Ratio et permet de justifier cette utilisation. Un travail futur pourrait être d'adapter la méthode du bootstrap au niveau du processus empirique avec information auxiliaire générale. Ce chapitre fait l'objet de l'article [3].

Table des matières

1	Préliminaires techniques	1
1.1	Convergence en loi	1
1.1.1	Théorème de Kolmogorov	1
1.1.2	Définition	2
1.2	Processus empirique	4
1.2.1	Notations et hypothèses	4
1.2.2	Définition du processus empirique	7
1.2.3	Transformation et approximation des quantiles	9
1.2.4	Inégalités de concentration	10
1.3	Approximation forte	12
1.3.1	Approximation forte : de Skorokhod à KMT	12
1.3.2	Approximation forte de Berthet-Mason	15
2	Information auxiliaire d'une partition	17
2.1	Introduction	18
2.1.1	Notations et définitions	18
2.1.2	Motivation	19
2.2	Résultats asymptotiques	21
2.2.1	Approximation forte	21
2.2.2	Convergence en loi	24
2.2.3	Variance et covariance asymptotiques	24
2.2.4	Loi du logarithme itéré	28
2.3	Résultats non-asymptotiques	28
2.3.1	Expression du biais	28
2.3.2	Divergence de Kullback-Leibler	29
2.4	Application : test de Kolmogorov-Smirnov	30
3	Méthode du Raking-Ratio	36
3.1	Introduction	36
3.1.1	Description de la méthode	36
3.1.2	Cadre de travail	38
3.1.3	Motivation	40

3.2	Résultats non-asymptotiques	41
3.2.1	Mesure empirique rakée	41
3.2.2	Processus empirique raké	44
3.3	Résultats asymptotiques	46
3.3.1	Processus gaussien raké	46
3.3.2	Approximation forte et applications	51
3.4	Résultats non-asymptotiques avec deux partitions	61
3.5	Résultat asymptotiques avec deux partitions	63
3.5.1	Expression du processus gaussien raké	63
3.5.2	Limite du processus gaussien raké	64
3.6	Exemples numériques	69
3.6.1	Calcul d'une moyenne rakée	69
3.6.2	Calcul de $S_{1,\text{pair}}(f)$ et $S_{2,\text{impair}}(f)$	70
3.6.3	Simulation du processus empirique raké	71
3.6.4	Fonctions de répartitions rakées.	72
3.7	Apprentissage de l'information auxiliaire	73
3.7.1	Introduction	73
3.7.2	Résultats principaux	74
3.7.3	Applications statistiques	79
3.7.4	Calcul de $\sigma_f^{(\infty)}$ dans un cas simple	83
4	Information auxiliaire générale	86
4.1	Introduction	86
4.2	Estimateur avec information auxiliaire générale	87
4.2.1	Littérature	87
4.2.2	Estimateur de Tarima et Pavlov	87
4.3	Information auxiliaire donnée par des espérances	89
4.3.1	Introduction	89
4.3.2	Résultats	91
4.3.3	Exemples	94
4.3.4	Généralisation des chapitres précédents	97
4.4	Information auxiliaire générale	99
4.4.1	Introduction	99
4.4.2	Résultats	102
4.4.3	Application : covariance connue	103
4.4.4	Application : espérance conditionnelle connue	106
4.4.5	Application : coefficient de corrélation connu	107
4.4.6	Régression linéaire et information auxiliaire	108
5	Processus empirique bootstrappé	114
5.1	Introduction	114
5.1.1	The classical bootstrap	114
5.1.2	The weighted bootstrap	117
5.1.3	More weighted bootstraps	118
5.2	Main results	120
5.2.1	The class \mathcal{F}	120
5.2.2	Strong approximation for one bootstrap	121
5.2.3	Strong approximation for several bootstraps	122
5.2.4	Bootstrap estimation of variance and distribution function	124
5.2.5	Rates of weak convergence	127

5.3	Raking-ratio results	127
5.3.1	Strong approximation of $\alpha_n^{*(N)}$	127
5.3.2	Strong approximation of $\alpha_n^{(N)*}$	129
5.4	Proofs	130
5.4.1	Decomposition of α_n^*	130
5.4.2	Proof of Propositions 5.2.1, 5.3.1 and 5.3.2	131
5.4.3	Construction of the limit Gaussian processes	134
5.4.4	Proof of Theorem 5.2.2 and 5.2.5	136
5.4.5	Proof of Theorem 5.3.3 and 5.3.4	141
5.4.6	Proof of Propositions 5.2.8, 5.2.9 and 5.2.10	144
5.4.7	Proof of Proposition 5.3.6	149

Liste des symboles

$\alpha_n(\mathcal{F})$ Processus empirique indexé par \mathcal{F}

$\alpha_n^*(\mathcal{F})$ Processus empirique bootstrappé

$\alpha_n^{\mathbf{g}}(\mathcal{F})$ Processus empirique avec l'information auxiliaire

$$P[\mathbf{g}] = (P(g_1), \dots, P(g_m)), \quad \mathbf{g} = (g_1, \dots, g_m)$$

$\alpha_n^{(N)}(\mathcal{F})$ Processus empirique raké

$\alpha_n^{*(N)}(\mathcal{F})$ Processus empirique bootstrappé puis raké

$\mathbb{G}(\mathcal{F}), \mathbb{G}^*(\mathcal{F})$ P -pont brownien indexé par \mathcal{F}

$\mathbb{G}^{(N)}(\mathcal{F}), \mathbb{G}^{*(N)}(\mathcal{F})$ P -pont brownien raké indexé par \mathcal{F}

$\mathbb{P}_n(\mathcal{F})$ Mesure empirique indexée par \mathcal{F}

$\mathbb{P}_n^{(N)}(\mathcal{F})$ Mesure empirique rakée

$\mathbb{P}_n^{*(N)}(\mathcal{F})$ Mesure empirique bootstrappée puis rakée

$\mathbb{P}_n^*(\mathcal{F})$ Mesure empirique bootstrappée

$\mathbf{a}_n^{\mathbf{g}}(\mathcal{F})$ Processus empirique avec l'information auxiliaire

$$\mathbf{g}[P] = (g_1(P), \dots, g_m(P)), \quad \mathbf{g} = (g_1, \dots, g_m)$$

$\mathbb{1}_A$ Fonction indicatrice de A

\mathfrak{S}_n Le groupe des permutations de degré n

$\sigma_{\mathcal{F}}^2$ Supremum des variances : $\sup_{f \in \mathcal{F}} \text{Var}(f(X))$

$LL(n)$ Logarithme itéré $L(L(n)), L(n) = \max(1, \log(n))$

$N(\mathcal{F}, \varepsilon, d)$ Nombre minimal de recouvrements de \mathcal{F} par des ε -boules pour la distance d

$N_{[\]}(\mathcal{F}, \varepsilon, d)$ Nombre minimal de recouvrements de \mathcal{F} par des ε -brackets pour la distance d

CHAPITRE 1

Préliminaires techniques

Comme énoncé dans l'introduction, l'objectif principal de cette thèse est d'étudier le comportement asymptotique et non-asymptotique du processus empirique avec informations auxiliaires. Il est nécessaire pour cela de rappeler dans un premier temps les résultats classiques du processus empirique. Ce chapitre regroupe les définitions et résultats qui seront utilisés dans la suite de ce manuscrit.

On rappelle à la section 1.1 les définitions de la convergence en loi d'une suite de variable aléatoire réelle. La section 1.2 rappelle la définition, les propriétés et résultats tels que des inégalités de concentration qui entourent le processus empirique, objet fondamental dans ce manuscrit. Elle est suivie de la section 1.3 qui traite de l'approximation forte, l'outil essentiel pour les résultats donnés dans cette thèse et donne la littérature qui touche cette notion.

1.1 Convergence en loi

Cette première section fait un rappel sur les définitions de la convergence en loi et les problèmes de mesurabilité que certaines de ces définitions peuvent entraîner. La sous-section 1.1.1 soulève et règle un premier problème de la définition de convergence en loi, lié à l'espace de probabilité sur lequel sont définies les variables aléatoires. La sous-section 1.1.2 donne deux approches possibles de la définition de convergence en loi. La première est classique mais pose quelques problèmes de mesurabilité. La deuxième définition, plus générale, règle ce problème.

1.1.1 Théorème de Kolmogorov

Motivation. Soient $(X_n)_{n \in \mathbb{N}^*}$ et X des variables aléatoires à valeurs dans un même espace probabilisable $(\mathcal{X}, \mathcal{A})$. Nous définirons plus loin la convergence faible ou convergence en loi de la suite (X_n) vers la variable X . Cette notion implique que les lois des variables X_n et X ne dépendent pas a priori d'un même espace de probabilité où seraient définies simultanément tous les X_n . Il se pourrait en effet que chaque variable X_n soit définie sur un espace de probabilité $(\Omega_n, \mathcal{T}_n, \mathbb{P}_n)$ différent, ce qui rend impossible l'étude même de la convergence en probabilité ou presque sûre, ces deux derniers types de convergence nécessitant de travailler

sur les mêmes $\omega \in \Omega$. Il convient donc de montrer comment répondre à cette problématique avant d'établir une définition de la convergence en loi.

Théorème d'extension de Kolmogorov. Le résultat suivant établi par Kolmogorov montre qu'il est possible de travailler sur un espace commun sans changer les lois des variables étudiées. Pour une esquisse de preuve, voir l'appendice 1 de Lamperti [55].

Théorème 1.1.1. *On note \mathfrak{S}_n le groupe symétrique d'ordre n . Soit T un intervalle et pour tout n -uplet $d_1, \dots, d_n \in T$, on se fixe une mesure de probabilité μ_{d_1, \dots, d_n} sur \mathbb{R}^n . On suppose que cette dernière vérifie pour tout $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ et $\tau \in \mathfrak{S}_n$,*

- $\mu_{d_1, \dots, d_n}(A_1 \times \dots \times A_n) = \mu_{d_{\tau(1)}, \dots, d_{\tau(n)}}(A_{\tau(1)} \times \dots \times A_{\tau(n)})$;
- $\mu_{d_1, \dots, d_n}(A_1 \times \dots \times A_{n-1} \times \mathbb{R}) = \mu_{d_1, \dots, d_{n-1}}(A_1 \times \dots \times A_{n-1})$.

Alors il existe un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ et un processus stochastique $X: T \times \Omega \rightarrow \mathbb{R}^n$ tels que pour tout $d_1, \dots, d_n \in T$ et $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(X_{d_1} \in A_1, \dots, X_{d_n} \in A_n) = \mu_{d_1, \dots, d_n}(A_1 \times \dots \times A_n).$$

Les deux conditions requises pour ce théorème sont satisfaites par les lois finies-dimensionnelles d'un processus stochastique $(X_t)_{t \in T}$ où T est un espace d'indice. La question de l'existence d'un processus donné est parfois délicate. Cependant les processus stochastiques qui sont étudiés dans cette thèse, notamment les processus gaussiens, sont bien représentables simultanément sur le même espace de probabilité grâce au théorème 1.1.1. Par conséquent il sera licite de parler de la convergence en loi sur un même espace de probabilité sans pour autant expliciter celui-ci.

1.1.2 Définition

Motivation. Nous allons à plusieurs reprises parler de la convergence en loi d'une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ vers une variable X . Nous allons donner dans un premier temps sa définition et dans un second temps une équivalence qui permettra de comprendre le lien théorique entre la convergence faible et l'approximation forte. Cette première définition ne sera pas suffisamment exigeante et peut poser des problèmes au niveau de la mesurabilité. On résout ce problème par une seconde approche qui est présentée.

Première définition. On dit qu'une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ définies sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ à valeurs dans un espace mesurable $(\mathcal{X}, \mathcal{A})$ converge en loi vers une variable aléatoire X si toute fonction continue bornée $\phi: \mathcal{X} \rightarrow \mathbb{R}$ vérifie

$$\lim_{n \rightarrow +\infty} \mathbb{E}[\phi(X_n)] = \mathbb{E}[\phi(X)].$$

Grâce au théorème d'extension de Kolmogorov (théorème 1.1.1), il est inutile d'expliciter dans quel espace vivent les variables X_n et X . On sait que l'on peut construire un espace commun sur lequel sont définies des copies de ces variables. Cette définition possède des équivalences comme le théorème porte-manteau ou le théorème de convergence de Lévy. L'équivalence qui nous intéresse met en relation la convergence en loi et la convergence presque sûre. On donne son énoncé dans le cas de variables aléatoires réelles.

Théorème 1.1.2 (Théorème de représentation de Skorokhod). *Soient $(X_n)_{n \in \mathbb{N}^*}$ des variables aléatoires réelles. Cette suite converge en loi vers une variable aléatoire réelle X si et seulement si il existe un espace de probabilité $(\Omega', \mathcal{T}', \mathbb{P}')$ et des variables aléatoires $(Y_n)_{n \in \mathbb{N}^*}, Y$ définies sur cet espace telles que $X \sim Y$, pour tout $n \in \mathbb{N}^*$, $X_n \sim Y_n$ et Y_n converge $(\Omega', \mathcal{T}', \mathbb{P}')$ -presque sûrement vers Y .*

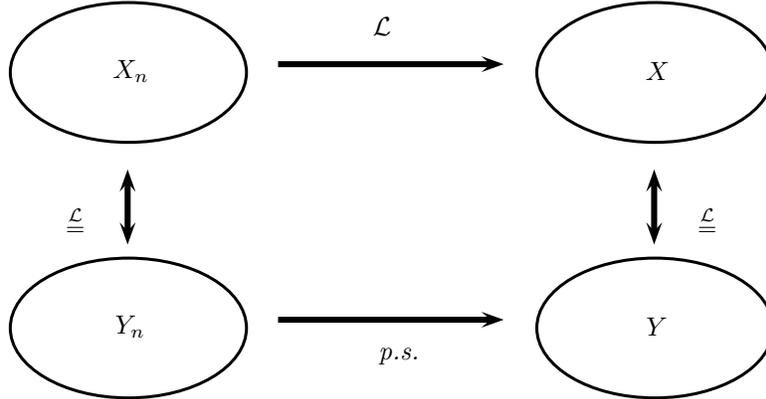


FIG. 1.1: Schéma du théorème de représentation de Skorokhod

La figure 1.1 est un schéma explicatif de ce résultat. Ce théorème sera l'outil qui nous permettra d'établir de manière efficace les convergences en loi. En effet, il suffira alors pour cela de construire un espace de probabilité convenable où on peut démontrer qu'il y a convergence presque sûre avec une vitesse de convergence explicite. Cet outil théorique est utile pour les études de convergence puisqu'il suffit de remplacer les variables X_n par des copies Y_n qui sont proches de leur limite. L'approximation forte qui sera présentée dans la partie suivante est un outil qui se fonde sur ce principe.

Problèmes de mesurabilité liés à la topologie. Cette façon d'introduire la convergence en loi n'est cependant pas suffisante. En effet, si les variables aléatoires X_n sont définies sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ à valeurs dans un espace mesurable $(\mathcal{X}, \mathcal{A})$, celles-ci doivent être des fonctions mesurables, autrement dit

$$\forall n \in \mathbb{N}, \quad \forall A \in \mathcal{A}, \quad X_n^{-1}(A) \in \mathcal{T}.$$

Cette hypothèse n'est pas vérifiée si on considère la variable aléatoire à valeurs dans un espace trop grand, par exemple sur certains espaces non séparables. Par exemple, le théorème de Donsker énonce que le processus empirique (dont la définition est redonnée après) converge en loi vers un pont brownien sur l'espace de Skorokhod muni de sa topologie usuelle. Mais cette convergence dépend de la topologie choisie. Par exemple, la convergence du processus empirique ne peut pas s'étudier avec cette définition sur l'espace $\mathcal{X} = \ell^\infty([0, 1])$ des fonctions réelles bornées définies sur $[0, 1]$ muni de la distance uniforme $\|\cdot\|_\infty$ car ce processus n'est pas mesurable par rapport à la σ -tribu engendrée par la distance uniforme. Pour montrer cela posons $(\Omega, \mathcal{T}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \lambda)$, où λ est la mesure de Lebesgue, $\mathcal{X} = \ell^\infty([0, 1])$ et $\mathcal{A} = \mathcal{B}([0, 1])$ la σ -tribu sur $[0, 1]$. On définit la variable aléatoire $X : \Omega \rightarrow \mathcal{X}$ définie par

$$\forall \omega \in \Omega, \quad \forall U \in [0, 1], \quad X(\omega)(U) = \mathbf{1}_{\{U \leq \omega\}}.$$

En notant H une partie non borélienne de $[0, 1]$, $B_{\|\cdot\|_\infty}(f, r) = \{h \in H : \|f - h\|_\infty < r\}$ la boule ouverte dans H centrée en $f \in H$ de rayon $r > 0$ et A un ouvert de l'espace $(\ell^\infty([0, 1]), \|\cdot\|_\infty)$ défini par

$$A = \bigcup_{t \in H} B_{\|\cdot\|_\infty}(\mathbf{1}_{\{\cdot \leq t\}}, 1/2),$$

on a que

$$H = \{w \in [0, 1] : X(\omega) \in A\} = X^{-1}(A),$$

donc X n'est pas mesurable. Dans ce cas, la convergence en loi introduite précédemment est dénuée de sens. La solution proposée dans Pollard [66] consiste à distinguer la topologie (pour la continuité) et la mesurabilité (tribu non borélienne), notamment en utilisant la tribu engendrée par les boules ouvertes. Une alternative usuelle consiste à changer la topologie uniforme en celle dite de Skorohod sur l'espace du même nom $D(\mathbb{R}, \|\cdot\|_\infty)$ (voir [14]).

Seconde définition. Pour résoudre ce problème de mesurabilité, il faut trouver une définition de la convergence faible qui gère ces éventuels problèmes de mesurabilité. Une manière de faire consiste à utiliser les intégrales extérieures (« outer integrals ») introduites par Hoffmann et Jørgensen [47] (voir également section 1.2 de [80]).

Définition 1.1.1 (Intégrale extérieure). *Si $(\Omega, \mathcal{T}, \mathbb{P})$ est un espace de probabilité et $X : \Omega \rightarrow \overline{\mathbb{R}}$ une variable aléatoire alors on définit l'intégrale extérieure de X comme*

$$\mathbb{E}^*[X] = \inf \{ \mathbb{E}[Y] : Y : \Omega \rightarrow \overline{\mathbb{R}} \text{ mesurable tel que } \mathbb{E}[Y] < +\infty \text{ et } Y \geq X \}.$$

Grâce à cette nouvelle définition, on introduit une nouvelle définition de la convergence en loi qui résout ce problème de mesurabilité, appelée convergence au sens de Hoffman-Jørgensen.

Définition 1.1.2. *On dit qu'une suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ définies sur $(\Omega, \mathcal{T}, \mathbb{P})$ converge en loi vers une variable X si pour toute fonction continue bornée $\phi : \mathbb{R} \rightarrow \mathbb{R}$,*

$$\lim_{n \rightarrow +\infty} \mathbb{E}^*[\phi(X_n)] = \mathbb{E}^*[\phi(X)].$$

Dans la suite, nous définirons et utiliserons une hypothèse qui nous permet d'éviter le recours aux intégrales extérieures.

1.2 Processus empirique

Le processus empirique est un objet fondamental en statistique théorique puisqu'il intervient dans la plupart des cas. Pour être général, nous travaillerons sur le processus empirique indexé par une classe de fonctions mesurables et à valeurs réelles. On fixe les notations et hypothèses des classes de fonctions dans la sous-section 1.2.1 et on donne la définition du processus empirique indexé par des classes de fonctions dans la sous-section 1.2.2. Le lien entre le processus empirique et le processus des quantiles est donné dans la sous-section 1.2.3. Certaines inégalités de concentration et d'autres résultats théoriques de la littérature concernant le processus empirique indexé par des classes de fonctions et les processus gaussiens sont rappelés dans la sous-section 1.2.4

1.2.1 Notations et hypothèses

Cadre d'étude. On travaillera au long de cette thèse avec des variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de même loi qu'une variable X que l'on note $P = \mathbb{P}^X$ toutes définies sur un même espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ à valeurs dans un espace mesurable $(\mathcal{X}, \mathcal{A})$. On note \mathcal{M} l'ensemble des fonctions mesurables de $(\mathcal{X}, \mathcal{A})$ à valeurs réelles que l'on dote de la semi-métrique d_P définie par $d_P(f, g) = (\int_{\mathcal{X}} (f - g)^2 dP)^{1/2}$. Pour $f, g \in \mathcal{M}$ on note par la suite $\|f\|_{Q,r} = (\int_{\mathcal{X}} f^r dQ)^{1/r}$ et $d_{Q,r}(f, g) = \|f - g\|_{Q,r}$. Pour une classe de fonctions $\mathcal{F} \subset \mathcal{M}$, nous notons

$$P(f) = \mathbb{E}[f(X)] = \int_{\mathcal{X}} f(x) dP,$$

$$\sigma_{\mathcal{F}}^2 = \sup_{f \in \mathcal{F}} \text{Var}(f(X)) = \sup_{f \in \mathcal{F}} [P(f^2) - P(f)^2].$$

Hypothèses de mesurabilité. L'objectif de cette thèse est d'établir des résultats les plus généraux possibles pour certains processus stochastiques indexés par des classes de fonctions \mathcal{F} à valeurs réelles et mesurables vérifiant des hypothèses simples et générales qui nous assurerons la bonne définition de convergence en loi. Les résultats de cette thèse feront intervenir régulièrement les deux hypothèses exposées ci-dessous.

Hypothèse (\mathcal{F} .i). $\exists M_{\mathcal{F}} > 0, \forall f \in \mathcal{F}, \|f\|_{\mathcal{X}} = \sup_{x \in \mathcal{X}} |f(x)| \leq M_{\mathcal{F}}$.

Hypothèse (\mathcal{F} .ii). \mathcal{F} est ponctuellement mesurable, autrement dit il existe un sous-ensemble dénombrable $\tilde{\mathcal{F}} \subset \mathcal{F}$ tel que toute fonction $f \in \mathcal{F}$ peut s'écrire comme la limite simple d'une suite d'éléments $(\tilde{f}_n)_{n \in \mathbb{N}}$ de $\tilde{\mathcal{F}}$, i.e. $\forall x \in \mathcal{X}, \lim_{n \rightarrow +\infty} \tilde{f}_n(x) = f(x)$.

L'hypothèse \mathcal{F} .i est une condition suffisante pour l'existence d'une enveloppe de la classe de fonction, i.e. d'une fonction intégrable $F : \mathcal{X} \rightarrow \mathbb{R}$ telle que

$$\forall f \in \mathcal{F}, \quad \forall x \in \mathcal{X}, \quad |f(x)| \leq F(x).$$

C'est une condition forte mais commode pour s'affranchir des problèmes techniques pour gérer les queues de distribution qui alourdiraient des preuves déjà longues. Une classe de fonctions \mathcal{F} vérifiant la condition \mathcal{F} .i admet donc pour enveloppe $F(x) = M_{\mathcal{F}}$. La seconde condition \mathcal{F} .ii est une condition suffisante pour s'assurer que le classe de fonctions \mathcal{F} soit P -mesurable (voir définition 2.3.3 de [80]). Cela permet d'assouplir nos hypothèses de mesurabilité puisqu'il suffit de reporter la condition de P -mesurabilité de \mathcal{F} sur la sous-classe dénombrable $\tilde{\mathcal{F}}$ de \mathcal{F} . Sans hypothèse de mesurabilité de ce type, les résultats du type loi des grands nombres ou théorème central limite peuvent échouer. Cette condition est, par exemple, évoquée dans l'exemple 2.3.4 de [80].

Hypothèse d'entropie. Les classes de fonctions sur lesquelles nous établissons des résultats doivent aussi vérifier une condition d'entropie, une façon de s'assurer que nous ne travaillons pas sur des classes de fonctions « trop grosses ». Nous devons faire des hypothèses qui font intervenir la notion de recouvrement. Cette notion permet de « mesurer » une classe infinie de fonction mesurables à valeurs réelles. Deux approches sont possibles : les classes à entropie uniforme et à entropie avec crochets.

Hypothèse sur les classes VC. Pour une classe de fonctions $\mathcal{F} \subset \mathcal{M}$, $\varepsilon > 0$ et d une distance sur \mathcal{M} , on note $N(\mathcal{F}, \varepsilon, d)$ le nombre minimal de boules de rayon ε par rapport à la distance d pour recouvrir \mathcal{F} . On note $H(\mathcal{F}, \varepsilon, dQ) = \log N(\mathcal{F}, \varepsilon, dQ)$ l'entropie de \mathcal{F} , c'est-à-dire le logarithme du nombre de recouvrements de \mathcal{F} . Une classe \mathcal{F} vérifie l'hypothèse VC si elle vérifie l'hypothèse suivante.

Hypothèse (VC). $\exists c_0 > 0, \nu_0 > 0$ et une enveloppe de fonction F de la classe \mathcal{F} tels que

$$\forall 0 < \varepsilon < 1, \quad \sup_Q N(\mathcal{F}, \varepsilon \|F\|_{Q,2}, d_Q) \leq c_0 \varepsilon^{-\nu_0},$$

où le supremum est pris parmi toutes les mesures de probabilités Q de $(\mathcal{X}, \mathcal{A})$ pour lesquelles $0 < Q(F^2) < +\infty$.

Il ne s'agit pas de la définition d'une VC-classe (au sens de Vapnik–Chervonenkis) mais d'une propriété de ces classes : les VC-classes sont des classes polynomiales, c'est-à-dire que le nombre de recouvrements d'une VC-classe est polynomial en son rayon ε . En plus d'être polynomial le recouvrement des VC-classes est uniforme à toute mesure de probabilité. Plus

précisément, il existe une constante universelle K telle que pour toute VC-classe \mathcal{F} et toute mesure de probabilité Q vérifiant $\|F\|_{Q,r} > 0$,

$$\forall 0 < \varepsilon < 1, \forall r \geq 1, N(\mathcal{C}, \varepsilon \|F\|_{Q,r}, d_{Q,r}) \leq K \cdot \text{VCD}(\mathcal{F}) (16e)^{\text{VCD}(\mathcal{F})} \left(\frac{1}{\varepsilon}\right)^{r(\text{VCD}(\mathcal{F})-1)},$$

où $\text{VCD}(\mathcal{F})$ est la dimension VC de l'ensemble des sous-graphes de \mathcal{F} (voir théorème 2.6.7 de [80]). Ainsi toute classe VC de dimension supérieure ou égale à 2 vérifie l'hypothèse VC avec $\nu_0 = 2(\text{VCD}(\mathcal{F})-1)$ et $c_0 = K \cdot \text{VCD}(\mathcal{F}) (16e)^{\text{VCD}(\mathcal{F})}$. Une classe vérifiant l'hypothèse VC n'est pas forcément une VC-classe. L'exemple suivant donne un exemple de classe vérifiant cette hypothèse.

Exemple 1.2.1. *D'après le Lemme 2.6.16 de [80], la classe de fonctions des translatés $\mathcal{F}_x = \{\psi(x-h) : h \in \mathbb{R}\}$ pour $x \in \mathbb{R}$ et $\psi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction monotone, est une classe VC de dimension 2. Elle vérifie en particulier l'hypothèse VC avec $\nu_0 = 2$ et $c_0 = 2(16e)^2 K \leq 3784K$.*

D'après le Lemme 2.6.17 de [80], la propriété d'être une VC-classe est stable pour de nombreuses opérations telles que la complémentarité, l'union, l'intersection, réciproque d'une fonction, produit cartésien, etc.

Hypothèse sur le recouvrement avec crochets. Bien que la définition de classe VC reste générale et concerne de nombreuses classes de fonctions intéressantes, de nombreux résultats ont été établis pour des classes de fonctions vérifiant l'hypothèse qui va suivre, notamment pour des classes de fonctions très régulières. Pour $\varepsilon > 0$ et d une distance sur \mathcal{M} on note $N_{[\]}(\mathcal{F}, \varepsilon, d)$ le nombre minimal de ε -crochets, c'est-à-dire d'éléments $[g^-, g^+] = \{f \in \mathcal{F} : g_- \leq f \leq g_+\}$ vérifiant $d(g_-, g_+) < \varepsilon$, nécessaires pour recouvrir \mathcal{F} . Les éléments g_- et g_+ n'appartiennent pas forcément à la classe de fonctions \mathcal{F} . On note $H_{[\]}(\mathcal{F}, \varepsilon, dQ) = \log N_{[\]}(\mathcal{F}, \varepsilon, dQ)$ l'entropie avec crochets de \mathcal{F} . Nous dirons qu'une classe \mathcal{F} vérifie l'hypothèse BR si elle vérifie l'hypothèse suivante :

Hypothèse (BR). $\exists b_0 > 0, 0 < r_0 < 1, \forall 0 < \varepsilon < 1, H_{[\]}(\mathcal{F}, \varepsilon, dP) \leq b_0^2 \varepsilon^{-2r_0}$.

On donne quelques exemples de classes de fonctions vérifiant cette dernière hypothèse.

Exemple 1.2.2. *D'après le Théorème 2.7.5 de [80], la classe \mathcal{F} des fonctions monotones $f : \mathbb{R} \rightarrow [0, 1]$ vérifie*

$$\sup_Q H_{[\]}(\mathcal{F}, \varepsilon, d_{Q,r}) \leq \frac{K}{\varepsilon},$$

où le supremum est pris sur l'ensemble des mesures de probabilités Q de $(\mathcal{X}, \mathcal{A})$, $r \geq 1$ et pour une certaine constante $K = K(r) > 0$. Elle satisfait en particulier BR avec $r_0 = 1/2$ et $b_0 > 0$.

Exemple 1.2.3. *Soit $d \geq 2$ et \mathcal{C}_d l'ensemble des sous-ensembles compacts et convexes d'un sous-ensemble borné \mathcal{B}_d de \mathbb{R}^d . D'après le corollaire 2.7.9 de [80],*

$$H_{[\]}(\mathcal{F}_d, d_{Q,r}) \leq \frac{K}{\varepsilon^{(d-1)r/2}},$$

où $\mathcal{F}_d = \{\mathbb{1}_C : C \in \mathcal{C}_d\}$, Q est une mesure de probabilité de Lebesgue et absolument continue et $K = K(\mathcal{B}_d, Q, d) > 0$ est une constante. En particulier, \mathcal{F}_2 vérifie BR avec $r_0 = 1/2$ et $b_0 = b_0(\mathcal{B}_2)$.

1.2.2 Définition du processus empirique

Processus empirique classique. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables i.i.d. à valeurs réelles définies sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ et de même loi $P = \mathbb{P}^X$. On note F la fonction de répartition de X et F_n sa fonction de répartition empirique, *i.e.*

$$\forall t \in \mathbb{R}, \quad F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq t\}},$$

qui converge presque sûrement et uniformément vers F d'après le théorème de Glivenko-Cantelli. Définissons le processus empirique réel α_n^X par

$$\forall t \in \mathbb{R}, \quad \alpha_n^X(t) = \sqrt{n}(F_n(t) - F(t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbb{1}_{\{X_i \leq t\}} - F(t)).$$

On devrait le noter $\alpha_n^{X_1, \dots, X_n}$ mais s'il n'y a pas d'ambiguïté ce processus sera simplement noté α_n . Une des convergences les plus importantes concernant α_n^X est la convergence faible, dite de Donsker. Le théorème de Donsker énonce que sur l'espace $D(\mathbb{R}, \|\cdot\|_\infty)$ ce processus converge faiblement vers un pont brownien, *i.e.* un processus $(P_t)_{t \in [0,1]}$ gaussien centré dont la fonction de covariance est donnée par

$$\forall 0 \leq s < t \leq 1, \quad \text{Cov}(P_t, P_s) = F(s)(1 - F(t)) = F(s \wedge t) - F(t)F(s).$$

Processus empirique indexé par des classes de fonctions. Généralisons les notions et les résultats de convergence en loi précédents à un processus empirique qui est indexé par des classes de fonctions mesurables, à valeurs réelles et vérifiant $\mathcal{F}.i$ et $\mathcal{F}.ii$. Soient $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables i.i.d. définies sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ et à valeur dans un espace mesurable $(\mathcal{X}, \mathcal{A})$. Nous donnons les définitions de la mesure et du processus empirique indexés par des classes de fonctions ainsi que des exemples.

Définition 1.2.1 (Mesure empirique). *On appelle mesure empirique indexée par une classe de fonctions \mathcal{F} la fonction linéaire $\mathbb{P}_n : \mathcal{F} \rightarrow [0, 1]$ définie par*

$$\forall f \in \mathcal{F}, \quad \mathbb{P}_n(f) = \frac{1}{n} \sum_{i=1}^n f(X_i).$$

Exemple 1.2.4 (Moyenne et variance empirique). *Si X_1, \dots, X_n sont des variables aléatoires réelles alors $\mathbb{P}_n(f_0) = (X_1 + \dots + X_n)/n$ correspond à la moyenne empirique des X_1, \dots, X_n et $\mathbb{P}_n(f_1) - \mathbb{P}_n^2(f_0)$ correspond à la variance empirique avec $f_0 = \text{id}$ et $f_1(x) = x^2$.*

Exemple 1.2.5 (Fonction de répartition). *La fonction de répartition peut s'écrire comme la mesure empirique indexée par la classe des fonctions indicatrices, *i.e.* $\mathcal{F} = \{\mathbb{1}_{]-\infty, t]}\} : t \in \mathbb{R}$.*

Le processus empirique indexé par une classe de fonctions correspond à la mesure empirique normalisée et indexée par une classe de fonctions centrées. On le note de la façon suivante.

Définition 1.2.2 (Processus empirique). *On appelle processus empirique indexé par la classe de fonctions \mathcal{F} vérifiant $\mathcal{F}.i$ et $\mathcal{F}.ii$ le processus $\alpha_n^X(\mathcal{F}) = \{\alpha_n(f) : f \in \mathcal{F}\}$ défini par*

$$\forall n \in \mathbb{N}^*, \quad \forall f \in \mathcal{F}, \quad \alpha_n^X(f) = \sqrt{n}(\mathbb{P}_n(f) - P(f)),$$

avec $\mathbb{P}_n(f)$ la mesure empirique indexée par \mathcal{F} et $P(f) = \mathbb{E}[f(X)]$. S'il n'y a pas d'ambiguïté, on le note $\alpha_n(\mathcal{F})$.

Exemple 1.2.6. Le processus empirique classique $t \mapsto \alpha_n^X(t)$ correspond au processus empirique $\alpha_n(\mathcal{F})$ indexé par la classe des fonctions indicatrices $\mathcal{F} = \{\mathbf{1}_{]-\infty, t]} : t \in \mathbb{R}\}$.

D'après le théorème central limite multidimensionnel, comme la classe de fonctions \mathcal{F} est incluse dans $L_2(P)$ alors le vecteur de variables aléatoires $(\alpha_1^X(f_1), \dots, \alpha_n^X(f_n))$ converge en loi vers $\mathcal{N}(0, \Sigma)$ où Σ est la matrice de covariance définie par

$$\forall i, j \leq n, \quad \Sigma_{i,j} = \text{Cov}(f_i, f_j) = \mathbb{E}[f_i f_j] - \mathbb{E}[f_i] \mathbb{E}[f_j].$$

S'il existe, le processus limite faible de $\alpha_n(\mathcal{F})$ dans $\ell^\infty(\mathcal{F})$ noté $\mathbb{G}(\mathcal{F}) = \{\mathbb{G}(f) : f \in \mathcal{F}\}$, est alors nécessairement centré et sa fonction de covariance est donnée par

$$\forall f, g \in \mathcal{F}, \quad \mathbb{E}[\mathbb{G}(f)\mathbb{G}(g)] = \text{Cov}(\alpha_n(f), \alpha_n(g)) = \mathbb{E}[fg] - \mathbb{E}[f]\mathbb{E}[g], \quad (1.1)$$

donc en particulier $\mathbb{G}(f)$ est une loi normale centrée et de variance $\text{Var}(f(X))$. Sans hypothèse supplémentaire sur \mathcal{F} , la convergence faible peut ne pas avoir lieu ou le processus $\mathbb{G}(\mathcal{F})$ peut ne pas être gaussien. Sous les hypothèses VC ou BR, \mathcal{F} est une classe de Donsker dans le sens où $\alpha_n(\mathcal{F})$ converge faiblement vers le processus gaussien $\mathbb{G}(\mathcal{F})$ suivant. Rappelons qu'un processus gaussien est caractérisé par ses lois finies-dimensionnelles.

Définition 1.2.3 (Processus de P -pont brownien). On appelle le P -pont brownien $\mathbb{G}(\mathcal{F})$ le processus gaussien centré, indexé par une classe de fonctions $\mathcal{F} \subset L_2(\mathcal{X}, \mathcal{A}, P)$ dont la covariance est donnée par (1.1).

Exemple 1.2.7 (Pont brownien classique). Le processus empirique $\alpha_n^U(t) = \sqrt{n}(F_n^U(t) - t)$ de variables aléatoires U_1, \dots, U_n i.i.d. et uniformes sur $[0, 1]$ converge en loi dans l'espace de Skorokhod $D(\mathbb{R}, \|\cdot\|_\infty)$ vers le pont brownien $(P^U(t))_{t \in [0,1]}$, processus gaussien centré dont la fonction de covariance est donnée par $\text{Cov}(P^U(t), P^U(s)) = t \wedge s - st$ où $s \wedge t = \min(s, t)$. Ce processus peut aussi être défini à partir d'un mouvement brownien $(B_t)_{t \geq 0}$ via $(P^U(t))_{t \in [0,1]} = (B_t - tB_1)_{t \in [0,1]}$. Le supremum du pont brownien $(P^U(t))_{t \in [0,1]}$ intervient dans le test d'adéquation de Kolmogorov-Smirnov. Grâce à la représentation du pont brownien par B_t et au principe de réflexion du mouvement brownien, il est possible d'expliciter la loi du supremum d'un pont brownien (voir section 12.3 de [35]). Pour tout $\varepsilon > 0$,

$$\mathbb{P} \left(\sup_{t \in [0,1]} |P^U(t)| \geq \varepsilon \right) = 2 \sum_{k \geq 1} (-1)^{k-1} e^{-2k^2 \varepsilon^2}.$$

Applications statistiques. Les fluctuations de $\mathbb{P}_n(\mathcal{F})$ par rapport à $P(\mathcal{F})$ jouent un rôle primordial en statistique. Ces écarts sont à la base des déviations de toutes les statistiques. En effet, les statistiques réelles $T(X_1, \dots, X_n)$ peuvent généralement s'écrire sous la forme

$$T(X_1, \dots, X_n) = \phi(\mathbb{P}_n),$$

avec ϕ une fonction à valeurs réelles, pas forcément régulière. L'étude du processus empirique α_n^X permet alors de contrôler les déviations de la statistique $T = \mathbb{E}[T(X_1, \dots, X_n)]$ par rapport à sa moyenne du fait que

$$T(X_1, \dots, X_n) = \phi(\mathbb{P}_n) = \phi \left(\mathbb{P} + \frac{\alpha_n^X}{\sqrt{n}} \right).$$

Exemple 1.2.8 (Minimisation d'un risque). Soit \mathcal{L} un ensemble de fonctions de perte. Pour estimer $\text{argmin}_{L \in \mathcal{L}} \mathbb{E}[L(X)]$ parmi les fonctions de pertes appartenant à la classe de fonctions \mathcal{L} , on calcule la statistique

$$T(X_1, \dots, X_n) = \text{argmin}_{L \in \mathcal{L}} \mathbb{P}_n(L) = \text{argmin}_{L \in \mathcal{L}} \frac{1}{n} \sum_{i=1}^n L(X_i).$$

Cela revient à étudier le processus empirique $\alpha_n^X(\mathcal{L})$ puisque

$$\begin{aligned} T(X_1, \dots, X_n) &= \operatorname{argmin}_{L \in \mathcal{L}} \left(\mathbb{E}[L(X)] + \frac{1}{n} \sum_{i=1}^n (L(X_i) - \mathbb{E}[L(X_i)]) \right) \\ &= \operatorname{argmin}_{L \in \mathcal{L}} \left(\mathbb{E}[L(X)] + \frac{1}{\sqrt{n}} \alpha_n^X(L) \right). \end{aligned}$$

1.2.3 Transformation et approximation des quantiles

Grâce à l'inverse généralisée d'une fonction de répartition F (appelée aussi fonction de quantile) définie par

$$\forall u \in]0, 1[, \quad F^{\leftarrow}(u) = \inf\{t \in \mathbb{R} : F(t) \geq u\},$$

on peut ramener l'étude de notre processus empirique à l'étude du processus empirique dans le cas des lois uniformes. C'est l'objet du résultat suivant.

Lemme 1.2.1 (Transformation des quantiles). *Soit une variable uniforme $U \sim \mathcal{U}([0, 1])$ alors $F^{\leftarrow}(U) \stackrel{\mathcal{L}}{=} X$. De plus, $F(X) \stackrel{\mathcal{L}}{=} U$ si et seulement si F est continue.*

Ainsi, pour étudier le processus empirique α_n^X on peut se ramener à l'étude du cas du processus empirique uniforme α_n^U . Le résultat suivant donne la condition suffisante pour que ce soit le cas.

Corollaire 1.2.2 (Transformation du processus empirique). *Supposons que U_1, \dots, U_n sont i.i.d. de loi uniforme $\mathcal{U}([0, 1])$ et que X soit une variable de fonction de répartition F . Alors $\alpha_n^U \stackrel{\mathcal{L}}{=} \alpha_n^X \circ F^{\leftarrow}$. De plus, si F est continue, alors $\alpha_n^X \stackrel{\mathcal{L}}{=} \alpha_n^U \circ F$.*

Exemple 1.2.9 (Pont brownien). *D'après le Corollaire 1.2.2 et l'Exemple 1.2.7, le processus empirique $\alpha_n^X(t) = \sqrt{n}(F_n(t) - F(t))$ de variables aléatoires X_1, \dots, X_n i.i.d. de fonction de répartition continue F converge en loi dans l'espace de Skorokhod vers le pont brownien $(P(t))_{t \in \mathbb{R}}$, processus gaussien centré dont la fonction de covariance est donnée par $\operatorname{Cov}(P(t), P(s)) = F(t \wedge s) - F(s)F(t)$. La relation entre les ponts browniens $(P(t))_{t \in \mathbb{R}}$ et $P^U(t)_{t \in [0, 1]}$ défini dans l'exemple 1.2.7 est $(P(t))_{t \in \mathbb{R}} = (P^U \circ F^X(t))_{t \in [0, 1]}$. On en déduit la loi du supremum du pont brownien suivant. Pour tout $\varepsilon > 0$,*

$$\mathbb{P} \left(\sup_{t \in \mathbb{R}} |P(t)| \geq \varepsilon \right) = 2 \sum_{k \geq 1} (-1)^{k-1} e^{-2k^2 \varepsilon^2}. \quad (1.2)$$

On peut effectuer la même réduction pour le processus des quantiles uniforme β_n^X défini pour tout $n \in \mathbb{N}^*$ et $u \in]0, 1[$ par

$$\beta_n^X(u) = \sqrt{n} (F_n^{-1}(u) - F^{\leftarrow}(u)),$$

où $F_n^{-1}(u)$ correspond à la i -ème plus grande valeur de l'échantillon X_1, \dots, X_n si $u \in](i-1)/n, i/n]$. Ce processus permet lui d'étudier les fluctuations des statistiques d'ordre, de leurs espacements ainsi que de leurs transformations.

Corollaire 1.2.3 (Transformation du processus des quantiles). *Supposons que U_1, \dots, U_n sont i.i.d. de loi uniforme $\mathcal{U}([0, 1])$ et que X soit une variable de fonction de répartition F . Alors $\beta_n^X \stackrel{\mathcal{L}}{=} \beta_n^U \circ F^{\leftarrow}$. De plus, si F est continue, $\beta_n^U \stackrel{\mathcal{L}}{=} \beta_n^X \circ F$.*

On a procédé jusqu'à présent à deux réductions, celle des deux processus α_n^X, β_n^X par α_n^U, β_n^U . Il existe une troisième réduction, celle de β_n^U par α_n^U . Grâce à l'étude du processus de Bahadur-Kiefer, Deheuvels et Mason ont pu obtenir des résultats sur l'écart de α_n^U et $-\beta_n^U$ (voir théorème 1A de [29]).

Théorème 1.2.4 (Théorème de Deheuvels-Mason). *Les processus empiriques et de quantiles empiriques uniformes vérifient*

$$\lim_{n \rightarrow +\infty} \frac{n^{1/4} \|\alpha_n^U + \beta_n^U\|_\infty}{\sqrt{\log n} \sqrt{\|\beta_n^U\|}} = 1 \quad p.s.$$

Comme corollaire de ce résultat, Deheuvels et Mason ont montré que l'écart entre les processus α_n et $-\beta_n$ était p.s. de l'ordre de $O(n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4})$ (voir corollaire 1A de [29]). En conclusion, l'étude du processus empirique uniforme est suffisant pour étudier tous les autres processus.

1.2.4 Inégalités de concentration

Inégalité de Borell-Sudakov. Dans les parties suivantes nous établirons des convergences presque sûres vers un processus gaussien centré indexé par une classe de fonction. Il est donc nécessaire de pouvoir étudier le supremum sur cette classe de fonctions d'un tel processus. Le résultat suivant nous permet d'établir une inégalité de concentration pour ce supremum indexé par un ensemble dénombrable. C'est un résultat puissant dans le sens où celui-ci ne requiert pas d'hypothèses très fortes. Nous travaillons avec un processus $(X_t)_{t \in T}$ séparable, c'est-à-dire avec un processus vérifiant la définition suivante.

Définition 1.2.4 (Processus séparable). *On dit qu'un processus $(X_t)_{t \in T}$ est séparable s'il possède une version séparable \tilde{X} , i.e. qu'il existe un ensemble $\tilde{T} \subset T$ dénombrable tel que pour tout $t \in T$ il existe $(t_k)_{k \in \mathbb{N}} \subset \tilde{T}$ vérifiant*

$$\lim_{k \rightarrow +\infty} \tilde{X}_{t_k} = X_t.$$

Le théorème suivant est un résultat de concentration valable pour les processus gaussiens séparables indexés par un ensemble quelconque.

Théorème 1.2.5. *Soit $(X_t)_{t \in T}$ un processus gaussien centré séparable vérifiant $\sigma_T^2 = \sup_{t \in T} \mathbb{E}[X_t^2] < +\infty$. Si on pose $\|X\|_T = \sup_{t \in T} |X_t|$ alors $\mathbb{E}[\|X\|_T] < +\infty$ et,*

$$\forall \lambda > 0, \quad \mathbb{P}(|\|X\|_T - \mathbb{E}[\|X\|_T]| > \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{2\sigma_T^2}\right),$$

$$\mathbb{P}(\|X\|_T > \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{8\mathbb{E}[\|X\|_T^2]}\right).$$

Ce résultat puissant est dû à Borell (*cf.* Théorème 5.2 de [18]). Différents énoncés sont donnés dans l'annexe A.2.1 de [80]. Ces inégalités ne sont pas vraies dans le cas où l'on travaille avec des processus non séparables.

Inégalité de Talagrand. Concernant le processus empirique, nous avons également quelques résultats de concentration. Comme pour l'inégalité de Borell, ils s'avèrent être des outils théoriques puissants pour étudier les déviations du supremum du processus empirique indexé par des classes de fonctions vérifiant certaines conditions d'entropie. Le premier résultat donne une inégalité de concentration de la déviation du processus empirique indexé par une classe VC (*cf.* théorème 1.3 de [78] ou théorème 2.14.9 de [80]).

Théorème 1.2.6. *Si \mathcal{F} est une classe de fonctions mesurables à valeurs réelles et vérifiant $\mathcal{F}.i$, $\mathcal{F}.ii$ et VC alors il existe $C = C(c_0) > 0$ tel que pour tout $t > 0$,*

$$\mathbb{P}(\|\alpha_n\|_{\mathcal{F}} > t) \leq \left(\frac{Ct}{2\sqrt{\nu_0}M_{\mathcal{F}}} \right)^{\nu_0} e^{-\frac{t^2}{2M_{\mathcal{F}}^2}},$$

où c_0, ν_0 sont donnés par l'hypothèse VC.

Il manque un résultat similaire dans la littérature dans le cas où l'hypothèse BR est vérifiée. Nous combinons les résultats de concentration connus pour établir un résultat similaire au théorème 1.2.6.

Théorème 1.2.7. *Si \mathcal{F} est une classe de fonctions mesurables à valeurs réelles et vérifiant $\mathcal{F}.i$, $\mathcal{F}.ii$ et BR alors il existe $C = C(M_{\mathcal{F}}, b_0, r_0, \sigma_{\mathcal{F}}^2), t_0 = t_0(M_{\mathcal{F}}, b_0, r_0) > 0$ tel que pour tout $t > t_0$,*

$$\mathbb{P}(\|\alpha_n\|_{\mathcal{F}} > t) \leq \exp(-Ct^2),$$

où b_0, r_0 sont données par l'hypothèse BR.

Démonstration. Puisque \mathcal{F} vérifie $\mathcal{F}.i$, il suffit d'étudier l'inégalité pour $t < 2M_{\mathcal{F}}\sqrt{n}$. Il existe d'après le théorème 2.14.25 de [80] ou corollaire 2 de [16], des constantes universelles $D_1, D_2 > 0$ telles que pour tout $t > 0$,

$$\mathbb{P}(\|\alpha_n\|_{\mathcal{F}} > D_1(\mu_n + t)) \leq \exp\left(-D_2 \min\left(\frac{t^2}{\sigma_{\mathcal{F}}^2}, \frac{t\sqrt{n}}{M_{\mathcal{F}}}\right)\right),$$

avec $\mu_n = \mathbb{E}[\|\alpha_n\|_{\mathcal{F}}]$ qu'il nous faut borner. D'après le théorème 2.14.2 de [80] ou corollaire 4.3 de [65],

$$\mu_n \leq \int_0^{M_{\mathcal{F}}} \sqrt{1 + H_{[1]}(\mathcal{F}, \varepsilon, dP)} d\varepsilon.$$

Puisque \mathcal{F} vérifie BR nécessairement $\mu_n \leq C'$ avec $C' = C'(M_{\mathcal{F}}, b_0, r_0) = M_{\mathcal{F}}(1 + b_0/(1 - r_0))$. On pose $t_0 = 2D_1C'$ et $t_1 = 2D_1\sigma_{\mathcal{F}}^2\sqrt{n}/M_{\mathcal{F}}$. Pour $t_0 < t < t_1$,

$$\min\left(\frac{t^2}{(2D_1\sigma_{\mathcal{F}})^2}, \frac{t\sqrt{n}}{2D_1M_{\mathcal{F}}}\right) = \frac{t^2}{(2D_1\sigma_{\mathcal{F}})^2},$$

et de fait,

$$\begin{aligned} \mathbb{P}(\|\alpha_n\|_{\mathcal{F}} > t) &= \mathbb{P}\left(\|\alpha_n\|_{\mathcal{F}} > \frac{t}{2} + \frac{t}{2}\right) \leq \mathbb{P}\left(\|\alpha_n\|_{\mathcal{F}} > D_1\left(\mu_n + \frac{t}{2D_1}\right)\right) \\ &\leq \exp\left(-\frac{D_2t^2}{4D_1^2\sigma_{\mathcal{F}}^2}\right). \end{aligned}$$

Pour $t_1 < t < 2M_{\mathcal{F}}\sqrt{n}$, on a

$$\min\left(\frac{t^2}{(2D_1\sigma_{\mathcal{F}})^2}, \frac{t\sqrt{n}}{2D_1M_{\mathcal{F}}}\right) = \frac{t\sqrt{n}}{2D_1M_{\mathcal{F}}} \geq \frac{t^2}{4D_1M_{\mathcal{F}}^2},$$

et on obtient

$$\mathbb{P}(\|\alpha_n\|_{\mathcal{F}} > t) \leq \exp\left(-\frac{D_2t^2}{4D_1M_{\mathcal{F}}^2}\right).$$

Le résultat est prouvé en posant $C = D_2/4D_1 \max(D_1\sigma_{\mathcal{F}}^2, M_{\mathcal{F}}^2)$. \square

Ce résultat entraîne en particulier que si \mathcal{F} vérifie $\mathcal{F}.i$, $\mathcal{F}.ii$ et BR ou VC alors presque sûrement $\|\alpha_n\|_{\mathcal{F}} = O(\sqrt{\log n})$.

Inégalité de Bousquet. L'inégalité suivante étudie les déviations du supremum du processus empirique par rapport à sa moyenne (cf. théorème 2.3 de [20] ou partie 12.4 de [19]).

Théorème 1.2.8. *Si \mathcal{F} est une classe de fonctions dénombrable vérifiant $\mathcal{F}.i$ alors pour tout $t \geq 0$,*

$$\mathbb{P} \left(\sup_{f \in \mathcal{F}} \alpha_n(f) - \mathbb{E}[\sup_{f \in \mathcal{F}} \alpha_n(f)] \geq \frac{M_{\mathcal{F}} t}{\sqrt{n}} \right) \leq \exp \left(-v h \left(\frac{t}{v} \right) \right),$$

avec $v = n\sigma_{\mathcal{F}}^2 + 2\sqrt{n}\mathbb{E}[\sup_{f \in \mathcal{F}} \alpha_n(f)]$ et $h(x) = (1+x)\log(1+x) - x$. En particulier pour tout $t \geq 0$,

$$\mathbb{P} \left(\sup_{f \in \mathcal{F}} \alpha_n(f) - \mathbb{E}[\sup_{f \in \mathcal{F}} \alpha_n(f)] \geq t \right) \leq \exp \left(-\frac{\sqrt{nt}^2}{2(2M_{\mathcal{F}}\mathbb{E}[\sup_{f \in \mathcal{F}} \alpha_n(f)] + \sqrt{n}\sigma_{\mathcal{F}}^2 + M_{\mathcal{F}}t/3)} \right).$$

Loi du logarithme itéré. Notons $LLn = L(L(n))$ avec $L(n) = \max(1, \log n)$. Nous rappelons le résultat d'Alexander [4] concernant la loi du logarithme itéré pour un processus empirique indexé par une classe de fonctions vérifiant une des conditions d'entropie.

Théorème 1.2.9. *Si \mathcal{F} est une classe de fonctions mesurables qui vérifie $\mathcal{F}.i$ et BR ou VC. Alors,*

$$\limsup_{n \rightarrow +\infty} \frac{\|\alpha_n\|_{\mathcal{F}}}{\sigma_{\mathcal{F}} \sqrt{2LLn}} \leq 1 \quad p.s.$$

Remarque 1.2.1. *Dans le cas réel, si la fonction de répartition F est continue alors l'inégalité est une égalité car d'après Chung et Smirnov (voir [42]),*

$$\limsup_{n \rightarrow +\infty} \frac{\sup_{t \in \mathbb{R}} |\alpha_n^X(t)|}{\sqrt{2LLn}} = \limsup_{n \rightarrow +\infty} \frac{\sup_{t \in [0,1]} |\alpha_n^U(t)|}{\sqrt{2LLn}} = \frac{1}{2} \quad p.s.$$

Le résultat concernant la condition BR résulte du théorème 2.12 et celui concernant la condition VC provient du théorème 2.13 basé sur le théorème 2.8 qui utilise dans sa preuve la conséquence du lemme 2.7 qui correspond en fait à l'hypothèse VC.

1.3 Approximation forte

Nous supposons dans cette partie que $(X_n)_{n \in \mathbb{N}^*}$ est une suite de variables i.i.d. de loi P et nous notons $S_n = \sum_{k=1}^n X_k$ la somme empirique. La sous-section 1.3.1 donne les éléments principaux de la littérature de l'approximation forte et les résultats importants. Une référence pour cet outil théorique est le livre de Csörgő et Révész [28]. Nous rappelons dans la sous-section 1.3.2 le résultat établi par Berthet et Mason, qui donne l'approximation forte du processus empirique indexé par des classes de fonctions sous des conditions d'entropie.

1.3.1 Approximation forte : de Skorokhod à KMT

Résultat de Skorokhod. L'approximation forte débute en 1961 avec le résultat de Skorokhod, connu sous le nom de *théorème de représentation de Skorokhod*. Il établit que si les X_i admettent un moment d'ordre deux, il existe un espace de probabilité, un temps d'arrêt τ défini sur cet espace et un mouvement brownien $(B_t)_{t \geq 0}$ tel que $S_n = \sum_{i=1}^n X_i$ ait la même loi que B_{τ} . Le théorème suivant énonce ce résultat.

Théorème 1.3.1 (Skorokhod, 1961). *Supposons que $(X_n)_{n \in \mathbb{N}^*}$ sont des variables aléatoires i.i.d. centrées réduites. Il existe un espace de probabilité, un mouvement brownien $(B_t)_{t \geq 0}$ et une suite $(\tau_i)_{i \in \mathbb{N}^*}$ de variables i.i.d. positives définies sur cet espace vérifiant*

1. $\forall n \in \mathbb{N}^*, B_{\tau_1 + \dots + \tau_n} \stackrel{\mathcal{L}}{=} S_n$;
2. $\forall n \in \mathbb{N}^*, \tau_1 + \dots + \tau_n$ est un temps d'arrêt ;
3. $\mathbb{E}[\tau_1] = 1$.

De plus si $\forall n \in \mathbb{N}^*, \mathbb{E}[X_1^{2n}] < \infty$ alors $\mathbb{E}[\tau_1^n] < \infty$.

Le théorème de Skorokhod est suffisant pour prouver le théorème de Donsker qui est un principe d'invariance faible. Ce théorème conduit à la notion de principe d'invariance forte qui consiste à construire sur le même espace de probabilité S_n et B_n de sorte que leurs trajectoires respectives soient les plus proches possibles (voir [77]).

Théorème 1.3.2 (Strassen, 1964). *Il existe un espace de probabilité, un mouvement brownien $(B_t)_{t \geq 0}$ et une suite $(X_n)_{n \in \mathbb{N}^*}$ de loi X définis sur cet espace vérifiant presque sûrement $|S_n - B_n| = o(\sqrt{n \log \log n})$.*

La borne proposée par Strassen est la meilleure que l'on puisse obtenir sans hypothèses supplémentaires sur la suite des X_n . Ce résultat provient de la proposition suivante [59] :

Proposition 1.3.3 (Major, 1979). *Pour toute suite de réels $(a_n)_{n \in \mathbb{N}}$ vérifiant $\lim_{n \rightarrow +\infty} a_n = +\infty$, il existe une suite de variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. centrées et réduites telle que pour tout mouvement brownien $(B_t)_{t \geq 0}$ on ait*

$$\limsup_{n \rightarrow +\infty} a_n \frac{|S_n - B_n|}{\sqrt{n \log \log n}} = +\infty \quad p.s.$$

Résultat de KMT. Bien que ce soit un résultat fort, le résultat de Strassen n'est pas suffisant pour démontrer des résultats comme le théorème de Donsker. C'est en 1975 qu'apparaît une meilleure approximation de la somme partielle par un mouvement brownien. Sous une hypothèse plus forte que l'existence du moment d'ordre 2, les mathématiciens Komlós, Major et Tusnády ont montré le résultat qui suit (voir [51, 52]).

Théorème 1.3.4 (KMT, 1975-1976). *Supposons que $(X_n)_{n \in \mathbb{N}^*}$ soient des variables centrées réduites et que la fonction génératrice $M_{X_1}(t) = \mathbb{E}[e^{tX_1}]$ de X_1 existe sur un voisinage de 0. Il existe alors un mouvement brownien $(B_t)_{t \geq 0}$ vérifiant pour tout $x > 0$ et $n \in \mathbb{N}^*$,*

$$\mathbb{P} \left(\max_{1 \leq k \leq n} |S_k - B_k| > C \log n + x \right) < K e^{-\lambda x},$$

où C, K, λ sont des constantes positives dépendant de la distribution de X_1 .

Corollaire 1.3.5. *Sous les hypothèses du théorème 1.3.4, presque sûrement $\max_{1 \leq k \leq n} |S_k - B_k| = O(\log n)$.*

Démonstration. Si l'on note $A_n = \{\max_{1 \leq k \leq n} |S_k - B_k| > C \log n + x\}$, le théorème 1.3.4 implique que $\sum_{n \geq 1} \mathbb{P}(A_n) < +\infty$. Le lemme de Borel-Cantelli appliqué à ce résultat montre que $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 0$, i.e. $\mathbb{P}(\liminf_{n \rightarrow +\infty} \bar{A}_n) = 1$. C'est-à-dire, avec probabilité 1 il existe $N = N(\omega) \in \mathbb{N}$ tel que pour tout $n \geq N$, $\max_{1 \leq k \leq n} |S_k - B_k| = O(\log n)$. \square

La borne proposée par KMT est la meilleure possible excepté dans le cas où X_1 serait lui-même une loi normale. Cette affirmation est confirmée par le résultat suivant (cf. théorème 2.3.2 de [28]) :

Proposition 1.3.6. *S'il existe un mouvement brownien $(B_t)_{t \geq 0}$ vérifiant presque sûrement $|S_n - B_n| = o(\log n)$ alors X_1 est une loi normale centrée réduite.*

Ce résultat est également optimal dans le sens où si l'hypothèse d'existence de M_{X_1} au voisinage de 0 est supprimée, alors la borne donnée par le théorème 1.3.4 n'est plus valide (voir relation 2.3.2 dans [28]) :

Proposition 1.3.7. *Si $M_{X_1}(t)$ n'est définie pour aucun $t > 0$ alors pour tout mouvement brownien $(B_t)_{t \geq 0}$,*

$$\limsup_{n \rightarrow +\infty} \frac{S_n - B_n}{\log n} = +\infty \quad p.s.$$

L'étude du comportement asymptotique de S_n permet celui du processus empirique puisque ce dernier se comporte asymptotiquement comme le processus interpolé

$$S_n(t) = S_{[nt]} + (nt - [nt])X_{[nt]+1},$$

où $[t]$ désigne la partie entière de t . L'approximation forte du processus empirique uniforme est donnée par le théorème de KMT (voir théorème 3 de [51]).

Théorème 1.3.8 (KMT). *Il existe des variables $(U_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi uniforme $U([0, 1])$ et une suite de pont brownien $(P_n)_{n \in \mathbb{N}^*}$ telles que pour tout $x \in \mathbb{R}$,*

$$\mathbb{P} \left(\sup_{t \in [0, 1]} |\alpha_n^U(t) - P_n(t)| > \frac{C \log n + x}{\sqrt{n}} \right) \leq K e^{-\lambda x},$$

avec C, K, λ des constantes positives universelles. En particulier presque sûrement,

$$\sup_{t \in [0, 1]} |\alpha_n^U(t) - P_n(t)| = O \left(\frac{\log n}{\sqrt{n}} \right).$$

Pour des valeurs explicites de K, λ, C , voir le théorème 1 de Bretagnolle et Massart [22]. Ce résultat est optimal car le fait de rajouter une constante n'entraîne plus l'approximation forte par des ponts browniens. C'est l'objet du théorème suivant (voir théorème 4.4.2 de [28]).

Théorème 1.3.9. *Soient des variables $(U_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi uniforme $U([0, 1])$ alors pour toute suite de pont brownien $(P_n(t))_{n \in \mathbb{N}^*}$ définies sur le même espace de probabilité que α_n^U on a que*

$$\mathbb{P} \left(\sup_{t \in [0, 1]} |\alpha_n^U(t) - P_n(t)| \geq \frac{\log n}{6\sqrt{n}} \right) \xrightarrow{n \rightarrow +\infty} 1.$$

Grâce au corollaire 1.2.2, on obtient l'approximation forte du processus empirique dans les cas autres que la loi uniforme. C'est l'objet du corollaire suivant.

Corollaire 1.3.10. *Soit X une variable aléatoire réelle ayant une fonction de répartition F continue. Alors il existe une suite de variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de même loi que X et une suite de ponts browniens $(P_n)_{n \in \mathbb{N}^*}$ telles que pour tout $x \in \mathbb{R}$,*

$$\mathbb{P} \left(\sup_{t \in \mathbb{R}} |\alpha_n^X(t) - P_n \circ F(t)| > \frac{C \log n + x}{\sqrt{n}} \right) \leq K e^{-\lambda x},$$

avec C, K, λ les constantes positives universelles définies par le théorème 1.3.8. En particulier, presque sûrement,

$$\sup_{t \in \mathbb{R}} |\alpha_n^X(t) - P_n \circ F(t)| = O(n^{-1/2} \log n).$$

1.3.2 Approximation forte de Berthet-Mason

Énoncé. En 1983, Dudley et Philipp ont établi une approximation forte pour les sommes de variables i.i.d. définies dans des espaces de Banach[36]. C'est en 2006 que Berthet et Mason proposèrent une généralisation de ce résultat, généralisant l'approximation forte au processus empirique indexé par des classes de fonctions soumises à des hypothèses d'entropie par des processus de P -ponts browniens. Le résultat de Berthet et Mason dans le cas des classes vérifiant la condition VC résulte du théorème 1 de [12]. Il est énoncé comme suit.

Théorème 1.3.11. *Sous les hypothèses $\mathcal{F}.i$, $\mathcal{F}.ii$ et VC, pour tout $(2 + 5\nu_0)/2 \leq \alpha \leq 2 + 5\nu_0$ et $\theta > 0$, il existe $C(\alpha, \theta) > 0$, des variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi P et une suite indépendante $(\mathbb{G}_n)_{n \in \mathbb{N}^*}$ de processus de P -pont browniens définis sur un même espace de probabilité vérifiant pour tout $n \in \mathbb{N}^*$,*

$$\mathbb{P} \left(\max_{1 \leq k \leq n} \left\| \sqrt{k} \alpha_k^X - \sum_{i=1}^k \mathbb{G}_i \right\|_{\mathcal{F}} > C(\alpha, \theta) n^{1/2 - \tau(\alpha)} (\log n)^{\tau_2} \right) \leq \frac{1}{n^\theta},$$

et presque sûrement,

$$\max_{1 \leq k \leq n} \left\| \sqrt{k} \alpha_k^X - \sum_{i=1}^k \mathbb{G}_i \right\|_{\mathcal{F}} = O \left(n^{1/2 - \tau(\alpha)} (\log n)^{\tau_2} \right),$$

où

$$\tau(\alpha) = \frac{\alpha/(2 + 5\nu_0) - 1/2}{1 + \alpha} > 0, \quad \tau_2 = \frac{4 + 5\nu_0}{4 + 10\nu_0}.$$

Remarque 1.3.1. *L'optimisation de la borne donne que presque sûrement*

$$\max_{1 \leq k \leq n} \left\| \sqrt{k} \alpha_k^X - \sum_{i=1}^k \mathbb{G}_i \right\|_{\mathcal{F}} = O \left(n^{1/2(1 - 1/(3 + 5\nu_0))} (\log n)^{\tau_2} \right).$$

L'approximation forte de $\alpha_n(\mathcal{F})$ dans le cas où \mathcal{F} vérifie BR résulte du théorème 2 de [12]. Il est énoncé comme suit.

Théorème 1.3.12. *Sous les hypothèses $\mathcal{F}.i$, $\mathcal{F}.ii$ et BR, pour tout $\kappa < 1/2$ et $\theta > 0$, il existe $C(\tau, \theta) > 0$, des variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi P et une suite indépendante $(\mathbb{G}_n)_{n \in \mathbb{N}^*}$ de processus de P -pont brownien définis sur un même espace de probabilité vérifiant pour tout $n \in \mathbb{N}^*$,*

$$\mathbb{P} \left(\max_{1 \leq k \leq n} \left\| \sqrt{k} \alpha_k^X - \sum_{i=1}^k \mathbb{G}_i \right\|_{\mathcal{F}} > \sqrt{n} C(\tau, \theta) (\log n)^{-\tau} \right) \leq (\log n)^{-\theta},$$

et presque sûrement,

$$\max_{1 \leq k \leq n} \left\| \sqrt{k} \alpha_k^X - \sum_{i=1}^k \mathbb{G}_i \right\|_{\mathcal{F}} = O \left(\sqrt{n} (\log n)^{-\tau} \right),$$

où $\tau = \kappa(1/2 - \kappa)/(1 - \kappa)$.

Remarque 1.3.2. *L'optimisation de la borne donne que presque sûrement*

$$\max_{1 \leq k \leq n} \left\| \sqrt{k} \alpha_k^X - \sum_{i=1}^k \mathbb{G}_i \right\|_{\mathcal{F}} = O \left(\sqrt{n} (\log n)^{-(1 - 1/\sqrt{2})^2} \right).$$

Reformulation de l'énoncé. Supposons qu'une classe \mathcal{F} donnée vérifie BR ou VC. On reformule les théorèmes 1.3.12 et 1.3.11 sous l'énoncé suivant.

Théorème 1.3.13. *Si \mathcal{F} vérifie les conditions $\mathcal{F}.i$, $\mathcal{F}.ii$, VC ou BR alors pour tout $\theta > 0$, il existe $C_\theta > 0$, une suite $(v_n)_{n \in \mathbb{N}^*}$ de la forme $v_n = n^{-\alpha}(\log n)^\beta$ si \mathcal{F} vérifie VC et $v_n = (\log n)^{-\beta}$ si \mathcal{F} vérifie BR, avec $\alpha, \beta > 0$, des variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi P et une suite indépendante $(\mathbb{G}_n)_{n \in \mathbb{N}^*}$ de processus de P-pont brownien définis sur un même espace de probabilité vérifiant pour tout $n \in \mathbb{N}^*$,*

$$\mathbb{P} \left(n^{-1/2} \max_{1 \leq k \leq n} \left\| \sqrt{k} \alpha_k^X - \sum_{i=1}^k \mathbb{G}_i \right\|_{\mathcal{F}} > C_\theta v_n \right) \leq u_n^{-\theta}, \quad (1.3)$$

et presque sûrement,

$$n^{-1/2} \max_{1 \leq k \leq n} \left\| \sqrt{k} \alpha_k^X - \sum_{i=1}^k \mathbb{G}_i \right\|_{\mathcal{F}} = O(v_n), \quad (1.4)$$

avec $u_n = 1/n$ si \mathcal{F} vérifie VC et $u_n = 1/\log n$ si \mathcal{F} vérifie BR.

Ce résultat sera souvent utilisé par la suite pour nous assurer des résultats aux preuves simples mais aux conséquences puissantes. N'oublions pas que dans le cas où la classe de fonction étudiée est la classe des fonctions indicatrices $\mathcal{F} = \{\mathbb{1}_{]-\infty, t]} : t \in \mathbb{R}\}$ et que les X_i sont des variables aléatoires réelles admettant une fonction de répartition continue alors on peut utiliser l'approximation KMT (théorème 1.3.8) pour améliorer la vitesse d'approximation.

Remarques. La suite v_n introduite dans le théorème précédent ne tend pas très rapidement vers 0 : ce qu'on gagne en généralité on le perd en vitesse. En effet, l'approximation forte KMT nous apprend que $\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} = O(n^{-1/2} \log n)$ dans le cas où $\mathcal{F} = \{\mathbb{1}_{]-\infty, t]} : t \in \mathbb{R}\}$ est la classe VC des fonctions indicatrices. L'approximation forte de Berthet-Mason nous apprend juste dans ce cas particulier que $\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} = O(n^{-\alpha}(\log n)^\beta)$ avec $\alpha \simeq 0.4375, \beta \simeq 0.6429$. Dans les cas BR et VC le passage de la relation (1.3) à la relation (1.4) se fait via un raisonnement par blocs et l'utilisation du lemme de Borel-Cantelli. Précisément, on prend la relation (1.3) pour tout n de la forme 2^l et le membre de droite devient le terme général d'une série convergente. Ce résultat est une conséquence du résultat suivant donnée par la proposition 1 et 2 de [12].

Théorème 1.3.14. *Si \mathcal{F} vérifie les conditions $\mathcal{F}.i$, $\mathcal{F}.ii$, VC ou BR alors $\forall \theta > 0$, il existe $C_\theta > 0$, une suite $(v_n)_{n \in \mathbb{N}^*}$ de la forme $v_n = n^{-\alpha}(\log n)^\beta$ si \mathcal{F} vérifie VC et $v_n = (\log n)^{-\beta}$ si \mathcal{F} vérifie BR, avec $\alpha, \beta > 0$, des variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi P et une suite de P-pont brownien $(\mathbb{G}_n)_{n \in \mathbb{N}^*}$ définis sur un même espace de probabilité vérifiant pour tout $n \in \mathbb{N}^*$,*

$$\mathbb{P} (\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} > C_\theta v_n) \leq \frac{1}{n^\theta},$$

et presque sûrement,

$$\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} = O(v_n).$$

Information auxiliaire d'une partition

Lors de l'étude d'un échantillon, il est monnaie courante de recueillir des informations qualitatives ou quantitatives sur les individus d'une population telles que le poids, la taille, l'âge, ... Par ailleurs, il arrive que les statisticiens aient à disposition des informations auxiliaires telles que la proportion de gens appartenant à une classe spécifique, c'est-à-dire la probabilité d'appartenir à un ensemble d'une partition. Par exemple, la proportion exacte d'une population dont l'âge, le poids ou la taille est comprise dans une certaine tranche. On peut imaginer qu'une telle information est apportée par un expert, par l'exploitation d'une grande base de données, par les résultats d'un sondage préliminaire de très grande taille, etc... On souhaite naturellement intégrer ces informations a priori dans l'étude de l'échantillon avec l'espoir de corriger des sous et sur-représentations d'individus appartenant à une certaine catégorie que le hasard entraînerait inéluctablement. Nous devons donc apporter des modifications au niveau du processus empirique afin d'utiliser cette information. Cela entraîne inévitablement un biais, que nous souhaitons être asymptotiquement nul, mais nous avons l'espoir d'augmenter la précision dans les estimations d'une large collection de statistiques, sous-entendu que l'information amenée permet de réduire asymptotiquement la variance et donc le risque quadratique des estimateurs.

Très peu de recherches porte sur l'utilisation d'une telle information auxiliaire. Le peu d'articles entourant ce sujet ne traite pas ce problème du point de vue de la théorie du processus empirique. La plupart de ces articles montrent que l'on peut améliorer l'estimation d'une seule statistique en utilisant l'information auxiliaire alors que nous souhaitons dans notre cas montrer qu'il y a amélioration d'une collection de statistiques. Nous rappelons la littérature qui entoure cette question.

Ce chapitre est dédié à l'étude de l'injection de l'information auxiliaire apportée par la connaissance de la probabilité d'appartenir à un ensemble d'une seule partition. Il est le cas particulier du chapitre suivant mais il regroupe les premières traces de recherche concernant l'étude du processus empirique avec information auxiliaire. Certains résultats sont par ailleurs plus précis que dans le chapitre suivant.

2.1 Introduction

2.1.1 Notations et définitions

Cadre de travail. Soient X, X_1, \dots, X_n des variables i.i.d. définies sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ à valeurs dans un espace mesurable $(\mathcal{X}, \mathcal{T}')$. On note $P = \mathbb{P}^X$ la loi de X et \mathbb{P}_n la mesure empirique associée à X_1, \dots, X_n . Dans la suite, \mathcal{F} désigne une classe de fonctions mesurables à valeurs réelles. Nous notons

$$\forall f \in \mathcal{F}, \quad \forall n \in \mathbb{N}^*, \quad P(f) = \mathbb{E}[f(X)], \quad \mathbb{P}_n(f) = \frac{1}{n} \sum_{i=1}^n f(X_i).$$

Nous supposons dans cette partie que l'on connaît la valeur de $P[\mathcal{A}] = (P(A_1), \dots, P(A_m))$ où les événements $(A_j)_{1 \leq j \leq m} \subset \mathcal{T}'$ forment une partition de \mathcal{X} . Il s'agit d'une information exacte que nous supposons apportée par le biais d'une source extérieure (expert, base de données, ...). Remarquons qu'il suffit de connaître $P(A)$ pour $A \in \mathcal{A}$ pour connaître la partition $\{A, A^C\}$ où $A^C = \mathcal{X} \setminus A$. Nous introduisons ci-après la mesure prenant en compte cette information auxiliaire.

Mesure empirique avec information auxiliaire. L'objectif statistique est de créer une version modifiée $\hat{\mathbb{P}}_n(\mathcal{F})$ du processus empirique $\mathbb{P}_n(\mathcal{F})$ qui vérifie en particulier $\hat{\mathbb{P}}_n(A_j) = P(A_j)$ pour tout $1 \leq j \leq m$. L'espoir est qu'en injectant cette information auxiliaire vraie, nous diminuons l'écart de $\mathbb{P}_n(\mathcal{F})$ à $P(\mathcal{F})$. Nous supposons que les probabilités connues ne sont pas triviales, c'est-à-dire $P(A_j) \notin \{0, 1\}$, $1 \leq j \leq m$.

Définition 2.1.1 (Mesure empirique avec information auxiliaire d'une partition). *On note $\hat{\mathbb{P}}_n^{\mathcal{A}}(\mathcal{F})$ la mesure empirique avec pour information auxiliaire la connaissance de $P[\mathcal{A}]$ et indexée par la classe de fonctions \mathcal{F} définie pour tout $n \in \mathbb{N}^*$ et $f \in \mathcal{F}$ par*

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

S'il n'y a pas d'ambiguïté sur la partition de l'information auxiliaire apportée, nous notons $\hat{\mathbb{P}}_n(\mathcal{F})$. Cette nouvelle mesure empirique affecte donc à $f(X_i)$ le poids aléatoire $P(A_j)/n \mathbb{P}_n(A_j)$ aux $X_i \in A_j$ et vérifie en particulier $\hat{\mathbb{P}}_n(A_j) = \hat{\mathbb{P}}_n(\mathbb{1}_{A_j}) = P(A_j)$. Cette mesure conserve la même propriété de linéarité que \mathbb{P}_n , i.e. $\hat{\mathbb{P}}_n(f + \lambda g) = \hat{\mathbb{P}}_n(f) + \lambda \hat{\mathbb{P}}_n(g)$. On remarque que cette mesure peut se voir comme la mesure empirique $\mathbb{P}_n(f)$ corrigée en \mathcal{A} puisque pour tout $f \in \mathcal{F}$,

$$\hat{\mathbb{P}}_n^{\mathcal{A}}(f) = \mathbb{P}_n(f) - \sum_{j=1}^m \frac{\mathbb{P}_n(f \mathbb{1}_{A_j})}{\mathbb{P}_n(A_j)} (\mathbb{P}_n(A_j) - P(A_j)). \quad (2.1)$$

Probabilité d'existence. Pour tout $n \in \mathbb{N}^*$, la nouvelle mesure empirique $\hat{\mathbb{P}}_n(\mathcal{F})$ est défini sur l'ensemble $B_n = \{\min_{j=1, \dots, m} \mathbb{P}_n(A_j) > 0\}$. La probabilité que la mesure empirique avec information auxiliaire ne soit pas définie est bornée par

$$P(B_n^C) = P(\Omega \setminus B_n) = P\left(\bigcup_{j=1}^m \bigcap_{i=1}^n \{X_i \notin A_j\}\right) \leq m \left(1 - \min_{j=1, \dots, m} P(A_j)\right)^n. \quad (2.2)$$

Avec l'hypothèse que $P(A_j) \notin \{0, 1\}$ pour tout $1 \leq j \leq m$, nécessairement la probabilité que la mesure $\hat{\mathbb{P}}_n(\mathcal{F})$ ne soit pas définie tend exponentiellement vers 0.

Processus empirique avec information auxiliaire. Naturellement on associe à cette nouvelle mesure le processus empirique avec information auxiliaire.

Définition 2.1.2 (Processus empirique avec information auxiliaire d'une partition). *On appelle $\hat{\alpha}_n^A(\mathcal{F})$ le processus empirique avec pour information auxiliaire la connaissance de $P[A]$ et indexé par la classe de fonctions \mathcal{F} , le processus défini pour tout $n \in \mathbb{N}^*$ et $f \in \mathcal{F}$ par*

$$\hat{\alpha}_n^A(f) = \sqrt{n}(\hat{\mathbb{P}}_n^A(f) - P(f)),$$

où $\hat{\mathbb{P}}_n^A(\mathcal{F})$ est la mesure empirique avec information auxiliaire apportée par \mathcal{A} indexée par \mathcal{F} .

Comme pour la mesure empirique avec information auxiliaire, s'il n'y a pas d'ambiguïté sur les variables aléatoires étudiées et sur la partition qui constitue l'information auxiliaire on note ce processus $\hat{\alpha}_n(\mathcal{F})$.

Remarque 2.1.1. *Il est également possible de définir le processus empirique indexé par des ensembles. Si on prend $f = \mathbf{1}_A$ on obtient que*

$$\hat{\mathbb{P}}_n(A) = \hat{\mathbb{P}}_n(\mathbf{1}_A) = \sum_{j=1}^m \frac{\mathbb{P}_n(A \cap A_j)}{\mathbb{P}_n(A_j)} P(A_j).$$

Le processus empirique avec information auxiliaire devient dans ce cas

$$\hat{\alpha}_n(A) = \hat{\alpha}_n(\mathbf{1}_A) = \sqrt{n} \left(\sum_{j=1}^m \frac{P(A_j)}{\mathbb{P}_n(A_j)} \mathbb{P}_n(A \cap A_j) - P(A \cap A_j) \right).$$

D'après la remarque précédente, on a en particulier que $\hat{\mathbb{P}}_n(A_j) = P(A_j)$ et $\hat{\alpha}_n(A_j) = 0$.

On dénote l'espérance conditionnelle de f sachant A par $P(f|A) = \mathbb{E}[f(X)|A] = P(f\mathbf{1}_A)/P(A)$. Avec cette notation on peut écrire que $\mathbb{P}_n(f) = \sum_{j=1}^m \mathbb{P}_n(A_j)\mathbb{P}_n(f|A_j)$ et d'autre part $\hat{\mathbb{P}}_n(f) = \sum_{j=1}^m P(A_j)\mathbb{P}_n(f|A_j)$. La mesure $\hat{\mathbb{P}}_n$ peut être vue comme une correction de la mesure \mathbb{P}_n .

2.1.2 Motivation

Approche existante. Il existe dans la littérature plusieurs méthodes, œuvrant dans ce but, appelées méthodes de redressement. Parmi ces méthodes, on retrouve le redressement par suppression, par extraction ou par repondération. C'est cette dernière qui est étudiée en détail sous un nouvel angle dans ce chapitre. En effet nous avons substitué les poids initiaux $1/n$ de chaque individu par des poids aléatoires, dépendant de l'ensemble de la partition \mathcal{A} dans lequel l'individu appartient. Cette approche diffère de ce qui a été fait dans la littérature. La seule notion se rapprochant le plus de notre étude est l'estimateur d'Horvitz-Thompson introduit par ces derniers en 1952 [49]. Dans notre étude, l'information auxiliaire est apportée par la probabilité d'appartenir à un ensemble d'une partition donnée de notre population et on considère que notre modèle est celui d'une superpopulation. L'information auxiliaire étudiée par Horvitz et Thompson est donnée par la probabilité d'appartenir à un sous-ensemble de la population et les individus n'ont pas la même probabilité d'appartenir à l'échantillon. Plus formellement, Horvitz et Thompson s'intéressent à l'estimation de

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i,$$

où x_i correspond à la valeur du i -ème individu. On note s un échantillon de la population x_1, \dots, x_N tiré aléatoirement et sans remise, $\pi_i = \mathbb{P}(i \in s)$ pour tout $1 \leq i \leq N$ la probabilité qu'un individu i appartienne à cet échantillon et ce modèle suppose que ces probabilités d'inclusion sont connues et possiblement inégales. L'estimateur non biaisé \bar{x} de \bar{X} proposé par Horvitz et Thompson est donné par

$$\bar{x} = \frac{1}{N} \sum_{i \in s} \frac{x_i}{\pi_i}.$$

Les principaux résultats établis sur l'estimateur de Horvitz-Thompson sont les suivants. Robinson [68] a montré que cet estimateur était consistant. Plus précisément, il a montré dans un premier temps que si les x_i étaient uniformément bornées alors

$$\bar{x} = \bar{X} + O(N^{-1}\epsilon^{1/2}) + O(N^{-1}\xi^{1/2}),$$

où $\epsilon = \sum_{i=1}^N (1 - \pi_i)/\pi_i$ et $\xi = \sum_{i=1}^N \sum_{j \neq i, \pi_i \pi_j < \pi_{ij}} (\pi_{ij} - \pi_i \pi_j)/\pi_i^2$ puis dans un second temps que si $\limsup_{N \rightarrow +\infty} \sum_{i=1}^N x_i^2/N < +\infty$ alors

$$\bar{x} = \bar{X} + O(N^{-1/2}\delta^{-1/2}) + O(N^{-1/2}\delta^{-1}\zeta),$$

où $\delta = \min_{i=1, \dots, N} \pi_i$, $\zeta = \max_{i=1, \dots, N} \sum_{j \neq i, \pi_i \pi_j < \pi_{ij}} (\pi_{ij} - \pi_i \pi_j)$. Erdős et Rényi [41] ont prouvé qu'il était asymptotiquement gaussien. Yates et Grundy [81] ainsi que Sen [70] ont proposé un estimateur de la variance de cet estimateur faisant intervenir les probabilités d'inclusion d'ordre 1 et 2. Le point de vue de Horvitz et Thompson ne permet pas de généraliser les résultats ci-dessus à une collection d'estimateurs. Pour cela, nous proposons d'adopter le point de vue des processus empiriques indexés par les fonctions dans lesquels nous exploitons la connaissance d'une information auxiliaire.

Convergence du processus stochastique $\hat{\alpha}_n(\mathcal{F})$. Une question à laquelle nous allons répondre dans ce chapitre est de connaître la convergence du processus $\hat{\alpha}_n(\mathcal{F})$. Nous allons montrer grâce à l'outil d'approximation forte que ce processus converge en loi dans $\ell^\infty(\mathcal{F})$ vers un processus gaussien centré $\hat{\mathbb{G}}(\mathcal{F})$ dont la variance sera uniformément sur la classe de fonctions plus faible que celle du P -pont brownien $\mathbb{G}(\mathcal{F})$. On s'intéresse ainsi non pas à un seul estimateur mais à une collection d'estimateurs. L'approximation forte nous permet même de quantifier la vitesse de convergence faible de $\hat{\alpha}_n(\mathcal{F})$ vers $\hat{\mathbb{G}}(\mathcal{F})$.

Problème de biais et variance. Le fait de substituer les poids de chaque individu par des poids aléatoires entraînent l'apparition d'un biais. Plus formellement,

$$\mathbb{E}[\hat{\mathbb{P}}_n(f)] = \sum_{j=1}^m P(A_j) \mathbb{E} \left[\frac{\mathbb{P}_n(f \mathbb{1}_{A_j})}{\mathbb{P}_n(A_j)} \right] \neq \mathbb{E}[f].$$

L'une de nos principales motivations est de montrer que, uniformément sur une grande classe de fonctions, ce biais introduit s'annule asymptotiquement et que la variance, donc le risque quadratique, diminue quand $n \rightarrow +\infty$. Nous allons montrer que c'est effectivement le cas en majorant le biais introduit par une borne tendant exponentiellement vers 0 et en quantifiant la diminution de variance asymptotique.

Organisation. Ce chapitre sera bref car le chapitre suivant étudie un processus qui généralise celui-ci. Nous donnons les résultats asymptotiques et non-asymptotiques du processus empirique avec information auxiliaire respectivement dans les sections 2.2 et 2.3. Nous finissons ce chapitre par une application de cette nouvelle théorie au test d'adéquation à une loi de Kolmogorov-Smirnov à la section 2.4.

2.2 Résultats asymptotiques

2.2.1 Approximation forte

Simplification du problème. Le processus empirique avec information auxiliaire $\hat{\alpha}_n(\mathcal{F})$ peut s'exprimer grâce au processus empirique classique $\alpha_n(\mathcal{F})$ de la manière suivante. Pour tout $n \in \mathbb{N}^*$,

$$\begin{aligned}\hat{\alpha}_n(f) &= \sqrt{n} \left[\sum_{j=1}^m \frac{P(A_j)}{\mathbb{P}_n(A_j)} \left(\frac{1}{n} \sum_{i=1}^n f(X_i) \mathbb{1}_{\{X_i \in A_j\}} - \mathbb{E}[f|A_j] \mathbb{P}_n(A_j) \right) \right] \\ &= \sqrt{n} \left[\sum_{j=1}^m \frac{P(A_j)}{\mathbb{P}_n(A_j)} \left(\mathbb{P}_n(f \mathbb{1}_{A_j}) - P(f \mathbb{1}_{A_j}) - \mathbb{E}[f|A_j] (\mathbb{P}_n(A_j) - P(A_j)) \right) \right] \\ &= \sum_{j=1}^m \frac{P(A_j)}{\mathbb{P}_n(A_j)} \left(\alpha_n(f \mathbb{1}_{A_j}) - \mathbb{E}[f|A_j] \alpha_n(A_j) \right).\end{aligned}$$

Puisque chaque $\mathbb{P}_n(A_j)$ converge presque sûrement vers $P(A_j)$, on s'attend à ce que $\hat{\alpha}_n(\mathcal{F})$ ait le même comportement asymptotique que le processus $\tilde{\alpha}_n(\mathcal{F})$ défini par

$$\begin{aligned}\tilde{\alpha}_n(f) &= \sum_{j=1}^m \alpha_n(f \mathbb{1}_{A_j}) - \mathbb{E}[f|A_j] \alpha_n(A_j) \\ &= \alpha_n(f) - \sum_{j=1}^m \mathbb{E}[f|A_j] \alpha_n(A_j) = \alpha_n \left(f - \sum_{j=1}^m \mathbb{E}[f|A_j] \mathbb{1}_{A_j} \right),\end{aligned}\tag{2.3}$$

car $\hat{\alpha}_n(f) = \sum_{j=1}^m \tilde{\alpha}_n(f \mathbb{1}_{A_j}) P(A_j) / \mathbb{P}_n(A_j)$. Nous allons montrer dans un premier temps l'approximation forte et par suite la convergence en loi de $\tilde{\alpha}_n(\mathcal{F})$ par un processus gaussien centré $\hat{\mathbb{G}}(\mathcal{F})$ dont on étudie la fonction de covariance dans la sous-section 2.2.3. Nous montrons ensuite que cette approximation forte par ce processus gaussien reste valable pour $\hat{\alpha}_n(\mathcal{F})$.

Hypothèses. Dans le cas où la classe de fonctions \mathcal{F} vérifie certaines conditions d'entropie, on est en mesure d'établir une approximation forte du processus avec information auxiliaire d'une partition. Nous supposons dans ce chapitre que \mathcal{F} vérifie BR ou VC. Si \mathcal{F} vérifie VC, on note $v_n = n^{-\alpha} (\log n)^\beta$, $\alpha = 1/(2 + 5\nu_0)$, $\beta = (4 + 5\nu_0)/(4 + 10\nu_0)$ et si \mathcal{F} vérifie BR on note $v_n = (\log n)^{-\gamma}$, $\gamma = (1 - r_0)/2r_0$. Posons la classe

$$\mathcal{G} = \left\{ f - \sum_{j=1}^m \mathbb{E}[f|A_j] \mathbb{1}_{A_j} : f \in \mathcal{F} \right\},$$

qui est de Donsker. Le lemme suivant établit que cette classe possède la même entropie que la classe de fonctions \mathcal{F} .

Lemme 2.2.1. *Si \mathcal{F} vérifie les conditions $\mathcal{F}.i$, $\mathcal{F}.ii$ alors il en va de même pour \mathcal{G} . Si VC (resp. BR) est vérifié alors \mathcal{G} vérifie également VC (resp. BR) avec la même puissance ν_0 (resp. r_0).*

Démonstration. Celle-ci vérifie la condition $\mathcal{F}.i$ avec la constante $M_{\mathcal{G}} = 2M_{\mathcal{F}}$. Le théorème de convergence dominée permet de montrer que si $\lim_{k \rightarrow +\infty} f_k = f$ avec $(f_k)_k \subset \tilde{\mathcal{F}}$ alors $\lim_{k \rightarrow +\infty} f_k - \sum_{j=1}^m \mathbb{E}[f_k|A_j] \mathbb{1}_{A_j} = f - \sum_{j=1}^m \mathbb{E}[f|A_j] \mathbb{1}_{A_j}$, ce qui entraîne que \mathcal{G} vérifie $\mathcal{F}.ii$

avec comme sous-ensemble dénombrable $\tilde{\mathcal{G}} = \{f - \sum_{j=1}^m \mathbb{E}[f|A_j] \mathbb{1}_{A_j} : f \in \tilde{\mathcal{F}}\}$. Pour $j = 1, \dots, m$ notons $\phi_j(f) = f - \mathbb{E}[f|A_j]$ et $\phi(f) = f - \sum_{j=1}^m \mathbb{E}[f|A_j] \mathbb{1}_{A_j} = \sum_{j=1}^m \phi_j(f) \mathbb{1}_{A_j}$. Alors

$$\begin{aligned} d_Q^2(\phi(f), \phi(g)) &= \int_{\mathcal{X}} (\phi(f) - \phi(g))^2 dQ = \int_{\mathcal{X}} \left(\sum_{j=1}^m (\phi_j(f) - \phi_j(g)) \mathbb{1}_{A_j} \right)^2 dQ \\ &= \sum_{j=1}^m \int_{A_j} (\phi_j(f) - \phi_j(g))^2 dQ = \sum_{j=1}^m \int_{A_j} (f - g - \mathbb{E}[f - g|A_j])^2 dQ \\ &\leq d_Q^2(f, g). \end{aligned}$$

Donc $d_Q(f, g) < \varepsilon$ entraîne $d_Q(\phi(f), \phi(g)) < \varepsilon$. Si \mathcal{F} peut être recouverte par $N(\mathcal{F}, \varepsilon, d_Q)$ boules de d_Q -rayon ε avec des centres g alors \mathcal{G} peut être recouvert par le nombre de boules de même rayon et centrés en les fonctions $\phi(g)$. On vient de montrer que si \mathcal{F} vérifie VC il en va de même pour \mathcal{G} . Si $g^- \leq f \leq g^+$ alors

$$h_j^- = (g^- - \mathbb{E}[g^+|A_j]) \mathbb{1}_{A_j} \leq \phi_j(f) \leq (g^+ - \mathbb{E}[g^-|A_j]) \mathbb{1}_{A_j} = h_j^+.$$

La taille d_P du crochet $[h_j^-, h_j^+]$ est

$$\begin{aligned} d_P^2(h_j^-, h_j^+) &= \int_{A_j} (g^+ - g^- + \mathbb{E}[g^+ - g^-|A_j])^2 dP \\ &= P((g^+ - g^-)^2) + P(A_j) \mathbb{E}[g^+ - g^-|A_j]^2 + 2\mathbb{E}[g^+ - g^-|A_j] P((g^+ - g^-) \mathbb{1}_{A_j}). \end{aligned}$$

Si $d_P(g^+, g^-) < \varepsilon$ alors l'inégalité de Hölder entraîne que $P((g^+ - g^-) \mathbb{1}_{A_j}) \leq \varepsilon \sqrt{P(A_j)}$ et $\mathbb{E}[g^+ - g^-|A_j] \leq \varepsilon / \sqrt{P(A_j)}$ d'où

$$d_P^2(h_j^-, h_j^+) \leq P((g^+ - g^-)^2 \mathbb{1}_{A_j}) + 3\varepsilon^2.$$

Ainsi, $\phi(f) = \sum_{j=1}^m \phi_j(f) \mathbb{1}_{A_j} \in [h^-, h^+]$ avec $h^\pm = \sum_{j=1}^m h_j^\pm$ qui satisfait

$$d_P^2(h^-, h^+) = \sum_{j=1}^m d_P^2(h_j^-, h_j^+) \leq d_P^2(g^-, g^+) + 3m\varepsilon^2 \leq 4m\varepsilon^2.$$

Il s'ensuit que $N_{[\]}(\mathcal{G}, \varepsilon, d_P) \leq N_{[\]}(\mathcal{F}, \varepsilon/2\sqrt{m}, d_P)$. On vient de montrer que si \mathcal{F} vérifie BR il en va de même pour \mathcal{G} avec les constantes r_0 et $2^{r_0} m b_0$. \square

Le lemme 2.2.1 nous permet d'utiliser l'approximation forte de Berthet-Mason sur la classe de fonctions \mathcal{G} .

Approximation forte. Le résultat qui suit montre que si la classe de fonctions \mathcal{F} le permet, le processus empirique $\tilde{\alpha}_n(\mathcal{F})$ peut être approché fortement par un P -pont brownien modifié $\hat{\mathbb{G}}^{\mathcal{A}}(\mathcal{F})$ défini pour tout $f \in \mathcal{F}$ par $\hat{\mathbb{G}}^{\mathcal{A}}(f) = \mathbb{G}(f) - \mathbb{G}^{\nabla, \mathcal{A}}(f)$, où $\mathbb{G}(\mathcal{F})$ est le P -pont brownien usuel et $\mathbb{G}^{\nabla, \mathcal{A}}(\mathcal{F})$ est définie pour tout $f \in \mathcal{F}$ par

$$\mathbb{G}^{\nabla, \mathcal{A}}(f) = \sum_{j=1}^m \mathbb{E}[f|A_j] \mathbb{G}(A_j).$$

On note respectivement $\hat{\mathbb{G}}(\mathcal{F}), \mathbb{G}^{\nabla}(\mathcal{F})$ les processus limites $\hat{\mathbb{G}}^{\mathcal{A}}(\mathcal{F}), \mathbb{G}^{\nabla, \mathcal{A}}(\mathcal{F})$ s'il n'y a pas de risque de confusion avec la partition \mathcal{A} étudiée.

Théorème 2.2.2. *Si \mathcal{F} vérifie les conditions $\mathcal{F}.i$, $\mathcal{F}.ii$ alors pour tout $\theta > 0$ il existe $C_\theta, n_\theta > 0$, des variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi P et une suite $(\mathbb{G}_n)_{n \in \mathbb{N}^*}$ de processus de P -pont brownien définis sur un même espace de probabilité vérifiant pour tout $n > n_\theta$,*

$$\mathbb{P} \left(\{ \|\hat{\alpha}_n - \hat{\mathbb{G}}_n\|_{\mathcal{F}} > C_\theta v_n \} \cup \{ \|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} > C_\theta v_n \} \right) \leq \frac{1}{n^\theta}, \quad (2.4)$$

et presque sûrement,

$$\max(\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}}, \|\hat{\alpha}_n - \hat{\mathbb{G}}_n\|_{\mathcal{F}}) = O(v_n). \quad (2.5)$$

Démonstration. Remarquons que $\|\hat{\alpha}_n - \hat{\mathbb{G}}_n\| \leq \|\hat{\alpha}_n - \tilde{\alpha}_n\|_{\mathcal{F}} + \|\tilde{\alpha}_n - \hat{\mathbb{G}}_n\|_{\mathcal{F}}$. D'après le lemme 2.2.1 on peut appliquer l'approximation forte de Berthet-Mason à \mathcal{G} . Il existe donc $C_0 > 0$, des variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi P et une suite $(\mathbb{G}_n)_{n \in \mathbb{N}^*}$ de processus de P -pont brownien définis sur un même espace de probabilité vérifiant pour tout $n \in \mathbb{N}^*$,

$$\mathbb{P} \left(\|\tilde{\alpha}_n - \hat{\mathbb{G}}_n\|_{\mathcal{F}} > C_0 v_n \right) \leq \mathbb{P}(\|\alpha_n - \mathbb{G}_n\|_{\mathcal{G}} > C_0 v_n) < \frac{1}{3n^\theta}. \quad (2.6)$$

Sous réserve que $\|\alpha_n\|_{\mathcal{F}} < \sqrt{n}\delta$ avec $\delta = \min_{j=1, \dots, m} P(A_j)$, on a

$$\begin{aligned} \|\hat{\alpha}_n - \tilde{\alpha}_n\|_{\mathcal{F}} &= \left\| \sum_{j=1}^m \left(\frac{P(A_j)}{\mathbb{P}_n(A_j)} - 1 \right) \alpha_n \left((f - \mathbb{E}[f|A_j]) \mathbb{1}_{A_j} \right) \right\|_{\mathcal{F}} \\ &\leq \frac{m \max_{j=1}^m |\alpha_n(A_j)|}{\sqrt{n}(\delta - \max_{j=1}^m |\alpha_n(A_j)|/\sqrt{n})} \|\alpha_n\|_{\mathcal{G}} \\ &\leq \frac{m}{\sqrt{n}(\delta - \|\alpha_n\|_{\mathcal{H}}/\sqrt{n})} \|\alpha_n\|_{\mathcal{H}}^2, \end{aligned} \quad (2.7)$$

où $\mathcal{H} = \mathcal{G} \cup \{\mathbb{1}_{A_j} : 1 \leq j \leq m\}$. D'après (2.7) et l'inégalité de Talagrand (théorèmes 1.2.6 et 1.2.7), il existe $D > 0$ et n_1 tel que pour tout $n > n_1$,

$$\begin{aligned} \mathbb{P}(\|\alpha_n\|_{\mathcal{H}} > \sqrt{n}\delta) &< \frac{1}{3n^\theta}, \\ \mathbb{P} \left(\|\hat{\alpha}_n - \tilde{\alpha}_n\|_{\mathcal{F}} > D \frac{\log n}{\sqrt{n}} \right) &\leq \mathbb{P} \left(\frac{m}{\sqrt{n}(\delta - \|\alpha_n\|_{\mathcal{H}}/\sqrt{n})} \|\alpha_n\|_{\mathcal{H}}^2 > D \frac{\log n}{\sqrt{n}} \right) \\ &\leq \mathbb{P} \left(\|\alpha_n\|_{\mathcal{H}}^2 > \frac{D\delta \log n}{2m} \right) + P \left(\|\alpha_n\|_{\mathcal{H}} > \frac{\sqrt{n}\delta}{2} \right) \\ &< \frac{1}{3n^\theta}. \end{aligned} \quad (2.8)$$

L'inégalité (2.4) est une conséquence de (2.6) et (2.8) avec $C_\theta > 0$ suffisamment grand pour vérifier $C_\theta v_n > C_0 v_n + D \log n / \sqrt{n}$ pour tout $n \geq n_\theta = n_1$. Pour $\theta > 1$, le lemme de Borell-Cantelli permet de conclure à (2.5) à partir de (2.4). \square

Remarque 2.2.1. *Dans la preuve du théorème 2.2.2, nous avons mis en évidence que presque sûrement $\|\hat{\alpha}_n - \tilde{\alpha}_n\|_{\mathcal{F}} = O(\log n / \sqrt{n})$.*

Remarque 2.2.2. *Si la classe de fonctions \mathcal{F} est la classe de fonctions indicatrices $\mathcal{F} = \{\mathbb{1}_{]-\infty, t]}\} : t \in \mathbb{R}\}$ alors on aurait pu appliquer l'approximation forte de KMT. Dans ce cas, on peut remplacer la suite v_n par $\log n / \sqrt{n}$.*

2.2.2 Convergence en loi

Distance de Lévy-Prokhorov. L'approximation forte de $\hat{\alpha}_n(\mathcal{F})$ par le P -pont brownien avec information auxiliaire $\hat{\mathbb{G}}(\mathcal{F})$ permet de dériver d'autres résultats tels que la convergence en loi. Mieux celle-ci donne une vitesse pour ce mode de convergence. On note d_{LP} la distance de Lévy-Prokhorov qui est définie de la manière suivante (voir section 11.3 de [35] pour une définition). Si (S, d) est un espace métrique et P, Q deux lois sur S alors

$$d_{LP}(P, Q) = \inf \{ \varepsilon > 0 : P(A) \leq Q(A^\varepsilon) + \varepsilon, \quad \forall A \text{ ensemble borélien} \},$$

où $A^\varepsilon = \{y \in S : \exists x \in A, d(x, y) \leq \varepsilon\}$. Cette distance caractérise la convergence en loi sur l'ensemble des lois de S .

Résultat. La proposition suivante borne la distance de Lévy-Prokhorov de $\hat{\alpha}_n(\mathcal{F})$ par rapport à $\hat{\mathbb{G}}(\mathcal{F})$. Elle implique en particulier que $\hat{\alpha}_n(\mathcal{F})$ converge en loi vers $\hat{\mathbb{G}}(\mathcal{F})$ dans $\ell^\infty(\mathcal{F})$.

Proposition 2.2.3. *Supposons que \mathcal{F} vérifie $\mathcal{F}.i$ et $\mathcal{F}.ii$. Il existe $C > 0$ tel que pour tout $n \in \mathbb{N}^*$, $d_{LP}(\hat{\alpha}_n, \hat{\mathbb{G}}) \leq C v_n$.*

Démonstration. On peut appliquer l'approximation forte du théorème 2.2.2 et notamment utiliser l'inégalité (2.4). Soit A un borélien de $\ell^\infty(\mathcal{F})$ alors

$$\begin{aligned} \mathbb{P}(\hat{\alpha}_n \in A) &\leq \mathbb{P}\left(\{\hat{\alpha}_n \in A\} \cap \{ \|\hat{\alpha}_n - \hat{\mathbb{G}}_n\|_{\mathcal{F}} < C_\theta v_n\} \cap B_n\right) \\ &\quad + \mathbb{P}\left(\{ \|\hat{\alpha}_n - \hat{\mathbb{G}}_n\|_{\mathcal{F}} > C_\theta v_n\} \cap B_n\right) + \mathbb{P}(B_n^C) \\ &\leq \mathbb{P}\left(\hat{\mathbb{G}}_n \in A^{C_\theta v_n}\right) + \frac{1}{n^\theta} + m(1 - \delta)^n, \end{aligned}$$

avec $\delta = \min_{j=1, \dots, m} P(A_j)$. Cette inégalité reste vraie en intervertissant $\hat{\alpha}_n$ par $\hat{\mathbb{G}}_n$. Puisque v_n est la suite convergeant le moins vite vers 0, il existe $n_1 > 0$ tel que pour tout $n > n_1$, $1/n^\theta + m(1 - \delta)^n < C_\theta v_n$. La proposition découle de la définition de la distance de Lévy-Prokhorov. \square

Nous montrons par la suite que le processus avec information auxiliaire, comme nous le souhaitons au départ, possède une variance plus faible que le pont brownien classique.

2.2.3 Variance et covariance asymptotiques

Diminution de variance et covariance asymptotiques. Le résultat suivant confirme la propriété de diminution de variance qu'on attend de notre processus empirique avec information auxiliaire.

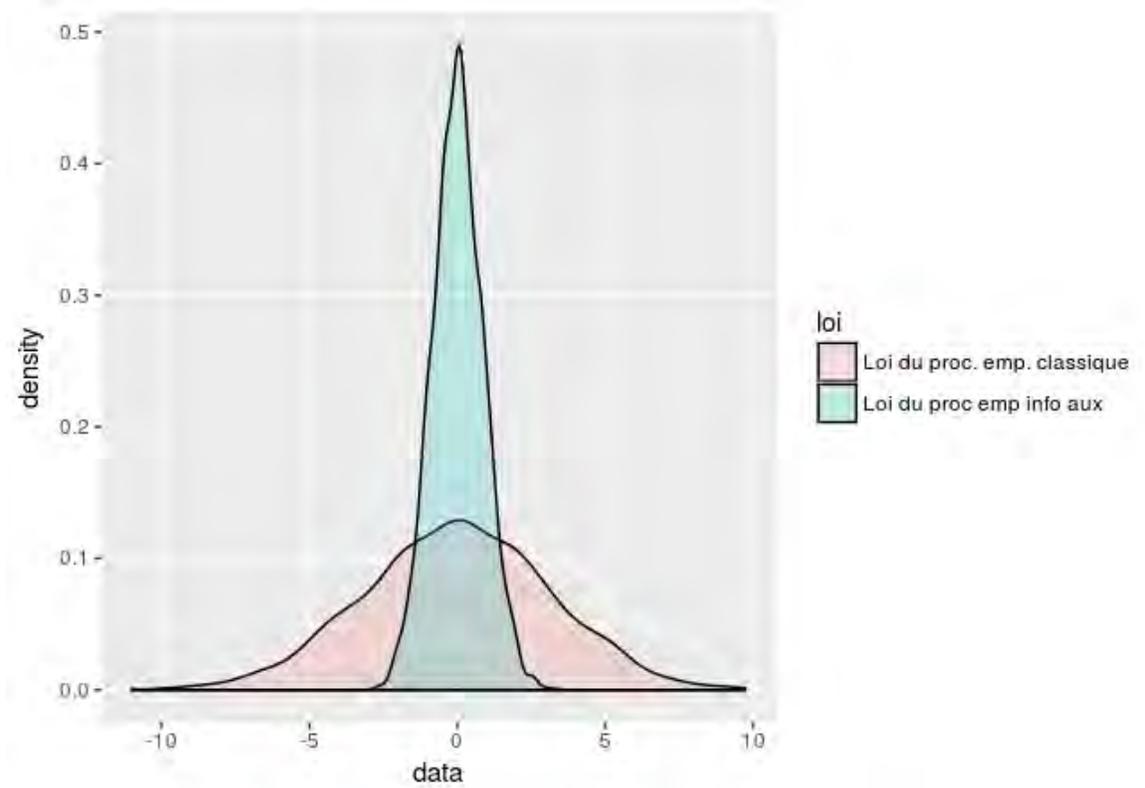
Proposition 2.2.4. *La variance du processus empirique avec information auxiliaire $\hat{\mathbb{G}}$ est plus faible que celle du processus empirique \mathbb{G} .*

Démonstration. Par convexité de $x \mapsto x^2$ et en utilisant (2.11), on obtient immédiatement que

$$\begin{aligned} \forall f \in \mathcal{F}, \quad \text{Var}(\hat{\mathbb{G}}(f)) &\leq \mathbb{E}[f^2(X)] - \left(\sum_{j=1}^m P(A_j) \mathbb{E}[f|A_j] \right)^2 \\ &= \mathbb{E}[f^2(X)] - \mathbb{E}[f(X)]^2 = \text{Var}(\mathbb{G}(f)). \end{aligned}$$

\square

Voici une simulation numérique représentant la proposition précédente. Sont en rose et en bleu respectivement la loi du processus empirique $\alpha_n(\mathcal{F})$ et la loi du processus $\hat{\alpha}_n(\mathcal{F})$.



Dans cette simulation, nous avons généré des lois normales, calculé leur moyenne et apporté comme information auxiliaire la probabilité pour ces lois normales d'appartenir à des intervalles disjoints. On peut apprécier dans cet exemple une réduction de variance importante : la variance du processus empirique avec information auxiliaire est ici 12 fois inférieure à celle du processus empirique classique. La proposition 2.2.4 nous apprend qu'injecter l'information auxiliaire permettait de réduire la variance du processus empirique sur une classe de fonctions donnée. On se demande alors s'il en va de même pour la covariance de ce processus. La proposition suivante affirme que l'ajout d'information auxiliaire permet également a fortiori de diminuer la covariance en un sens à définir.

Proposition 2.2.5. *Pour tout $f_1, \dots, f_k \in \mathcal{F}$ si on note Σ la matrice de covariance de $(\mathbb{G}(f_1), \dots, \mathbb{G}(f_k))$ et $\hat{\Sigma}$ celle de $(\hat{\mathbb{G}}(f_1), \dots, \hat{\mathbb{G}}(f_k))$ alors $\Sigma - \hat{\Sigma}$ est semi-définie positive.*

Remarque 2.2.3. *En particulier, toute combinaison linéaire $\sum_{i=1}^k a_i \hat{\mathbb{G}}(f_i)$ sera nécessairement de variance plus faible que la combinaison $\sum_{i=1}^k a_i \mathbb{G}(f_i)$.*

Démonstration. La différence de matrice $\Sigma - \hat{\Sigma}$ correspond d'après (2.10) à la matrice de covariance de $(\mathbb{G}^\nabla(f_1), \dots, \mathbb{G}^\nabla(f_k))$ d'où le résultat. \square

Expression. Le processus empirique avec information auxiliaire converge vers une modification du pont brownien qui possède une variance plus faible que le P -pont brownien

initial. La covariance du processus limite $\widehat{\mathbb{G}}(\mathcal{F})$ est la même que celle du processus dominant $\tilde{\alpha}_n(\mathcal{F})$ défini par l'équation (2.3). Étudions les covariances des processus $\mathbb{G}(\mathcal{F})$ et $\mathbb{G}^\nabla(\mathcal{F})$ intervenant dans l'écriture du processus limite $\widehat{\mathbb{G}}$. Pour tout $f, g \in \mathcal{F}$,

$$\begin{aligned} \text{Cov}(\mathbb{G}^\nabla(f), \mathbb{G}^\nabla(g)) &= \text{Cov}\left(\sum_{j=1}^m \mathbb{E}[f|A_j]\mathbb{G}(A_j), \sum_{j=1}^m \mathbb{E}[g|A_j]\mathbb{G}(A_j)\right) \\ &= \sum_{j_1, j_2 \leq m} \mathbb{E}[f|A_{j_1}]\mathbb{E}[g|A_{j_2}] (P(A_{j_1} \cap A_{j_2}) - P(A_{j_1})P(A_{j_2})) \\ &= \sum_{j=1}^m P(A_j) \mathbb{E}[f|A_j]\mathbb{E}[g|A_j] - \left(\sum_{j_1=1}^m P(A_{j_1})\mathbb{E}[f|A_{j_1}]\right) \left(\sum_{j_2=1}^m P(A_{j_2})\mathbb{E}[g|A_{j_2}]\right) \\ &= \sum_{j=1}^m P(A_j) \mathbb{E}[f|A_j]\mathbb{E}[g|A_j] - \mathbb{E}[f(X)]\mathbb{E}[g(X)]. \end{aligned}$$

De même,

$$\begin{aligned} \text{Cov}(\mathbb{G}(f), \mathbb{G}^\nabla(g)) &= \text{Cov}\left(\mathbb{G}(f), \sum_{j=1}^m \mathbb{E}[g|A_j]\mathbb{G}(A_j)\right) = \sum_{j=1}^m \mathbb{E}[g|A_j] (\mathbb{E}[f\mathbf{1}_{A_j}] - \mathbb{E}[f(X)]P(A_j)) \\ &= \sum_{j=1}^m P(A_j) \mathbb{E}[f|A_j]\mathbb{E}[g|A_j] - \mathbb{E}[f(X)]\mathbb{E}[g(X)] \\ &= \text{Cov}(\mathbb{G}^\nabla(f), \mathbb{G}^\nabla(g)), \end{aligned} \tag{2.9}$$

qui est en général différent de $\text{Cov}(\mathbb{G}(f), \mathbb{G}(g))$. On est donc à même de donner la fonction de covariance du processus limite $\widehat{\mathbb{G}}(\mathcal{F})$.

$$\begin{aligned} \forall f, g \in \mathcal{F}, \quad \text{Cov}(\widehat{\mathbb{G}}(f), \widehat{\mathbb{G}}(g)) &= \text{Cov}(\mathbb{G}(f), \mathbb{G}(g)) - \text{Cov}(\mathbb{G}^\nabla(f), \mathbb{G}^\nabla(g)) \\ &= \mathbb{E}[f(X)g(X)] - \sum_{j=1}^m P(A_j) \mathbb{E}[f|A_j]\mathbb{E}[g|A_j]. \end{aligned} \tag{2.10}$$

Donc la variance est donnée par

$$\begin{aligned} \forall f \in \mathcal{F}, \quad \text{Var}(\widehat{\mathbb{G}}(f)) &= \text{Var}(\mathbb{G}(f)) - \text{Var}(\mathbb{G}^\nabla(f)) \\ &= \mathbb{E}[f^2(X)] - \sum_{j=1}^m P(A_j) \mathbb{E}[f|A_j]^2. \end{aligned} \tag{2.11}$$

La réduction de variance peut s'exprimer également comme

$$\begin{aligned} \text{Var}(\mathbb{G}(f)) - \text{Var}(\widehat{\mathbb{G}}(f)) &= \sum_{j=1}^m P(A_j) \mathbb{E}[f|A_j]^2 - \mathbb{E}[f(X)]^2 \\ &= \sum_{j=1}^m \frac{\mathbb{E}[f(X)\mathbf{1}_{A_j}]^2}{P(A_j)} - \mathbb{E}[f(X)] \sum_{j=1}^m \mathbb{E}[f(X)\mathbf{1}_{A_j}] \\ &= \sum_{j=1}^m \mathbb{E}[f(X)\mathbf{1}_{A_j}] (\mathbb{E}[f|A_j] - \mathbb{E}[f(X)]). \end{aligned}$$

On remarque que l'ajout de l'information auxiliaire est inutile si toutes les moyennes conditionnelles $\mathbb{E}[f|A_i]$ sont égales à la moyenne $\mathbb{E}[f]$: il faut donc au minimum une information

auxiliaire qui propose des ensembles où la moyenne dans ces ensembles diffère de la moyenne de la population.

Qualité de l'information auxiliaire. Une question naturelle se pose alors, celle de savoir qu'est-ce qui fait qu'une information sera meilleure qu'une autre. La réécriture de la formule (2.11) permet d'apporter un début de réponse. Pour tout $f \in \mathcal{F}$,

$$\text{Var}(\widehat{\mathbb{G}}(f)) = \mathbb{E}[f^2] - \sum_{j=1}^m P(A_j) (\mathbb{E}[f^2|A_j] - \text{Var}(f|A_j)) = \sum_{j=1}^m P(A_j) \text{Var}(f|A_j).$$

En d'autres termes, plus les données sont dispersées autour de la moyenne conditionnelle, moins l'information est bonne. Dans la continuité de la recherche de qualité d'une information auxiliaire, une autre question se pose naturellement dans l'étude de la diminution de la variance du processus empirique avec information auxiliaire d'une partition. La variance est-elle diminuée de manière plus efficace si le nombre de partitions est plus important ? Autrement dit si deux experts utilisent une information auxiliaire et que l'un d'eux dispose de plus d'information, est-ce que ce dernier sera plus efficace que l'autre ? Il semble intuitif d'apporter une réponse affirmative à cette question mais la formulation d'un tel résultat ne doit pas tenir compte uniquement du nombre de partitions apportées par l'information auxiliaire mais du rapport entre ces deux partitionnements. La proposition suivante montre que si l'on dispose de deux informations auxiliaires et que si l'une de ces informations auxiliaires est le raffinement de l'autre, il y a nécessairement diminution de la variance.

Proposition 2.2.6. *Soient $m' \geq m$ et $(B_j)_{1 \leq j \leq m'} \subset \mathcal{T}'$ une partition de \mathcal{X} constituant une autre information auxiliaire. Supposons que B_1, \dots, B_m un raffinement de A_1, \dots, A_m , i.e. $\forall j = 1, \dots, m', \exists i = 1, \dots, m$ tel que $B_j \subset A_i$. Alors*

$$\forall f \in \mathcal{F}, \quad \text{Var}(\widehat{\mathbb{G}}^{\mathcal{B}}(f)) \leq \text{Var}(\widehat{\mathbb{G}}^{\mathcal{A}}(f)),$$

où $\widehat{\mathbb{G}}^{\mathcal{B}}$ est le processus gaussien avec l'information auxiliaire apportée par B_1, \dots, B_m .

Démonstration. D'après (2.11) il suffit de montrer que

$$\sum_{j=1}^{m'} P(B_j) \mathbb{E}[f|B_j]^2 \geq \sum_{j=1}^m P(A_j) \mathbb{E}[f|A_j]^2.$$

Pour $j = 1, \dots, m$, on note $S_j = \{j' = 1, \dots, m' : B_{j'} \subset A_j\}$ alors il suffit de montrer que

$$\forall j = 1, \dots, m, \quad P(A_j) \mathbb{E}[f|A_j]^2 \leq \sum_{j' \in S_j} P(B_{j'}) \mathbb{E}[f|B_{j'}]^2.$$

Par convexité de $x \mapsto x^2$, pour tout $j = 1, \dots, m$

$$\begin{aligned} \sum_{j' \in S_j} P(B_{j'}) \mathbb{E}[f|B_{j'}]^2 &= P(A_j) \left(\sum_{j' \in S_j} \frac{P(B_{j'})}{P(A_j)} \mathbb{E}[f|B_{j'}]^2 \right) \\ &\geq P(A_j) \left(\sum_{j' \in S_j} \frac{P(B_{j'})}{P(A_j)} \mathbb{E}[f|B_{j'}] \right)^2 \\ &= \frac{1}{P(A_j)} \mathbb{E}[f \mathbf{1}_{A_j}]^2 = P(A_j) \mathbb{E}[f|A_j]^2. \end{aligned}$$

□

2.2.4 Loi du logarithme itéré

Le résultat suivant implique que la mesure avec information auxiliaire $\widehat{\mathbb{P}}_n(\mathcal{F})$ vérifie de la même manière que $\mathbb{P}_n(\mathcal{F})$ une loi du logarithme itéré.

Théorème 2.2.7. *Si \mathcal{F} vérifie F.i, VC ou BR alors presque sûrement*

$$\limsup_{n \rightarrow +\infty} \sqrt{\frac{n}{2\sigma_{\mathcal{F}}^2 LL(n)}} \|\widehat{\mathbb{P}}_n - P\|_{\mathcal{F}} \leq mM_{\mathcal{F}} + \sigma_{\mathcal{F}}.$$

Démonstration. On pose $u_n = \sqrt{n/2LL(n)}$ et $\varepsilon > 0$. D'après la loi du logarithme appliquée à \mathbb{P}_n , il existe un rang $n_0 \in \mathbb{N}^*$ tel que pour tout $n \geq n_0$ on ait $u_n \|\mathbb{P}_n - P\|_{\mathcal{F}} \leq \sigma_{\mathcal{F}}(1 + \varepsilon)$. Alors,

$$\begin{aligned} u_n \|\widehat{\mathbb{P}}_n - P\|_{\mathcal{F}} &\leq u_n \|\widehat{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} + u_n \|\mathbb{P}_n - P\|_{\mathcal{F}} \\ &\leq u_n \|\widehat{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} + \sigma_{\mathcal{F}}(1 + \varepsilon). \end{aligned}$$

En utilisant le fait que pour tout ensemble A , $\|\mathbb{P}_n(f\mathbf{1}_A)\|_{\mathcal{F}} \leq M_{\mathcal{F}}\mathbb{P}_n(A)$ on obtient que

$$\begin{aligned} \|\widehat{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} &\leq \sum_{j=1}^m \frac{\|\mathbb{P}_n(f\mathbf{1}_{A_j})\|_{\mathcal{F}}}{\mathbb{P}_n(A_j)} |\mathbb{P}_n(A_j) - P(A_j)| \\ &\leq mM_{\mathcal{F}} \max_{j=1, \dots, m} |\mathbb{P}_n(A_j) - P(A_j)|. \end{aligned}$$

D'après la loi du logarithme itéré classique, il existe un rang $n_1 \in \mathbb{N}^*$ tel que pour tout $n \geq n_1$ on ait $u_n \max_{j=1, \dots, m} |\mathbb{P}_n(A_j) - P(A_j)| \leq 1 + \varepsilon$. Finalement, pour tout $n \geq \max(n_0, n_1)$ on obtient que $u_n \|\widehat{\mathbb{P}}_n - P\|_{\mathcal{F}} \leq (mM_{\mathcal{F}} + \sigma_{\mathcal{F}})(1 + \varepsilon)$. \square

2.3 Résultats non-asymptotiques

2.3.1 Expression du biais

Même si le processus étudié est asymptotiquement sans biais, ce n'est pas le cas pour n fini. En effet, on a apporté des modifications au niveau du processus empirique, qui est un objet sans biais, afin de prendre en compte l'information auxiliaire. Le biais du processus empirique avec information auxiliaire est donné par la proposition suivante.

Proposition 2.3.1. *Pour tout $\forall n \in \mathbb{N}^*$, $\mathbb{E}[\widehat{\alpha}_n(f)] = -\sqrt{n} \sum_{j=1}^m \mathbb{E}[f(X)\mathbf{1}_{\{X \in A_j\}}]P(\overline{A}_j)^n$.*

Pour le démontrer, nous aurons besoin du résultat suivant :

Lemme 2.3.2. *Si $X \sim \mathcal{B}(n; p)$ alors $\mathbb{E}[1/(1+X)] = (1 - (1-p)^{n+1})/((n+1)p)$.*

Démonstration. Soit $X \sim \mathcal{B}(n; p)$ alors

$$\begin{aligned} \mathbb{E}\left[\frac{1}{1+X}\right] &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \times \frac{1}{k+1} = \frac{1}{n+1} \sum_{k=0}^n \binom{n+1}{k+1} p^k (1-p)^{n-k} \\ &= \frac{1}{(n+1)} \sum_{k=1}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} = \frac{1 - (1-p)^{n+1}}{(n+1)p}. \end{aligned}$$

\square

Preuve de la proposition 2.3.1. En reprenant la définition du processus empirique avec information auxiliaire et avec les propriétés de l'espérance conditionnelle on a que

$$\begin{aligned}\mathbb{E}[\widehat{\alpha}_n(f)] &= \mathbb{E} \left[\sqrt{n} \left(\sum_{j=1}^m \frac{P(A_j)}{\sum_{i=1}^n \mathbb{1}_{\{X_i \in A_j\}}} \sum_{i=1}^n f(X_i) \mathbb{1}_{\{X_i \in A_j\}} - \mathbb{E}[f(X)] \right) \right] \\ &= \sqrt{n} \left(\sum_{j=1}^m P(A_j)^2 \left(\sum_{i=1}^n \mathbb{E} \left[\frac{f(X_i)}{\sum_{k=1}^n \mathbb{1}_{\{X_k \in A_j\}}} \middle| X_i \in A_j \right] \right) - \mathbb{E}[f(X)] \right) \\ &= \sqrt{n} \left(\sum_{j=1}^m P(A_j) \mathbb{E}[f(X) \mathbb{1}_{\{X \in A_j\}}] \left(\sum_{i=1}^n \mathbb{E} \left[\frac{1}{1 + B_j} \right] \right) - \mathbb{E}[f(X)] \right),\end{aligned}$$

où $B_j \sim \mathcal{B}(n-1; P(A_j))$ est une variable indépendante de X_i (par indépendance des X). D'après le lemme précédent, $\mathbb{E}[1/(1+B_j)] = (1 - P(\bar{A}_j)^n)/(nP(A_j))$. Alors,

$$\begin{aligned}\mathbb{E}[\widehat{\alpha}_n(f)] &= \sqrt{n} \left(\sum_{j=1}^m \mathbb{E}[f(X) \mathbb{1}_{\{X \in A_j\}}] (1 - P(\bar{A}_j)^n) - \mathbb{E}[f(X)] \right) \\ &= \sqrt{n} \sum_{j=1}^m (\mathbb{E}[f(X) \mathbb{1}_{\{X \in A_j\}}] (1 - P(\bar{A}_j)^n) - \mathbb{E}[f(X) \mathbb{1}_{\{X \in A_j\}}]) \\ &= -\sqrt{n} \sum_{j=1}^m \mathbb{E}[f(X) \mathbb{1}_{\{X \in A_j\}}] P(\bar{A}_j)^n.\end{aligned}$$

□

Si la classe de fonctions \mathcal{F} vérifie l'hypothèse $\mathcal{F}.i$ avec la constante $M_{\mathcal{F}}$ alors le biais est majoré pour tout $f \in \mathcal{F}$ par

$$|\mathbb{E}[\widehat{\alpha}_n(f)]| \leq M_{\mathcal{F}} \sqrt{nm} \left(1 - \min_{j=1, \dots, m} P(A_j) \right)^n.$$

Puisqu'on a supposé que les $P(A_j) \notin \{0, 1\}$, nous retrouvons que le processus empirique est asymptotiquement sans biais et que ce biais converge exponentiellement vers 0.

2.3.2 Divergence de Kullback-Leibler

La divergence de Kullback-Leibler est une mesure entre deux distributions de probabilités P et Q . Dans le cas où \mathcal{X} est un espace discret, cette mesure est définie par

$$d_{\text{KL}}(P||Q) = \sum_{i \in \mathcal{X}} P(\{i\}) \log \left(\frac{P(\{i\})}{Q(\{i\})} \right).$$

Dans le cas où P, Q sont des probabilités admettant pour densités respectives p, q par rapport à une mesure μ , la divergence de Kullback-Leibler est définie par

$$d_{\text{KL}}(P||Q) = \int_{\mathcal{X}} p(x) \log \left(\frac{p(x)}{q(x)} \right) d\mu(x)$$

dans le cas de distributions. La divergence de Kullback-Leibler entre la mesure empirique avec information auxiliaire $\hat{\mathbb{P}}_n$ et la mesure empirique \mathbb{P}_n est donnée par :

$$\begin{aligned} d_{\text{KL}}(\mathbb{P}_n || \hat{\mathbb{P}}_n) &= \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^m \mathbb{1}_{\{X_i \in A_j\}} \log \left(\frac{1/n}{P(A_j)/(n\mathbb{P}_n(A_j))} \right) \\ &= \sum_{j=1}^m \mathbb{P}_n(A_j) \log \left(\frac{\mathbb{P}_n(A_j)}{P(A_j)} \right). \end{aligned}$$

La mesure empirique avec information auxiliaire minimise la divergence de Kullback-Leibler par rapport à la mesure empirique sur l'ensemble des mesures vérifiant l'information auxiliaire. En d'autres termes, elle est la projection de la mesure empirique sur cet ensemble par la divergence de Kullback-Leibler. Ceci est justifiée par le résultat suivant :

Proposition 2.3.3. *La mesure empirique $\hat{\mathbb{P}}_n$ minimise la divergence de Kullback-Leibler $d_{\text{KL}}(\mathbb{P}_n || Q)$ entre \mathbb{P}_n et l'ensemble des probabilités discrètes Q sur X_1, \dots, X_n qui coïncident avec P sur les $A_j, 1 \leq j \leq m$.*

Démonstration. Soit Q une probabilité discrète sur X_1, \dots, X_n affectant à chaque X_i un poids $Q_i, 1 \leq i \leq n$ et coïncidant avec la mesure P sur les $A_j, 1 \leq j \leq m$, i.e. $Q(A_j) = P(A_j)$. Alors

$$d_{\text{KL}}(\mathbb{P}_n || Q) = -\log n - \frac{1}{n} \sum_{i=1}^n \log(Q_i) = -\log n - \frac{1}{n} \sum_{j=1}^m n\mathbb{P}_n(A_j) \sum_{i \in I_j} \frac{1}{n\mathbb{P}_n(A_j)} \log(Q_i),$$

où $I_j = \{1 \leq i \leq n : X_i \in A_j\}$ qui est de cardinal $n\mathbb{P}_n(A_j)$. Avec la concavité du logarithme et le fait que $\sum_{i \in I_j} Q_i = P(A_j)$ on a que

$$\begin{aligned} d_{\text{KL}}(\mathbb{P}_n || Q) &\geq -\log n - \sum_{j=1}^m \mathbb{P}_n(A_j) \log \left(\frac{1}{n\mathbb{P}_n(A_j)} \sum_{i \in I_j} Q_i \right) \\ &\geq \sum_{j=1}^m \mathbb{P}_n(A_j) \log \left(\frac{\mathbb{P}_n(A_j)}{P(A_j)} \right) = d_{\text{KL}}(\mathbb{P}_n || \hat{\mathbb{P}}_n). \end{aligned}$$

□

2.4 Application : test de Kolmogorov-Smirnov

Description du test. Le test de Kolmogorov-Smirnov consiste à tester si des variables aléatoires X_1, \dots, X_n i.i.d. de fonction de répartition continue F^X ont la même fonction de répartition F qu'une loi continue donnée. Plus formellement, l'hypothèse nulle est donnée par

$$(H_0) : F^X = F, \quad (H_1) : F^X \neq F.$$

Ce test exploite la loi asymptotique du supremum du processus empirique uniforme, c'est-à-dire la loi de $\sup_{t \in [0,1]} |\alpha_n^U(t)|$ quand $n \rightarrow +\infty$. La statistique du test est

$$D_n = \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| = \frac{1}{\sqrt{n}} \sup_{t \in \mathbb{R}} |\alpha_n^X(\mathbb{1}_{\cdot \leq t})|,$$

où F_n est la fonction de répartition empirique. L'écart entre la fonction de répartition empirique et la vraie fonction de répartition correspond au processus empirique

$$\alpha_n^X(t) = \alpha_n^X(\mathbb{1}_{\cdot \leq t}) = \sqrt{n}(\mathbb{P}_n^X(\mathbb{1}_{\cdot \leq t}) - F^X(t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbb{1}_{X_i \leq t} - F^X(t)).$$

Sous (H_0) , le processus $(\alpha_n^X(t))_{t \in \mathbb{R}}$ converge en loi dans l'espace de Skorokhod vers le pont brownien $(P(t))_{t \in \mathbb{R}} = (P^U \circ F^X(t))_{t \in \mathbb{R}}$, qui est un processus gaussien centré dont la fonction de covariance est $\text{Cov}(P(t), P(s)) = F(t \wedge s) - F(s)F(t)$ comme rappelée dans l'exemple 1.2.9. En conséquence, la statistique $\sqrt{n}D_n$ converge vers le supremum d'un pont brownien standard. Sous (H_1) , la statistique $\sqrt{n}D_n$ diverge vers $+\infty$ presque sûrement car $\alpha_n^X(t) + \sqrt{n}(F^X - F)(t)$ se comporte, avec n suffisamment grand, comme $P(t) + \sqrt{n}(F^X - F)(t)$.

Motivation. La question est de savoir si l'on peut améliorer asymptotiquement ce test en ajoutant comme information auxiliaire la probabilité que les variables aléatoires X_i soit comprises dans des intervalles qui forment une partition de son support. On pourrait imaginer par exemple qu'une enquête préliminaire à moindre coût a été menée sur un large échantillon mais que la précision de cette étude n'a permis de récupérer que l'intervalle de valeur dans lequel appartenait les variables et non pas les valeurs directement. On aimerait pouvoir exploiter les résultats de cette enquête préliminaire pour améliorer l'enquête qui récolterait directement les valeurs. Asymptotiquement, nous verrons qu'améliorer ce test revient à faire en sorte que la probabilité pour le processus gaussien avec information auxiliaire $(\hat{P}(t))_{t \in [0,1]}$ de dévier devienne plus faible que celle du pont brownien $(P(t))_{t \in [0,1]}$. Nous allons établir dans cette partie la loi exacte du processus gaussien avec information auxiliaire limite et quantifier le gain d'amélioration.

Notation. Posons $-\infty = a_0 < a_1 < \dots < a_{m-1} < a_m = +\infty$ et $A_i = [a_{i-1}, a_i]$ pour tout $j = 1, \dots, m$. On suppose que l'information auxiliaire est donnée par la connaissance des $P(A_i) = P(a_{i-1} < X < a_i) = F^X(a_i) - F^X(a_{i-1})$. La mesure empirique avec information auxiliaire est donnée pour tout $t \in \mathbb{R}$ par

$$\begin{aligned} \hat{F}_n(t) &= \sum_{j=1}^m \frac{P(A_j)}{\mathbb{P}_n(A_j)} \mathbb{P}_n(A_j \cap]-\infty, t]) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{A_j \cap]-\infty, t]}(X_i) \left(\sum_{j=1}^m \frac{nP(A_j)}{\sum_{i=1}^n \mathbb{1}_{A_j}(X_i)} \right). \end{aligned}$$

Le processus empirique avec information auxiliaire $(\hat{\alpha}_n(t))_{t \in \mathbb{R}}$ défini par

$$\begin{aligned} \hat{\alpha}_n(t) &= \sqrt{n}(\hat{F}_n(t) - F^X(t)) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\mathbb{1}_{A_j \cap]-\infty, t]}(X_i) \left(\sum_{j=1}^m \frac{nP(A_j)}{\sum_{i=1}^n \mathbb{1}_{A_j}(X_i)} \right) - F^X(t) \right), \end{aligned}$$

s'annule aux points a_j pour tout $0 \leq j \leq m$ puisque $\hat{F}_n(a_j) = P(A_j)$. D'après la proposition 2.2.3 et l'équation (2.10), le processus $(\hat{\alpha}_n(t))_{t \in \mathbb{R}}$ converge en loi dans l'espace de Skorokhod vers un processus gaussien centré $(\hat{P}(t))_{t \in \mathbb{R}}$ dont la fonction de covariance est

donnée pour $s, t \in \mathbb{R}$ par

$$\begin{aligned} \text{Cov}(\widehat{P}(t), \widehat{P}(s)) &= F^X(s \wedge t) - \sum_{j=1}^m \frac{1}{P(A_j)} \mathbb{E}[\mathbb{1}_{[0,t] \cap A_j}] \mathbb{E}[\mathbb{1}_{[0,s] \cap A_j}] \\ &= F^X(s \wedge t) - \sum_{j=1}^m \frac{1}{F^X(a_j) - F^X(a_{j-1})} \mathbb{E}[\mathbb{1}_{[0,t] \cap [a_{j-1}, a_j]}] \mathbb{E}[\mathbb{1}_{[0,s] \cap [a_{j-1}, a_j]}]. \end{aligned}$$

La statistique du test amélioré sera donnée par

$$\widehat{D}_n = \sup_{t \in \mathbb{R}} |\widehat{F}_n(t) - F(t)|,$$

et en particulier sous (H_0) , la statistique $\sqrt{n}\widehat{D}_n$ converge en loi vers la variable $\sup_{t \in \mathbb{R}} |\widehat{P}(t)|$.

Loi du processus gaussien limite. Le comportement du processus $(\widehat{P}(t))_{t \in \mathbb{R}}$ avec ce type d'information auxiliaire est celui de plusieurs ponts browniens indépendants s'annulant aux points a_j pour $1 \leq j \leq m$. Le résultat suivant donne la loi du supremum du pont brownien avec information auxiliaire donnée par les $P(A_j)$.

Proposition 2.4.1. *Pour tout $\varepsilon > 0$,*

$$\mathbb{P} \left(\sup_{t \in \mathbb{R}} |\widehat{P}(t)| \geq \varepsilon \right) = 1 - \prod_{j=1}^m \left(1 - 2 \sum_{k \geq 1} (-1)^{k-1} e^{-2k^2 \varepsilon^2 / (F^X(a_j) - F^X(a_{j-1}))} \right). \quad (2.12)$$

Démonstration. Soient $s, t \in \mathbb{R}$ avec $t \in A_{j_1}, s \in A_{j_2}$. Si $j_1 \neq j_2$ alors $\text{Cov}(\widehat{P}(t), \widehat{P}(s)) = 0$, il y a donc indépendance des processus $(\widehat{P}(t))_{t \in A_j}$ restreints sur chaque bloc formé par les A_j . Si $j_1 = j_2$ alors

$$\text{Cov}(\widehat{P}(t), \widehat{P}(s)) = F^X(s \wedge t) - F^X(a_{j-1}) - \frac{(F^X(t) - F^X(a_{j-1}))(F^X(s) - F^X(a_{j-1}))}{F^X(a_j) - F^X(a_{j-1})}.$$

On vient de montrer que $(\widehat{P}(t))_{t \in A_j} \stackrel{\mathcal{L}}{=} (Y^{(j)}(t))_{t \in A_j}$ avec

$$\begin{aligned} (Y^{(j)}(t))_{t \in A_j} &\stackrel{\mathcal{L}}{=} \left(B(F^X(t) - F^X(a_{j-1})) - \frac{F^X(t) - F^X(a_{j-1})}{F^X(a_j) - F^X(a_{j-1})} B(F^X(a_j) - F^X(a_{j-1})) \right)_{t \in A_j} \\ &\stackrel{\mathcal{L}}{=} \sqrt{F^X(a_j) - F^X(a_{j-1})} \left(B \left(\frac{F^X(t) - F^X(a_{j-1})}{F^X(a_j) - F^X(a_{j-1})} \right) - \frac{F^X(t) - F^X(a_{j-1})}{F^X(a_j) - F^X(a_{j-1})} B(1) \right)_{t \in A_j} \\ &\stackrel{\mathcal{L}}{=} \sqrt{F^X(a_j) - F^X(a_{j-1})} (P(t))_{t \in \mathbb{R}}, \end{aligned}$$

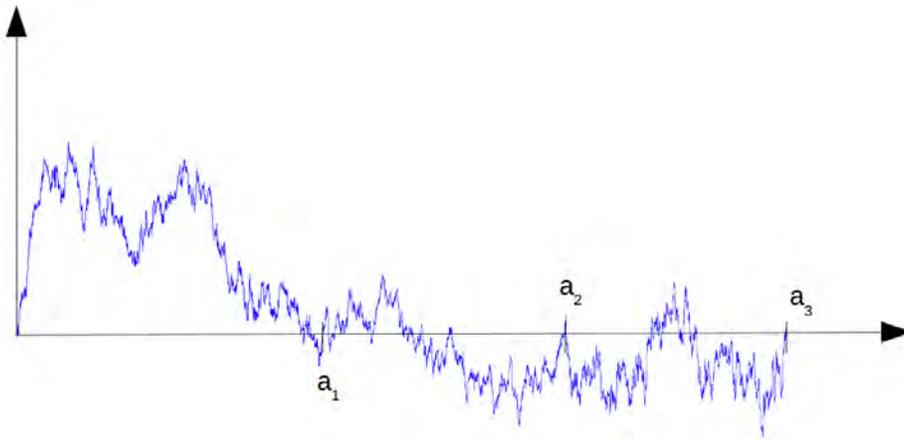
où $(B_t)_{t \geq 0}$ désigne le mouvement brownien. Le processus limite $(\widehat{P}(t))_{t \in \mathbb{R}}$ est donc la concaténation de m ponts browniens indépendants. Par indépendance du processus sur chaque A_j et en utilisant la loi du supremum d'un pont brownien rappelée par (1.2), la loi du supremum

du processus $(\widehat{P}(t))_{t \in \mathbb{R}}$ est donnée par

$$\begin{aligned} \mathbb{P} \left(\sup_{t \in \mathbb{R}} |\widehat{P}(t)| \geq \varepsilon \right) &= 1 - \prod_{j=1}^m \left(1 - \mathbb{P} \left(\sup_{t \in A_j} |\widehat{P}(t)| \geq \varepsilon \right) \right) \\ &= 1 - \prod_{j=1}^m \left(1 - \mathbb{P} \left(\sup_{t \in A_j} |Y^{(j)}(t)| \geq \varepsilon \right) \right) \\ &= 1 - \prod_{j=1}^m \left(1 - \mathbb{P} \left(\sup_{t \in \mathbb{R}} |P(t)| \geq \frac{\varepsilon}{\sqrt{F^X(a_j) - F^X(a_{j-1})}} \right) \right), \end{aligned}$$

qui conduit à (2.12) grâce à (1.2). □

FIG. 2.1: Illustration de $(\widehat{P}(t))_{t \in \mathbb{R}}$



Remarque 2.4.1. La fonction $\varepsilon \mapsto 2 \sum_{k \geq 1} (-1)^{k-1} e^{-2k^2 \varepsilon^2}$ étant décroissante, la proposition précédente implique en particulier que pour tout $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P} \left(\sup_{t \in \mathbb{R}} |\widehat{P}(t)| \geq \varepsilon \right) &\leq \min_{j=1, \dots, m} \left(2 \sum_{k \geq 1} (-1)^{k-1} e^{-2k^2 \left(\varepsilon / \sqrt{F^X(a_j) - F^X(a_{j-1})} \right)^2} \right) \\ &= 2 \sum_{k \geq 1} (-1)^{k-1} e^{-2k^2 \left(\varepsilon / \min_{j=1, \dots, m} \sqrt{F^X(a_j) - F^X(a_{j-1})} \right)^2} \\ &\leq 2 \sum_{k \geq 1} (-1)^{k-1} e^{-2k^2 \varepsilon^2} = \mathbb{P} \left(\sup_{t \in \mathbb{R}} |P(t)| \geq \varepsilon \right). \end{aligned}$$

Amélioration du test de Kolmogorov-Smirnov. On note θ_4 la fonction theta de Jacobi, c'est-à-dire la fonction définie par

$$\theta_4(z, q) = \sum_{n=-\infty}^{+\infty} (-1)^n q^{n^2} e^{2niz}.$$

D'après (1.2) les lois du supremum des processus $(P(t))_{t \in \mathbb{R}}$ et $(\hat{P}(t))_{t \in \mathbb{R}}$ peuvent se réécrire avec cette fonction comme suit. Pour tout $\varepsilon > 0$,

$$\begin{aligned}\Phi_1(\varepsilon) &= \mathbb{P}\left(\sup_{t \in \mathbb{R}} |P(t)| > \varepsilon\right) = 1 - \theta_4(0, e^{-2\varepsilon^2}), \\ \Phi_2(\varepsilon) &= \mathbb{P}\left(\sup_{t \in \mathbb{R}} |\hat{P}(t)| > \varepsilon\right) = 1 - \prod_{j=1}^m \theta_4\left(0, e^{-2\varepsilon^2/(F^X(a_j) - F^X(a_{j-1}))}\right).\end{aligned}\quad (2.13)$$

Les fonctions Φ_1 et Φ_2 sont des bijections décroissantes de \mathbb{R}^+ dans $[0, 1]$ vérifiant $\Phi_2(\varepsilon) \leq \Phi_1(\varepsilon)$ pour tout $\varepsilon > 0$. Avec un logiciel de calcul il est possible de calculer l'inverse de ces deux fonctions. La fonction Φ_2 est plus difficile à inverser du fait de son écriture. On pourra néanmoins utiliser le fait que d'après la remarque 2.4.1,

$$\Phi_2(\varepsilon) \leq \Phi_1\left(\varepsilon / \min_{j=1, \dots, m} \sqrt{F^X(a_j) - F^X(a_{j-1})}\right),$$

alors pour tout $\alpha \in (0, 1)$,

$$\Phi_2^{-1}(\alpha) \leq \Phi_1^{-1}(\alpha) \min_{j=1, \dots, m} \sqrt{F^X(a_j) - F^X(a_{j-1})} \leq \Phi_1^{-1}(\alpha).$$

Le test de Kolmogorov-Smirnov qui consistait à rejeter l'hypothèse nulle quand $\sqrt{n}D_n > \Phi_1^{-1}(\alpha)$ au seuil α peut être remplacé par le test qui rejette l'hypothèse nulle quand la nouvelle statistique avec information auxiliaire \hat{D}_n vérifie $\sqrt{n}\hat{D}_n > \Phi_2^{-1}(\alpha)$.

Exemple. Supposons que l'on connaisse les a_j tels que $F^X(a_j) = j/m$, c'est-à-dire que l'on connaît les quantiles de X d'ordre j/m avec $1 \leq j < m$. Dans ce cas on a en particulier que pour tout $\varepsilon > 0$,

$$\Phi_2(\varepsilon) = 1 - \left(1 - 2 \sum_{k \geq 1} (-1)^{k-1} e^{-2k^2 \varepsilon^2}\right)^m = 1 - \theta_4(0, e^{-2m\varepsilon^2})^m.$$

Les tableaux 2.1 et 2.2 donnent les valeurs des seuils $\Phi_1^{-1}(\alpha)$ et $\Phi_2^{-1}(\alpha)$ pour quelques valeurs de α et de m . Malheureusement, il n'existe pas de lien simple et exploitable entre $\Phi_1^{-1}(\alpha)$ et $\Phi_2^{-1}(\alpha)$.

Résultat. La proposition suivante montre qu'en exploitant l'information auxiliaire apportée par la connaissance de la probabilité des ensembles A_j alors ce nouveau test améliore le test de Kolmogorov-Smirnov.

Proposition 2.4.2. *Supposons que (H_1) est vérifiée. Alors pour tout $\alpha > 0$, il existe presque sûrement $n_1 \geq n_2$ tel que pour tout $n > n_1$, $|D_n| > \Phi_1^{-1}(\alpha)$ et pour tout $n > n_2$, $|\hat{D}_n| > \Phi_2^{-1}(\alpha)$.*

Démonstration. D'après la loi des supremums de $(P(t))_{t \in \mathbb{R}}$ et $(\hat{P}(t))_{t \in \mathbb{R}}$ donnée par (2.13),

$$\begin{aligned}\Phi_1(\varepsilon) &= \mathbb{P}\left(\sup_{t \in \mathbb{R}} |P(t)| > \varepsilon\right) \leq 2e^{-2\varepsilon^2}, \\ \Phi_2(\varepsilon) &= \mathbb{P}\left(\sup_{t \in \mathbb{R}} |\hat{P}(t)| > \varepsilon\right) \leq 2e^{-2M\varepsilon^2},\end{aligned}$$

α	0.05	0.1	0.15	0.2	0.3	0.4	0.5
	1.3581 $\Phi_1^{-1}(\alpha)$	1.22385	1.13795	1.07275	0.973063	0.894764	0.827574

TAB. 2.1: Valeurs de $\Phi_1^{-1}(\alpha)$

$\alpha \backslash m$	2	3	4	5	10	20	50	100
0.05	1.04514	0.891674	0.794988	0.726492	0.546282	0.408055	0.275243	0.203332
0.1	0.956931	0.82256	0.736796	0.675535	0.512414	0.385418	0.261851	0.194283
0.15	0.900493	0.778424	0.6997	0.643098	0.490959	0.371144	0.253451	0.188628
0.2	0.857519	0.744828	0.671484	0.618445	0.4747	0.360361	0.24713	0.184382
0.3	0.791284	0.692985	0.627948	0.580425	0.449691	0.343828	0.23748	0.177919
0.4	0.738507	0.651528	0.593099	0.549986	0.429702	0.330656	0.229826	0.172809
0.5	0.692445	0.615152	0.562455	0.523196	0.412108	0.319087	0.223128	0.168349

TAB. 2.2: Valeurs de $\Phi_2^{-1}(\alpha)$

où $M = \max_{j=1, \dots, m} (F^X(a_j) - F^X(a_{j-1}))$. D'après Borel-Cantelli, presque sûrement pour n suffisamment grand,

$$\sup_{t \in \mathbb{R}} |P(t)| \leq \sqrt{\log n}, \quad \sup_{t \in \mathbb{R}} |\hat{P}(t)| \leq \sqrt{\log n / M}.$$

Cela entraîne que

$$|D_n| \geq \sqrt{n} \sup_{t \in \mathbb{R}} |(F^X - F)(t)| - \sqrt{\log n}, \quad |\hat{D}_n| \geq \sqrt{n} \sup_{t \in \mathbb{R}} |(F^X - F)(t)| - \sqrt{\frac{\log n}{M}}.$$

Le résultat est vérifié pour n_1, n_2 vérifiant

$$\begin{aligned} \sqrt{\frac{n}{\log n}} &\geq \frac{1}{\sup_{t \in \mathbb{R}} |F^X(t) - F(t)|} \left(1 + \frac{\Phi_1^{-1}(\alpha)}{\sqrt{\log n}} \right), \\ \sqrt{\frac{n}{\log n}} &\geq \frac{1}{\sup_{t \in \mathbb{R}} |F^X(t) - F(t)|} \left(\frac{1}{M} + \frac{\Phi_2^{-1}(\alpha)}{\sqrt{\log n}} \right). \end{aligned}$$

Compte tenu du fait que $m > 1$ et $\Phi_2^{-1}(\alpha) < \Phi_1^{-1}(\alpha)$ on peut prendre $n_1 \geq n_2$. \square

Méthode du Raking-Ratio

La partie précédente a été l'occasion d'étudier le comportement du processus empirique lorsqu'on injecte à celui-ci l'information donnée par la probabilité des ensembles d'une seule partition. La transformation du processus empirique qui a été étudiée consistait à substituer les poids $1/n$ de chaque individu d'un échantillon de taille n par des poids aléatoires de telle sorte que la somme des poids des individus appartenant à l'un des ensembles d'une partition corresponde à la probabilité connue qu'un individu appartienne à cet ensemble. Ce procédé ne peut pas s'appliquer si on connaît la probabilité d'ensembles de plusieurs partitions puisqu'il n'y a aucune raison que l'on connaisse la probabilité des intersections possibles parmi ces ensembles. Or c'est justement une situation courante que le statisticien a à sa disposition la connaissance de la probabilité d'ensembles de plusieurs partitions (étude préliminaire sur un échantillon plus grand et indépendant, big data, connaissance d'un expert, ...). Une question naturelle est de savoir comment combiner toute cette information, le statisticien pouvant ne pas avoir à sa disposition les probabilités des intersections des événements ou bien les événements n'étant pas forcément indépendants. Des méthodes statistiques ont été mises au point afin de transformer le processus empirique dans le but d'utiliser toute l'information connue. On présente et étudie dans cette partie la méthode du Raking-Ratio.

3.1 Introduction

3.1.1 Description de la méthode

Motivation. La méthode du Raking-Ratio ou méthode de ratissage est une technique itérative couramment utilisée dans la pratique (en statistiques, informatique, économie, ...) et qui a pour but d'exploiter au niveau d'un échantillon l'information auxiliaire donnée par la connaissance de plusieurs lois marginales d'une variable discrète multivariée. Historiquement et malgré la littérature autour de ce sujet, cette méthode ne semble pas avoir été traitée du point de vue des processus empiriques. Dans le but d'étudier la méthode du Raking-Ratio il est nécessaire de voir que cette dernière fait justement intervenir le processus empirique en lui faisant subir itérativement des modifications afin de combiner l'information auxiliaire. Nous nommerons ce processus modifié, processus empirique associé à la méthode du Raking-

Ratio ou bien processus empirique raké. Quand la taille de l'échantillon sera suffisamment grande, nous pourrons approcher fortement le processus empirique raké par un processus gaussien dont on souhaite connaître les principales propriétés : variance, covariance, biais, ... Les propriétés du processus empirique associé à la méthode du Raking-Ratio pour une taille d'échantillon suffisamment grande seront proches de celles du processus gaussien limite avec une erreur que l'on souhaite majorer par une borne convergeant rapidement vers 0. On souhaite également établir des résultats non-asymptotiques pour le processus empirique raké. L'étude de ce processus se fera sur une classe de fonctions afin de conserver un cadre très général. Nous montrerons que les estimateurs qui font intervenir le processus empirique voient leur variance diminuer si on substitue le processus empirique par le processus empirique raké.

L'algorithme. L'algorithme associé à la méthode du Raking-Ratio a été introduit par Deming et Stephan [30]. L'idée de départ était de déterminer la projection d'un tableau de fréquence initiale, c'est-à-dire la projection de la mesure empirique, au sens des moindres carrés, vers une mesure satisfaisant les marges connues. La partie 5 de leur papier détaille le fonctionnement de cet algorithme : celui-ci part d'un tableau initial de fréquence qui affecte un poids $1/n$ à chaque donnée et rectifie les effectifs de ce tableau au tour-à-tour afin de vérifier une marginale donnée. Nous dirons alors par la suite que nous avons « raké » par rapport à une marge donnée. On vérifie les marges données périodiquement, une marge seulement étant vérifiée à la fois à chaque tour. L'algorithme se stabilise après un nombre acceptable d'étapes vers un tableau vérifiant toutes les marges connues. Nous donnons un exemple numérique dans le paragraphe suivant.

Exemple de l'algorithme. Supposons que l'on observe dans un échantillon la table de fréquences suivante, dont les cellules contiennent les valeurs de la mesure empirique $\mathbb{P}_n(\mathbb{1}_{A_i^{(1)} \cap A_j^{(2)}})$ pour $i = 1, 2$ et $j = 1, 2, 3$.

	$A_1^{(2)}$	$A_2^{(2)}$	$A_3^{(2)}$	Total	Total attendu
$A_1^{(1)}$	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	
Total Attendu	0.31	0.4	0.29		

Les différences entre les fréquences observées et les fréquences attendues sont dues à l'effet du hasard. L'algorithme rectifie dans un premier temps par exemple les fréquences connaissant les marges des lignes. Concrètement dans notre cas, pour que la somme des fréquences de la première ligne (catégorie $A_1^{(1)}$) fasse 0.52, on multiplie chaque case par $0.52/0.55$ et de la même manière on multiplie chaque case de la seconde ligne par $0.48/0.45$. Le tableau devient (en arrondissant) :

	$A_1^{(2)}$	$A_2^{(2)}$	$A_3^{(2)}$	Total	Total Attendu
$A_1^{(1)}$	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	
Total Attendu	0.31	0.4	0.29		

On effectue les mêmes opérations mais par rapport aux marges des colonnes. On multiplie ainsi les cases des colonnes respectivement par $0.31/0.299$, $0.4/0.446$ et $0.29/0.255$. On obtient alors la correction suivante :

	$A_1^{(2)}$	$A_2^{(2)}$	$A_3^{(2)}$	Total	Total Attendu
$A_1^{(1)}$	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	
Total Attendu	0.31	0.4	0.29		

Ces deux dernières étapes sont répétées jusqu'à ce que le tableau se stabilise. Au bout de 7 itérations seulement l'algorithme se stabilise et on obtient le tableau final suivant (en arrondissant à 10^{-3}) :

	$A_1^{(2)}$	$A_2^{(2)}$	$A_3^{(2)}$	Total	Total Attendu
$A_1^{(1)}$	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	
Total Attendu	0.31	0.4	0.29		

Les effectifs finaux vérifient les deux marges connues. Il est possible d'appliquer ce genre d'algorithme avec plus que deux marges connues. Cette méthode est élémentaire : il est facile de programmer un tel calcul avec des logiciels simples comme un tableur. Étant couramment utilisé en statistiques, cet outil est déjà implémenté dans les principaux langages statistiques : SAS (CALMAR), R (icarus).

Limite. Deming et Stephan [30] affirment que cette méthode qu'ils nomment « iterative proportions » fait converger un tableau de fréquences vers la solution des moindres carrés sur l'ensemble des tableaux de fréquences vérifiant les contraintes de marge. Cette affirmation est erronée. En effet, Stephan [76] montre que la méthode proposée dans le papier d'origine donne seulement une approximation de la solution des moindres carrés et non pas la solution exacte. Il propose une modification de l'algorithme pour corriger cela. Lewis [57] et Brown [23] ont étudié la minimisation avec la distance de Shannon-Kullback dans le cas de variables de Bernoulli. Brown utilise en particulier l'algorithme du Raking-Ratio afin d'atteindre cette minimisation. La méthode du Raking-Ratio dans le cas où on rake périodiquement deux partitions peut être vu comme une matrice stochastique double dont on normalise les lignes et les colonnes au tour-à-tour afin de respecter alternativement l'information apportée par ces partitions. Sinkhorn [74, 75] montre que l'on peut décomposer de manière unique ces matrices avec deux matrices diagonales et prouve que l'algorithme converge nécessairement si les coefficients sont strictement positifs. Plus concrètement, Ireland et Kullback [50] apportent les arguments nécessaires pour justifier que cet algorithme converge vers l'unique projection de la mesure empirique au sens de la divergence de Kullback-Leibler sur l'ensemble des mesures discrètes de l'échantillon vérifiant les contraintes apportées par l'information auxiliaire.

3.1.2 Cadre de travail

Notation. On note $(\mathcal{X}, \mathcal{A})$ un espace mesurable. Pour $N \in \mathbb{N}^*$ on note $(m_N)_{N \in \mathbb{N}^*}$ le nombre d'ensembles donnés par une partition et $\mathcal{A}^{(N)} = \{A_1^{(N)}, \dots, A_{m_N}^{(N)}\} \subset \mathcal{A}$ les partitions de \mathcal{X} qui nous servent d'informations auxiliaires. Autrement dit, on connaît les probabilités $P(A_j^{(N)})$ pour tout $N \in \mathbb{N}^*$ et $1 \leq j \leq m_N$. Par convention, on note $m_0 = 1$ et $\mathcal{A}^{(0)} = \{\mathcal{X}\}$. Nous supposons que la probabilité d'appartenir à un ensemble d'une partition n'est pas nulle, autrement dit

$$\forall N \in \mathbb{N}^*, \quad \delta_{(N)} = \min_{k=1, \dots, N} \delta_k > 0 \quad \text{où} \quad \delta_N = \min_{j \leq m_N} P(A_j^{(N)}).$$

Supposons que nous travaillons avec un tableau de fréquence. L'algorithme du Raking-Ratio appliqué à des marges successives consiste à modifier tour-à-tour les fréquences afin de les rapprocher au mieux des marges connues. Notons $f^{(N)}(A)$ les fréquences d'un ensemble $A \in \mathcal{A}$ après la N ème modification de l'algorithme. Celui-ci consiste à multiplier au tour N les éléments de la catégorie $A_j^{(N)}$ par le rapport $P(A_j^{(N)})/f^{(N)}(A_j^{(N)})$. Au tour $N + 1$ la méthode du Raking-Ratio applique la relation de récurrence suivante :

$$f^{(N+1)}(A) = \sum_{j=1}^{m_{N+1}} f^{(N)}(A \cap A_j^{(N+1)}) \frac{P(A_j^{(N+1)})}{f^{(N)}(A_j^{(N+1)})}. \quad (3.1)$$

Pour généraliser cette méthode du point de vue de la théorie des processus empiriques, nous allons créer une suite de mesures empiriques $\mathbb{P}_n^{(N)}(\mathcal{F}) = \{\mathbb{P}_n^{(N)}(f) : f \in \mathcal{F}\}$ indexées par la classe de fonctions \mathcal{F} qui correspond aux opérations effectuées successivement. Nous l'appelons mesure empirique associée à la méthode du Raking-Ratio ou bien mesure empirique rakée. Si \mathcal{F} vérifie l'hypothèse VC ou BR alors la classe de fonction

$$\mathcal{G} = \mathcal{F} \cup \left\{ f \mathbb{1}_{A_j^{(N)}} : f \in \mathcal{F}, N \in \mathbb{N}^*, 1 \leq j \leq m_N \right\}$$

vérifie la même condition d'entropie avec des nouvelles constantes c_0 ou b_0 . Nous supposons donc sans perte de généralité pour la suite de cette partie que $\mathcal{F} = \mathcal{G}$. La mesure initiale doit correspondre au tableau initial, *i.e.* les fréquences observées. On définit donc la mesure initiale par la mesure empirique classique, c'est-à-dire $\mathbb{P}_n^{(0)} = \mathbb{P}_n$. En s'inspirant de (3.1), la mesure empirique rakée est définie par récurrence pour tout $f \in \mathcal{F}$ par

$$\mathbb{P}_n^{(N+1)}(f) = \sum_{j \leq m_{N+1}} \mathbb{P}_n^{(N)}(f \mathbb{1}_{A_j^{(N+1)}}) \frac{P(A_j^{(N+1)})}{\mathbb{P}_n^{(N)}(A_j^{(N+1)})}. \quad (3.2)$$

Pour $A \in \mathcal{A}$, nous utilisons la notation suivante $\mathbb{P}_n^{(N)}(A) = \mathbb{P}_n^{(N)}(\mathbb{1}_A)$. Au tour N cette mesure vérifie l'information auxiliaire donnée par la N ème marge $\mathcal{A}^{(N)}$ puisque

$$\forall N \in \mathbb{N}^*, \quad \forall j = 1, \dots, m_N, \quad \mathbb{P}_n^{(N)}(A_j^{(N)}) = P(A_j^{(N)}). \quad (3.3)$$

Dans le cas $N = 1$, on retrouve la mesure empirique avec l'information auxiliaire apportée par une partition étudiée dans la partie précédente. Le Raking-Ratio est donc une généralisation de notre étude précédente. On définit le processus empirique associé à la méthode du Raking-Ratio $\alpha_n(\mathcal{F}) = \{\alpha_n^{(N)}(f) : f \in \mathcal{F}\}$ pour tout $f \in \mathcal{F}$ par

$$\alpha_n^{(N)}(f) = \sqrt{n}(\mathbb{P}_n^{(N)}(f) - P(f)). \quad (3.4)$$

En particulier, $\alpha_n^{(0)} = \alpha_n$. Nous notons également pour $A \in \mathcal{A}$, $\alpha_n^{(N)}(A) = \alpha_n^{(N)}(\mathbb{1}_A)$. D'après (3.3), on a immédiatement qu'au tour N ce processus vérifie l'information auxiliaire en s'annulant en les fonctions indicatrices des ensembles appartenant à la partition $\mathcal{A}^{(N)}$, *i.e.*

$$\forall N \in \mathbb{N}^*, \quad \forall j = 1, \dots, m_N, \quad \alpha_n^{(N)}(A_j^{(N)}) = 0.$$

Mélange de processus empiriques conditionnels. On introduit pour tout $A \in \mathcal{A}$ tel que $\mathbb{P}_n^{(N)}(A) > 0$, la notation $\mathbb{E}_n^{(N)}[f|A] = \mathbb{P}_n^{(N)}(f\mathbb{1}_A)/\mathbb{P}_n^{(N)}(A)$. On peut alors écrire que

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

avec $\alpha_{n,j}^{(N-1)}(f) = \sqrt{n} \left(\mathbb{E}_n^{(N-1)}[f|A_j^{(N)}] - \mathbb{E}[f|A_j^{(N)}] \right)$. Chaque processus $\alpha_{n,j}^{(N-1)}$ correspond au processus empirique $\alpha_n^{(N-1)}$ conditionné sur l'ensemble $A_j^{(N)}$ de la partition $\mathcal{A}^{(N)}$ à raker au tour N . La combinaison linéaire de ces derniers avec les poids $P(A_j^{(N)})$ constitue $\alpha_n^{(N)}$.

3.1.3 Motivation

Biais et variance non explicite. La mesure initiale avant application du Raking-Ratio est la mesure empirique. Il en résulte que les mesures $\mathbb{P}_n^{(N)}$ apportées par la méthode dévient de la vraie mesure de probabilité et ce de manière de plus en plus complexe à chaque tour de la méthode. De fait, le vecteur de biais et la matrice de covariance des poids aléatoires $\mathbb{P}_n^{(N)}(\{X_i\})$ que l'algorithme affecte au N -ème tour à chaque individu sont délicats à exprimer pour n fini. Cette remarque nous conforte dans l'idée qu'il faille étudier cette méthode d'un point de vue asymptotique afin d'obtenir l'expression exacte du biais et de la covariance asymptotique. Il est nécessaire d'obtenir aussi une borne d'erreur entre le biais, variance asymptotique et non-asymptotique en fonction de la taille d'échantillon n de limite nulle.

Problème concernant le biais. De la même manière que dans le cas de l'information auxiliaire apportée par une partition, le fait de modifier le processus empirique initial afin que celui-ci utilise notre information a priori lui rajoute du biais. En effet, les processus $\alpha_{n,j}^{(N-1)}$ ne sont pas centrés du fait de la présence des facteurs aléatoires $1/\mathbb{P}_n^{(N-1)}(A_j^{(N)})$ dans la définition de ceux-ci. De manière plus formelle pour $f \neq \mathbb{1}_{A_j^{(N-1)}}$, $j = 1, \dots, m_{N-1}$, on a

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

Nous espérons d'une part avoir un biais asymptotiquement nul et d'autre part compenser l'introduction de ce biais par une diminution de la variance. Par la définition (3.2), nous voyons que le biais se propage car l'information auxiliaire donnée par les $(P(A_j^{(N+1)}))_{j=1, \dots, m_{N+1}}$ est appliquée à la mesure biaisée $\mathbb{P}_n^{(N)}$ au lieu de la mesure empirique non biaisée \mathbb{P}_n . Pour des classes de fonctions raisonnables, nous allons montrer que le biais converge vers 0 avec une vitesse plus rapide que $1/\sqrt{n}$.

Problème concernant la variance. L'algorithme produit à la fin des estimateurs de fréquence vérifiant les marges imposées. L'objectif est d'obtenir des estimateurs avec une variance plus faible que l'estimateur classique fourni par la mesure empirique. Certains ont donné une expression de la variance après un nombre fini d'étapes. Dans cette littérature on retrouve Bracktone et Rao en 1979[21] avec l'étude de 4 étapes, Konijn en 1981[53] et Choudhry et Lee en 1987[27] avec 2 étapes ou Bankier en 1986 avec un nombre arbitraire d'étapes. L'outil classique est la linéarisation de Taylor. D'autres ont donné une estimation

de la variance asymptotique complexe à utiliser en pratique (*cf.* Binder et Théberge en 1988[15]). Finalement, Deville et Särndal [32, 33] ont étudié à population finie une méthode appelée « calibration » faisant intervenir une fonction de distance que le statisticien peut choisir : l'une de ces distances permet de retrouver la méthode du Raking-Ratio. Ils ont montré qu'asymptotiquement le choix de cette distance importait peu et ont pu établir un estimateur de la variance asymptotique. Aucun résultat cité jusqu'à présent ne donne une expression exacte et simple de la variance asymptotique au N -ème tour de l'algorithme. Un des objectifs de cette partie est de trouver l'expression exacte de la variance asymptotique de $\alpha_n^{(N)}$ et de montrer que pour n suffisamment grand, $\text{Var}(\alpha_n^{(N)}(f)) \leq \text{Var}(\alpha_n^{(0)}(f))$. Si certaines conditions simples et naturelles sont respectées, nous montrons également que, asymptotiquement, $\text{Var}(\alpha_n^{(N_1)}(f)) \leq \text{Var}(\alpha_n^{(N_2)}(f))$ pour certains $N_1 > N_2$. Néanmoins pour des petites valeurs de n le risque quadratique risque d'augmenter, *i.e.* $\mathbb{E}[(\mathbb{P}_n^{(N)}(f) - P(f))^2] > \mathbb{E}[(\mathbb{P}_n^{(0)}(f) - P(f))^2]$ à cause de l'introduction du biais. Nous montrons que ce n'est pas le cas asymptotiquement.

Organisation. Nous étudions les propriétés de la mesure et du processus empirique rakés dans un premier temps. La partie 3.2 regroupe les propriétés et résultats non-asymptotiques de ces objets tandis que la partie 3.3 étudie leur comportement asymptotique. Nous étudierons le cas spécial où on applique la méthode du Raking-Ratio sur deux partitions en boucle. Nous renforçons les propriétés non-asymptotique de la mesure et du processus empirique rakés dans la partie 3.4 et nous donnerons plus de précision du comportement asymptotique dans la partie 3.5. La section 3.6 contient quelques simulations numériques ainsi que des résultats pour mettre en pratique les résultats établis à la partie 3.5. Nous finirons par étudier dans la section 3.7 le comportement de la mesure et du processus empirique raké dans le cas où l'information auxiliaire n'est pas exacte mais est donnée par une estimation provenant d'une source auxiliaire.

3.2 Résultats non-asymptotiques

3.2.1 Mesure empirique rakée

Probabilité de définition. Posons pour tout $n, N_0 \in \mathbb{N}^*$ l'ensemble

$$B_{n, N_0} = \left\{ \min_{0 \leq N \leq N_0} \min_{1 \leq j \leq m_N} \mathbb{P}_n(A_j^{(N)}) > 0 \right\}. \quad (3.5)$$

La probabilité que $\alpha_n^{(N_0)}$ n'est pas défini est donnée par

$$P(B_{n, N_0}^C) = P \left(\bigcup_{N=1}^{N_0} \bigcup_{j=1}^{m_N} \bigcap_{i=1}^n \{X_i \notin A_j^{(N)}\} \right) \leq \sum_{N=1}^{N_0} m_N (1 - \delta_N)^n.$$

Pour $N_0 = 1$ cette probabilité est celle de la relation (2.2).

Poids des individus. La mesure associée à la méthode du Raking-Ratio affecte à chaque $X_i \in A_j^{(N)}$ le poids aléatoire

$$\mathbb{P}_n^{(N)}(\{X_i\}) = \mathbb{P}_n^{(N-1)}(\{X_i\}) \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})}.$$

La proposition suivante donne le poids que la mesure empirique associée à la méthode du Raking-Ratio $\mathbb{P}_n^{(N)}(\mathcal{F})$ attribue pour chaque $f(X_i)$ en fonction des ensembles auxquels X_i appartient.

Proposition 3.2.1. *Soit $N \in \mathbb{N}^*$. Si $X_i \in A_{j_1}^{(1)} \cap \dots \cap A_{j_N}^{(N)}$ avec $1 \leq j_i \leq m_i$ pour tout $1 \leq i \leq N$ alors $\mathbb{P}_n^{(N)}$ affecte à $f(X_i)$ le poids $\frac{1}{n} \prod_{k=1}^N P(A_{j_k}) / \mathbb{P}_n^{(k-1)}(A_{j_k}^{(k)})$.*

Démonstration. Le cas $N = 1$ correspond au processus empirique avec information auxiliaire donnée par une partition qui affecte un poids aléatoire $P(A_j^{(1)})/n\mathbb{P}_n(A_j^{(1)})$ aux $X_i \in A_j^{(1)}$. Supposons que cette proposition soit vraie jusqu'au rang N alors

$$\begin{aligned} \mathbb{P}_n^{(N+1)}(f) &= \sum_{j_{N+1}=1}^{m_{N+1}} \frac{P(A_{j_{N+1}}^{(N+1)})}{\mathbb{P}_n^{(N)}(A_{j_{N+1}}^{(N+1)})} \mathbb{P}_n^{(N)}(f \mathbb{1}_{A_{j_{N+1}}^{(N+1)}}) \\ &= \sum_{j_{N+1}=1}^{m_{N+1}} \frac{P(A_{j_{N+1}}^{(N+1)})}{\mathbb{P}_n^{(N)}(A_{j_{N+1}}^{(N+1)})} \sum_{\substack{j_1 \leq m_1 \\ \vdots \\ j_N \leq m_N}} \sum_{i=1}^n f(X_i) \left(\frac{1}{n} \prod_{k=1}^N \frac{P(A_{j_k}^{(k)})}{\mathbb{P}_n^{(k-1)}(A_{j_k}^{(k)})} \right) \mathbb{1}_{X_i \in A_{j_1}^{(1)} \cap \dots \cap A_{j_{N+1}}^{(N+1)}} \\ &= \sum_{i=1}^n f(X_i) \sum_{\substack{j_1 \leq m_1 \\ \vdots \\ j_{N+1} \leq m_{N+1}}} \left(\frac{1}{n} \prod_{k=1}^{N+1} \frac{P(A_{j_k}^{(k)})}{\mathbb{P}_n^{(k-1)}(A_{j_k}^{(k)})} \right) \mathbb{1}_{X_i \in A_{j_1}^{(1)} \cap \dots \cap A_{j_{N+1}}^{(N+1)}}. \end{aligned}$$

Donc $\mathbb{P}_n^{(N+1)}(\mathcal{F})$ affecte à chaque $f(X_i)$ le poids $\frac{1}{n} \prod_{k=1}^{N+1} P(A_{j_k}^{(k)}) / \mathbb{P}_n^{(k-1)}(A_{j_k}^{(k)})$ si $X_i \in A_{j_1}^{(1)} \cap \dots \cap A_{j_{N+1}}^{(N+1)}$. Par principe de récurrence, la proposition est vraie pour tout $N \in \mathbb{N}^*$. \square

Projection de Kullback-Leibler. La proposition suivante généralise la proposition 2.3.3 dans le sens où $\mathbb{P}_n^{(N)}(\mathcal{F})$ est la projection de $\mathbb{P}_n^{(N-1)}(\mathcal{F})$ sur l'ensemble des mesures vérifiant l'information auxiliaire donnée par les $P(A_j^{N+1})$, $1 \leq j \leq m_{N+1}$.

Proposition 3.2.2. *Soit $N \in \mathbb{N}^*$. La mesure empirique $\mathbb{P}_n^{(N)}$ minimise la divergence de Kullback-Leibler $d_{KL}(\mathbb{P}_n^{(N-1)} || Q)$ entre $\mathbb{P}_n^{(N-1)}$ et l'ensemble des probabilités discrètes Q sur X_1, \dots, X_n qui coïncident avec P sur les $A_j^{(N)}$, $j \leq m_N$.*

Démonstration. La divergence de Kullback-Leibler entre $\mathbb{P}_n^{(N-1)}$ et $\mathbb{P}_n^{(N)}$ est donnée d'après la proposition 3.2.1 par

$$\begin{aligned} d_{KL}(\mathbb{P}_n^{(N-1)} || \mathbb{P}_n^{(N)}) &= \sum_{i=1}^n \mathbb{P}_n^{(N-1)}(\{X_i\}) \log \left(\frac{\mathbb{P}_n^{(N-1)}(\{X_i\})}{\mathbb{P}_n^{(N)}(\{X_i\})} \right) \\ &= \sum_{i=1}^n \mathbb{P}_n^{(N-1)}(\{X_i\}) \sum_{\substack{j_1 \leq m_1 \\ \vdots \\ j_N \leq m_N}} \mathbb{1}_{X_i \in A_{j_1}^{(1)} \cap \dots \cap A_{j_N}^{(N)}} \log \left(\frac{\frac{1}{n} \prod_{k=1}^{N-1} P(A_{j_k}) / \mathbb{P}_n^{(k-1)}(A_{j_k}^{(k)})}{\frac{1}{n} \prod_{k=1}^N P(A_{j_k}) / \mathbb{P}_n^{(k-1)}(A_{j_k}^{(k)})} \right) \\ &= \sum_{i=1}^n \mathbb{P}_n^{(N-1)}(\{X_i\}) \sum_{j_N=1}^{m_N} \mathbb{1}_{X_i \in A_{j_N}^{(N)}} \log \left(\frac{\mathbb{P}_n^{(N-1)}(A_{j_N}^{(N)})}{P(A_{j_N}^{(N)})} \right) \\ &= \sum_{j_N=1}^{m_N} \mathbb{P}_n^{(N-1)}(A_{j_N}^{(N)}) \log \left(\frac{\mathbb{P}_n^{(N-1)}(A_{j_N}^{(N)})}{P(A_{j_N}^{(N)})} \right). \end{aligned}$$

Soit Q une mesure de probabilité discrète sur X_1, \dots, X_n vérifiant $Q(A_j^{(N)}) = P(A_j^{(N)})$ pour tout $1 \leq j \leq m_N$. Par concavité du logarithme,

$$\begin{aligned} d_{\text{KL}}(\mathbb{P}_n^{(N-1)} || Q) &= - \sum_{i=1}^n \sum_{j_N=1}^{m_N} \mathbb{P}_n^{(N-1)}(A_{j_N}^{(N)}) \frac{\mathbb{P}_n^{(N-1)}(X_i)}{\mathbb{P}_n^{(N-1)}(A_{j_N}^{(N)})} \mathbb{1}_{X_i \in A_{j_N}} \log \left(\frac{Q(X_i)}{\mathbb{P}_n^{(N-1)}(X_i)} \right) \\ &\geq - \sum_{j_N=1}^{m_N} \mathbb{P}_n^{(N-1)}(A_{j_N}^{(N)}) \log \left(\sum_{i=1}^n \frac{Q(X_i)}{\mathbb{P}_n^{(N-1)}(A_{j_N}^{(N)})} \mathbb{1}_{X_i \in A_{j_N}} \right) \\ &\geq \sum_{j_N=1}^{m_N} \mathbb{P}_n^{(N-1)}(A_{j_N}^{(N)}) \log \left(\frac{\mathbb{P}_n^{(N-1)}(A_{j_N}^{(N)})}{P(A_{j_N}^{(N)})} \right) = d_{\text{KL}}(\mathbb{P}_n^{(N-1)} || \mathbb{P}_n^{(N)}). \end{aligned}$$

□

Loi du logarithme itéré. La proposition suivante donne une loi du logarithme itéré pour la mesure empirique rakké $\mathbb{P}_n^{(N)}(\mathcal{F})$.

Proposition 3.2.3 (LLI de $\mathbb{P}_n^{(N)}$). *Soit \mathcal{F} vérifiant $\mathcal{F}.i$, $\mathcal{F}.ii$, BR ou VC. Alors pour tout $N_0 \in \mathbb{N}$,*

$$\limsup_{n \rightarrow +\infty} \sqrt{\frac{n}{2LL(n)}} \max_{0 \leq N \leq N_0} \|\mathbb{P}_n^{(N)} - P\|_{\mathcal{F}} \leq \sigma_{\mathcal{F}} \prod_{N=1}^{N_0} (1 + M_{\mathcal{F}}/\delta_N) \quad p.s.$$

Remarque 3.2.1. *La constante qui intervient dans le résultat précédent explose si la plus petite des probabilités d'appartenir à un ensemble est très faible ou bien si le nombre de partitions est trop importante. Cette borne n'est certainement pas optimale sauf dans le cas trivial où $N_0 = 0$ qui correspond à la loi des logarithmes itérés classique. Dans le cas $N_0 = 1$, on ne retrouve pas la même borne que celle donnée par le théorème 2.2.7.*

Démonstration. Pour $N_0 = 0$, la proposition est vraie d'après la LLI (cf. théorème 1.2.9). On pose $u_n = \sqrt{n/2\sigma_{\mathcal{F}}^2 LL(n)}$, $b_N = \prod_{k=1}^N (1 + M_{\mathcal{F}}/\delta_k)$ et $\varepsilon > 0$. Supposons pour $N_0 > 0$ fixé qu'il existe p.s. un rang $n_0 \in \mathbb{N}^*$ tel que pour tout $n \geq n_0$, $u_n \max_{0 \leq N \leq N_0} \|\mathbb{P}_n^{(N)} - P\|_{\mathcal{F}} \leq b_{N_0}(1 + \varepsilon)$. Donc pour tout $n \geq n_0$,

$$\begin{aligned} u_n \|\mathbb{P}_n^{(N_0+1)} - P\|_{\mathcal{F}} &\leq u_n \|\mathbb{P}_n^{(N_0+1)} - \mathbb{P}_n^{(N_0)}\|_{\mathcal{F}} + u_n \|\mathbb{P}_n^{(N_0)} - P\|_{\mathcal{F}} \\ &\leq u_n \|\mathbb{P}_n^{(N_0+1)} - \mathbb{P}_n^{(N_0)}\|_{\mathcal{F}} + b_{N_0}(1 + \varepsilon). \end{aligned}$$

Puisque $\mathbb{P}_n^{(N_0)}$ est une mesure de probabilité pour tout ensemble A on a $\|\mathbb{P}_n^{(N_0)}(f \mathbb{1}_A)\|_{\mathcal{F}} \leq M_{\mathcal{F}} \mathbb{P}_n^{(N_0)}(A)$ d'où

$$\begin{aligned} \|\mathbb{P}_n^{(N_0+1)} - \mathbb{P}_n^{(N_0)}\|_{\mathcal{F}} &= \left\| \sum_{j=1}^{m_{N_0+1}} \mathbb{P}_n^{(N_0)}(f \mathbb{1}_{A_j^{(N_0+1)}}) \left(\frac{P(A_j^{(N_0+1)})}{\mathbb{P}_n^{(N_0)}(A_j^{(N_0+1)})} - 1 \right) \right\|_{\mathcal{F}} \\ &\leq \sum_{j=1}^{m_{N_0+1}} \frac{\|\mathbb{P}_n^{(N_0)}(f \mathbb{1}_{A_j^{(N_0+1)}})\|_{\mathcal{F}}}{\mathbb{P}_n^{(N_0)}(A_j^{(N_0+1)})} |\mathbb{P}_n^{(N_0)}(A_j^{(N_0+1)}) - P(A_j^{(N_0+1)})| \\ &\leq M_{\mathcal{F}} m_{N_0+1} \|\mathbb{P}_n^{(N_0)} - P\|_{\mathcal{F}}. \end{aligned}$$

Enfin en utilisant le fait que $m_{N_0+1} \leq 1/\delta_{N_0+1}$, on obtient pour tout $n \geq n_0$,

$$u_n \|\mathbb{P}_n^{(N_0+1)} - \mathbb{P}_n^{(N_0)}\|_{\mathcal{F}} \leq \frac{M_{\mathcal{F}}}{\delta_{N_0+1}} b_{N_0}(1 + \varepsilon)$$

ce qui prouve l'hérédité de notre hypothèse puisqu'on a montré que

$$u_n \|\mathbb{P}_n^{(N_0+1)} - P\|_{\mathcal{F}} \leq (1 + M_{\mathcal{F}}/\delta_{N_0+1})b_N(1 + \varepsilon) = b_{N_0+1}(1 + \varepsilon).$$

□

3.2.2 Processus empirique raké

Nous réécrivons le processus empirique $\alpha_n^{(N)}(\mathcal{F})$ pour les preuves qui vont suivre :

$$\begin{aligned} \alpha_n^{(N+1)}(f) &= \sum_{j \leq m_{N+1}} \frac{P(A_j^{(N+1)})}{\mathbb{P}_n^{(N)}(A_j^{(N+1)})} \alpha_n^{(N)}(f \mathbb{1}_{A_j^{(N+1)}}) - \frac{\mathbb{E}[f \mathbb{1}_{A_j^{(N+1)}}]}{\mathbb{P}_n^{(N)}(A_j^{(N+1)})} \alpha_n^{(N)}(A_j^{(N+1)}) \\ &= \sum_{j \leq m_{N+1}} \frac{P(A_j^{(N+1)})}{\mathbb{P}_n^{(N)}(A_j^{(N+1)})} \left(\alpha_n^{(N)}(f \mathbb{1}_{A_j^{(N+1)}}) - \mathbb{E}[f | A_j^{(N+1)}] \alpha_n^{(N)}(A_j^{(N+1)}) \right) \end{aligned}$$

Pour $N \in \mathbb{N}^*$, notons $\Delta_N = \prod_{k=1}^N \delta_k$, $S_N = \sum_{k=1}^N m_k$. Le résultat suivant donne une inégalité de concentration pour $\alpha_n^{(N)}(\mathcal{F})$. Elle borne la probabilité que le supremum $\|\alpha_n\|_{\mathcal{F}}$ dévie d'une valeur donnée.

Proposition 3.2.4. *Si \mathcal{F} vérifie $\mathcal{F}.i$ et $\mathcal{F}.ii$ alors pour tout $n, N_0 \in \mathbb{N}^*$ et $t > 0$,*

$$\mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\alpha_n^{(N)}\|_{\mathcal{F}} \geq t \right) \leq \frac{2^{N_0} N_0}{\Delta_{N_0}} \mathbb{P} \left(\|\alpha_n\|_{\mathcal{F}} \geq \frac{t \Delta_{N_0}}{(1 + M_{\mathcal{F}} + t/\sqrt{n})^{N_0}} \right) + S_{N_0} (1 - \delta_{(N_0)})^n. \quad (3.6)$$

En particulier, pour $t \leq 2M_{\mathcal{F}}\sqrt{n}$,

$$\mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\alpha_n^{(N)}\|_{\mathcal{F}} \geq t \right) \leq \frac{2^{N_0} N_0}{\Delta_{N_0}} \mathbb{P} \left(\|\alpha_n\|_{\mathcal{F}} \geq \frac{t \Delta_{N_0}}{(1 + 3M_{\mathcal{F}})^{N_0}} \right) + S_{N_0} (1 - \delta_{(N_0)})^n.$$

Si \mathcal{F} vérifie VC alors il existe $C_1 = C_1(c_0, \nu_0, N_0, M_{\mathcal{F}})$, $C_2 = C_2(N_0, M_{\mathcal{F}}) > 0$ tels que pour tout $0 < t < 2M_{\mathcal{F}}\sqrt{n}$,

$$\mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\alpha_n^{(N)}\|_{\mathcal{F}} \geq t \right) \leq C_1 t^{\nu_0} \exp(-C_2 t^2) + S_{N_0} (1 - \delta_{(N_0)})^n, \quad (3.7)$$

où c_0, ν_0 sont les constantes données par l'hypothèse VC. Si \mathcal{F} vérifie BR alors il existe $t_0 = t_0(M_{\mathcal{F}}, b_0, r_0)$, $D_1 = D_1(N_0)$, $D_2 = D_2(M_{\mathcal{F}}, N_0, b_0, r_0, \sigma_{\mathcal{F}}^2) > 0$ tels que pour tout $t_0 < t$,

$$\mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\alpha_n^{(N)}\|_{\mathcal{F}} \geq t \right) \leq D_1 \exp(-D_2 t^2) + S_{N_0} (1 - \delta_{(N_0)})^n. \quad (3.8)$$

Remarque 3.2.2. *La preuve de cette proposition et donc la borne qui en résulte ne prend pas en compte le fait que l'on puisse raker périodiquement sur un nombre fini et petit de partitions. Pour une étude plus fine dans le cas du ratissage à deux partitions, nous renvoyons à la partie 3.4. Cette proposition suggère donc de fixer N_0 , de prendre des grandes valeurs pour n et de travailler sur une classe de fonctions \mathcal{F} avec une petite entropie. En particulier pour les classes vérifiant l'hypothèse VC ou BR, cette proposition indique que presque sûrement $\|\alpha_n^{(N_0)}\|_{\mathcal{F}}$ est de l'ordre de $C\sqrt{\log n}$. Les bornes de concentration de mesure sont plus difficiles à obtenir dû au mélange des processus $\alpha_{n,j}^{(N-1)}$ impliquant des coefficients aléatoires non bornés.*

Démonstration. Nous travaillons sur l'ensemble B_{n,N_0} défini par (3.5). Par soucis de simplification, les événements suivants sont implicitement en intersection avec cet ensemble. On peut écrire pour tout $N \geq 1$,

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

avec $|\mathbb{E}[f|A_j^{(N)}]| \leq M$. Par suite,

$$\begin{aligned} \mathbb{P}(\|\alpha_n^{(N)}\|_{\mathcal{F}} \geq t) &\leq P \left(\sum_{j=1}^{m_N} P(A_j^{(N)}) \|\alpha_{n,j}^{(N-1)}\|_{\mathcal{F}} \geq \sum_{j=1}^{m_N} P(A_j^{(N)}) t \right) \\ &\leq \sum_{j=1}^{m_N} \mathbb{P}(\|\alpha_{n,j}^{(N-1)}\|_{\mathcal{F}} \geq t). \end{aligned}$$

Tous les termes de la dernière somme satisfont pour $K \leq P(A_j^{(N)})$ et $K' \leq P(A_j^{(N)}) - K$,

$$\begin{aligned} &\mathbb{P} \left(\frac{1}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} \left\| \alpha_n^{(N-1)}(f) - \mathbb{E}[f|A_j^{(N)}] \alpha_n^{(N-1)}(A_j^{(N)}) \right\|_{\mathcal{F}} \geq t \right) \\ &\leq \mathbb{P}((1 + M_{\mathcal{F}}) \|\alpha_n^{(N-1)}\|_{\mathcal{F}} \geq Kt) + \mathbb{P}(\mathbb{P}_n^{(N-1)}(A_j^{(N)}) \leq K) \\ &\leq \mathbb{P} \left(\|\alpha_n^{(N-1)}\|_{\mathcal{F}} \geq \frac{Kt}{1 + M_{\mathcal{F}}} \right) + \mathbb{P}(\alpha_n^{(N-1)}(A_j^{(N)}) \leq -K' \sqrt{n}) \\ &\leq 2\mathbb{P} \left(\|\alpha_n^{(N-1)}\|_{\mathcal{F}} \geq \frac{Kt}{1 + M_{\mathcal{F}}} \right), \end{aligned} \tag{3.9}$$

la dernière borne étant satisfaite si $K' \sqrt{n} \geq tK/(1 + M_{\mathcal{F}})$. On pose

$$\beta = \frac{1}{1 + M_{\mathcal{F}} + t/\sqrt{n}} \in]0, 1[, \quad K = \beta \delta_N, \quad K' = \delta_N(1 - \beta).$$

Puisque $\beta < 1$ pour tout $N \geq 1$ on a $K' > 0$ et en utilisant le fait que $\beta \leq (1 + M_{\mathcal{F}})/(1 + M_{\mathcal{F}} + t/\sqrt{n})$,

$$\frac{t}{\sqrt{n}(1 + M_{\mathcal{F}})} \leq \frac{\delta_N(1 - \beta)}{\delta_N \beta} = \frac{K'}{K}.$$

Nous avons montré pour tout $N \geq 1$,

$$\mathbb{P}(\|\alpha_n^{(N)}\|_{\mathcal{F}} \geq t) \leq \frac{2}{\delta_N} \mathbb{P} \left(\|\alpha_n^{(N-1)}\|_{\mathcal{F}} \geq \frac{\beta \delta_N t}{1 + M_{\mathcal{F}}} \right).$$

En appliquant (3.9) à nouveau avec pour t la valeur $\beta \delta_N t/(1 + M_{\mathcal{F}}) \leq t$ puis en itérant on obtient que

$$\mathbb{P}(\|\alpha_n^{(N)}\|_{\mathcal{F}} \geq t) \leq \frac{2^N}{\Delta_N} \mathbb{P} \left(\|\alpha_n^{(0)}\|_{\mathcal{F}} \geq \frac{t \Delta_N}{(1 + M + t/\sqrt{n})^N} \right). \tag{3.10}$$

Puisque la dernière borne croît avec N on en conclut que

$$\mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\alpha_n^{(N)}\|_{\mathcal{F}} \geq t \right) \leq \sum_{N=1}^{N_0} \mathbb{P}(\|\alpha_n^{(N)}\|_{\mathcal{F}} \geq t) \leq N_0 \mathbb{P}(\|\alpha_n^{(N_0)}\|_{\mathcal{F}} \geq t). \tag{3.11}$$

Les inégalités (3.10), (3.11) et en prenant en compte le fait que B_{n,N_0}^C soit de probabilité plus petite que $S_{N_0}(1 - \delta_{(N_0)})^n$ entraînent (3.6). Sous l'hypothèse VC, il suffit d'appliquer l'inégalité de Talagrand pour les classes vérifiant VC (*cf.* théorème 1.2.6) à (3.6) en posant

$$C_1 = \frac{2^{N_0} N_0}{\Delta_{N_0}} \left(\frac{C_1 \Delta_{N_0}}{(1 + 3M_{\mathcal{F}})^{N_0}} \right)^{\nu_0}, \quad C_2 = \frac{1}{2} \left(\frac{t \Delta_{N_0}}{(1 + 3M_{\mathcal{F}})^{N_0}} \right)^2,$$

afin d'obtenir (3.7). Sous l'hypothèse BR, on obtient (3.8) en appliquant l'inégalité de Talagrand pour les classes vérifiant BR (*cf.* théorème 1.2.7 et en posant

$$D_1 = \frac{2^{N_0} N_0}{\Delta_{N_0}}, \quad D_2 = C \left(\frac{\Delta_{N_0}}{(1 + M_{\mathcal{F}})^{N_0}} \right)^2,$$

avec $C = C(M_{\mathcal{F}}, b_0, r_0, \sigma_{\mathcal{F}}^2) > 0$. □

3.3 Résultats asymptotiques

3.3.1 Processus gaussien raké

Définition et motivation. À N fixé, nous allons établir la convergence en loi de $\alpha_n^{(N)}$ quand $n \rightarrow +\infty$ vers le processus gaussien $\mathbb{G}^{(N)}(\mathcal{F})$ défini par récurrence pour tout $f \in \mathcal{F}$ par

$$\mathbb{G}^{(0)} = \mathbb{G} \quad \text{et} \quad \mathbb{G}^{(N+1)}(f) = \mathbb{G}^{(N)}(f) - \sum_{j \leq m_{N+1}} \mathbb{E}[f | A_j^{(N+1)}] \mathbb{G}^{(N)}(A_j^{(N+1)}). \quad (3.12)$$

Par construction, $\mathbb{G}^{(N)}(\mathcal{F})$ est un processus gaussien centré indexé par \mathcal{F} . Nous le désignons par processus gaussien raké ou bien P -pont brownien raké. Cette partie sera consacrée à l'étude des propriétés de ce processus : expression formelle, formule de variance et covariance et propriété de réduction de variance. Asymptotiquement le processus empirique raké $\alpha_n^{(N)}(\mathcal{F})$ possédera ces propriétés établies.

Expression de $\mathbb{G}^{(N)}(\mathcal{F})$. En effectuant les calculs, on peut trouver l'expression du processus gaussien $\mathbb{G}^{(N)}(\mathcal{F})$ défini par récurrence par (3.12). L'expression donnée par la proposition suivante est complexe et on devra adapter notre notation pour en faciliter l'étude.

Proposition 3.3.1. *Soit $\mathbb{G}^{(N)}$ le processus gaussien centré défini par (3.12). Alors,*

$$\forall N \geq 1, \quad \mathbb{G}^{(N)}(f) = \mathbb{G}(f) - \sum_{\substack{1 \leq i \leq N \\ 1 \leq j_i \leq m_i}} \phi^{(N)}(f, i, j_i) \mathbb{G}(A_{j_i}^{(i)})$$

avec

$$\begin{aligned} \phi^{(N)}(f, i, j_i) &= \mathbb{E}[f | A_{j_i}^{(i)}] \\ &+ \sum_{\substack{1 \leq L \leq N-i \\ i < l_1 < l_2 < \dots < l_L \leq N}} (-1)^L \sum_{\substack{k_1 \leq m_{l_1} \\ \vdots \\ k_L \leq m_{l_L}}} P(A_{k_1}^{(l_1)} | A_{j_i}^{(i)}) P(A_{k_2}^{(l_2)} | A_{k_1}^{(l_1)}) \dots P(A_{k_L}^{(l_L)} | A_{k_{L-1}}^{(l_{L-1})}) \mathbb{E}[f | A_{k_L}^{(l_L)}] \end{aligned} \quad (3.13)$$

Démonstration. Pour $N = 1$ le résultat est vrai car $\forall j \leq m_1, \phi^{(N)}(f, 1, j) = \mathbb{E}[f|A_j^{(1)}]$ et donc $\mathbb{G}^{(1)} = \widehat{\mathbb{G}}^{\mathcal{A}^{(1)}}$. Supposons que le résultat soit vrai pour un certain rang $N \geq 1$. Alors,

$$\begin{aligned} \mathbb{G}^{(N+1)}(f) &= \mathbb{G}^{(N)}(f) - \sum_{j=1}^{m_{N+1}} \mathbb{E}[f|A_j^{(N+1)}] \mathbb{G}^{(N)}(A_j^{(N+1)}) \\ &= \mathbb{G}(f) - \sum_{\substack{1 \leq i \leq N \\ 1 \leq j_i \leq m_i}} \mathbb{G}(A_{j_i}^{(i)}) \left(\phi^{(N)}(f, i, j_i) - \sum_{j=1}^{m_{N+1}} \mathbb{E}[f|A_j^{(N+1)}] \phi^{(N)}(\mathbb{1}_{A_j^{(N+1)}}(i, j_i)) \right) \\ &\quad - \sum_{j=1}^{m_{N+1}} \mathbb{E}[f|A_j^{(N+1)}] \mathbb{G}(A_j^{(N+1)}). \end{aligned}$$

La proposition est vraie au rang $N + 1$ puisque d'une part $\mathbb{E}[f|A_j^{(N+1)}] = \phi^{(N+1)}(f, N + 1, j)$ et d'autre part on peut vérifier par (3.13) que

$$\phi^{(N)}(f, i, j_i) - \sum_{j=1}^{m_{N+1}} \mathbb{E}[f|A_j^{(N+1)}] \phi^{(N)}(\mathbb{1}_{A_j^{(N+1)}}(i, j_i)) = \phi^{(N+1)}(f, i, j_i).$$

Par principe de récurrence, la proposition est vraie pour tout $N \geq 1$. \square

Afin d'alléger l'expression (3.13) on peut adopter une notation matricielle que l'on conserve par la suite. On note

$$\begin{aligned} \forall 1 \leq j \leq N, \quad \mathbb{E}^{(j)}[f] &= \left(\mathbb{E}[f|A_1^{(j)}], \dots, \mathbb{E}[f|A_{m_j}^{(j)}] \right)^t, \\ \mathbb{G}[\mathcal{A}^{(i)}] &= \left(\mathbb{G}(A_1^{(i)}), \dots, \mathbb{G}(A_{m_i}^{(i)}) \right)^t, \end{aligned}$$

et $\mathbf{P}_{\mathcal{A}^{(k)}|\mathcal{A}^{(l)}} \in \mathcal{M}_{m_l, m_k}(\mathbb{R})$ la matrice stochastique définie par

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

L'expression de $\mathbb{G}^{(N)}$ s'exprime alors par

$$\mathbb{G}^{(N)}(f) = \mathbb{G}(f) - \sum_{1 \leq i \leq N} \Phi_i^{(N)}(f)^t \cdot \mathbb{G}[\mathcal{A}^{(i)}], \quad (3.14)$$

avec

$$\Phi_i^{(N)}(f) = \mathbb{E}^{(i)}[f] + \sum_{\substack{1 \leq L \leq N-i \\ i < l_1 < l_2 < \dots < l_L \leq N}} (-1)^L \mathbf{P}_{\mathcal{A}^{(l_1)}|\mathcal{A}^{(i)}} \mathbf{P}_{\mathcal{A}^{(l_2)}|\mathcal{A}^{(l_1)}} \dots \mathbf{P}_{\mathcal{A}^{(l_L)}|\mathcal{A}^{(l_{L-1})}} \mathbb{E}^{(l_L)}[f].$$

Variance de $\mathbb{G}^{(N)}(\mathcal{F})$. Même si la formule qui va suivre n'est pas exploitable, on peut exprimer la variance du processus $\mathbb{G}^{(N)}(\mathcal{F})$ par rapport à celle du processus $\widehat{\mathbb{G}}^{\cap_{i=1}^N \mathcal{A}^{(i)}}(\mathcal{F})$ le pont brownien exploitant information auxiliaire l'information complète, c'est-à-dire la probabilité de toutes les intersections $A_{i_1}^{(1)} \cap \dots \cap A_{i_N}^{(N)}$ pour tout $1 \leq i_1 \leq m_1, \dots, 1 \leq i_N \leq m_N$ dont on n'a pas connaissance. On peut écrire que $\mathbb{G}^{(N)} = \widehat{\mathbb{G}}^{\cap_{i=1}^N \mathcal{A}^{(i)}} - \mathbb{G}^{\nabla \nabla, \mathcal{A}^{(1)}, \dots, \mathcal{A}^{(N)}}$ avec

$$\mathbb{G}^{\nabla \nabla, \mathcal{A}^{(1)}, \dots, \mathcal{A}^{(N)}}(f) = \sum_{\substack{j_1 \leq m_1 \\ j_N \leq m_N}} \mathbb{G}(A_{j_1}^{(1)} \cap \dots \cap A_{j_N}^{(N)}) \left(\sum_{i=1}^N \phi(i, j_i) - \mathbb{E}[f|A_{j_1}^{(1)} \cap \dots \cap A_{j_N}^{(N)}] \right).$$

Les processus $\widehat{\mathbb{G}}^{\cap_{i=1}^N \mathcal{A}^{(i)}}$ et $\mathbb{G}^{\nabla \nabla, \mathcal{A}^{(1)}, \dots, \mathcal{A}^{(N)}}$ sont indépendants puisque pour tout $f \in \mathcal{F}$,

$$\begin{aligned} & \text{Cov}(\widehat{\mathbb{G}}^{\cap_{i=1}^N \mathcal{A}^{(i)}}(f), \mathbb{G}^{\nabla \nabla, \mathcal{A}^{(1)}, \dots, \mathcal{A}^{(N)}}(f)) \\ &= \sum_{\substack{j_1 \leq m_1 \\ \dots \\ j_N \leq m_N}} K_{j_1, \dots, j_N} \text{Cov} \left(\mathbb{G}(f) - \sum_{\substack{j'_1 \leq m_1 \\ \dots \\ j'_N \leq m_N}} \mathbb{E}[f | A_{j'_1}^{(1)} \cap \dots \cap A_{j'_N}^{(N)}] \mathbb{G}(A_{j'_1}^{(1)} \cap \dots \cap A_{j'_N}^{(N)}), \mathbb{G}(A_{j_1}^{(1)} \cap \dots \cap A_{j_N}^{(N)}) \right) \\ &= 0, \end{aligned}$$

où $K_{j_1, \dots, j_N} = \sum_{i=1}^N \phi(i, j_i) - \mathbb{E}[f | A_{j_1}^{(1)} \cap \dots \cap A_{j_N}^{(N)}]$. La variance du processus avec la méthode du Raking Ratio généralisée est alors donnée par :

$$\text{Var}(\mathbb{G}^{(N)}(f)) = \text{Var}(\widehat{\mathbb{G}}^{\cap_{i=1}^N \mathcal{A}^{(i)}}(f)) + \text{Var}(\mathbb{G}^{\nabla \nabla, \mathcal{A}^{(1)}, \dots, \mathcal{A}^{(N)}}(f)).$$

En particulier, on voit que la méthode du Raking Ratio fait toujours moins bien que si on avait l'information complète. Nous pouvons aussi nous poser la question de savoir si la variance décroît nécessairement à chaque itération de l'algorithme, *i.e.* a-t-on forcément $\text{Var}(\mathbb{G}^{(N+1)}(f)) \leq \text{Var}(\mathbb{G}^{(N)}(f))$ pour tout $f \in \mathcal{F}$, $N \in \mathbb{N}$? La réponse est négative. Prenons par exemple la matrice des probabilités

$P(A_i \cap B_j)$	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	

et les espérances suivantes

$\mathbb{E}[f A_i \cap B_j]$	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	

et fixons $\text{Var}(f | A_i \cap B_j) = 0.5$. Alors on obtient les variances théoriques suivantes : $\text{Var}(\mathbb{G}(f)) \simeq 0.734$, $\text{Var}(\mathbb{G}^{(1)}(f)) \simeq 0.563$, $\text{Var}(\mathbb{G}^{(2)}(f)) \simeq 0.569$ et $\text{Var}(\mathbb{G}^{(3)}(f)) \simeq 0.402$. La variance ne décroît pas à chaque étape du Raking-Ratio puisque $\text{Var}(\mathbb{G}^{(2)}(f)) > \text{Var}(\mathbb{G}^{(1)}(f))$. Une autre question est de savoir maintenant si on fait mieux que le P -pont brownien $\mathbb{G}(\mathcal{F})$ à chaque tour dans le sens où on aurait $\text{Var}(\mathbb{G}^{(N)}(f)) \leq \text{Var}(\mathbb{G}(f))$ pour tout $f \in \mathcal{F}$ et $N \in \mathbb{N}$? Dans l'exemple précédent, $\mathbb{G}^{(N)}(\mathcal{F})$ semble au moins avoir une variance plus faible que $\mathbb{G}(\mathcal{F})$. Nous allons montrer que cela est toujours vrai. On note $\text{Cov}(\mathbb{G}[\mathcal{A}^{(i)}])$ la matrice de variance du vecteur $\mathbb{G}[\mathcal{A}^{(i)}]$, *i.e.*

$$\begin{aligned} (\text{Cov}(\mathbb{G}[\mathcal{A}^{(i)}]))_{j,k} &= \text{Cov}(\mathbb{G}(A_j^{(i)}), \mathbb{G}(A_k^{(i)})) = P(A_j \cap A_k) - P(A_j)P(A_k) \\ &= \text{diag}(P[\mathcal{A}^{(k)}]) - P[\mathcal{A}^{(k)}]^t P[\mathcal{A}^{(k)}], \end{aligned}$$

où $P[\mathcal{A}^{(k)}] = (P(A_1^{(k)}), \dots, P(A_{m_k}^{(k)}))^t$. On note également $\text{Cov}(\mathbb{G}[\mathcal{A}^{(i)}], \mathbb{G}(f))$ le vecteur constitué des covariances des éléments de $\mathbb{G}[\mathcal{A}^{(i)}]$ avec $\mathbb{G}(f)$, *i.e.*

$$\begin{aligned} \text{Cov}(\mathbb{G}[\mathcal{A}^{(i)}], \mathbb{G}(f)) &= \left(\text{Cov}(\mathbb{G}(A_1^{(i)}), \mathbb{G}(f)), \dots, \text{Cov}(\mathbb{G}(A_{m_i}^{(i)}), \mathbb{G}(f)) \right)^t \\ &= \left(P(A_1^{(i)})(\mathbb{E}[f | A_1^{(i)}] - \mathbb{E}[f]), \dots, P(A_{m_i}^{(i)})(\mathbb{E}[f | A_{m_i}^{(i)}] - \mathbb{E}[f]) \right)^t. \end{aligned}$$

Ces éléments vérifient le lemme suivant :

Lemme 3.3.2. *Pour tout $1 \leq k, l \leq N$ et $f \in \mathcal{F}$ on a*

$$\text{Cov} \left(\mathbb{G}[\mathcal{A}^{(k)}], \mathbb{G}(f) \right) = \text{Var} \left(\mathbb{G}[\mathcal{A}^{(k)}] \right) \cdot \mathbb{E}^{(k)}[f], \quad (3.15)$$

$$\text{Cov} \left(\mathbb{G}[\mathcal{A}^{(k)}], \mathbb{G}[\mathcal{A}^{(l)}] \right) = \text{Var} \left(\mathbb{G}[\mathcal{A}^{(k)}] \right) \cdot \mathbf{P}_{\mathcal{A}^{(l)}|\mathcal{A}^{(k)}}, \quad (3.16)$$

$$\Phi_k^{(N)}(f) = \mathbb{E}^{(k)}[f] - \sum_{k < l \leq N} \mathbf{P}_{\mathcal{A}^{(l)}|\mathcal{A}^{(k)}} \Phi_l^{(N)}(f). \quad (3.17)$$

Démonstration. La j -ème coordonnée du vecteur $\text{Var} \left(\mathbb{G}[\mathcal{A}^{(k)}] \right) \cdot \mathbb{E}^{(k)}[f]$ est

$$\begin{aligned} P(A_j^{(k)})\mathbb{E}[f|A_j^{(k)}] - \sum_{i \leq m_k} P(A_i^{(k)})P(A_j^{(k)})\mathbb{E}[f | A_i^{(k)}] &= \mathbb{E}[f\mathbf{1}_{A_j^{(k)}}] - P(A_j^{(k)}) \sum_{1 \leq i \leq m_k} \mathbb{E}[f\mathbf{1}_{A_i^{(k)}}] \\ &= \text{Cov} \left(\mathbb{G}(A_j^{(k)}), \mathbb{G}(f) \right), \end{aligned}$$

ce qui prouve (3.15). De la même manière la coordonnée (i, j) de la matrice $\text{Var} \left(\mathbb{G}[\mathcal{A}^{(k)}] \right) \cdot \mathbf{P}_{\mathcal{A}^{(l)}|\mathcal{A}^{(k)}}$ est

$$\begin{aligned} P(A_i^{(k)})P(A_j^{(l)} | A_i^{(k)}) - \sum_{m \leq m_k} P(A_i^{(k)})P(A_m^{(k)})P(A_j^{(l)} | A_m^{(k)}) \\ = P(A_j^{(l)} \cap A_i^{(k)}) - P(A_i^{(k)}) \sum_{1 \leq m \leq m_k} P(A_j^{(l)} \cap A_m^{(k)}) \\ = \text{Cov} \left(\mathbb{G}(A_i^{(k)}), \mathbb{G}(A_j^{(l)}) \right), \end{aligned}$$

ce qui prouve (3.16). Par définition des fonctions $\Phi_j^{(N)}$ données par (3.14), on trouve que

$$\begin{aligned} \sum_{i < j \leq N} \mathbf{P}_{\mathcal{A}^{(j)}|\mathcal{A}^{(i)}} \Phi_j^{(N)}(f) &= \sum_{i < j \leq N} \mathbf{P}_{\mathcal{A}^{(j)}|\mathcal{A}^{(i)}} \mathbb{E}^{(j)}[f] \\ &\quad + \sum_{\substack{i < j \leq N \\ 1 \leq L \leq N-j \\ j < l_1 < \dots < l_L \leq N}} (-1)^L \mathbf{P}_{\mathcal{A}^{(j)}|\mathcal{A}^{(i)}} \mathbf{P}_{\mathcal{A}^{(l_1)}|\mathcal{A}^{(j)}} \dots \mathbf{P}_{\mathcal{A}^{(l_L)}|\mathcal{A}^{(l_{L-1})}} \mathbb{E}^{(l_L)}[f] \\ &= \sum_{\substack{1 \leq L \leq N-i \\ i < l_1 < \dots < l_L \leq N}} (-1)^{L+1} \mathbf{P}_{\mathcal{A}^{(l_1)}|\mathcal{A}^{(i)}} \dots \mathbf{P}_{\mathcal{A}^{(l_L)}|\mathcal{A}^{(l_{L-1})}} \mathbb{E}^{(l_L)}[f] \\ &= -\Phi_i^{(N)}(f) + \mathbb{E}^{(i)}[f], \end{aligned}$$

ce qui prouve (3.17). □

Proposition 3.3.3. *Pour tout $N \in \mathbb{N}^*$ et $(f, g) \in \mathcal{F}^2$,*

$$\text{Var}(\mathbb{G}^{(N)}(f)) = \text{Var}(\mathbb{G}(f)) - \sum_{i=1}^N \Phi_i^{(N)}(f)^t \cdot \text{Var}(\mathbb{G}[\mathcal{A}^{(i)}]) \cdot \Phi_i^{(N)}(f) \quad (3.18)$$

$$\text{Cov}(\mathbb{G}^{(N)}(f), \mathbb{G}^{(N)}(g)) = \text{Cov}(\mathbb{G}(f), \mathbb{G}(g)) - \sum_{i=1}^N \text{Cov} \left(\Phi_i^{(N)}(f)^t \cdot \mathbb{G}[\mathcal{A}^{(i)}], \Phi_i^{(N)}(g)^t \cdot \mathbb{G}[\mathcal{A}^{(i)}] \right). \quad (3.19)$$

En particulier, $\text{Var}(\mathbb{G}^{(N)}(f)) \leq \text{Var}(\mathbb{G}(f))$ pour tout $f \in \mathcal{F}$.

Remarque 3.3.1. La réduction de risque asymptotique est quantifiée par la combinaison des propositions 3.3.3 et 3.3.10. Soit $\varepsilon > 0$, il existe $n_0 \in \mathbb{N}^*$ tel que pour tout $n > n_0$ et pour tout $f \in \mathcal{F}$ le nouveau risque après avoir rakié N fois est donné par

$$\begin{aligned}\mathbb{E}\left[(\mathbb{P}_n^{(N)}(f) - P(f))^2\right] &= \frac{\sigma_{\mathcal{F}}^2}{n}(\Delta(f) + e(f)v_n), \\ \Delta(f) &= \frac{\text{Var}(\mathbb{G}^{(N)}(f))}{\text{Var}(f(X))} \in [0, 1], \\ \sup_{f \in \mathcal{F}} |e(f)| &< (1 + \varepsilon)\sqrt{\frac{8}{\pi}}C_\theta,\end{aligned}$$

où $v_n \rightarrow 0$ et C_θ sont données par le théorème 3.3.7. Le risque de réduction a donc lieu puisque $\Delta(f) < 1$ pour des valeurs de n suffisamment grandes.

Démonstration. D'après (3.14) et le lemme 3.3.2, on peut écrire que

$$\begin{aligned}\text{Var}(\mathbb{G}^{(N)}(f)) - \text{Var}(\mathbb{G}(f)) &= \sum_{i=1}^N \Phi_i^{(N)}(f)^t \cdot \text{Var}(\mathbb{G}[\mathcal{A}^{(i)}]) \cdot \Phi_i^{(N)}(f) - 2 \sum_{i=1}^n \Phi_i^{(N)}(f)^t \cdot \text{Cov}(\mathbb{G}[\mathcal{A}^{(i)}], \mathbb{G}(f)) \\ &\quad + 2 \sum_{\substack{1 \leq i \leq N \\ i < j \leq N}} \Phi_i^{(N)}(f)^t \cdot \text{Cov}(\mathbb{G}[\mathcal{A}^{(i)}], \mathbb{G}[\mathcal{A}^{(j)}]) \cdot \Phi_j^{(N)}(f) \\ &= \sum_{i=1}^N \Phi_i^{(N)}(f)^t \cdot \text{Var}(\mathbb{G}[\mathcal{A}^{(i)}]) \cdot \left(\Phi_i^{(N)}(f) - 2\mathbb{E}^{(i)}[f] + 2 \sum_{i < j \leq N} \mathbf{P}_{\mathcal{A}^{(j)}|\mathcal{A}^{(i)}} \Phi_j^{(N)}(f) \right).\end{aligned}$$

On obtient (3.18) en remarquant que le dernier terme en facteur correspond à $-\Phi_i^{(N)}(f)$ d'après le lemme 3.3.2. La forme polaire permet de passer de (3.18) à (3.19). \square

La méthode du Raking Ratio diminue la structure de covariance du P -pont brownien initial $\mathbb{G}(\mathcal{F})$ dans le sens où pour tout $f_1, \dots, f_k \in \mathcal{F}$, si on note Σ la matrice de covariance de $(\mathbb{G}(f_1), \dots, \mathbb{G}(f_k))$ et $\Sigma^{(N)}$ celle de $(\mathbb{G}^{(N)}(f_1), \dots, \mathbb{G}^{(N)}(f_k))$ alors $\Sigma - \Sigma^{(N)}$ est semi-définie positive. En effet cette différence de matrices correspond à la matrice de covariance de $(\mathbb{G}^{\nabla, (N)}(f_1), \dots, \mathbb{G}^{\nabla, (N)}(f_k))$ où $\mathbb{G}^{\nabla, (N)}(f) = \sum_{i=1}^N \Phi_i^{(N)}(f)^t \cdot \mathbb{G}[\mathcal{A}^{(i)}]$. Finalement, nous avons vu que le processus gaussien limite ne diminue pas nécessairement sa variance à chaque tour mais il existe néanmoins un cas naturel où celle-ci subit une diminution. Il s'agit du bouclage qui consiste à répéter cette méthode sur un même cycle de partitions. Ce résultat est donné de manière généralisée par le lemme suivant.

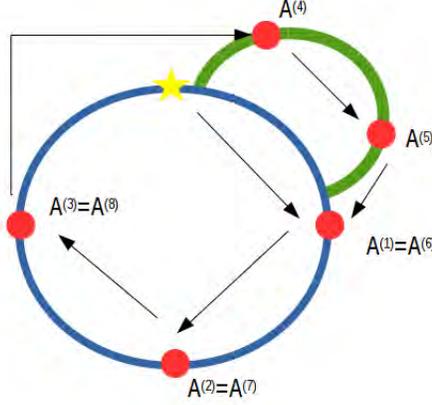
Lemme 3.3.4 (Lemme de bouclage). Soient $N_0, N_1 \in \mathbb{N}$ tels que $N_1 \geq 2N_0$ et supposons que

$$\forall 0 \leq i < N_0, \quad \mathcal{A}^{(N_0-i)} = \mathcal{A}^{(N_1-i)}.$$

Alors $\forall f \in \mathcal{F}$, $\text{Var}(\mathbb{G}^{(N_1)}(f)) \leq \text{Var}(\mathbb{G}^{(N_0)}(f))$.

Démonstration. Avec l'hypothèse de bouclage et la formule de variance donnée par la pro-

FIG. 3.1: Illustration du lemme de bouclage



position 3.3.3 on a nécessairement que $\forall 0 \leq i < N_0, \Phi_{N_0-i}^{(N_0)} = \Phi_{N_1-i}^{(N_1)}$. Alors

$$\begin{aligned}
 & \text{Var}(\widehat{\mathbb{G}}^{(N_0)}(f)) - \text{Var}(\widehat{\mathbb{G}}^{(N_1)}(f)) \\
 &= \sum_{i=1}^{N_1} \Phi_i^{(N_1)}(f)^t \cdot \text{Var}(\mathbb{G}[\mathcal{A}^{(i)}]) \cdot \Phi_i^{(N_1)}(f) - \sum_{i=1}^{N_0} \Phi_i^{(N_0)}(f)^t \cdot \text{Var}(\mathbb{G}[\mathcal{A}^{(i)}]) \cdot \Phi_i^{(N_0)}(f) \\
 &= \sum_{i=1}^{N_1} \Phi_i^{(N_1)}(f)^t \cdot \text{Var}(\mathbb{G}[\mathcal{A}^{(i)}]) \cdot \Phi_i^{(N_1)}(f) - \sum_{i=N_1-N_0+1}^{N_1} \Phi_i^{(N_1)}(f)^t \cdot \text{Var}(\mathbb{G}[\mathcal{A}^{(i)}]) \cdot \Phi_i^{(N_1)}(f) \\
 &= \sum_{i=1}^{N_1-N_0} \Phi_i^{(N_1)}(f)^t \cdot \text{Var}(\mathbb{G}[\mathcal{A}^{(i)}]) \cdot \Phi_i^{(N_1)}(f) \geq 0
 \end{aligned}$$

□

La figure 3.1 illustre un cas où la variance $\text{Var}(\mathbb{G}^{(i)}(f)), i \in \{6, 7, 8\}$ est plus faible que $\text{Var}(\mathbb{G}^{(j)}(f)), j = 1, 2, 3$ et ce pour tout $f \in \mathcal{F}$.

3.3.2 Approximation forte et applications

Approximation forte. Avec des hypothèses fortes sur la classe de fonctions \mathcal{F} il est possible de donner une vitesse de convergence asymptotique du processus $\alpha_n^{(N)}(\mathcal{F})$ vers $\mathbb{G}^{(N)}(\mathcal{F})$. On note pour $N \in \mathbb{N}^*$ et $j = 1, \dots, m_N$,

$$\phi_{(N),j}(f) = (f - \mathbb{E}[f|A_j^{(N)}])\mathbb{1}_{A_j^{(N)}}, \quad \phi_{(N)}(f) = \sum_{j=1}^{m_N} \phi_{(N),j}(f),$$

avec la convention $\phi_{(0)}(f) = f$. De manière immédiate $\mathbb{E}[\phi_{(N)}(f)] = \mathbb{E}[\phi_{(N),j}(f)] = 0$ et puisque $\mathcal{A}^{(N)}$ est une partition de \mathcal{X} , on a que

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

De plus les propriétés de l'espérance conditionnelle entraîne que

$$\sigma_{\phi_{(N),j}(\mathcal{F})}^2 \leq \sigma_{\phi_{(N)}(\mathcal{F})}^2 \leq \sigma_{\mathcal{F}}^2. \quad (3.20)$$

Avec ces nouvelles notations, on peut alors écrire que

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

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

Par itérations successives, on a nécessairement que

$$\alpha_n^{(N)}(f) = \alpha_n^{(0)}(\phi_{(1)} \circ \cdots \circ \phi_{(N)}(f)) + F_n^{(N)}(f), \quad (3.21)$$

$$F_n^{(N)}(f) = \sum_{k=1}^{N-1} \Gamma_n^{(k)}(\phi_{(k+1)} \circ \cdots \circ \phi_{(N)}(f)).$$

Il sera nécessaire d'étudier les classes de fonctions des itérés renversées suivantes

$$\mathcal{G}_{(N)} = \phi_{(1)} \circ \cdots \circ \phi_{(N)}(\mathcal{F})$$

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

avec la convention $\phi_{(k+1)} \circ \cdots \circ \phi_{(N)} = \text{id}$ si $k = N$ et $\mathcal{G}_{(0)} = \mathcal{H}_{(0)} = \mathcal{F}$. En itérant (3.20), il vient $\sigma_{\mathcal{H}_0}^2 \leq \sigma_{\mathcal{G}_0}^2 \leq \sigma_{\mathcal{F}}^2$. On pose de la même manière $\mathcal{G}_0 = \cup_{0 \leq N \leq N_0} \mathcal{G}_{(N)}$ et $\mathcal{H}_0 = \cup_{0 \leq N \leq N_0} \mathcal{H}_{(N)}$. Le lemme suivant montre que les classes introduites précédemment conservent les mêmes conditions d'entropie que \mathcal{F} .

Lemme 3.3.5. *Si \mathcal{F} vérifie $\mathcal{F}.i$, $\mathcal{F}.ii$ alors $\mathcal{G}_{(N)}, \mathcal{H}_{(N)}$ (respectivement $\mathcal{G}_0, \mathcal{H}_0$) vérifient également $\mathcal{F}.i$ avec pour borne $(2M_{\mathcal{F}})^N / \Delta_N$ (respectivement $(2M_{\mathcal{F}})^{N_0} / \Delta_{N_0}$) ainsi que $\mathcal{F}.ii$. Si \mathcal{F} vérifie BR ou VC il en va de même pour $\mathcal{G}_{(N)}, \mathcal{H}_{(N)}, \mathcal{G}_0, \mathcal{H}_0$.*

Remarque 3.3.2. *Dans la preuve de ce lemme, on constate que l'indice r_0 (pour l'hypothèse BR) est préservé, ce qui n'est pas le cas de l'indice v_0 (pour l'hypothèse VC). Cela aurait pour conséquence de détériorer la vitesse d'approximation forte des classes $\mathcal{G}_{(N)}$ et $\mathcal{H}_{(N)}$ par rapport à la vitesse d'approximation forte v_n de la classe \mathcal{F} . Néanmoins si $d_P(f, g) < \varepsilon$ avec $f, g \in \mathcal{F}$ alors*

$$d_P^2(\phi_{(N),j}(f), \phi_{(N),j}(g)) = \int_{A_j^{(N)}} (f - g - \mathbb{E}[f - g | A_j^{(N)}])^2 dP \leq \int_{A_j^{(N)}} (f - g)^2 dP,$$

$$d_P^2(\phi_{(N)}(f), \phi_{(N)}(g)) = \int_X \left(\sum_{j=1}^{m_N} \phi_{(N),j}(f) - \phi_{(N),j}(g) \right)^2 dP$$

$$= \sum_{j=1}^{m_N} d_Q^2(\phi_{(N),j}(f), \phi_{(N),j}(g)) \leq \sum_{j=1}^{m_N} \int_{A_j^{(N)}} (f - g)^2 dP \leq \varepsilon^2.$$

En prenant en compte cette inégalité et avec l'article de Berthet-Mason [12] il est possible de conserver la vitesse v_n de l'approximation forte de la classe \mathcal{F} .

Démonstration. Si \mathcal{F} est uniformément bornée par $M_{\mathcal{F}}$ alors pour $N \leq N_0$ on a,

$$\sup_{f \in \mathcal{F}} \sup_{x \in \mathcal{X}} |\phi_{(N)}(f(x))| = \sup_{f \in \mathcal{F}} \max_{1 \leq j \leq m_N} \sup_{x \in \mathcal{X}} |\phi_{(N),j}(f(x))| \leq M_{\mathcal{F}} \left(1 + \frac{1}{\delta_N} \right) \leq \frac{2M_{\mathcal{F}}}{\delta_N}.$$

Il s'ensuit immédiatement que $\mathcal{G}_{(N)}, \mathcal{H}_{(N)}$ (respectivement $\mathcal{G}_0, \mathcal{H}_0$) sont uniformément bornés par $(2M_{\mathcal{F}})^N/\Delta_N$ (respectivement par $(2M_{\mathcal{F}})^{N_0}/\Delta_{N_0}$). Soit $(f_m)_m \subset \tilde{\mathcal{F}}$ convergeant simplement vers $f \in \mathcal{F}$. À partir de

$$\lim_{m \rightarrow +\infty} \mathbf{1}_{A_j^{(N)}}(X) f_m(X) = \mathbf{1}_{A_j^{(N)}}(X) f(X), \quad P(f \mathbf{1}_{A_j^{(N)}}) \leq P(f_m) \leq M_{\mathcal{F}},$$

on en déduit par convergence dominée que $\lim_{m \rightarrow +\infty} \mathbb{E}[f_m | A_j^{(N)}] = \mathbb{E}[f | A_j^{(N)}]$. Par conséquence $\phi_{(N),j}(f_m)$ converge simplement vers $\phi_{(N),j}(f)$ et $\phi_{(N)}(f_m) = \sum_{j=1}^{m_N} \phi_{(N),j}(f_m)$ vers $\phi_{(N)}(f) = \sum_{j=1}^{m_N} \phi_{(N),j}(f)$. En itérant ce raisonnement de N à 1 on obtient que $\mathcal{G}_{(N)}, \mathcal{H}_{(N)}$ vérifie la condition $\mathcal{F}.ii$ avec pour sous-classes dénombrables $\tilde{\mathcal{G}}_{(N)}$ et $\tilde{\mathcal{H}}_{(N)}$ définies par

$$\tilde{\mathcal{G}}_{(N)} = \phi_{(1)} \circ \dots \circ \phi_{(N)}(\tilde{\mathcal{F}}), \quad \tilde{\mathcal{H}}_{(N)} = \bigcup_{1 \leq k \leq N} \bigcup_{1 \leq j \leq m_k} \phi_{(k),j} \circ \phi_{(k+1)} \circ \dots \circ \phi_{(N)}(\tilde{\mathcal{F}}).$$

Si \mathcal{F} vérifie VC, il en va de même pour $\mathcal{G}_{(N)}$ et $\mathcal{H}_{(N)}$. En effet d'après la proposition 3.3.1, pour tout $f \in \mathcal{F}$, $\phi_{(1)} \circ \dots \circ \phi_{(N)}(f) = f - \sum_{\substack{1 \leq i \leq N \\ 1 \leq j_i \leq m_i}} \phi^{(N)}(f, i, j_i) \mathbf{1}_{A_{j_i}^{(i)}}$, avec $\phi^{(N)}(f, i, j_i)$ définie par (3.13). Si $d_Q(f, f_Q) < \varepsilon$ et $d_P(f, f_P) < \varepsilon$ avec $f_P, f_Q \in \mathcal{F}$ alors $(\phi^{(N)}(f, i, j_i) - \phi^{(N)}(f_P, i, j_i))^2 \leq 2^{N-i+1} \varepsilon^2$, ce qui entraîne que

$$\begin{aligned} & d_Q^2 \left(\phi_{(1)} \circ \dots \circ \phi_{(N)}(f), f_Q - \sum_{\substack{1 \leq i \leq N \\ 1 \leq j_i \leq m_i}} \phi^{(N)}(f_P, i, j_i) \mathbf{1}_{A_{j_i}^{(i)}} \right) \\ & \leq 2d_Q^2(f, f_Q) + 2d_Q^2 \left(\sum_{\substack{1 \leq i \leq N \\ 1 \leq j_i \leq m_i}} \phi^{(N)}(f, i, j_i) \mathbf{1}_{A_{j_i}^{(i)}}, \sum_{\substack{1 \leq i \leq N \\ 1 \leq j_i \leq m_i}} \phi^{(N)}(f_P, i, j_i) \mathbf{1}_{A_{j_i}^{(i)}} \right) \\ & \leq 2\varepsilon^2 + 4 \sum_{\substack{1 \leq i \leq N \\ 1 \leq j_i \leq m_i}} Q(A_{j_i}^{(i)}) (\phi^{(N)}(f, i, j_i) - \phi^{(N)}(f_P, i, j_i))^2 \leq 2^{N+3} \varepsilon^2. \end{aligned}$$

Si \mathcal{F} peut être recouvert par au plus \mathcal{N} boules de rayon ε par rapport à la distance d_Q centrées en des fonctions $(f_{Q,i})_{1 \leq i \leq \mathcal{N}}$ et par au plus \mathcal{N} boules de rayon ε par rapport à la distance d_P centrées en les fonctions $(f_{P,i})_{1 \leq i \leq \mathcal{N}}$ alors $\mathcal{G}_{(N)}$ peut être recouvert par au plus \mathcal{N}^2 boules de rayon $2^{(N+3)/2} \varepsilon$ et centrées en les fonctions $\left(f_{Q,k} - \sum_{\substack{1 \leq i \leq N \\ 1 \leq j_i \leq m_i}} \phi^{(N)}(f_{P,l}, i, j_i) \mathbf{1}_{A_{j_i}^{(i)}} \right)_{1 \leq k, l \leq \mathcal{N}}$.

Toutes les fonctions $\left(\left(f_{Q,k} - \sum_{\substack{1 \leq i \leq N \\ 1 \leq j_i \leq m_i}} \phi^{(N)}(f_{P,l}, i, j_i) \mathbf{1}_{A_{j_i}^{(i)}} \right) \mathbf{1}_{A_r^{(s)}} \right)_{1 \leq k, l \leq \mathcal{N}, 1 \leq s \leq N, 1 \leq r \leq m_s}$ suffisent à recouvrir $\mathcal{H}_{(N)}$, c'est-à-dire qu'il faut au plus $S_N \mathcal{N}^2$ boules de rayon $2^{(N+3)/2} \varepsilon$ pour recouvrir cette classe. Les classes $\mathcal{G}_0, \mathcal{H}_0$ vérifient également VC avec la même puissance ν_0 et les constantes $c_0(N_0+1), c_0 \sum_{N=0}^{N_0} S_N$. De la même manière, supposons que \mathcal{F} vérifie BR et montrons que ces classes vérifient également cette hypothèse. Supposons que $f_- < f < f_+$ avec $d_P(f_+, f_-) < \varepsilon$. On pose

$$h_{(N),j}^- = (f_- - \mathbb{E}[f_+ | A_j^{(N)}]) \mathbf{1}_{A_j^{(N)}} \leq \phi_{(N),j}(f) \leq (f_+ - \mathbb{E}[f_- | A_j^{(N)}]) \mathbf{1}_{A_j^{(N)}} = h_{(N),j}^+,$$

et la distance d_P du crochet $[h_{(N),j}^-, h_{(N),j}^+]$ est donnée par

$$\begin{aligned} d_P^2(h_{(N),j}^-, h_{(N),j}^+) &= \int_{A_j^{(N)}} \left(f_+ - f_- + \mathbb{E}[f_+ - f_- | A_j^{(N)}] \right)^2 dP \\ &= \mathbb{E}[(f_+ - f_-)^2 \mathbb{1}_{A_j^{(N)}}] + P(A_j^{(N)}) \mathbb{E}[f_+ - f_- | A_j^{(N)}]^2 \\ &\quad + 2\mathbb{E}[f_+ - f_- | A_j^{(N)}] \mathbb{E}[(f_+ - f_-) \mathbb{1}_{A_j^{(N)}}]. \end{aligned}$$

D'après l'inégalité de Hölder, $\mathbb{E}[(f_+ - f_-) \mathbb{1}_{A_j^{(N)}}] \leq \varepsilon \sqrt{P(A_j^{(N)})}$ et $\mathbb{E}[f_+ - f_- | A_j^{(N)}] \leq \varepsilon / \sqrt{P(A_j^{(N)})}$ d'où

$$d_P^2(h_{(N),j}^-, h_{(N),j}^+) \leq \mathbb{E}[(f_+ - f_-)^2 \mathbb{1}_{A_j^{(N)}}] + 3\varepsilon^2,$$

et par suite $\phi_{(N)}(f) = \sum_{j=1}^{m_N} \phi_{(N),j}(f) \in [h_{(N)}^-, h_{(N)}^+]$ où $h_{(N)}^\pm = \sum_{j=1}^{m_N} h_{(N),j}^\pm$ satisfait

$$d_P^2(h_{(N)}^-, h_{(N)}^+) = \sum_{j=1}^{m_N} d_P^2(h_{(N),j}^-, h_{(N),j}^+) \leq d_P^2(f_-, f_+) + 3m_N \varepsilon^2 \leq 4m_N \varepsilon^2.$$

Il s'ensuit que

$$\begin{aligned} N_{\lfloor \cdot \rfloor}(\phi_{(N)}(\mathcal{F}), \varepsilon, d_P) &\leq N_{\lfloor \cdot \rfloor}(\mathcal{F}, \varepsilon/2\sqrt{m_N}, d_P), \\ N_{\lfloor \cdot \rfloor}(\mathcal{G}_{(N)}, \varepsilon, d_P) &\leq N_{\lfloor \cdot \rfloor}(\mathcal{F}, \varepsilon/2^N \sqrt{M_N}, d_P). \end{aligned}$$

Pour couvrir $\phi_{(k),j} \circ \phi_{(k+1)} \circ \dots \circ \phi_{(N)}(\mathcal{F})$ on a besoin d'au plus $m_k N_{\lfloor \cdot \rfloor}(\mathcal{F}, \varepsilon/2^{N-k} \sqrt{m_{k+1} \dots m_N}, d_P)$ crochets. On a prouvé alors que

$$\begin{aligned} N_{\lfloor \cdot \rfloor}(\mathcal{G}_0, \varepsilon, d_P) &\leq (N_0 + 1) N_{\lfloor \cdot \rfloor}(\mathcal{F}, \varepsilon/2^{N_0} \sqrt{M_{N_0}}, d_P), \\ N_{\lfloor \cdot \rfloor}(\mathcal{H}_0, \varepsilon, d_P) &\leq S_{N_0} N_{\lfloor \cdot \rfloor}(\mathcal{F}, \varepsilon/2^{N_0} \sqrt{M_{N_0}}, d_P). \end{aligned}$$

Par conséquence $\mathcal{G}_0, \mathcal{H}_0$ satisfont BR avec la puissance r_0 et la constante $2^{r_0 N_0} M_{N_0}^{r_0} b_0$. \square

Au vue de (3.21) on voit que $\alpha_n^{(0)}(\mathcal{G}_{(N)})$ est le processus contribuant principalement à $\alpha_n^{(N)}(\mathcal{F})$. Le processus $F_n^{(N)}$ agit comme un processus d'erreur, il est donc nécessaire de l'étudier en détail :

Lemme 3.3.6. *Soit \mathcal{F} vérifiant les conditions $\mathcal{F}.i$, $\mathcal{F}.ii$, BR ou VC. Il existe $C_0 < +\infty$ tel que presque sûrement pour n suffisamment grand,*

$$\max_{0 \leq N \leq N_0} \|F_n^{(N)}\|_{\mathcal{F}} \leq \frac{C_0 LL(n)}{\sqrt{n}}.$$

De plus, pour tout $\theta, \zeta > 0$, il existe une suite v_n de la forme $v_n = n^{-\alpha} (\log n)^\beta$ si \mathcal{F} vérifie VC et $v_n = (\log n)^{-\beta}$ si \mathcal{F} vérifie BR (avec $\alpha, \beta > 0$) et des variables $(X_i)_{i \in \mathbb{N}^*}$ i.i.d. de loi P vérifiant pour tout n suffisamment grand,

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

Démonstration. Grâce au lemme 3.3.5 on peut appliquer la proposition 3.2.3 aux classes \mathcal{H}_0 et $\mathcal{H}_{(N)}$. En prenant également en compte (2.2.6), pour n suffisamment grand on a presque sûrement que

$$\begin{aligned} \max_{1 \leq N \leq N_0} \max_{1 \leq j \leq m_N} |\alpha_n^{(N-1)}(A_j^{(N)})| &\leq b_n = 2\sigma_{\mathcal{F}}\kappa_{N_0}\sqrt{LL(n)}, \\ \max_{1 \leq N \leq N_0} \max_{1 \leq k \leq N} \max_{1 \leq j \leq m_k} |\alpha_n^{(k-1)}(\phi_{(k),j} \circ \phi_{(k+1)} \circ \dots \circ \phi_{(N)}(f))| &\leq \max_{1 \leq k \leq N_0} \|\alpha_n^{(k-1)}\|_{\mathcal{H}_0} \leq b_n, \end{aligned}$$

où $\kappa_{N_0} = \prod_{N=1}^{N_0} (1 + M_{\mathcal{F}}/\delta_N)$. Avec des valeurs de n suffisamment grande pour avoir $b_n/\sqrt{n}\delta_{(N_0)} \leq 1/2$ on obtient la majoration des erreurs $q_n(j, N) = \frac{1}{1 - \alpha_n^{(N-1)}(A_j^{(N)})/(\sqrt{n}P(A_j^{(N)}))}$ 1 suivante

$$\max_{1 \leq N \leq N_0} \max_{1 \leq j \leq m_N} |q_n(j, N)| \leq \frac{2b_n}{\sqrt{n}\delta_{(N_0)}}.$$

Ce qui implique

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

On conclut la première inégalité en prenant $C_0 = 8\sigma_{\mathcal{F}}^2\kappa_{N_0}^2 S_{N_0}/\delta_{(N_0)}$. Pour établir la seconde inégalité on travaille sur l'événement B_{n, N_0} défini par (3.5). Il existe un rang n_1 à partir duquel $S_{N_0}(1 - \delta_{(N_0)})^n \leq 1/4n^\theta$ pour tout $n > n_1$. Soit $\varepsilon > 0$ tel que $n^{2\varepsilon} = o(\sqrt{n}v_n)$ qui vérifie alors pour n suffisamment grand $\zeta v_n > 2S_{N_0}n^{2\varepsilon}/\delta_{(N_0)}\sqrt{n}$. Alors pour des valeurs de n suffisamment grandes,

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

Par l'inégalité de Talagrand donnée par la proposition 3.2.4, quelque soit la condition d'entropie respectée, la dernière probabilité peut être bornée par $1/8n^\theta$ pour n suffisamment grand. \square

Nous avons les outils suffisants pour démontrer le résultat d'approximation forte suivant.

Théorème 3.3.7. *Si \mathcal{F} vérifie les conditions $\mathcal{F}.i$, $\mathcal{F}.ii$, BR ou VC alors pour tout $\theta > 0$, il existe $C_\theta > 0$, une suite v_n de la forme $v_n = n^{-\alpha}(\log n)^\beta$ si \mathcal{F} vérifie VC et $v_n = (\log n)^{-\beta}$ si \mathcal{F} vérifie BR (avec $\alpha, \beta > 0$), des variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi P et une suite (\mathbb{G}_n) de P -ponts browniens indépendants définis sur un même espace de probabilité vérifiant pour tout n suffisamment grand,*

$$\mathbb{P}\left(\max_{0 \leq N \leq N_0} \|\alpha_n^{(N)} - \mathbb{G}_n^{(N)}\|_{\mathcal{F}} > C_\theta v_n\right) \leq \frac{1}{n^\theta}, \quad (3.22)$$

et presque sûrement

$$\max_{0 \leq N \leq N_0} \|\alpha_n^{(N)} - \mathbb{G}_n^{(N)}\|_{\mathcal{F}} = O(v_n), \quad (3.23)$$

où $\mathbb{G}_n^{(N)}(\mathcal{F})$ est une version de $\mathbb{G}^{(N)}(\mathcal{F})$ construite à partir de $\mathbb{G}_n^{(0)} = \mathbb{G}_n$ via (3.12).

Démonstration. On se fixe $0 \leq N \leq N_0$. D'après le lemme 3.3.5 on peut appliquer le résultat de Berthet-Mason à la classe \mathcal{G}_0 . Avec un certain $D_\theta > 0$ et $n_\theta > 0$ on peut construire sur un même espace de probabilité des variables $(X_n)_n$ indépendantes et de loi P et une suite $\mathbb{G}_n^{(0)}(\mathcal{G}_0)$ de versions de $\mathbb{G}^{(0)}(\mathcal{G}_0)$ tel que pour tout $n \geq n_\theta$,

$$\mathbb{P}\left(\|\alpha_n^{(0)} - \mathbb{G}_n^{(0)}\|_{\mathcal{G}_0} > D_\theta v_n\right) \leq \frac{1}{2n^\theta}, \quad (3.24)$$

Par construction de la suite $(\mathbb{G}_n^{(N)})_N$ et en utilisant (3.21) on peut écrire que

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

En utilisant (3.24), le lemme 3.3.6 et en posant $C_\theta = D_\theta + \zeta$, on obtient pour n suffisamment grand (3.22). Le lemme de Borell-Cantelli appliqué à (3.22) avec $\theta > 1$ entraîne (3.23). \square

Avec le résultat précédent on a nécessairement que le processus empirique raké $\alpha_n^{(N)}(\mathcal{F})$ converge en loi dans $\ell^\infty(\mathcal{F})$ vers le P -pont brownien raké N fois $\mathbb{G}^{(N)}(\mathcal{F})$ défini par (3.12). Nous pouvons même grâce à l'approximation forte étendre cette convergence au vecteur aléatoire constitué des processus empiriques rakés à tous les tours avant N_0 vers le vecteur des P -pont browniens rakés correspondant.

Proposition 3.3.8. *Si \mathcal{F} vérifie $\mathcal{F}.i$, BR ou VC alors pour tout $N_0 \in \mathbb{N}$, le vecteur $(\alpha_n^{(0)}(\mathcal{F}), \dots, \alpha_n^{(N_0)}(\mathcal{F}))$ converge en loi dans $\ell^\infty(\mathcal{F} \rightarrow \mathbb{R}^{N_0+1})$ quand $n \rightarrow +\infty$ vers le vecteur $(\mathbb{G}^{(0)}(\mathcal{F}), \dots, \mathbb{G}^{(N_0)}(\mathcal{F}))$.*

L'approximation forte du processus empirique raké nous permet de faciliter les études de convergence faible de statistiques faisant intervenir $\alpha_n^{(N)}(\mathcal{F})$ en substituant ce dernier par $\mathbb{G}_n^{(N)}(\mathcal{F})$. Elle nous permet également d'exploiter les propriétés limites de ces statistiques à travers celles de $\mathbb{G}_n^{(N)}(\mathcal{F})$. Des exemples sont donnés dans les deux paragraphes qui suivent.

Inégalité de Berry-Esseen. Une conséquence possible et immédiate du résultat d'approximation forte donné par le théorème 3.3.7 est l'établissement d'une borne du type Berry-Esseen pour la fonction de répartition du processus empirique raké. On note Φ la fonction de répartition de la loi normale centrée réduite et \mathcal{L} un ensemble de fonctions ϕ lipschitziennes définies sur $\ell^\infty(\mathcal{F})$ dont la constante de Lipschitz est bornée par $C_1 < +\infty$ et telles que $\phi(\mathbb{G}^{(N)})$ ait une densité bornée par $C_2 < +\infty$ pour tout $0 \leq N \leq N_0$ et $\phi \in \mathcal{L}$.

Proposition 3.3.9. *Soit $N_0 \in \mathbb{N}$ et $\mathcal{F}_0 \subset \mathcal{F}$ tel que*

$$\sigma_0^2 = \inf\{\text{Var}(\mathbb{G}^{(N)}(f)) : f \in \mathcal{F}_0, 0 \leq N \leq N_0\} > 0.$$

Si \mathcal{F} vérifie les conditions $\mathcal{F}.i$, $\mathcal{F}.ii$, BR ou VC alors presque sûrement il existe $n_0 \in \mathbb{N}^*$ tel que pour tout $n \geq n_0$,

$$\max_{0 \leq N \leq N_0} \sup_{\phi \in \mathcal{L}, f \in \mathcal{F}_0} \sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\phi(\alpha_n^{(N)}) \leq x \right) - \mathbb{P} \left(\phi(\mathbb{G}^{(N)}) \leq x \right) \right| \leq C_0 C_1 C_2 v_n, \quad (3.25)$$

avec $C_0 > C$ et $C > 0$ donné par le théorème 3.3.7. En particulier, pour $n \geq n_0$,

$$\begin{aligned} & \max_{0 \leq N \leq N_0} \sup_{f \in \mathcal{F}_0} \sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\alpha_n^{(N)}(f) \leq x \right) - \mathbb{P} \left(\mathbb{G}^{(N)}(f) \leq x \right) \right| \\ &= \max_{0 \leq N \leq N_0} \sup_{f \in \mathcal{F}_0} \sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\sqrt{n} \frac{\mathbb{P}_n^{(N)}(f) - P(f)}{\sqrt{\text{Var}(\mathbb{G}^{(N)}(f))}} \leq x \right) - \Phi(x) \right| \leq \frac{C}{\sqrt{2\pi}\sigma_0} v_n. \end{aligned} \quad (3.26)$$

Remarque 3.3.3. Les fonctions $f \in \mathcal{F}$ surdéterminées par l'information auxiliaire apportée par les $P(\mathcal{A}^{(N)})$ ont des faibles variances $\text{Var}(\mathbb{G}^{(N)}(f))$ et doivent être exclues de la classe \mathcal{F}_0 . La proposition 3.3.9 est particulièrement utile dans le cas où \mathcal{F} vérifie VC puisque v_n décroît polynomialement et autorise de travailler avec des grandes valeurs de C_1, C_2 .

Démonstration. On applique le théorème 3.3.7 : pour tout $0 \leq N \leq N_0$, $\phi \in \mathcal{L}$ et $n \geq n_0$,

$$\begin{aligned} \mathbb{P} \left(\varphi(\alpha_n^{(N)}) \leq x \right) &\leq \frac{1}{n^2} + \mathbb{P} \left(\left\{ \varphi(\alpha_n^{(N)}) \leq x \right\} \cap \left\{ \left\| \alpha_n^{(N)} - \mathbb{G}_n^{(N)} \right\|_{\mathcal{F}} < C v_n \right\} \right) \\ &\leq \frac{1}{n^2} + \mathbb{P} \left(\varphi(\mathbb{G}_n^{(N)}) \leq x + C C_1 v_n \right) \\ &\leq \frac{1}{n^2} + \mathbb{P} \left(\varphi(\mathbb{G}_n^{(N)}) \leq x \right) + C C_1 C_2 v_n, \end{aligned}$$

et

$$\begin{aligned} & \mathbb{P} \left(\varphi(\mathbb{G}_n^{(N)}) \leq x - C C_1 v_n \right) \\ &\leq \frac{1}{n^2} + \mathbb{P} \left(\left\{ \varphi(\mathbb{G}_n^{(N)}) \leq x - C C_1 v_n \right\} \cap \left\{ \left\| \alpha_n^{(N)} - \mathbb{G}_n^{(N)} \right\|_{\mathcal{F}} < C v_n \right\} \right) \\ &\leq \frac{1}{n^2} + \mathbb{P} \left(\varphi(\alpha_n^{(N)}) \leq \varphi(\mathbb{G}_n^{(N)}) + C C_1 v_n \leq x \right), \end{aligned}$$

de telle sorte que

$$\mathbb{P} \left(\varphi(\alpha_n^{(N)}) \leq x \right) \geq \mathbb{P} \left(\varphi(\mathbb{G}_n^{(N)}) \leq x \right) - C C_1 C_2 v_n - \frac{1}{n^2}.$$

L'inégalité (3.25) est montrée pour $C_0 > C$ et $n \geq n_0$ suffisamment grand pour avoir $1/n^2 \leq (C_0 - C)C_1 C_2 v_n$. L'inégalité (3.26) est une conséquence de la première inégalité en prenant le cas spécial où $\mathcal{L} = \{\phi_f : f \in \mathcal{F}_0\}$ avec les fonctions lipschitziennes $\phi_f(g) = g(f)$ pour $g \in \ell^\infty(\mathcal{F})$, dont les constantes de Lipschitz sont bornées par 1 qui vérifie que $\phi_f(\mathbb{G}_n^{(N)}) = \mathbb{G}_n^{(N)}$ et qui ont une densité gaussienne bornée par

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

□

Nous donnons une application de ce résultat dans l'exemple suivant :

Exemple 3.3.1 (Matrices de covariances rakées). Soient $d \in \mathbb{N}^*$ et $f_1, \dots, f_d \in \mathcal{F}$. On note $\text{Cov}(Y)$ la matrice de covariance du vecteur aléatoire $Y = (f_1(X), \dots, f_d(X))$ que nous supposons centré par soucis de simplicité. À la place de la covariance empirique $\text{Cov}_n(Y) = n^{-1} \sum_{i=1}^n Y_i^t Y_i$ on considère la version rakée

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

On note $\|\cdot\|$ la norme de Frobenius et on pose

$$\phi_Y(\alpha_n^{(N)}) = \sqrt{n} \left\| \text{Cov}_n^{(N)}(Y) - \text{Cov}(Y) \right\|.$$

En d'autres termes,

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

On observe dans le contexte de la proposition 3.3.9 que la fonction ϕ_Y est $(\|\cdot\|_{\mathcal{F}}, \|\cdot\|)$ -Lipschitz de constante $C_1 = d$. Clairement $\phi_Y(\mathbb{G}^{(N)})$ possède une densité bornée car $\phi_Y^2(\mathbb{G}^{(N)})$ est une forme quadratique de composants gaussiens et possède donc une distribution de χ^2 modifiée. Prendre un ensemble fini de ϕ_Y nous assurerait l'hypothèse $C_2 < +\infty$. Plus généralement, on suppose que (f_1, \dots, f_d) appartient à un ensemble infini $\mathcal{L}_d \subset \mathcal{F}^d$ de petite entropie vérifiant des conditions de régularité telle que l'on ait $C_2 < +\infty$ et $\{f_i f_j : f_i, f_j \in \mathcal{F}\}$ qui vérifie BR. La plus grande constante C_2 des ensembles $(\mathcal{L}_d)_{d \leq d_0}$ reste valable pour $\mathcal{L} = \cup_{d \leq d_0} \mathcal{L}_d$. Alors la proposition 3.3.9 garantit que

$$\max_{\substack{0 \leq N \leq N_0 \\ d \leq d_0}} \sup_{\substack{(f_1, \dots, f_d) \in \mathcal{L}_d \\ x > 0}} \left| \mathbb{P} \left(\phi_Y(\alpha_n^{(N)}) \leq x \right) - \mathbb{P} \left(\phi_Y(\mathbb{G}^{(N)}) \leq x \right) \right| \leq d_0 C_0 C_1 C_2 v_n,$$

avec $C_0 > C$ et $C > 0$ donné par le théorème 3.3.7. Or pour tout $N \leq N_0, d \leq d_0, (f_1, \dots, f_d) \in \mathcal{L}_d$ et $x > 0$,

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

par la propriété de réduction de variance de la proposition 3.3.3 et notamment de la remarque 3.3.1. Asymptotiquement nous avons donc $\mathbb{P}(\phi_Y(\alpha_n^{(N)}) \leq x) < \mathbb{P}(\phi_Y(\alpha_n^{(0)}) \leq x) - \varepsilon$ uniformément sur les Y tels que $\mathbb{P}(\phi_Y(\mathbb{G}^{(N)}) \leq x) < \mathbb{P}(\phi_Y(\mathbb{G}^{(0)}) \leq x) - 2\varepsilon$ pour tout $\varepsilon > 0$ fixé.

Biais et estimation de la variance. La puissance de l'approximation forte donnée par le théorème 3.3.7 nous permet d'établir le résultat suivant concernant le biais et l'estimation de la variance du processus empirique raké :

Proposition 3.3.10. Si \mathcal{F} vérifie $\mathcal{F}.i$, $\mathcal{F}.ii$, VC ou BR alors pour tout $N_0 \in \mathbb{N}$,

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \frac{\sqrt{n}}{v_n} \max_{0 \leq N \leq N_0} \sup_{f \in \mathcal{F}} \left| \mathbb{E}[\mathbb{P}_n^{(N)}(f)] - P(f) \right| &\leq C, \\ \limsup_{n \rightarrow +\infty} \frac{n}{v_n} \sup_{f, g \in \mathcal{F}} \left| \text{Cov}(\mathbb{P}_n^{(N)}(f), \mathbb{P}_n^{(N)}(g)) - \frac{1}{n} \text{Cov}(\mathbb{G}^{(N)}(f), \mathbb{G}^{(N)}(g)) \right| &\leq \sqrt{\frac{8}{\pi}} C \sigma_{\mathcal{F}}, \\ \limsup_{n \rightarrow +\infty} \frac{n}{v_n} \sup_{f \in \mathcal{F}} \left| \text{Var}(\mathbb{P}_n^{(N)}(f)) - \frac{1}{n} \text{Var}(\mathbb{G}^{(N)}(f)) \right| &\leq \sqrt{\frac{8}{\pi}} C \sigma_{\mathcal{F}}, \end{aligned}$$

où $C > 0$ et $(v_n)_{n \in \mathbb{N}^*}$ sont donnés par le théorème 3.3.7 et $\sigma_{\mathcal{F}}^2 = \sup_{f \in \mathcal{F}} \text{Var}(f(X))$.

Démonstration. (i). Le théorème 3.3.7 implique pour tout $f \in \mathcal{F}$,

$$\left(\mathbb{P}_n^{(N)} - P\right)(f) = \frac{1}{\sqrt{n}}\mathbb{G}_n^{(N)}(f) + \frac{1}{\sqrt{n}}\mathbb{R}_n^{(N)}(f), \quad (3.27)$$

où $\mathbb{G}_n^{(N)}(\mathcal{F})$ est une suite de versions du processus gaussien $\mathbb{G}^{(N)}(\mathcal{F})$ définie par (3.12) et la suite aléatoire $r_n^{(N)} = \|\mathbb{R}_n^{(N)}\|_{\mathcal{F}}$ vérifie

$$r_n^{(N)} \leq \|\mathbb{G}_n^{(N)}\|_{\mathcal{F}} + \|\alpha_n^{(N)}\|_{\mathcal{F}} \leq \|\mathbb{G}_n^{(N)}\|_{\mathcal{F}} + 2M_{\mathcal{F}}\sqrt{n}, \quad \lim_{n \rightarrow +\infty} \frac{r_n^{(N)}}{v_n} \leq C \text{ p.s.}$$

Nous devons faire attention à l'espérance, la variance et la covariance de l'erreur d'approximation $R_n^{(N)}(\mathcal{F})$. Puisque $\mathbb{G}_n^{(N)}(\mathcal{F})$ est un processus gaussien centré, le biais de $\mathbb{P}_n^{(N)}(\mathcal{F})$ est borné par

$$\sup_{f \in \mathcal{F}} \frac{\sqrt{n}}{v_n} \left| \mathbb{E}[\mathbb{P}_n^{(N)}(f)] - P(f) \right| = \sup_{f \in \mathcal{F}} \left| \frac{1}{v_n} \mathbb{E}[\mathbb{R}_n^{(N)}(f)] \right| \leq \mathbb{E} \left[\frac{r_n^{(N)}}{v_n} \right]. \quad (3.28)$$

On se fixe des paramètres $K > 0, \theta_0 > 1$ aussi grands que l'on souhaite et on pose $a_n = \sqrt{K \log n}$. Pour $\gamma > 1, \varepsilon > 0, \theta > 1$ et $k \in \mathbb{N}^*$ on pose les événements

$$A_n = \{r_n^{(N)} \leq (C_\theta + \varepsilon)v_n\}, \quad B_n = \{\|\mathbb{G}_n^{(N)}\|_{\mathcal{F}} \leq a_n\}, \quad C_{n,k} = \{\gamma^{k-1}a_n \leq \|\mathbb{G}_n^{(N)}\|_{\mathcal{F}} \leq \theta^k a_n\}.$$

D'après le théorème d'approximation forte 3.3.7 appliqué au paramètre θ_0 , $\mathbb{P}(A_n^C) < 1/n^{\theta_0}$. Pour n suffisamment grand, $v_n > a_n/\sqrt{n}$ et par suite

$$\begin{aligned} \frac{1}{v_n} \mathbb{E}[r_n^{(N)}] &= \mathbb{E} \left[\frac{r_n^{(N)}}{v_n} \mathbf{1}_{A_n} \right] + \mathbb{E} \left[\frac{r_n^{(N)}}{v_n} \mathbf{1}_{A_n^C \cap B_n} \right] + \mathbb{E} \left[\frac{r_n^{(N)}}{v_n} \mathbf{1}_{A_n^C \cap B_n^C} \right] \\ &\leq (C_\theta + \varepsilon) + \frac{a_n + 2M_{\mathcal{F}}\sqrt{n}}{v_n} \mathbb{P}(A_n^C) + \mathbb{E} \left[\frac{r_n^{(N)}}{v_n} \mathbf{1}_{B_n^C} \right] \\ &\leq C_\theta + 2\varepsilon + \sum_{k=1}^{+\infty} \mathbb{E} \left[\frac{r_n^{(N)}}{v_n} \mathbf{1}_{C_{n,k}} \right]. \end{aligned}$$

D'après le théorème 3.3.3, $\mathbb{G}^{(N)}(\mathcal{F})$ vérifie sous BR ou VC,

$$\begin{aligned} \mathbb{E}[\|\mathbb{G}^{(N)}\|_{\mathcal{F}}] &< +\infty, \quad \sup_{f \in \mathcal{F}} \text{Var}(\mathbb{G}^{(N)}(f)) \leq \sigma_{\mathcal{F}}^2 < +\infty, \\ \mathbb{E}[\|\mathbb{G}^{(N)}\|_{\mathcal{F}}^2] &\leq C_{\mathcal{F}}^2 = \sigma_{\mathcal{F}}^2 + \mathbb{E}[\|\mathbb{G}^{(N)}\|_{\mathcal{F}}]^2 < +\infty, \end{aligned}$$

ce qui entraîne d'après l'inégalité de Borell-Sudakov, pour toute version $\mathbb{G}_n^{(N)}$ de $\mathbb{G}^{(N)}$,

$$\mathbb{P}(\|\mathbb{G}_n^{(N)}\|_{\mathcal{F}} > \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{8C_{\mathcal{F}}^2}\right). \quad (3.29)$$

Ainsi en utilisant (3.29) on obtient pour $\theta > 1$ et des valeurs de n suffisamment grandes pour avoir $v_n > 4M_{\mathcal{F}}/\sqrt{n} > 2a_n/n$,

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

et la série suivante converge vers une somme arbitrairement petite,

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

où $\delta < K/8C_{\mathcal{F}}^2 - 1$. Il s'ensuit que (3.28) est finalement bornée par C_θ .

(ii). En utilisant (3.27) et la décomposition biais-variance, le risque quadratique est contrôlé par

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

D'après la première étape, le premier terme est de l'ordre de $C_\theta^2 v_n^2/n$. Il nous reste à borner les deux derniers termes. Soit $\varepsilon > 0$, $s_n^{(N)} = (r_n^{(N)})^2$ et n suffisamment grand pour avoir $v_n > a_n/\sqrt{n}$, $a_n = K\sqrt{\log n} < \sqrt{n}$. On obtient pour $\theta_0 = 2$,

$$\begin{aligned} \frac{1}{v_n^2} \sup_{f \in \mathcal{F}} \mathbb{E} \left[\mathbb{R}_n^{(N)}(f)^2 \right] &\leq \mathbb{E} \left[\frac{s_n^{(N)}}{v_n^2} \mathbf{1}_{A_n} \right] + \mathbb{E} \left[\frac{s_n^{(N)}}{v_n^2} \mathbf{1}_{A_n^c \cap B_n} \right] + \mathbb{E} \left[\frac{s_n^{(N)}}{v_n^2} \mathbf{1}_{A_n^c \cap B_n^c} \right] \\ &\leq (C_\theta + \varepsilon)^2 + \left[\frac{a_n + 2M_{\mathcal{F}}\sqrt{n}}{v_n} \right]^2 \mathbb{P}(A_n^c) + \mathbb{E} \left[\frac{s_n^{(N)}}{v_n^2} \mathbf{1}_{B_n^c} \right] \\ &\leq (C_\theta + 2\varepsilon)^2 + \left(\frac{3M_{\mathcal{F}}\sqrt{n}}{\log n} \right)^2 \frac{1}{n^2} + \sum_{k=1}^{+\infty} \mathbb{E} \left[\frac{s_n^{(N)}}{v_n^2} \mathbf{1}_{C_{n,k}} \right] \\ &\leq (C_\theta + 3\varepsilon)^2 + \sum_{k=1}^{+\infty} \theta^{2k} n^2 \mathbb{P} \left(\|\mathbb{G}_n^{(N)}\|_{\mathcal{F}} > \theta^{k-1} a_n \right) \leq (C_\theta + 4\varepsilon)^2, \end{aligned}$$

la série étant bornée par son premier terme $n^2 \exp(-a_n^2/8C_{\mathcal{F}}^2)$ multiplié par une série convergente, en utilisant à nouveau (3.29) avec $K > 16C_{\mathcal{F}}^2$. Nous avons borné notre premier terme par

$$\limsup_{n \rightarrow +\infty} \frac{1}{v_n^2} \sup_{f \in \mathcal{F}} \text{Var} \left(\mathbb{R}_n^{(N)}(f) \right) \leq \limsup_{n \rightarrow +\infty} \frac{1}{v_n^2} \mathbb{E}[s_n^{(N)}] \leq C_\theta^2.$$

Concernant le terme avec la covariance il vient

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

où

$$T_D(f) = \mathbb{E} \left[\left| \mathbb{G}_n^{(N)}(f) \right| \frac{r_n^{(N)}}{v_n} \mathbf{1}_D \right], \quad D \in \{A_n, A_n^c \cap B_n, A_n^c \cap B_n^c\}.$$

Or,

$$T_{A_n}(f) \leq \mathbb{E} \left[\left| \mathbb{G}_n^{(N)}(f) \right| (C_\theta + \varepsilon) \mathbf{1}_{A_n} \right] \leq (C_\theta + \varepsilon) \sigma_{\mathcal{F}} \mathbb{E}[\|\mathcal{N}(0, 1)\|].$$

En utilisant le fait que $\mathbb{P}(A_n^C) < 1/n^2$ on voit que

$$T_{A_n^C \cap B_n}(f) \leq \mathbb{E} \left[a_n \left(\frac{2M\sqrt{n} + a_n}{v_n} \right) 1_{A_n^C \cap B_n} \right] \leq a_n \left(\frac{3M_{\mathcal{F}}\sqrt{n}}{v_n} \right) \frac{1}{n^2} \leq \varepsilon.$$

Enfin pour $g_n^{(N)} = \|\mathbb{G}_n^{(N)}\|_{\mathcal{F}}$, K et n suffisamment grand, il vient d'après (3.29),

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

La borne établie ci-dessus ne dépend pas de f . Au final,

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

(iii). Étendons l'étape précédente au cas de la covariance. Avec la première étape, pour n suffisamment grand nous avons que

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

Avec les bornes établies à l'étape 2, on a

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

□

3.4 Résultats non-asymptotiques avec deux partitions

Motivation. Dans le cas où on applique le Raking-Ratio itérativement et périodiquement sur un nombre fini de partitions, la loi du logarithme itéré établie par la proposition 3.2.3 est insatisfaisante pour montrer que la mesure $\mathbb{P}_n^{(N)}(\mathcal{F})$ converge quand $N \rightarrow +\infty$. En effet, la borne donnée exige de prendre un n très grand pour compenser des valeurs de N même

petites, à cause de la dépendance exponentielle en N issue de cette proposition. Or comme nous avons précédemment dit que l'algorithme se stabilisait dans ce cas en un nombre fini et très petit de tours, nous avons l'espoir de trouver une borne plus satisfaisante qui prend en compte le fait que l'on ratisse périodiquement.

Notation. Nous allons étudier le cas du ratissage de deux partitions $\mathcal{A} = \{A_1, \dots, A_{m_1}\}$ et $\mathcal{B} = \{B_1, \dots, B_{m_2}\}$, le cas de plusieurs partitions se montre de manière similaire. On peut exprimer la mesure empirique associée à la méthode du Raking-Ratio avec des suites $(a_{n,i}^{(N)})_{1 \leq i \leq m_1}, (b_{n,j}^{(N)})_{1 \leq j \leq m_2}$ de la manière suivante. Pour tout $1 \leq i \leq m_1$ et $1 \leq j \leq m_2$, $a_{n,i}^{(0)} = \mathbb{P}_n(A_i)$, $b_{n,j}^{(0)} = \mathbb{P}_n(B_j)$, si N est pair alors $a_{n,i}^{(N+1)} = P(A_i) / \sum_{j=1}^{m_2} b_{n,j}^{(N)} \mathbb{P}_n(\mathbb{1}_{A_i \cap B_j})$, $b_{n,j}^{(N+1)} = b_{n,j}^{(N)}$, et si N est impair alors $a_{n,i}^{(N+1)} = a_{n,i}^{(N)}$, $b_{n,j}^{(N+1)} = P(B_j) / \sum_{i=1}^{m_1} a_{n,i}^{(N)} \mathbb{P}_n(\mathbb{1}_{A_i \cap B_j})$. Pour tout $N \in \mathbb{N}^*$ la mesure empirique $\mathbb{P}_n^{(N)}(\mathcal{F})$ s'exprime donc avec les coefficients introduits précédemment par

$$\mathbb{P}_n^{(N)}(f) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} a_{n,i}^{(N)} b_{n,j}^{(N)} \mathbb{P}_n(f \mathbb{1}_{A_i \cap B_j}).$$

Elle vérifie comme attendu $\mathbb{P}_n^{(2N-1)}(A_i) = P(A_i)$ et $\mathbb{P}_n^{(2N)}(B_j) = P(B_j)$ pour tout $N \in \mathbb{N}^*$.

Résultats. Ireland et Kullback[50] ont montré que les suites $a_{n,i}^{(N)}$ et $b_{n,j}^{(N)}$ convergeaient quand $N \rightarrow +\infty$ vers des réels $a_{n,i}^{(\infty)}$ et $b_{n,j}^{(\infty)}$ pour tout $1 \leq i \leq m_1$ et $1 \leq j \leq m_2$. Puisque la mesure \mathbb{P}_n converge presque sûrement vers P pour les indicatrices des ensembles $A_i \cap B_j$ et que P vérifie naturellement l'information auxiliaire, nécessairement les réels $a_{n,i}^{(\infty)}, b_{n,j}^{(\infty)}$ sont de limite 1 quand $n \rightarrow +\infty$. On pose $\mathbb{P}_n^{(\infty)}(f) = \sum_{i \leq m_1, j \leq m_2} a_{n,i}^{(\infty)} b_{n,j}^{(\infty)} \mathbb{P}_n(f \mathbb{1}_{A_i \cap B_j})$ la mesure associée au Raking-Ratio après stabilisation *i.e.* quand $N \rightarrow +\infty$. Le résultat suivant montre que la mesure empirique rakée $\mathbb{P}_n^{(N)}$ converge quand $N \rightarrow +\infty$ vers la vraie mesure de probabilité P .

Proposition 3.4.1. *Pour tout $N, n \in \mathbb{N}^*$,*

$$\begin{aligned} \|\mathbb{P}_n^{(N)} - \mathbb{P}_n^{(\infty)}\|_{\mathcal{F}} &\leq M_{\mathcal{F}} \max_{i,j} |a_{n,i}^{(N)} b_{n,j}^{(N)} - a_{n,i}^{(\infty)} b_{n,j}^{(\infty)}|, \\ \|\mathbb{P}_n^{(\infty)} - P\|_{\mathcal{F}} &\leq \frac{1}{\sigma_{\mathcal{F}}} \sqrt{\frac{LL(n)}{2n}} + M_{\mathcal{F}} \max_{i,j} |a_{n,i}^{(\infty)} b_{n,j}^{(\infty)} - 1|, \end{aligned}$$

où les maximums sont pris pour $1 \leq i \leq m_1$ et $1 \leq j \leq m_2$.

Démonstration. Pour tout $n, N \in \mathbb{N}^*$,

$$\begin{aligned} \|\mathbb{P}_n^{(N)} - \mathbb{P}_n^{(\infty)}\|_{\mathcal{F}} &= \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} (a_{n,i}^{(N)} b_{n,j}^{(N)} - a_{n,i}^{(\infty)} b_{n,j}^{(\infty)}) \mathbb{P}_n(f \mathbb{1}_{A_i \cap B_j}) \right| \\ &\leq M_{\mathcal{F}} \max_{i,j} |a_{n,i}^{(N)} b_{n,j}^{(N)} - a_{n,i}^{(\infty)} b_{n,j}^{(\infty)}| \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \mathbb{P}_n(A_i \cap B_j) \\ &\leq M_{\mathcal{F}} \max_{i,j} |a_{n,i}^{(N)} b_{n,j}^{(N)} - a_{n,i}^{(\infty)} b_{n,j}^{(\infty)}|, \\ \|\mathbb{P}_n^{(\infty)} - P\|_{\mathcal{F}} &\leq \|\mathbb{P}_n^{(\infty)} - \mathbb{P}_n\|_{\mathcal{F}} + \|\mathbb{P}_n - P\|_{\mathcal{F}}. \end{aligned}$$

D'après la loi du logarithme itéré donnée par la proposition 3.2.3, $\|\mathbb{P}_n - P\|_{\mathcal{F}} \leq \sqrt{LL(n)/2n\sigma_{\mathcal{F}}^2}$.
De plus,

$$\begin{aligned} \|\mathbb{P}_n^{(\infty)} - \mathbb{P}_n\|_{\mathcal{F}} &= \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} (a_{n,i}^{(\infty)} b_{n,j}^{(\infty)} - 1) \mathbb{P}_n(f \mathbf{1}_{A_i \cap B_j}) \right| \\ &\leq M_{\mathcal{F}} \max_{i,j} |a_{n,i}^{(\infty)} b_{n,j}^{(\infty)} - 1|. \end{aligned}$$

□

3.5 Résultat asymptotiques avec deux partitions

3.5.1 Expression du processus gaussien raké

Motivation Nous allons étudier plus précisément le cas particulier où on applique le Raking Ratio en boucle sur deux ensembles \mathcal{A} et \mathcal{B} . Dans le cas particulier du mélange de deux information auxiliaire, l'expression (3.13) se simplifie.

Notation. On conserve la même notation que la partie précédente : on note $(\mathbf{P}_{\mathcal{A}|\mathcal{B}})_{i,j} = P(A_j|B_i)$, $(\mathbf{P}_{\mathcal{B}|\mathcal{A}})_{i,j} = P(B_j|A_i)$ et

$$\begin{aligned} \mathbb{G}[\mathcal{A}] &= (\mathbb{G}(A_1), \dots, \mathbb{G}(A_{m_1}))^t, & \mathbb{E}^{(1)}[f] &= (\mathbb{E}[f|A_1], \dots, \mathbb{E}[f|A_{m_1}])^t, \\ \mathbb{G}[\mathcal{B}] &= (\mathbb{G}(B_1), \dots, \mathbb{G}(B_{m_2}))^t, & \mathbb{E}^{(2)}[f] &= (\mathbb{E}[f|B_1], \dots, \mathbb{E}[f|B_{m_2}])^t. \end{aligned}$$

L'expression du processus associé à la méthode du Raking Ratio dans le cas de deux informations auxiliaires est donnée par la proposition suivante. Nous conviendrons pour la proposition suivante qu'une somme dont la borne supérieure est négative est nulle et qu'une puissance négative d'une des deux matrices $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ ou $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ est nulle.

Proposition 3.5.1. *Soit $N \in \mathbb{N}$. Si N est pair alors*

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

et si N est impair alors

$$\begin{aligned} \mathbb{G}^{(N)}(f) &= \mathbb{G}(f) - \left[\sum_{k=0}^{\frac{N-3}{2}} (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k \cdot (\mathbb{E}^{(1)}[f] - \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbb{E}^{(2)}[f]) + (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^{(N-1)/2} \mathbb{E}^{(1)}[f] \right]^t \mathbb{G}[\mathcal{A}] \\ &\quad - \left[\sum_{k=0}^{\frac{N-3}{2}} (\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k \cdot (\mathbb{E}^{(2)}[f] - \mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbb{E}^{(1)}[f]) \right]^t \mathbb{G}[\mathcal{B}]. \end{aligned}$$

Démonstration. Nous montrons ce résultat par récurrence double. Pour $N = 0$ et $N = 1$ la proposition est vraie compte tenu du fait que $\mathbb{G}^{(0)} = \mathbb{G}$ par définition et $\mathbb{G}^{(1)} = \hat{\mathbb{G}}$. Supposons

que la propriété soit vérifiée pour un rang $N, N + 1$ avec N pair. Alors

$$\begin{aligned}\mathbb{G}^{(N+2)}(f) &= \mathbb{G}^{(N+1)}(f) - \mathbb{E}^{(2)}[f]^t \cdot \mathbb{G}^{(N+1)}[\mathcal{B}], \\ \mathbb{G}^{(N+3)}(f) &= \mathbb{G}^{(N+2)}(f) - \mathbb{E}^{(1)}[f]^t \cdot \mathbb{G}^{(N+2)}[\mathcal{A}],\end{aligned}$$

où

$$\begin{aligned}\mathbb{G}^{(N+1)}[\mathcal{B}] &:= (\mathbb{G}^{(N+1)}(B_1), \dots, \mathbb{G}^{(N+1)}(B_{m_2}))^t, \\ \mathbb{G}^{(N+2)}[\mathcal{A}] &:= (\mathbb{G}^{(N+2)}(A_1), \dots, \mathbb{G}^{(N+2)}(A_{m_1}))^t.\end{aligned}$$

Grâce à la notation matricielle, on peut dire que

$$\begin{aligned}\mathbb{G}^{(N+1)}[\mathcal{B}] &= \mathbb{G}[\mathcal{B}] - \left[\sum_{k=0}^{\frac{N}{2}-1} (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k (\mathbf{P}_{\mathcal{B}|\mathcal{A}} - \mathbf{P}_{\mathcal{B}|\mathcal{A}}\text{Id}_{m_1}) + (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^{\frac{N}{2}}\mathbf{P}_{\mathcal{B}|\mathcal{A}} \right]^t \mathbb{G}[\mathcal{A}] \\ &\quad - \left[\sum_{k=0}^{\frac{N}{2}-1} (\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k (\text{Id}_{m_2} - \mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}) \right]^t \mathbb{G}[\mathcal{B}] \\ &= - \left[(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^{N/2}\mathbf{P}_{\mathcal{B}|\mathcal{A}} \right]^t \mathbb{G}[\mathcal{A}] + \left[(\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^{N/2} \right]^t \mathbb{G}[\mathcal{B}].\end{aligned}$$

En réunissant les termes qui factorisent $\mathbb{G}[\mathcal{A}]$ et $\mathbb{G}[\mathcal{B}]$ dans les expressions de $\mathbb{G}^{(N+1)}(f)$ et $\mathbb{G}^{(N+1)}[\mathcal{B}]$ on retrouve l'expression voulue au rang $N + 2$. On fait de même pour le rang $N + 3$. \square

On note pour tout $N \geq 0$ les matrices $S_{1,\text{pair}}^{(N)}(f), S_{1,\text{impair}}^{(N)}(f) \in \mathcal{M}_{m_1,1}(\mathbb{R})$ et $S_{2,\text{pair}}^{(N)}(f), S_{2,\text{impair}}^{(N)}(f) \in \mathcal{M}_{m_2,1}(\mathbb{R})$ définies par

$$S_{1,\text{pair}}^{(N)}(f) = \sum_{k=0}^N (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k \cdot (\mathbb{E}^{(1)}[f] - \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbb{E}^{(2)}[f]), \quad (3.30)$$

$$S_{2,\text{impair}}^{(N)}(f) = \sum_{k=0}^N (\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k \cdot (\mathbb{E}^{(2)}[f] - \mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbb{E}^{(1)}[f]), \quad (3.31)$$

$$S_{2,\text{pair}}^{(N)}(f) = S_{2,\text{impair}}^{(N)} + (\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^{N+1}\mathbb{E}^{(2)}[f], \quad (3.32)$$

$$S_{1,\text{impair}}^{(N)}(f) = S_{1,\text{pair}}^{(N)} + (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^{N+1}\mathbb{E}^{(1)}[f]. \quad (3.33)$$

Le lien entre ces matrices et le processus itératif est donné par la proposition 3.5.1. En conservant la convention adoptée plus haut, pour tout $N \in \mathbb{N}^*$,

$$\mathbb{G}^{(2N)}(f) = \mathbb{G}(f) - \left(S_{1,\text{pair}}^{(N-1)}(f) \right)^t \mathbb{G}[\mathcal{A}] - \left(S_{2,\text{pair}}^{(N-2)}(f) \right)^t \mathbb{G}[\mathcal{B}], \quad (3.34)$$

$$\mathbb{G}^{(2N+1)}(f) = \mathbb{G}(f) - \left(S_{1,\text{impair}}^{(N-1)}(f) \right)^t \mathbb{G}[\mathcal{A}] - \left(S_{2,\text{impair}}^{(N-1)}(f) \right)^t \mathbb{G}[\mathcal{B}]. \quad (3.35)$$

3.5.2 Limite du processus gaussien raké

Matrices limites. Intéressons-nous à la convergence des matrices définies ci-dessus et leur limite éventuelle. Par soucis de simplicité, nous ne noterons pas nécessairement que ces matrices dépendent de f .

Lemme 3.5.2. *Les matrices $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ et $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ sont stochastiques et leur mesure invariante est respectivement $P[\mathcal{B}] := (P(B_1), \dots, P(B_{m_2}))^t$ et $P[\mathcal{A}] := (P(A_1), \dots, P(A_{m_1}))^t$.*

Démonstration. Les matrices $\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ et $\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ étant stochastiques il en va de même par stabilité du produit pour les matrices étudiées. Par symétrie, il en va de même pour $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$. Vérifions que $P[\mathcal{A}]$ est mesure invariante de la matrice $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$:

$$\begin{aligned} (P[\mathcal{A}]^t \cdot \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})_k &= \sum_{i \leq m_1} P(A_i) \cdot (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})_{i,k} = \sum_{i \leq m_1} P(A_i) \cdot \left(\sum_{l \leq m_2} (\mathbf{P}_{\mathcal{B}|\mathcal{A}})_{i,l} (\mathbf{P}_{\mathcal{A}|\mathcal{B}})_{l,k} \right) \\ &= \sum_{i \leq m_1} P(A_i) \cdot \left(\sum_{l \leq m_2} P(B_l|A_i)P(A_k|B_l) \right) \\ &= \sum_{l \leq m_2} P(B_l)P(A_k|B_l) = P(A_k). \end{aligned}$$

De la même manière, on montre que $P[\mathcal{B}]$ est la mesure invariante de la matrice $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$. \square

Nous ferons l'hypothèse supplémentaire dans tout le raisonnement qui suit :

Hypothèse (d'ergodicité). *Les matrices $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ et $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ sont ergodiques.*

Remarque 3.5.1. *L'hypothèse d'ergodicité nous permettra de montrer la convergence de la méthode de Raking Ratio. Elle est en particulier vérifiée si les matrices $\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ et $\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ sont strictement positives. Une condition suffisante qui peut correspondre à certains cas d'études est que les événements $A_i \cap B_j$ soient tous de probabilité non nulle.*

Puisque les matrices $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ et $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ sont stochastiques d'après le lemme 3.5.2 il existe des matrices $U, V \in \text{GL}_{m_1}(\mathbb{R})$, $U', V' \in \text{GL}_{m_2}(\mathbb{R})$ et $A \in \mathcal{M}_{m_1-1}$, $A' \in \mathcal{M}_{m_2-1}$ vérifiant $V = U^{-1}$, $V' = U'^{-1}$ et

$$\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}} = U \begin{pmatrix} 1 & \mathbf{0}_{\mathcal{M}_{1,m_1-1}} \\ \mathbf{0}_{\mathcal{M}_{m_1-1,1}} & A \end{pmatrix} V, \quad \mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}} = U' \begin{pmatrix} 1 & \mathbf{0}_{\mathcal{M}_{1,m_2-1}} \\ \mathbf{0}_{\mathcal{M}_{m_2-1,1}} & A' \end{pmatrix} V'. \quad (3.36)$$

On peut supposer que les matrices U, U' contiennent des 1 sur leur première colonne (puisque c'est un vecteur propre associé à la valeur propre 1) et par conséquence les matrices V, V' contiennent respectivement sur leur première ligne les vecteurs $P[\mathcal{A}], P[\mathcal{B}]$ (puisque d'après le lemme 3.5.2 ces vecteurs sont les vecteurs propres à gauche associées à 1 des matrices concernées).

Lemme 3.5.3. *Les matrices A, A' vérifient*

$$\lim_{k \rightarrow +\infty} A^k = \mathbf{0}_{\mathcal{M}_{m_1-1}}, \quad \lim_{k \rightarrow +\infty} A'^k = \mathbf{0}_{\mathcal{M}_{m_2-1}}.$$

Démonstration. On note $A = D + N$ et $A' = D' + N'$ la décomposition de Dunford des matrices A, A' . Il existe des matrices Δ, Δ' diagonales, des matrices P, P' inversibles tels que

$$D = P\Delta P^{-1}, \quad D' = P'\Delta' P'^{-1}.$$

En particulier, les éléments diagonaux de Δ et Δ' sont de module strictement inférieur à 1 (puisque c'est le cas des valeurs propres de A, A' et donc de D, D') donc

$$\lim_{k \rightarrow +\infty} \Delta^k = \mathbf{0}, \quad \lim_{k \rightarrow +\infty} \Delta'^k = \mathbf{0},$$

ce qui entraîne que

$$\lim_{k \rightarrow +\infty} D^k = \mathbf{0}, \quad \lim_{k \rightarrow +\infty} D'^k = \mathbf{0}.$$

Enfin, on note n et n' les indices de nilpotence de N et N' . Alors, puisque les matrices D, D' commutent respectivement avec les matrices N, N' :

$$A^k = \sum_{j=0}^{n-1} \binom{k}{j} N^j D^{k-j}, \quad A'^k = \sum_{j=0}^{n-1} \binom{k}{j} N'^j D'^{k-j}.$$

Les matrices A^k, A'^k s'écrivent comme une somme finie de termes convergeant vers la matrice nulle quand k tend vers l'infini. Ces matrices convergent donc vers la matrice nulle. \square

Lemme 3.5.4. *Les matrices stochastiques $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ et $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ vérifient :*

$$\begin{aligned} \lim_{k \rightarrow +\infty} (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k \cdot (\mathbb{E}^{(1)}[f] - \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbb{E}^{(2)}[f]) &= \mathbf{0}_{\mathcal{M}_{m_1,1}}, \\ \lim_{k \rightarrow +\infty} (\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k \cdot (\mathbb{E}^{(2)}[f] - \mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbb{E}^{(1)}[f]) &= \mathbf{0}_{\mathcal{M}_{m_2,1}}. \end{aligned}$$

Démonstration. D'après le lemme 3.5.2 et par ergodicité les matrices $((\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k)_k$ et $((\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k)_k$ convergent respectivement vers $\begin{pmatrix} P[\mathcal{A}]^t \\ \vdots \\ P[\mathcal{A}]^t \end{pmatrix}$ et $\begin{pmatrix} P[\mathcal{B}]^t \\ \vdots \\ P[\mathcal{B}]^t \end{pmatrix}$. D'une part, $P[\mathcal{A}]^t \cdot \mathbb{E}^{(1)}[f] = \sum_{j_1 \leq m_1} P(A_{j_1})\mathbb{E}[f|A_{j_1}] = \mathbb{E}[f]$ et d'autre part

$$P[\mathcal{A}]^t \cdot \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbb{E}^{(2)}[f] = \sum_{j_1 \leq m_1} P(A_{j_1})(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbb{E}^{(2)}[f])_{j_1} = \sum_{\substack{j_1 \leq m_1 \\ j_2 \leq m_2}} P(A_{j_1} \cap B_{j_2})\mathbb{E}[f|A_{j_1} \cap B_{j_2}] = \mathbb{E}[f].$$

La première relation est donc montrée. La seconde se démontre *mutatis mutandis*. \square

Remarque 3.5.2. *En notation matricielle et d'après le lemme 3.5.3 cela revient à dire que*

$$\begin{aligned} U \begin{pmatrix} 1 & \mathbf{0}_{\mathcal{M}_{m_1, m_1-1}} \\ \mathbf{0}_{\mathcal{M}_{m_1-1, 1}} & \mathbf{0}_{\mathcal{M}_{m_1-1}} \end{pmatrix} V \cdot (\mathbb{E}^{(1)}[f] - \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbb{E}^{(2)}[f]) &= \mathbf{0}_{\mathcal{M}_{m_1,1}}, \\ U' \begin{pmatrix} 1 & \mathbf{0}_{\mathcal{M}_{m_1, m_2-1}} \\ \mathbf{0}_{\mathcal{M}_{m_2-1, 1}} & \mathbf{0}_{\mathcal{M}_{m_2-1}} \end{pmatrix} V' \cdot (\mathbb{E}^{(2)}[f] - \mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbb{E}^{(1)}[f]) &= \mathbf{0}_{\mathcal{M}_{m_2,1}}. \end{aligned}$$

On est donc à même de déterminer la convergence des matrices intervenant dans la méthode du Raking-Ratio définies plus haut. En particulier les matrices associées à $\mathbb{G}[\mathcal{A}]$ et celles associées à $\mathbb{G}[\mathcal{B}]$ ne sont pas les mêmes :

Proposition 3.5.5. *Les matrices $S_{i, \text{pair}}^{(N)}(f)$ et $S_{i, \text{impair}}^{(N)}(f)$ pour $i = \{1, 2\}$ définies par les relations (3.30), (3.32), (3.33) et (3.31) sont convergentes. Si on note $S_{i, \text{pair}}(f)$ et $S_{i, \text{impair}}(f)$ les limites des matrices $S_{i, \text{pair}}^{(N)}(f)$ et $S_{i, \text{impair}}^{(N)}(f)$ pour $i = \{1, 2\}$ alors on a les relations*

$$S_{1, \text{impair}}(f) = S_{1, \text{pair}}(f) + \mathbb{E}^{(0,1)}[f], \quad S_{2, \text{pair}}(f) = S_{2, \text{impair}}(f) + \mathbb{E}^{(0,2)}[f],$$

avec $\mathbb{E}^{(0,i)}[f] := (\mathbb{E}[f], \dots, \mathbb{E}[f])^t \in \mathcal{M}_{m_i, 1}(\mathbb{R})$ pour $i = 1, 2$. De plus, il existe $0 \leq \lambda_1, \lambda_2 < 1$, une constante universelle $K > 0$ tels que pour toutes normes vectorielles $\|\cdot\|_{m_1}, \|\cdot\|_{m_2}$ sur \mathbb{R}^{m_1} et \mathbb{R}^{m_2} , pour tout $i = 1, 2$ et $N \in \mathbb{N}^*$,

$$\begin{aligned} \sup_{f \in \mathcal{F}} \left\| S_{i, \text{pair}}^{(N)}(f) - S_{i, \text{pair}}(f) \right\|_{m_1} &\leq KM_{\mathcal{F}} \lambda_1^N, \\ \sup_{f \in \mathcal{F}} \left\| S_{i, \text{impair}}^{(N)}(f) - S_{i, \text{impair}}(f) \right\|_{m_2} &\leq KM_{\mathcal{F}} \lambda_2^N. \end{aligned} \tag{3.37}$$

Démonstration. Les matrices $\text{Id}_{m_1-1} - A$ et $\text{Id}_{m_2-1} - A$ sont inversibles (puisque 1 est une valeur propre de multiplicité 1 pour les matrices $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ et $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$) d'où

$$\begin{aligned} \sum_{k=0}^N (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k &= U \begin{pmatrix} N+1 & \mathbf{0}_{\mathcal{M}_{1,m_1-1}} \\ \mathbf{0}_{\mathcal{M}_{m_1-1,1}} & \sum_{k=0}^N A^k \end{pmatrix} V \\ &= U \begin{pmatrix} N+1 & \mathbf{0}_{\mathcal{M}_{1,m_1-1}} \\ \mathbf{0}_{\mathcal{M}_{m_1-1,1}} & (\text{Id}_{m_1-1} - A)^{-1} \cdot (\text{Id}_{m_1-1} - A^{N+1}) \end{pmatrix} V, \\ \sum_{k=0}^N (\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k &= U' \begin{pmatrix} N+1 & \mathbf{0}_{\mathcal{M}_{1,m_2-1}} \\ \mathbf{0}_{\mathcal{M}_{m_2-1,1}} & (\text{Id}_{m_2-1} - A')^{-1} \cdot (\text{Id}_{m_2-1} - A'^{N+1}) \end{pmatrix} V'. \end{aligned}$$

En utilisant le lemme 3.5.4 (et notamment la remarque 3.5.2 de ce lemme), on a que

$$\begin{aligned} S_{1,\text{pair}}^{(N)}(f) &= U \begin{pmatrix} 0 & \mathbf{0}_{\mathcal{M}_{1,m_1-1}} \\ \mathbf{0}_{\mathcal{M}_{m_1-1,1}} & (\text{Id}_{m_1-1} - A)^{-1} \cdot (\text{Id}_{m_1-1} - A^{N+1}) \end{pmatrix} V \cdot (\mathbb{E}^{(1)}[f] - \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbb{E}^{(2)}[f]), \\ S_{2,\text{impair}}^{(N)}(f) &= U' \begin{pmatrix} 0 & \mathbf{0}_{\mathcal{M}_{1,m_2-1}} \\ \mathbf{0}_{\mathcal{M}_{m_2-1,1}} & (\text{Id}_{m_2-1} - A')^{-1} \cdot (\text{Id}_{m_2-1} - A'^{N+1}) \end{pmatrix} V' \cdot (\mathbb{E}^{(2)}[f] - \mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbb{E}^{(1)}[f]). \end{aligned}$$

Les deux matrices ci-dessus convergent d'après le lemme 3.5.3 respectivement vers les matrices

$$S_{1,\text{pair}}(f) = U \begin{pmatrix} 0 & \mathbf{0}_{\mathcal{M}_{1,m_1-1}} \\ \mathbf{0}_{\mathcal{M}_{m_1-1,1}} & (\text{Id}_{m_1-1} - A)^{-1} \end{pmatrix} V \cdot (\mathbb{E}^{(1)}[f] - \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbb{E}^{(2)}[f]), \quad (3.38)$$

$$S_{2,\text{impair}}(f) = U' \begin{pmatrix} 0 & \mathbf{0}_{\mathcal{M}_{1,m_2-1}} \\ \mathbf{0}_{\mathcal{M}_{m_2-1,1}} & (\text{Id}_{m_2-1} - A')^{-1} \end{pmatrix} V' \cdot (\mathbb{E}^{(2)}[f] - \mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbb{E}^{(1)}[f]). \quad (3.39)$$

Les convergences des matrices $S_{1,\text{impair}}^{(N)}$ et $S_{2,\text{pair}}^{(N)}$ vers des matrices $S_{1,\text{impair}}$ et $S_{2,\text{pair}}$ et les relations entre les différentes matrices limites données par la proposition se déduisent du fait que, comme rappelé dans la preuve du lemme 3.5.4, les matrices $((\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k)_k$ et $((\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k)_k$ convergent respectivement vers $\begin{pmatrix} P[\mathcal{A}]^t \\ \vdots \\ P[\mathcal{A}]^t \end{pmatrix}$ et $\begin{pmatrix} P[\mathcal{B}]^t \\ \vdots \\ P[\mathcal{B}]^t \end{pmatrix}$ ce qui entraîne que

$$\lim_{k \rightarrow +\infty} (\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}})^k \cdot \mathbb{E}^{(1)}[f] = \mathbb{E}^{(0,1)}[f], \quad \lim_{k \rightarrow +\infty} (\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}})^k \cdot \mathbb{E}^{(2)}[f] = \mathbb{E}^{(0,2)}[f].$$

Montrons maintenant que la première concentration donnée par (3.37) est vraie, la suivante se démontre *mutatis mutandis*. Soit $\lambda_1 = \rho(A) + \varepsilon$ avec $\varepsilon > 0$ suffisamment petit pour avoir $\lambda_1 < 1$. Il existe alors une norme $\|\cdot\|$ sur \mathbb{R}^{m_1-1} vérifiant $\|A\| \leq \lambda_1$. On note $\|\cdot\|'$ la norme qui à $x = (x_1, \dots, x_{m_1}) \in \mathbb{R}^{m_1}$ associe $|x_1| + \|(x_2, \dots, x_{m_1})^t\|$ et $\|\cdot\|'$ sa norme subordonnée. Cette norme matricielle vérifie en particulier pour toute matrice $B \in \mathcal{M}_{m_1-1}$, $\|\begin{pmatrix} 0 & \mathbf{0}_{\mathcal{M}_{1,m_1-1}} \\ \mathbf{0}_{\mathcal{M}_{m_1-1,1}} & B \end{pmatrix}\|' = \|B\|$. Alors,

$$\begin{aligned} \|S_{1,\text{pair}}^{(N)}(f) - S_{1,\text{pair}}(f)\|' &\leq \left\| U \begin{pmatrix} 0 & \mathbf{0}_{\mathcal{M}_{1,m_1-1}} \\ \mathbf{0}_{\mathcal{M}_{m_1-1,1}} & -(\text{Id}_{m_1-1} - A)^{-1}A^{N+1} \end{pmatrix} V \right\|' \|\mathbb{E}^{(1)}[f] - \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbb{E}^{(2)}[f]\|' \\ &\leq \|(\text{Id}_{m_1-1} - A)^{-1}\| \cdot \|A\|^{N+1} \cdot K_1 \|\mathbb{E}^{(1)}[f] - \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbb{E}^{(2)}[f]\|_\infty, \end{aligned}$$

où $K_1 > 0$ vérifie $\|\cdot\|' \leq K_1 \|\cdot\|_\infty$. Le premier facteur ne dépend ni de N , ni de \mathcal{F} . Le second facteur est majoré par λ_1^{N+1} tandis que le dernier facteur peut être majoré par $2K_1 M_{\mathcal{F}}$. Finalement,

$$\sup_{f \in \mathcal{F}} \|S_{1,\text{pair}}^{(N)}(f) - S_{1,\text{pair}}(f)\| \leq 2 \|(\text{Id}_{m_1-1} - A)^{-1}\| K_1 M_{\mathcal{F}} \lambda_1^{N+1}.$$

L'inégalité (3.37) résulte de l'équivalence des normes. \square

Processus gaussien limite. Nous allons donner et démontrer le résultat principal de cette partie. Il justifie la convergence de la méthode de Raking Ratio et donne l'expression du processus limite quand le nombre d'itérations tend vers l'infini. Malgré la différence des limites des matrices 1 et 2 donnée par la parité, le processus limite est le même. Autrement dit, la différence est compensée en termes de processus.

Théorème 3.5.6. *La suite de processus $(\mathbb{G}^{(N)})_N$ converge presque sûrement dans $\ell^\infty(\mathcal{F})$ quand N tend vers l'infini vers le processus gaussien centré $\mathbb{G}^{(\infty)}$ indexé par la classe de fonctions \mathcal{F} et défini par*

$$\mathbb{G}^{(\infty)}(f) = \mathbb{G}(f) - S_{1,\text{pair}}(f)^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,\text{impair}}(f)^t \cdot \mathbb{G}[\mathcal{B}].$$

De plus, il existe $K > 0$ tel que

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

où $0 < \lambda_1, \lambda_2 < 1$ sont donnés par la proposition 3.5.5.

Remarque 3.5.3. *Ce théorème peut être vu comme l'équivalent stochastique de la vitesse déterministe obtenue par Franklin et Lorentz [43] vis-à-vis de l'algorithme de Sinkhorn. Une recherche plus poussée qui combinerait les deux résultats pourrait renforcer ces résultats.*

Remarque 3.5.4. *Les matrices $\mathbf{P}_{\mathcal{A}|\mathcal{B}}, \mathbf{P}_{\mathcal{B}|\mathcal{A}}$ ainsi que les vecteurs $\mathbb{E}[f|\mathcal{A}], \mathbb{E}[f|\mathcal{B}]$ ne sont pas connus sans information supplémentaire. Ils peuvent être néanmoins estimés uniformément sur \mathcal{F} quand $n \rightarrow +\infty$ afin d'évaluer les distributions de $\mathbb{G}^{(N)}$ et $\mathbb{G}^{(\infty)}$ et de pouvoir établir des tests statistiques ou des estimateurs. Enfin, les constantes K, λ_1, λ_2 dépendent des matrices stochastiques $\mathbf{P}_{\mathcal{A}|\mathcal{B}}, \mathbf{P}_{\mathcal{B}|\mathcal{A}}$: elles peuvent donc être estimées avec une vitesse en $1/\sqrt{n}$ en probabilité et fournissent une évaluation de la borne d'erreur de $d_{LP}(\mathbb{G}^{(N)}, \mathbb{G}^{(\infty)})$.*

Démonstration. D'après la proposition 3.5.5, le processus empirique associé à la méthode du Raking Ratio converge respectivement dans les cas pair et impair vers les processus gaussiens centrés $\mathbb{G}_{\text{pair}}^{(\infty)}, \mathbb{G}_{\text{impair}}^{(\infty)}$ indexés par \mathcal{F} et définis par

$$\begin{aligned} \mathbb{G}_{\text{pair}}^{(\infty)}(f) &= \mathbb{G}(f) - S_{1,\text{pair}}(f)^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,\text{pair}}(f)^t \cdot \mathbb{G}[\mathcal{B}] \\ &= \mathbb{G}(f) - S_{1,\text{pair}}(f)^t \cdot \mathbb{G}[\mathcal{A}] - (S_{2,\text{impair}}(f) + \mathbb{E}^{(0,2)}[f])^t \cdot \mathbb{G}[\mathcal{B}] \\ &= \mathbb{G}^{(\infty)}(f) - \mathbb{E}^{(0,2)}[f]^t \cdot \mathbb{G}[\mathcal{B}] \\ \mathbb{G}_{\text{impair}}^{(\infty)}(f) &= \mathbb{G}(f) - S_{1,\text{impair}}(f)^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,\text{impair}}(f)^t \cdot \mathbb{G}[\mathcal{B}] \\ &= \mathbb{G}(f) - (S_{1,\text{pair}}(f) + \mathbb{E}^{(0,1)}[f])^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,\text{impair}}(f)^t \cdot \mathbb{G}[\mathcal{B}] \\ &= \mathbb{G}^{(\infty)}(f) - \mathbb{E}^{(0,1)}[f]^t \cdot \mathbb{G}[\mathcal{A}]. \end{aligned}$$

Ces processus sont les mêmes car $\mathbb{E}^{(0,1)}[f]^t \cdot \mathbb{G}[\mathcal{A}] = \mathbb{E}[f] \sum_{j \leq m_1} \mathbb{G}(A_j) = \mathbb{E}[f] \mathbb{G}(1) = 0$ et $\mathbb{E}^{(0,2)}[f]^t \cdot \mathbb{G}[\mathcal{B}] = 0$ d'où $\mathbb{G}_{\text{pair}}^{(\infty)} = \mathbb{G}_{\text{impair}}^{(\infty)}$. En utilisant (3.34) et (3.35), on obtient presque

sûrement les inégalités

$$\begin{aligned}
\|\mathbb{G}^{(2N)} - \mathbb{G}_{\text{pair}}^{(\infty)}\|_{\mathcal{F}} &= \left\| (S_{1,\text{pair}}^{(N-1)} - S_{1,\text{pair}})^t \mathbb{G}[\mathcal{A}] + (S_{2,\text{pair}}^{(N-2)} - S_{2,\text{pair}})^t \mathbb{G}[\mathcal{B}] \right\|_{\mathcal{F}} \\
&\leq m_1 \max_{i=1,\dots,m_1} |\mathbb{G}(A_i)| \|S_{1,\text{pair}}^{(N-1)} - S_{1,\text{pair}}\|_{\mathcal{F}} \\
&\quad + m_2 \max_{i=1,\dots,m_2} |\mathbb{G}(B_i)| \|S_{2,\text{pair}}^{(N-2)} - S_{2,\text{pair}}\|_{\mathcal{F}} \\
&\leq CZ \max(\lambda_1, \lambda_2)^{N-2}, \\
\|\mathbb{G}^{(2N+1)} - \mathbb{G}_{\text{impair}}^{(\infty)}\|_{\mathcal{F}} &\leq CZ \max(\lambda_1, \lambda_2)^{N-1},
\end{aligned}$$

où $C = 2 \max(K, K') M_{\mathcal{F}}(m_1 + m_2)$, $Z = \max_{i=1,\dots,m_1} |\mathbb{G}(A_i)| + \max_{i=1,\dots,m_2} |\mathbb{G}(B_i)|$ et $0 < K, K', 0 < \lambda_1, \lambda_2 < 1$ sont donnés par la proposition 3.5.5. \square

En conclusion $\mathbb{G}^{(\infty)}$ est le processus limite quand on fait tendre $n \rightarrow +\infty$ puis $N \rightarrow +\infty$. Même si n est grand, il serait judicieux de prendre des valeurs faibles pour N afin de s'assurer d'avoir des bornes de majoration satisfaisantes.

3.6 Exemples numériques

3.6.1 Calcul d'une moyenne rakée

Exemple d'une moyenne rakée. La manière usuelle de calculer la moyenne de X_1, \dots, X_n est d'additionner les données X_i multipliées par les poids $w_i = 1/n$. Si on a une information auxiliaire $P[\mathcal{A}^{(N)}] = (P(A_1^{(N)}), \dots, P(A_{m_N}^{(N)}))$ pour $1 \leq N \leq N_0$ on vaut changer itérativement les poids initiaux w_i par des nouveaux poids $w_i^{(N)}$ tels que $\sum_{i=1}^n w_i^{(N)}$ et

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

pour tout $1 \leq N \leq N_0$ et $1 \leq j \leq m_N$. Il faut avoir en mémoire que cela n'implique pas que $\sum_{i=1}^n w_i^{(N_1)} \mathbf{1}_{A_j}^{(N_2)}(X_i) = P(A_j^{(N_2)})$ avec $N_1 \neq N_2$ et $1 \leq j \leq N_2$. Pour cet exemple on prend $N_0 = 2$, $\mathcal{A}^{(2)} = \{A_1, A_2, A_3\}$, $\mathcal{B} = \{B_1, B_2\}$ et on génère des variables aléatoires normales X_i avec des variances fixées $\sigma^2 = 0.1$ et tel que la probabilité et l'espérance conditionnelle sont données par le tableau suivant :

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

TAB. 3.1: Probabilités des ensembles

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

TAB. 3.2: Espérance conditionnelle

En particulier,

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

On génère $n = 10$ valeurs et on obtient les données suivantes :

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

TAB. 3.3: Variables aléatoires générées

Dans ce cas la moyenne usuelle correspond à la somme de tous les X_i sur 10, c'est-à-dire qu'on assigne un poids $1/n = 0.1$ à chaque X_i et on obtient $\mathbb{P}_n(X) \simeq 0.055$. Quand on rake une fois on assigne les poids 0.15, 0.07, 0.1 aux individus appartenant respectivement aux ensembles A_1, A_2, A_3 . La moyenne rakée pour $N = 1$ vaut

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

Quand l'algorithme est stabilisé les poids finaux sont donnés par le tableau suivant :

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

TAB. 3.4: Poids rakes finaux

Notons que la croix dans le tableau signifie que l'on n'a pas généré de variable aléatoire appartenant à $A_1 \cap B_2$ dû à la faible valeur de n . La moyenne rakée finale est $\mathbb{P}_n^{(\infty)}(X) \simeq 0.212$ qui est plus proche de $P(X)$ que la moyenne usuelle $\mathbb{P}_n(X)$.

3.6.2 Calcul de $S_{1,\text{pair}}(f)$ et $S_{2,\text{impair}}(f)$

Pour déterminer de manière numérique les limites $S_{1,\text{pair}}(f)$ et $S_{2,\text{impair}}(f)$ il est possible de faire une décomposition spectrale des matrices stochastiques $\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}$ et $\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ et de supprimer la valeur 1 de la matrice contenant les valeurs propres. Plus précisément :

Proposition 3.6.1. *On note ψ l'application qui, à une matrice A admettant pour décomposition spectrale $A = P \begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix} P^{-1}$ avec $B \in \text{GL}_n(\mathbb{R})$ vérifiant $\text{Id}_n - B \in \text{GL}_n(\mathbb{R})$, lui supprime la valeur propre 1, i.e. lui associe la matrice $P \begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix} P^{-1}$. Alors,*

$$\begin{aligned} S_{1,\text{pair}}(f) &= \psi \left((\text{Id} - \psi(\mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbf{P}_{\mathcal{A}|\mathcal{B}}))^{-1} \right) \cdot (\mathbb{E}^{(1)}[f] - \mathbf{P}_{\mathcal{B}|\mathcal{A}}\mathbb{E}^{(2)}[f]), \\ S_{2,\text{impair}}(f) &= \psi \left((\text{Id} - \psi(\mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbf{P}_{\mathcal{B}|\mathcal{A}}))^{-1} \right) \cdot (\mathbb{E}^{(2)}[f] - \mathbf{P}_{\mathcal{A}|\mathcal{B}}\mathbb{E}^{(1)}[f]). \end{aligned}$$

Démonstration. Avec les relations (3.36), (3.38) et (3.39), il suffit de vérifier que

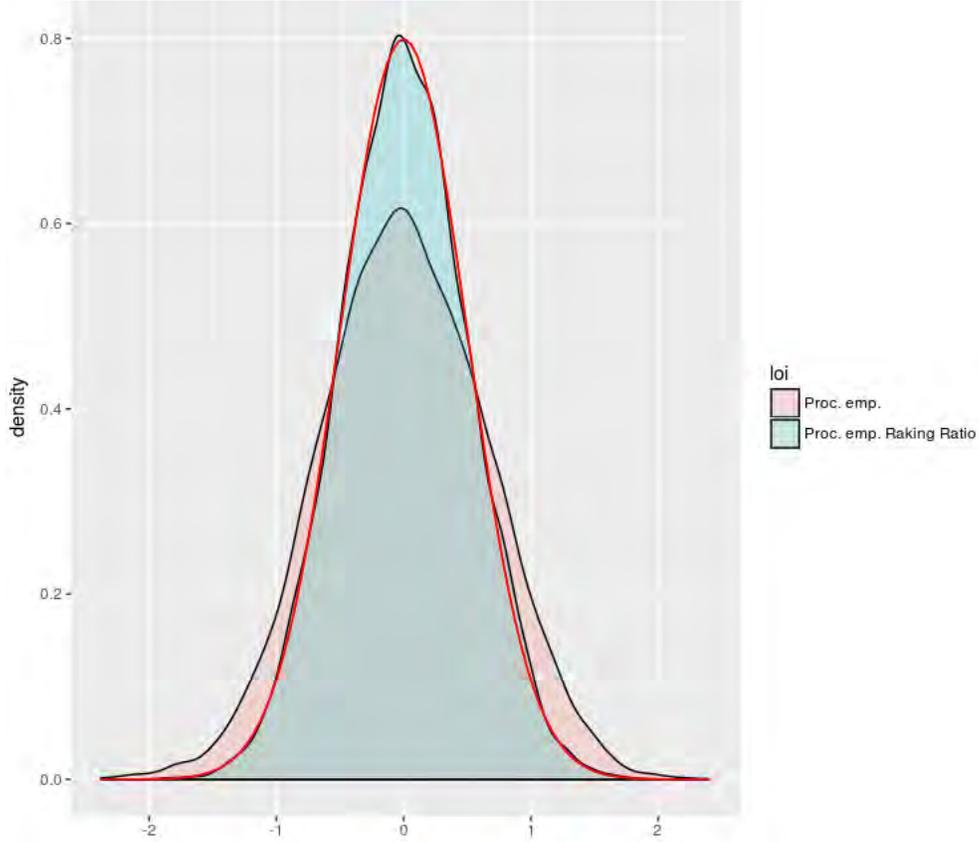
$$\begin{aligned} \psi \left(\left(\text{Id} - \psi \begin{pmatrix} 1 & \mathbf{0}_{\mathcal{M}_{1,m_1-1}} \\ \mathbf{0}_{\mathcal{M}_{m_1-1,1}} & A \end{pmatrix} \right)^{-1} \right) &= \begin{pmatrix} 0 & \mathbf{0}_{\mathcal{M}_{1,m_1-1}} \\ \mathbf{0}_{\mathcal{M}_{m_1-1,1}} & (\text{Id}_{m_1-1} - A)^{-1} \end{pmatrix}, \\ \psi \left(\left(\text{Id} - \psi \begin{pmatrix} 1 & \mathbf{0}_{\mathcal{M}_{1,m_2-1}} \\ \mathbf{0}_{\mathcal{M}_{m_2-1,1}} & A' \end{pmatrix} \right)^{-1} \right) &= \begin{pmatrix} 0 & \mathbf{0}_{\mathcal{M}_{1,m_2-1}} \\ \mathbf{0}_{\mathcal{M}_{m_2-1,1}} & (\text{Id}_{m_2-1} - A')^{-1} \end{pmatrix}. \end{aligned}$$

Ces relations sont immédiates par définition de ψ . □

Il est possible d'estimer les matrices $S_{1,\text{pair}}(f)$, $S_{2,\text{impair}}(f)$ en utilisant le résultat précédent avec les estimés empiriques de $\mathbf{P}_{\mathcal{B}|\mathcal{A}}$, $\mathbf{P}_{\mathcal{A}|\mathcal{B}}$ et de $\mathbb{E}^{(1)}[f]$, $\mathbb{E}^{(2)}[f]$.

3.6.3 Simulation du processus empirique raké

Un moyen de tester numériquement nos résultats théoriques est de générer des variables aléatoires tout en ayant la main sur l'information auxiliaire. Nous avons pris deux partitions $\mathcal{A} = \{A_1, \dots, A_{m_1}\}$ et $\mathcal{B} = \{B_1, \dots, B_{m_2}\}$ dont nous pouvons contrôler la taille et la probabilité d'appartenir aux ensembles de ces partitions. La simulation génère des variables normales de même variance paramétrable, appartenant avec probabilité $P(A_i \cap B_j)$ à l'ensemble $A_i \cap B_j$ et de moyenne dépendant de l'ensemble auquel la variable appartient. L'objectif de cette simulation était de vérifier que la loi asymptotique des estimateurs rakés, par exemple de la moyenne ou de la variance des X_i , suivait bien la loi limite donnée par le théorème 3.5.6 quelque soit les paramètres d'entrée du programme. Par la méthode de Monte-Carlo on voit numériquement d'une part le processus empirique raké $\alpha_n^{(N)}$ converger vers une gaussienne pour des valeurs n grandes et d'autre part la variance estimée tendre vers la variance théorique que l'on peut calculer à partir des données d'entrée modifiables. La figure suivante illustre ce comportement en générant $N = 10000$ échantillons de taille $n = 700$. La courbe rose représente la loi du processus empirique $\alpha_n(\text{id})$ autrement dit la moyenne des X_i centrée et normalisée. La courbe bleue représente la loi de $\alpha_n^{(N)}(\text{id})$ pour un N qu'on a fixé à l'avance. On observe une diminution de variance en ayant appliqué la méthode du Raking-Ratio comme le prouve la proposition 3.3.3. La courbe rouge correspond à la densité de $\text{Var}(\mathbb{G}^{(N)}(\text{id}))$ calculable grâce à la proposition 3.5.1 puisque les probabilités d'ensemble des partitions et l'espérance des X_i sur ces ensembles ont été fixées par notre programme.



3.6.4 Fonctions de répartition rakées.

Soit (X, Y) un vecteur gaussien centré de matrice de covariance $\begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}$. On considère l'estimation jointe et rakée des deux fonctions de répartition F_X, F_Y . On apporte comme information auxiliaire la valeur de ces fonctions aux points de l'intervalle $[-2, 2]$ avec un pas de 0.5. La classe \mathcal{F} sur laquelle nous travaillons est composée des fonctions de la forme $f_t^X(x, y) = \mathbb{1}_{]-\infty, t]}(x), f_t^Y(x, y) = \mathbb{1}_{]-\infty, t]}(y)$ pour $t \in \mathbb{R}$ donc l'hypothèse VC est respectée. Pour $Z = X, Y$ on pose $F_{Z,n}^{(N)}(t) = \sum_{Z_i \leq t} \mathbb{P}_n^{(N)}(\{Z_i\})$ la fonction de répartition empirique rakée N fois et on écrit $Z_{(1)} \leq \dots \leq Z_{(n)}$ les statistiques d'ordre. Afin d'exploiter au mieux l'information, nous utilisons $N = 2m - 1$ pour étudier $F_{X,n}^{(N)}$ et $N = 2m$ pour étudier $F_{Y,n}^{(N)}$: si $\mathbb{1}_{]-\infty, t]}$ est une information auxiliaire on a d'après cette convention que $F_{X,n}^{(N)}(f) = F_X(t)$ et $F_{Y,n}^{(N)}(t) = F_Y(t)$. On considère la statistique

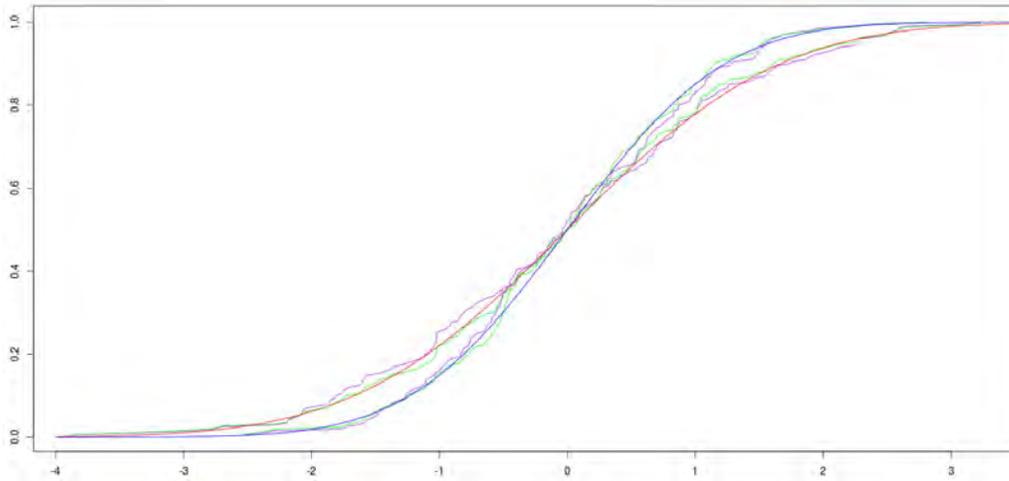
$$d_{Z,n}^{(N)} = \sum_{i=1}^{n-1} (Z_{(i+1)} - Z_{(i)}) \left| F_{Z,n}^{(N)}(Z_{(i+1)}) - F_Z(Z_{(i+1)}) \right|,$$

qui approxime sur $[Z_{(1)}, Z_{(n)}]$ la distance L^1 entre $F_{Z,n}^{(N)}$ et F_Z . On note $\#_{Z,n}^{(N)}$ la proportion aléatoire de couples de l'échantillon qui ont été mieux estimés par la méthode du Raking-Ratio, c'est-à-dire la proportion d'individus où $F_{Z,n}^{(N)}$ est plus proche de F_Z que $F_{Z,n}^{(0)}$. Le ta-

bleau suivant fournit des estimations par Monte-Carlo de $D_{Z,n}^{(N)} = \mathbb{E}[d_{Z,n}^{(N)}]$ et $p_{Z,n}^{(N)} = \mathbb{E}[\#_{Z,n}^{(N)}]$ à partir de 1000 simulations basées sur des échantillons de taille $n = 200$:

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

Ce tableau montre quelques améliorations, spécialement pour $N = 10$. Pour des valeurs n trop petites, il ne semble pas toujours pertinent d'attendre la stabilisation de l'algorithme que l'on dénote par $N = \infty$. Nos résultats théoriques nous apportent des garanties uniquement pour des petites valeurs de N et des grandes valeurs de n . On peut observer graphiquement que $F_{Z,n}^{(N)}$ passe forcément par les points connus. La figure suivante représente les fonctions de répartition empirique rakées en vert et non rakées en violet. Les courbes rouges et bleus représentent respectivement F_X, F_Z .



3.7 Apprentissage de l'information auxiliaire

3.7.1 Introduction

Motivation. Supposons que le statisticien ne dispose pas de la vraie probabilité des ensembles d'une partition donnée, mais dispose d'une source d'informations lui permettant d'avoir une estimation de cette probabilité plus précisément que s'il utilisait son propre échantillon. Cette source peut être de différents types : enquête préliminaire sur un large échantillon de personnes, traitement d'une base de données, achat de données supplémentaires à moindre coût, connaissances d'un expert ... Nous supposons dans notre modèle que seule l'estimation de l'information auxiliaire est transmise par la source. Cette hypothèse assure une rapidité d'acquisition des données et permet une pluralité de sources d'informations et donc une diversité de partitions. Il s'agit d'une situation courante dans le domaine des statistiques, car les technologies actuelles, telles que la transmission en continu de données, permettent la collecte et la transmission de telles informations en temps réel. Le statisticien peut utiliser ces informations acquises comme information auxiliaire, même si elle constitue une estimation de la véritable information. La méthode du Raking-Ratio permet de combiner des informations partagées de plusieurs sources. La principale question statistique de cette

partie est de savoir si le statisticien peut toujours appliquer la méthode du Raking-Ratio en utilisant l'estimation des probabilités d'inclusion plutôt que la vraie comme information auxiliaire. Nous montrerons que la réponse à cette question est positive à condition de contrôler la taille minimale des échantillons des différentes sources d'informations auxiliaires.

Notation. Pour $N > 0$ soit $\mathbb{P}'_N[\mathcal{A}^{(N)}] = (\mathbb{P}'_N(A_1^{(N)}), \dots, \mathbb{P}'_N(A_{m_N}^{(N)}))$ un vecteur aléatoire suivant une loi multinomiale avec n_N essais et de probabilités d'événements $P[\mathcal{A}^{(N)}] = (P(A_1^{(N)}), \dots, P(A_{m_N}^{(N)}))$. Ce vecteur aléatoire correspond à l'estimation des informations auxiliaires de la N -ème source basée sur un échantillon de taille $n_N = n_N(n) \gg n$ pas nécessairement indépendant de X_1, \dots, X_n . Nous étudions le comportement asymptotique du processus empirique rakié qui utiliserait $\mathbb{P}'_N[\mathcal{A}^{(N)}]$ comme information auxiliaire au lieu de $P[\mathcal{A}^{(N)}]$. En définissant la séquence $(n_N)_{N \geq 1}$, nous supposons que ces informations peuvent être estimées par différentes sources qui n'auraient pas nécessairement la même taille d'échantillon mais conserveraient quand même une taille d'échantillon supérieure à n . On note $\tilde{\mathbb{P}}_n^{(N)}(\mathcal{F}) = \{\tilde{\mathbb{P}}_n^{(N)}(f) : f \in \mathcal{F}\}$ la mesure empirique rakiée avec information auxiliaire apprise définie récursivement par $\tilde{\mathbb{P}}_n^{(0)} = \mathbb{P}_n$ et pour tout $N > 0, f \in \mathcal{F}$,

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

Cette mesure empirique satisfait l'information auxiliaire apprise puisque

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

Nous définissons $\tilde{\alpha}_n^{(N)}(\mathcal{F}) = \{\tilde{\alpha}_n^{(N)}(f) : f \in \mathcal{F}\}$ le processus empirique rakié avec informations auxiliaires apprises définies pour tout $f \in \mathcal{F}$ par

$$\tilde{\alpha}_n^{(N)}(f) = \sqrt{n}(\tilde{\mathbb{P}}_n^{(N)}(f) - P(f)). \quad (3.40)$$

Pour cette partie, nous adopterons les notations qui suivent. On note $K_{\mathcal{F}} = \max(1, M_{\mathcal{F}})$ et pour $N_0 > 0$,

$$p_{(N_0)} = \min_{1 \leq N \leq N_0} \min_{1 \leq j \leq m_N} P(A_j^{(N)}), \quad m_{(N_0)} = \sup_{0 \leq N \leq N_0} m_N, \quad n_{(N_0)} = \min_{1 \leq N \leq N_0} n_N > n.$$

3.7.2 Résultats principaux

Préliminaires. On suppose dans cette partie que \mathcal{F} vérifie $\mathcal{F}.i$ et $\mathcal{F}.ii$. Nous fixons $N_0 > 0$ et nous notons $\Lambda_n, \Lambda'_n > 0$ les supremums suivants

$$\Lambda_n = \max \left(\sup_{0 \leq N \leq N_0} \|\tilde{\alpha}_n^{(N)}\|_{\mathcal{F}}, \sup_{0 \leq N \leq N_0} \|\alpha_n^{(N)}\|_{\mathcal{F}} \right),$$

$$\Lambda'_n = \sup_{1 \leq N \leq N_0} \sup_{1 \leq j \leq m_N} |\alpha'_N(A_j^{(N)})|,$$

où $\alpha'_N(A_j^{(N)}) = \sqrt{n_N}(\mathbb{P}'_N(A_j^{(N)}) - P(A_j^{(N)}))$. Immédiatement, par l'inégalité de Hoeffding on a que pour tout $\lambda > 0$,

$$\mathbb{P}(\Lambda'_n > \lambda) \leq 2N_0 m_{(N_0)} \exp(-2\lambda^2). \quad (3.41)$$

On donne les décompositions de $\alpha_n^{(N)}(\mathcal{F})$ et $\tilde{\alpha}_n^{(N)}(\mathcal{F})$ qui seront utilisées dans les preuves qui vont suivre. En utilisant la définition de $\alpha_n^{(N)}(\mathcal{F})$ donnée par (3.4) on obtient que

$$\begin{aligned}\alpha_n^{(N)}(f) &= \sqrt{n} \left(\sum_{j=1}^{m_N} \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} \mathbb{P}_n^{(N-1)}(f \mathbb{1}_{A_j^{(N)}}) - P(f \mathbb{1}_{A_j^{(N)}}) \right) \\ &= \sum_{j=1}^{m_N} \frac{1}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} \left(P(A_j^{(N)}) \alpha_n^{(N-1)}(f \mathbb{1}_{A_j^{(N)}}) - P(f \mathbb{1}_{A_j^{(N)}}) \alpha_n^{(N-1)}(A_j^{(N)}) \right).\end{aligned}\quad (3.42)$$

De la même manière en utilisant la définition de $\tilde{\alpha}_n^{(N)}(\mathcal{F})$ donnée par (3.40) on obtient que

$$\tilde{\alpha}_n^{(N)}(f) = \sum_{j=1}^{m_N} \frac{\mathbb{P}'_N(A_j^{(N)})}{\tilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} \tilde{\alpha}_n^{(N-1)}(f \mathbb{1}_{A_j^{(N)}}) - \frac{P(f \mathbb{1}_{A_j^{(N)}})}{\tilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} \left(\tilde{\alpha}_n^{(N-1)}(A_j^{(N)}) - \sqrt{\frac{n}{n_N}} \alpha'_N(A_j^{(N)}) \right).\quad (3.43)$$

Inégalité de Talagrand. La proposition suivante borne la probabilité que le supremum $\|\tilde{\alpha}_n^{(N)}\|_{\mathcal{F}}$ dévie d'une valeur donnée. Il s'agit du pendant de la proposition 3.2.4.

Proposition 3.7.1. *Pour tout $N_0 \in \mathbb{N}$, $n > 0$ et $t > 0$, sous l'événement B_{n, N_0} on a*

$$\begin{aligned}\mathbb{P} \left(\sup_{0 \leq N \leq N_0} \|\tilde{\alpha}_n^{(N)}\|_{\mathcal{F}} > t \right) \\ \leq N_0 \mathbb{P} \left(\|\tilde{\alpha}_n^{(0)}\|_{\mathcal{F}} > \frac{tp_{(N_0)}^{N_0}}{4^{N_0} m_{(N_0)}^{N_0} K_{\mathcal{F}}^{N_0} (1 + t/\sqrt{n})^{N_0}} \right) + 2N_0^3 m_{(N_0)} \exp \left(-\frac{n_{(N_0)} p_{(N_0)}^2 t^2}{2nm_{(N_0)}^2 K_{\mathcal{F}}^2} \right).\end{aligned}\quad (3.44)$$

Sous l'hypothèse VC et l'événement B_{n, N_0} on a pour tout $t > 0$,

$$\mathbb{P} \left(\sup_{0 \leq N \leq N_0} \|\tilde{\alpha}_n^{(N)}\|_{\mathcal{F}} > t \right) \leq D_1 t^{\nu_0} \exp(-D_2 t^2) + 2N_0^3 m_{(N_0)} \exp \left(-\frac{n_{(N_0)} p_{(N_0)}^2 t^2}{2nm_{(N_0)}^2 K_{\mathcal{F}}^2} \right), \quad (3.45)$$

où $D_1, D_2 > 0$ sont définis par (3.50). Sous BR et l'événement B_{n, N_0} il existe $t_0, C > 0$ tels que pour tout $t_0 < t < C\sqrt{n}$,

$$\mathbb{P} \left(\sup_{0 \leq N \leq N_0} \|\tilde{\alpha}_n^{(N)}\|_{\mathcal{F}} > t \right) \leq D_3 \exp(-D_4 t^2) + 2N_0^3 m_{(N_0)} \exp \left(-\frac{n_{(N_0)} p_{(N_0)}^2 t^2}{2nm_{(N_0)}^2 K_{\mathcal{F}}^2} \right), \quad (3.46)$$

où $D_3, D_4 > 0$ sont définis par (3.52).

Remarque 3.7.1. *La proposition 3.7.1 prouve que si \mathcal{F} vérifie VC ou BR alors presque sûrement $\|\tilde{\alpha}_n^{(N)}\|_{\mathcal{F}} = O(\sqrt{\log n})$.*

Démonstration. Nous prouvons (3.44), (3.45) et (3.46) respectivement aux étapes 1, 2 et 3.

Étape 1. Soit $0 \leq N \leq N_0$. Grâce à (3.43) on peut écrire que

$$\begin{aligned} \mathbb{P}(\|\tilde{\alpha}_n^{(N)}\|_{\mathcal{F}} > t) &\leq \mathbb{P}\left(\frac{K_{\mathcal{F}}m_{(N)}\left(2\|\tilde{\alpha}_n^{(N-1)}\|_{\mathcal{F}} + \sqrt{\frac{n}{n_{(N)}}}\Lambda'_n\right)}{p_{(N)} - \|\tilde{\alpha}_n^{(N-1)}\|_{\mathcal{F}}/\sqrt{n}} > t\right) \\ &\leq \mathbb{P}\left(\Lambda'_n > \sqrt{\frac{n_{(N)}}{n}} \frac{tp_{(N)}}{2m_{(N)}K_{\mathcal{F}}}\right) + \mathbb{P}\left(\|\tilde{\alpha}_n^{(N-1)}\|_{\mathcal{F}} > \frac{tp_{(N)}}{4m_{(N)}K_{\mathcal{F}}(1+t/\sqrt{n})}\right) \\ &\leq \mathbb{P}\left(\Lambda'_n > \sqrt{\frac{n_{(N_0)}}{n}} \frac{tp_{(N_0)}}{2m_{(N_0)}K_{\mathcal{F}}}\right) + \mathbb{P}\left(\|\tilde{\alpha}_n^{(N-1)}\|_{\mathcal{F}} > \frac{tp_{(N)}}{4m_{(N)}K_{\mathcal{F}}(1+t/\sqrt{n})}\right). \end{aligned} \quad (3.47)$$

Avec (3.41) et par induction de (3.47), on trouve que

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

La borne de la dernière inégalité croît avec N ce qui entraîne (3.44). Puisque

$$\tilde{\alpha}_n^{(0)}(\mathcal{F}) = \alpha_n(\mathcal{F}) = \alpha_n^{(0)}(\mathcal{F}), \quad (3.48)$$

on peut appliquer l'inégalité de Talagrand pour contrôler la probabilité que $\|\tilde{\alpha}_n^{(0)}\|_{\mathcal{F}}$ dévie comme décrit dans les deux étapes ci-après.

Étape 2. D'après le théorème 1.2.6 si \mathcal{F} vérifie VC, il existe une constante $D = D(c_0) > 0$ telle que pour $t > 0$,

$$\mathbb{P}\left(\|\tilde{\alpha}_n^{(0)}\|_{\mathcal{F}} > t\right) \leq \left(\frac{Dt}{M_{\mathcal{F}}\sqrt{\nu_0}}\right)^{\nu_0} \exp\left(\frac{-2t^2}{M_{\mathcal{F}}^2}\right). \quad (3.49)$$

Les inégalités (3.44) et (3.49) entraînent (3.45) pour tout $0 < t$ avec $D_1, D_2 > 0$ définis par

$$D_1 = N_0 \left(\frac{Dp_{(N_0)}^{N_0}}{\nu_0 4^{N_0} m_{(N_0)}^{N_0} K_{\mathcal{F}}^{N_0+1}}\right)^{\nu_0}, \quad D_2 = \frac{p_{(N_0)}^{2N_0}}{72^{N_0} m_{(N_0)}^{2N_0} K_{\mathcal{F}}^{3N_0+1}}. \quad (3.50)$$

Étape 3. D'après le théorème 1.2.7, si \mathcal{F} vérifie BR il existe $D = D(M_{\mathcal{F}}, b_0, r_0, \sigma_{\mathcal{F}}^2) > 0, t_0 = t_0(M_{\mathcal{F}}, b_0, r_0) > 0$ tels que pour tout $t_0 < t$,

$$\mathbb{P}\left(\|\tilde{\alpha}_n^{(0)}\|_{\mathcal{F}} > t\right) \leq \exp(-Dt^2). \quad (3.51)$$

Alors (3.44) et (3.51) implique (3.46) où $D_3, D_4 > 0$ sont définis par

$$D_3 = N_0, \quad D_4 = \frac{Dp_{(N_0)}^{2N_0}}{8^{N_0} m_{(N_0)}^{2N_0} K_{\mathcal{F}}^{2N_0} (1 + 2\sigma_{\mathcal{F}}^2/M_{\mathcal{F}})^{2N_0}}. \quad (3.52)$$

□

Approximation forte. Le résultat suivant établit l'approximation forte de $\tilde{\alpha}_n^{(N)}(\mathcal{F})$ par $\mathbb{G}^{(N)}(\mathcal{F})$.

Théorème 3.7.2. Soit $N_0 \in \mathbb{N}$. Il existe $d_0, n_0 > 0$, une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables i.i.d. de loi P et une suite $(\mathbb{G}_n)_{n \in \mathbb{N}^*}$ de versions de \mathbb{G} définis sur un même espace de probabilité tels que pour tout $n > n_0$,

$$\mathbb{P} \left(\sup_{0 \leq N \leq N_0} \|\tilde{\alpha}_n^{(N)} - \mathbb{G}_n^{(N)}\|_{\mathcal{F}} > d_0 \left(v_n + \sqrt{\frac{n \log n}{n_{(N_0)}}} \right) \right) < \frac{1}{n^2}, \quad (3.53)$$

où $\mathbb{G}_n^{(N)}$ est une version de $\mathbb{G}^{(N)}$ construite à partir de $\mathbb{G}_n^{(0)} = \mathbb{G}_n$ via (3.12).

Remarque 3.7.2. D'après le lemme de Borel-Cantelli, presque sûrement pour n suffisamment grand,

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

La suite v_n dans le borne précédente correspond à la déviation de $\alpha_n^{(N)}(\mathcal{F})$ par rapport à $\mathbb{G}_n^{(N)}(\mathcal{F})$ alors que $\sqrt{n \log n / n_{(N_0)}}$ représente la déviation de $\tilde{\alpha}_n^{(N)}(\mathcal{F})$ par rapport à $\alpha_n^{(N)}(\mathcal{F})$. Sous réserve que la taille d'échantillon des sources est suffisamment grande, le théorème 3.7.2 implique que le vecteur aléatoire $(\tilde{\alpha}_n^{(0)}(\mathcal{F}), \dots, \tilde{\alpha}_n^{(N_0)}(\mathcal{F}))$ converge en loi vers $(\mathbb{G}^{(0)}(\mathcal{F}), \dots, \mathbb{G}^{(N_0)}(\mathcal{F}))$ dans $\ell^\infty(\mathcal{F} \rightarrow \mathbb{R}^{N_0+1})$ de la même manière que $(\alpha_n^{(0)}(\mathcal{F}), \dots, \alpha_n^{(N_0)}(\mathcal{F}))$.

Démonstration. D'après la proposition 3.7.1, l'inégalité (3.41) et les propositions 3.2.4 et 3.7.1, il existe $D > 0$ tel que

$$\mathbb{P} \left(\{\Lambda_n > D\sqrt{\log n}\} \cup \{\Lambda'_n > D\sqrt{\log n}\} \right) \leq \frac{1}{3n^2}. \quad (3.54)$$

D'après le théorème 3.3.7, on peut définir sur un même espace de probabilité une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables i.i.d. et de loi P ainsi qu'une suite $(\mathbb{G}_n)_{n \in \mathbb{N}^*}$ de versions de \mathbb{G} vérifiant la propriété suivante. Il existe $n_1, d_1 > 0$ tels que pour tout $n > n_1$,

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

où $\mathbb{G}_n^{(N)}$ est une version de $\mathbb{G}^{(N)}$ construite à partir de $\mathbb{G}_n^{(0)} = \mathbb{G}_n$ via (3.12). Pour montrer (3.53) il reste à prouver, grâce à (3.48), que pour tout n suffisamment et pour un certain $d_0 > 0$,

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

Soit $1 \leq N \leq N_0$. La décomposition de $\alpha_n^{(N)}$ et $\tilde{\alpha}_n^{(N)}$ donnée respectivement par (3.42) et (3.43) entraînent que

$$\begin{aligned} \tilde{\alpha}_n^{(N)}(f) - \alpha_n^{(N)}(f) &= \sum_{j=1}^{m_N} \frac{\mathbb{P}'_N(A_j^{(N)})}{\tilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} (\tilde{\alpha}_n^{(N-1)}(f \mathbf{1}_{A_j^{(N)}}) - \alpha_n^{(N-1)}(f \mathbf{1}_{A_j^{(N)}})) \\ &+ \alpha_n^{(N-1)}(f \mathbf{1}_{A_j^{(N)}}) \left(\frac{\mathbb{P}'_N(A_j^{(N)})}{\tilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} - \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} \right) \\ &- P(f \mathbf{1}_{A_j^{(N)}}) \left(\frac{\tilde{\alpha}_n^{(N-1)}(A_j^{(N)})}{\tilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} - \frac{\alpha_n^{(N-1)}(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} \right) + \sqrt{\frac{n}{n_N}} \frac{P(f \mathbf{1}_{A_j^{(N)}})}{\tilde{\mathbb{P}}_n^{(N-1)}(A_j^{(N)})} \alpha'_N(A_j^{(N)}). \end{aligned} \quad (3.55)$$

Avec (3.48) pour $N = 1$ on a en particulier que

$$\tilde{\alpha}_n^{(1)}(f) - \alpha_n^{(1)}(f) = \sum_{j=1}^{m_1} \alpha_n(f \mathbf{1}_{A_j^{(1)}}) \left(\frac{\mathbb{P}'_{n_1}(A_j^{(1)}) - P(A_j^{(1)})}{\mathbb{P}_n(A_j^{(1)})} \right) + \sqrt{\frac{n}{n_1}} \frac{P(f \mathbf{1}_{A_j^{(1)}})}{\mathbb{P}_n(A_j^{(1)})} \alpha'_{n_1}(A_j^{(1)}),$$

qui est uniformément bornée par

$$\|\tilde{\alpha}_n^{(1)} - \alpha_n^{(1)}\|_{\mathcal{F}} \leq \frac{m_{(N)} K_{\mathcal{F}} \Lambda'_n}{p_{(N)} - \Lambda_n / \sqrt{n}} \sqrt{\frac{n}{n_{(N)}}} (1 + \Lambda_n / \sqrt{n}). \quad (3.56)$$

Let $C_{n,N} = 4m_{(N)} K_{\mathcal{F}} / (p_{(N)} - \Lambda_n / \sqrt{n})^2$. L'équation (3.55) entraîne aussi que

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

Par récurrence sur la dernière inégalité et en remarquant que pour tout $n > 0$, $m_{(N)} K_{\mathcal{F}} / (p_{(N)} - \Lambda_n / \sqrt{n})^2 \geq 1$, on a que

$$\|\tilde{\alpha}_n^{(N)}(f) - \alpha_n^{(N)}(f)\|_{\mathcal{F}} \leq C_{n,N}^{N-1} \|\tilde{\alpha}_n^{(1)} - \alpha_n^{(1)}\|_{\mathcal{F}} + (N-1) C_{n,N}^{N-1} \left(\frac{\Lambda_n^2}{\sqrt{n}} + \frac{\Lambda'_n(\Lambda_n + \sqrt{n})}{\sqrt{n_{(N)}}} \right),$$

alors l'inégalité (3.56) implique immédiatement que

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

Puisque la borne de la dernière inégalité croît avec N on trouve que pour tout $t > 0$,

$$\begin{aligned} & \mathbb{P} \left(\sup_{1 \leq N \leq N_0} \|\tilde{\alpha}_n^{(N)}(f) - \alpha_n^{(N)}(f)\|_{\mathcal{F}} > t \right) \\ & \leq \mathbb{P} \left(\frac{C_0}{(p_{(N_0)} - \Lambda_n / \sqrt{n})^{2N_0}} \left(\frac{\Lambda_n^2}{\sqrt{n}} + \frac{\Lambda'_n(\Lambda_n + \sqrt{n})}{\sqrt{n_{(N_0)}}} \right) > t \right), \end{aligned} \quad (3.57)$$

avec $C_0 = N_0^2 (4m_{(N_0)} K_{\mathcal{F}} N_0)^{N_0} > 0$. Il existe $n_2 > 0$ tel que pour tout $n > n_2$ on ait $D\sqrt{\log n/n} \leq p_{(N_0)}/2 \leq 1/2$. Pour $n > n_2$ on a d'après (3.54) et (3.57) que

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

En utilisant (3.54) à nouveau, la dernière inégalité implique

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

pour tout $n > n_2$ et

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

Par définition de v_n , il existe $d_2 > \max(d_1, 4^{N_0+1}C_0D/p_{(N_0)}^{2N_0})$ et $n_3 > 0$ tels que pour tout $n > n_3$,

$$d_2 \left(v_n + \sqrt{\frac{n \log n}{n_{(N_0)}}} \right) > d_1 v_n + t_n.$$

Alors (3.53) est prouvé pour $d_0 = d_2$ et $n_0 = \max(n_0, n_1, n_3)$. \square

3.7.3 Applications statistiques

Motivation. Tout test statistique utilisant le processus empirique peut être modifié pour utiliser l'information auxiliaire afin de renforcer ce test. Il suffit de remplacer dans l'expression de la statistique du test le processus $\alpha_n(\mathcal{F})$ par $\alpha_n^{(N)}(\mathcal{F})$ si nous avons la véritable information auxiliaire ou par $\tilde{\alpha}_n^{(N)}(\mathcal{F})$ si nous avons une estimation de cette information. Les deux paragraphes suivants donnent un exemple d'application dans le cas du test de la moyenne et du test du χ^2 . Dans ces deux cas, on transforme la statistique de ces tests et on garde la même décision de rejet. Dans le premier cas nous montrons que le nouveau test statistique a le même risque α mais une puissance améliorée. Dans le second cas, nous prouvons que le risque de première espèce diminue et que sous (H_1) la nouvelle statistique diverge vers l'infini comme la statistique du test du χ^2 usuel.

Test de la moyenne. On note $\sigma_f^{(N)} = \text{Var}(\mathbb{G}^{(N)}(f))$ la variance du processus gaussien raké $\mathbb{G}^{(N)}$ évalué en $f \in \mathcal{F}$. Ce test sert à comparer la moyenne d'un échantillon à une valeur donnée lorsque la variance de l'échantillon est connue. L'hypothèse nulle est $(H_0) : P(f) = P_0(f)$, pour une certaine fonction $f \in \mathcal{F}$ et une mesure de probabilité $P_0 \in \ell^\infty(\mathcal{F})$. La statistique usuelle de ce test est

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

Sous (H_0) , la statistique Z_n suit asymptotiquement la distribution normale standard. Nous rejetons l'hypothèse nulle au niveau α lorsque $|Z_n| > t_\alpha$, $t_\alpha = \Phi(1 - \alpha/2)$ avec Φ la fonction probit. Notre objectif est de savoir si les statistiques suivantes

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

améliore le test de la moyenne. Puisque la loi P est inconnue, les variances σ_f et $\sigma_f^{(N)}$ pour $N \geq 1$ sont généralement inconnues. Néanmoins, un estimateur consistant de ces variances peut être utilisé pour calculer $Z_n, Z_n^{(N)}$ et $\tilde{Z}_n^{(N)}$. Le paragraphe suivant donnera un exemple concret de cette dernière remarque. Faire cela ne changera pas le comportement asymptotique des variables $Z_n, Z_n^{(N)}, \tilde{Z}_n^{(N)}$ que l'hypothèse (H_0) soit vérifiée ou non. Les tests statistiques $|Z_n| > t_\alpha, |Z_n^{(N)}| > t_\alpha, |\tilde{Z}_n^{(N)}| > t_\alpha$ ont le même risque de première espèce puisque sous (H_0) ces variables convergent toutes vers $\mathcal{N}(0, 1)$. La proposition ci-dessous montre que le test est effectivement amélioré quand on utilise les statistiques avec information auxiliaire. Elle montre en particulier que le ratio du risque β du test usuel sur celui du test amélioré diverge vers l'infini.

Proposition 3.7.3. *Supposons que $\text{Var}(\mathbb{G}^{(N)}(f)) < \text{Var}(\mathbb{G}(f))$ et que $n \log(n) = o(n_{(N_0)})$. Sous (H_1) , pour tout $\alpha \in (0, 1)$ il existe une suite $U_n \underset{n \rightarrow +\infty}{\sim} n(P(f) - P_0(f))^2 \left(1/\sigma_f^{(N)} - 1/\sigma_f\right)$ telle que pour n suffisamment grand,*

$$\frac{\mathbb{P}(|Z_n| \leq t_\alpha)}{\mathbb{P}(|Z_n^{(N)}| \leq t_\alpha)} \geq \exp(U_n), \quad \frac{\mathbb{P}(|Z_n| \leq t_\alpha)}{\mathbb{P}(|\tilde{Z}_n^{(N)}| \leq t_\alpha)} \geq \exp(U_n). \quad (3.58)$$

Démonstration. D'après le théorème 3.3.7 et 3.7.2, on peut construire une suite de variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi P et $z_n \sim \mathcal{N}(0, 1)$ tels que pour $n > n_1$ pour un certain $n_1 > 0$, $\mathbb{P}(Z_n) \leq 1/n^2$ avec

$$Z_n = \{|\alpha_n(f)/\sigma_f - z_n| > u_n\} \cup \{|\alpha_n^{(N)}(f)/\sigma_f - z_n| > u_n\},$$

où u_n est une suite de limite nulle. L'approximation forte implique que

$$\lim_{n \rightarrow +\infty} \frac{\mathbb{P}(|Z_n| \leq t_\alpha)}{\mathbb{P}(|z_n + M_n/\sigma_f| \leq t_\alpha)} = 1, \quad \lim_{n \rightarrow +\infty} \frac{\mathbb{P}(|Z_n^{(N)}| \leq t_\alpha)}{\mathbb{P}(|z_n + M_n/\sigma_f^{(N)}| \leq t_\alpha)} = 1, \quad (3.59)$$

avec $M_n = \sqrt{n}(P(f) - P_0(f))$. Si on note f_{μ, σ^2} la fonction de densité de $\mathcal{N}(\mu, \sigma^2)$ alors

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

ce qui entraîne que

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

Pour n suffisamment grand, il existe une suite $U_n \underset{n \rightarrow +\infty}{\sim} M_n^2 \left(1/\sigma_f^{(N)} - 1/\sigma_f\right)$ telle que

$$\frac{\mathbb{P}(|z_n + M_n/\sigma_f| \leq t_\alpha)}{\mathbb{P}(|z_n + M_n/\sigma_f^{(N)}| \leq t_\alpha)} \geq \exp(U_n). \quad (3.60)$$

Alors (3.59) et (3.60) entraîne (3.58). \square

Test de la moyenne dans un cas simple. Pour calculer $Z_n^{(N)}$ or $\tilde{Z}_n^{(N)}$ on a besoin de l'expression de $\sigma_f^{(N)}$. Pour illustrer comment l'obtenir nous travaillons dans un cas simple, quand l'information auxiliaire est donnée par les probabilités de deux partitions ayant chacune deux ensembles. Plus formellement, pour $k \in \mathbb{N}^*$ on définit $\mathcal{A}^{(2k-1)} = \mathcal{A} = \{A, A^C\}$ et $\mathcal{A}^{(2k)} = \mathcal{B} = \{B, B^C\}$. En utilisant la proposition 3.3.3 on peut exprimer simplement $\sigma_f^{(N)}$ pour $N = 1, 2$. Pour des soucis de simplification on note

$$\begin{aligned} p_A &= P(A), \quad p_{\bar{A}} = P(A^C), \quad p_B = P(B), \quad p_{\bar{B}} = P(B^C), \\ p_{AB} &= P(A \cap B), \quad \Delta_A = \mathbb{E}[f|A] - \mathbb{E}[f], \quad \Delta_B = \mathbb{E}[f|B] - \mathbb{E}[f], \end{aligned} \quad (3.61)$$

alors,

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

où $\mathbf{P}_{\mathcal{A}|\mathcal{B}}, \mathbf{P}_{\mathcal{B}|\mathcal{A}}$ sont les matrices stochastiques exprimées simplement avec (3.67), $\mathbb{E}[f|\mathcal{A}], \mathbb{E}[f|\mathcal{B}]$ est les vecteurs conditionnels données par (3.68) et $\text{Var}(\mathbb{G}[\mathcal{A}]), \text{Var}(\mathbb{G}[\mathcal{B}])$ sont les matrices de covariance de $\mathbb{G}[\mathcal{A}] = (\mathbb{G}(A), \mathbb{G}(A^C))$ et $\mathbb{G}[\mathcal{B}] = (\mathbb{G}(B), \mathbb{G}(B^C))$, c'est-à-dire les matrices données par (3.69). D'après le théorème 3.5.6, le processus gaussien raké $\mathbb{G}^{(N)}$ converge presque sûrement quand $N \rightarrow +\infty$ vers le processus gaussien $\mathbb{G}^{(\infty)}$ qui a une expression explicite. La stabilisation de la méthode du Raking-Ratio dans le cas de deux partitions quand $N \rightarrow +\infty$ est rapide puisque la distance de Levy-Prokhorov entre $\mathbb{G}^{(N)}$ et $\mathbb{G}^{(\infty)}$ est presque sûrement bornée par $O(N\lambda^{N/2})$ pour un certain $\lambda \in (0, 1)$. On note $\mathbb{P}_n^{(\infty)}(\mathcal{F})$ la mesure empirique rakée après stabilisation de l'algorithme du Raking-Ratio et $\sigma_f^{(\infty)} = \text{Var}(\mathbb{G}^{(\infty)}(f))$ la variance asymptotique. On définit la statistique suivante

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

D'après la proposition 3.7.3, le test statistique fondé sur la décision de rejet $|Z_n^{(\infty)}| > t_\alpha$ a la même risque α que le Z -test usuel fondé sur la décision $|Z_n| > t_\alpha$ mais il est plus puissant quand n tend vers l'infini. Dans le cas de deux partitions, on peut donner une expression simple et explicite de la variance asymptotique. En utilisant les notations introduites par (3.61) on a

$$\sigma_f^{(\infty)} = \sigma_f^2 - \frac{p_{APB}(p_A \Delta_A^2 + p_B \Delta_B^2 - p_{APB}(\Delta_A - \Delta_B)^2 - 2p_{AB}\Delta_A\Delta_B)}{p_{APB}p_{\bar{A}}p_{\bar{B}} - (p_{AB} - p_{APB})^2}. \quad (3.62)$$

Le calcul de cette variance est faite la sous-section suivante. Si on n' pas les valeurs données par (3.61) on peut utiliser leur estimateur consistant pour évaluer les valeurs de $\sigma_f^{(\infty)}$. Si $\Delta_A = \Delta_B = 0$ alors naturellement l'information auxiliaire est inutile puisque dans ce cas $\sigma_f^{(\infty)} = \sigma_f$, donc il n'y a pas de réduction du risque quadratique. Si A est indépendant de B alors $p_{AB} = p_{APB}$ et

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

Test du χ^2 . Le test d'ajustement du χ^2 consiste à savoir si les données de l'échantillon correspondent à une distribution hypothétique lorsque nous avons une variable catégorielle. Soit $\mathcal{B} = \{B_1, \dots, B_m\}$ une partition de \mathcal{X} . L'hypothèse nulle est

$$(H_0) : P[\mathcal{B}] = P_0[\mathcal{B}], \quad (3.63)$$

où $P[\mathcal{B}] = (P(B_1), \dots, P(B_m))$, $P_0[\mathcal{B}] = (P_0(B_1), \dots, P_0(B_m))$ avec P_0 une certaine mesure de probabilité. La statistique usuelle de ce test est

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

Sous (H_0) , la statistique T_n suit asymptotiquement une loi du χ^2 avec $m-1$ degrés de liberté. L'hypothèse (H_0) est rejetée au seuil α si $Z_n > t_\alpha^{(m)}$, $t_\alpha^{(m)} = \Phi_m(1-\alpha)$ où Φ_m est la fonction de quantile d'une loi $\chi^2(m)$. L'objectif est de savoir si les statistiques suivantes

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

améliore d'une quelconque manière le test du χ^2 . La proposition suivante montre que c'est effectivement le cas.

Proposition 3.7.4. *Sous (H_0) et pour tout $\alpha > 0$,*

$$\lim_{n \rightarrow +\infty} \mathbb{P}(T_n^{(N)} > t_\alpha^{(m)}) \leq \lim_{n \rightarrow +\infty} \mathbb{P}(T_n > t_\alpha^{(m)}) = \alpha, \quad (3.64)$$

et si $n \log n = o(n_{(N)})$ alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\tilde{T}_n^{(N)} > t_\alpha^{(m)}) \leq \alpha. \quad (3.65)$$

Sous (H_1) et pour tout $\alpha > 0$, presque sûrement il existe $n_0 > 0$ tel que pour tout $n > n_0$,

$$\min(|T_n|, |T_n^{(N)}|, |\tilde{T}_n^{(N)}|) > t_\alpha^{(m)}. \quad (3.66)$$

Démonstration. Notons $\mathcal{C} = (C_1, \dots, C_m) = (\mathbb{1}_{B_1}/\sqrt{P(B_1)}, \dots, \mathbb{1}_{B_m}/\sqrt{P(B_m)})$. On traite le cas (H_0) avec **(i)** et le cas (H_1) avec **(ii)**.

(i). Sous (H_0) , $T_n = \alpha_n[\mathcal{C}] \cdot \alpha_n[\mathcal{C}]^T$, $T_n^{(N)} = \alpha_n^{(N)}[\mathcal{C}] \cdot \alpha_n^{(N)}[\mathcal{C}]^T$ et $\tilde{T}_n^{(N)} = \tilde{\alpha}_n^{(N)}[\mathcal{C}] \cdot \tilde{\alpha}_n^{(N)}[\mathcal{C}]^T$. La statistique $\alpha_n[\mathcal{C}]$ converge en loi vers une loi normale multivariée $Y \sim \mathcal{N}(\mathbf{0}, \Sigma)$ alors que les statistiques $\alpha_n^{(N)}[\mathcal{C}]$, $\tilde{\alpha}_n^{(N)}[\mathcal{C}]$ convergent en loi vers $Y^{(N)} \sim \mathcal{N}(\mathbf{0}, \Sigma^{(N)})$ d'après les théorèmes 3.3.7 et 3.7.2. D'après la proposition 3.3.3, $\Sigma - \Sigma^{(N)}$ est semi-définie positive ce qui implique que pour tout $\alpha > 0$,

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

et par conséquence (3.64), (3.65) par convergence en loi.

(ii). Sous (H_1) , il existe $i \in \{1, \dots, m\}$ tel que $P_0(B_i) \neq P(B_i)$ impliquant

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

Par le lemme de Borel-Cantelli et (3.54) avec probabilité 1 il existe $n_1 > 0$ tel que pour tout $n > n_1$, $\Lambda_n < D\sqrt{\log n}$ ce qui entraîne que

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

Puisque $\lim_{n \rightarrow +\infty} t_n = +\infty$, pour tout $\alpha \in (0, 1)$ il existe $n_2 > 0$ tel que $t_n > t_\alpha$ pour tout $n > n_2$. L'inégalité (3.66) est vérifiée pour $n_0 = \max(n_1, n_2)$. \square

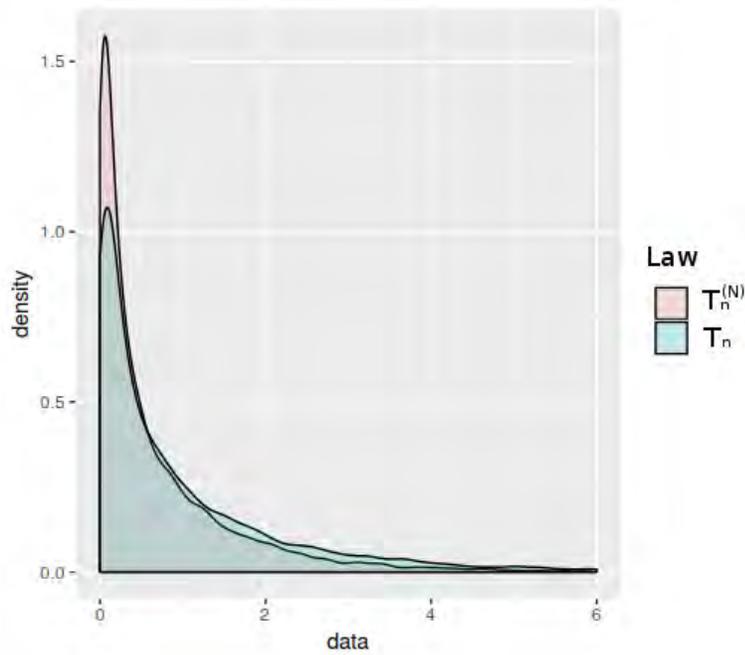


FIG. 3.2: Loi de T_n et $T_n^{(1)}$

La figure 3.2 est un exemple numérique de la proposition 3.7.4 sous (H_0) . Nous avons simulé une table de contingence avec des probabilités fixées $P[\mathcal{B}]$, $P[\mathcal{A}]$ et nous avons appliqué le test du χ^2 avec l'hypothèse nulle (3.63). Par Monte-Carlo, nous avons simulé la loi de T_n avec $n = 1000$ ainsi que la loi de $T_n^{(1)}$ avec l'information auxiliaire donnée par $P[\mathcal{A}]$.

3.7.4 Calcul de $\sigma_f^{(\infty)}$ dans un cas simple

Dans le cas simple, c'est-à-dire le cas où on travaille avec $\mathcal{A}^{(2k-1)} = \mathcal{A} = \{A, A^C\}$ et $\mathcal{A}^{(2k)} = \mathcal{B} = \{B, B^C\}$ pour $k \geq 1$ alors on peut exprimer plus simplement les matrices et

vecteurs intervenant dans l'expression de $\mathbb{G}^{(\infty)}$. On a

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

$$\mathbf{P}_{\mathcal{B}|\mathcal{A}} = \begin{pmatrix} P(B|A) & P(B^C|A) \\ P(B|A^C) & P(B^C|A^C) \end{pmatrix} = \begin{pmatrix} p_{AB}/p_A & 1 - p_{AB}/p_A \\ (p_B - p_{AB})/p_{\bar{A}} & 1 - (p_B - p_{AB})/p_{\bar{A}} \end{pmatrix}, \quad (3.67)$$

$$\mathbb{E}[f|\mathcal{A}] = (\mathbb{E}[f|A], \mathbb{E}[f|A^C]), \quad \mathbb{E}[f|\mathcal{B}] = (\mathbb{E}[f|B], \mathbb{E}[f|B^C]), \quad (3.68)$$

$$\text{Var}(\mathbb{G}[A]) = p_{A\bar{A}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad \text{Var}(\mathbb{G}[B]) = p_{B\bar{B}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad (3.69)$$

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

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

Les valeurs propres de $\mathbf{P}_{\mathcal{A}|\mathcal{B}} \cdot \mathbf{P}_{\mathcal{B}|\mathcal{A}}$ et $\mathbf{P}_{\mathcal{B}|\mathcal{A}} \cdot \mathbf{P}_{\mathcal{A}|\mathcal{B}}$ sont 1 et $T_1 = T_2 = (p_{AB} - p_{A\bar{A}}p_B)^2 / p_{A\bar{A}}p_{\bar{A}}p_Bp_{\bar{B}}$. Leurs vecteurs propres associés à T_1 et T_2 sont respectivement $(p_{\bar{B}}/p_B, -1)^t$ et $(p_{\bar{A}}/p_A, -1)^t$ ce qui entraîne que

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

Dans le cas de deux marginales, $\mathbb{G}^{(N)}$ converge presque sûrement vers $\mathbb{G}^{(\infty)}(f) = \mathbb{G}(f) - S_{1,pair}(f)^t \cdot \mathbb{G}[\mathcal{A}] - S_{2,impair}(f)^t \cdot \mathbb{G}[\mathcal{B}]$ où

$$\begin{aligned} S_{1,pair}(f) &= U_1 \begin{pmatrix} 0 & 0 \\ 0 & (1 - T_1)^{-1} \end{pmatrix} \cdot U_1^{-1} \cdot V_1(f) = C_{1,pair}(f) \begin{pmatrix} -p_{\bar{A}}p_B \\ p_{A\bar{A}}p_B \end{pmatrix}, \\ C_{1,pair}(f) &= \frac{\mathbb{E}[f|B](p_{AB} - p_{A\bar{A}}p_B) - \mathbb{E}[f|A]p_{A\bar{A}}p_{\bar{B}} - \mathbb{E}[f](p_{AB} - p_A)}{p_{A\bar{A}}p_Bp_{\bar{A}}p_{\bar{B}} - (p_{AB} - p_{A\bar{A}}p_B)^2}, \\ S_{2,impair}(f) &= U_2 \begin{pmatrix} 0 & 0 \\ 0 & (1 - T_2)^{-1} \end{pmatrix} \cdot U_2^{-1} \cdot V_2(f) = C_{2,impair}(f) \begin{pmatrix} -p_{A\bar{A}}p_{\bar{B}} \\ p_{A\bar{A}}p_B \end{pmatrix}, \\ C_{2,impair}(f) &= \frac{\mathbb{E}[f|A](p_{AB} - p_{A\bar{A}}p_B) - \mathbb{E}[f|B]p_{\bar{A}}p_B - \mathbb{E}[f](p_{AB} - p_B)}{p_{A\bar{A}}p_Bp_{\bar{A}}p_{\bar{B}} - (p_{AB} - p_{A\bar{A}}p_B)^2}. \end{aligned}$$

Par linéarité de $f \mapsto \mathbb{G}(f)$ et le fait que $\mathbb{G}(a) = 0$ pour toute constante $a \in \mathbb{R}$ on peut écrire que

$$\mathbb{G}^{(\infty)}(f) = \mathbb{G}(f + p_B C_{1,pair}(f) \mathbb{1}_A + p_A C_{2,impair}(f) \mathbb{1}_B),$$

ce qui entraîne que

$$\begin{aligned} \sigma_f^{(\infty)} &= \text{Var}(\mathbb{G}^{(\infty)}(f)) \\ &= \text{Var}(f) + \text{Var}(p_B C_{1,pair}(f) \mathbb{1}_A + p_A C_{2,impair}(f) \mathbb{1}_B) \\ &\quad + 2\text{Cov}(f, p_B C_{1,pair}(f) \mathbb{1}_A + p_A C_{2,impair}(f) \mathbb{1}_B) \\ &= \text{Var}(f) + p_A p_{\bar{A}} p_B^2 C_{1,pair}^2(f) + p_A^2 p_B p_{\bar{B}} C_{2,impair}^2(f) \\ &\quad + 2p_A p_B C_{1,pair}(f) C_{2,impair}(f) (p_{AB} - p_A p_B) \\ &\quad + 2p_A p_B (C_{1,pair}(f) \Delta_A + C_{2,impair}(f) \Delta_B) \end{aligned}$$

Avec quelques calculs on trouve la simple expression de $\sigma_f^{(\infty)}$ donnée par (3.62).

4.1 Introduction

Jusqu'à présent nous avons traité le cas d'une information auxiliaire donnée par une ou plusieurs partitions (chapitres 2 et 3). Cela n'est pas suffisant car l'information auxiliaire peut être plus complexe que la probabilité qu'un individu appartienne à un ensemble d'une partition donnée. Les deux chapitres précédents ne peuvent pas répondre à la problématique de pouvoir traiter une information auxiliaire générale. Il faut créer un outil capable de s'adapter à tout type d'information et donc dans un premier temps définir ce qu'est une information auxiliaire générale.

En premier lieu nous avons pensé à définir une information auxiliaire par la connaissance de l'espérance d'une fonction, c'est-à-dire formellement la connaissance de $\mathbb{E}[f(X)]$ pour une certaine fonction f mesurable et à valeur réelle. Bien qu'elle concerne des cas intéressants, cette définition n'est pas encore satisfaisante puisque certains types d'information auxiliaire que nous évoquerons ne sont pas concernés par cette définition. Dans un second temps, nous avons pensé à l'information auxiliaire comme la connaissance de l'image de la loi $P = \mathbb{P}^X$ par une fonction vérifiant une hypothèse de différentiabilité. Cette définition généralise la définition précédente et englobe les cas manquants à celle-ci. En résumé, l'objectif de cette partie est de créer un outil généralisant toutes les notions précédentes et qui assure de baisser la variance quelque soit l'information apportée.

L'étude générale de l'information auxiliaire sera aussi l'occasion de traiter certains cas spéciaux sur lesquels nous pouvons tomber lors d'une étude statistique. Par exemple, nous évoquerons le cas où une information auxiliaire ne correspond pas à l'espérance d'une fonction mesurable à valeurs réelles. Parmi les exemples proposés, nous traiterons le cas de l'information auxiliaire donnée par la connaissance d'une covariance de deux variables, d'une espérance conditionnelle ou du coefficient de corrélation.

Nous rappelons à la section 4.2 la littérature principale ayant traité l'information auxiliaire d'un point de vue général. Nous définissons et traitons les deux points de vue de l'information auxiliaire générale évoquées ci-dessus dans les sections 4.3 et 4.4.

4.2 Estimateur avec information auxiliaire générale

4.2.1 Littérature

Dmitriev et Tarasenko[34] ont étudié quand l'information auxiliaire est donnée par la connaissance de $P(g_1), \dots, P(g_m)$ ou quand une approximation de cette information est disponible. Ils ont déterminé la projection de la mesure empirique qui minimise la divergence de Kullback-Leibler sur l'ensemble des mesures de probabilité vérifiant cette information. Zhang a étudié l'information auxiliaire donnée par des espérances de fonctions supposées nulles. Il a en particulier montré que les M -estimateurs et les estimateurs de quantile avec information auxiliaire possédaient une variance asymptotique plus faible que les estimateurs classiques [82], que la fonction de répartition empirique avec ce type d'information auxiliaire était uniformément plus petite que la fonction de répartition empirique [83] et a étudié le comportement asymptotique de la fonction quantile avec information auxiliaire [84]. Cette vision de l'information auxiliaire correspond à celle que nous étudions à la section 4.3. Tarima et Pavlov [79] ont étudié de manière générale comment modifier une statistique pour utiliser une information auxiliaire générale donnée par des sources d'information différentes. Ils mettent en évidence que la corrélation entre l'estimateur d'intérêt et l'information auxiliaire intervient dans l'utilisation d'une l'information auxiliaire. Si cette corrélation ainsi que la variance de l'information auxiliaire est connue du statisticien, il est assuré d'avoir l'estimateur le plus efficace en termes de variance, quelque soit la taille de l'échantillon. À l'inverse, si celles-ci ne sont pas connues, le statisticien devra recourir à une méthode de plug-in et sera assuré asymptotiquement d'avoir une réduction de variance optimale. À de nombreux points de vue cet article est très général mais il fait appel à de nombreuses hypothèses de normalité qui sont néanmoins vérifiées dans la plupart des cas pratiques. De plus leur résultat est de type convergence en loi et ne propose pas de vitesse de convergence. En somme, les définitions du processus empirique avec information auxiliaire de cette partie seront des cas particuliers de l'article de Tarima et Pavlov mais les résultats seront plus forts.

4.2.2 Estimateur de Tarima et Pavlov

Notation. De par sa généralité, les notations pour introduire les résultats de Tarima et Pavlov sont lourdes. Nous les rappelons ici et nous ferons le lien avec les objets statistiques introduits dans la suite. Avec un échantillon X_1, \dots, X_n on estime un paramètre $\Theta = (\theta_1, \dots, \theta_S)$ avec un estimateur non biaisé $\hat{\Theta} = (\hat{\theta}_1, \dots, \hat{\theta}_S)$ où $S \in \mathbb{N}^*$ désigne le nombre de statistiques d'intérêts. L'information auxiliaire est donnée par des estimateurs non biaisés $\tilde{\mathcal{B}}_i = (\tilde{\beta}_{i1}, \dots, \tilde{\beta}_{iJ_i})^t$ provenant de I sources d'information différentes pas forcément indépendantes, $i = 1, \dots, I$ et J_i désigne le nombre d'informations apportées par la i ème source d'information. Notre modèle suppose que les $\tilde{\mathcal{B}}_i$ estiment de manière précise $\mathcal{B}_i = (\beta_{i1}, \dots, \beta_{iJ_i})^t$. Le statisticien estime ces derniers avec des estimateurs non biaisés $\hat{\mathcal{B}}_i$. On note $\mathcal{B} = (\mathcal{B}_1^t, \dots, \mathcal{B}_I^t)^t$, $\hat{\mathcal{B}} = (\hat{\mathcal{B}}_1^t, \dots, \hat{\mathcal{B}}_I^t)^t$ et $\tilde{\mathcal{B}} = (\tilde{\mathcal{B}}_1^t, \dots, \tilde{\mathcal{B}}_I^t)^t$. Si X, Y sont des vecteurs aléatoires, on note $(\text{Cov}(X, Y))_{i,j} = \text{Cov}(X_i, Y_j)$. Les vecteurs aléatoires $\hat{\Theta}, \hat{\mathcal{B}}$ dépendent de la taille d'échantillon n .

Estimateur optimal. L'idée de cet article est d'étudier une transformation de l'estimateur $\hat{\Theta}$ afin d'exploiter l'information auxiliaire. L'objectif est de déterminer la valeur de la matrice Λ qui minimise la variance de l'estimateur non biaisé

$$\hat{\Theta}^\Lambda = \hat{\Theta} + \Lambda \cdot (\hat{\mathcal{B}} - \tilde{\mathcal{B}}).$$

Si la matrice $\mathbb{V} = \text{Var}(\hat{\mathcal{B}}) + \text{Var}(\tilde{\mathcal{B}})$ est inversible alors la solution à cette optimisation est de prendre la valeur

$$\Lambda_0 = -\text{Cov}(\hat{\Theta}, \hat{\mathcal{B}}) \cdot \mathbb{V}^{-1}. \quad (4.1)$$

Si la matrice \mathbb{V} n'est pas inversible il suffit de considérer la valeur $\Lambda_0^+ = -\text{Cov}(\hat{\Theta}, \hat{\mathcal{B}}) \cdot \mathbb{V}^+$, où $A^+ = \lim_{\delta \rightarrow 0} (A^t A + \delta \text{Id})^{-1} A^t$ est la pseudo-inverse (ou inverse généralisée) d'une matrice carrée A . Sauf remarque de notre part nous supposons que la matrice de covariance \mathbb{V} est inversible. L'estimateur optimal de Θ devient

$$\hat{\Theta}^0 = \hat{\Theta} + \Lambda_0 \cdot (\hat{\mathcal{B}} - \tilde{\mathcal{B}}) = \hat{\Theta} - \text{Cov}(\hat{\Theta}, \hat{\mathcal{B}}) \cdot \mathbb{V}^{-1} \cdot (\hat{\mathcal{B}} - \tilde{\mathcal{B}}).$$

La matrice de covariance de cette nouvelle statistique est

$$\text{Var}(\hat{\Theta}^0) = \text{Var}(\hat{\Theta}) - \text{Cov}(\hat{\Theta}, \hat{\mathcal{B}}) \cdot \mathbb{V}^{-1} \cdot \text{Cov}(\hat{\Theta}, \hat{\mathcal{B}})^t. \quad (4.2)$$

Puisque l'inverse d'une matrice semi-définie positive est semi-définie positive,

$$\text{Var}(\hat{\Theta}^0) \leq \text{Var}(\hat{\Theta}). \quad (4.3)$$

La pseudo-inverse d'une matrice semi-définie positive est semi-définie positive (voir corollaire 3 de [58]), donc cette propriété est aussi vérifiée quand on utilise la matrice Λ_0^+ , autrement dit,

$$\text{Var}(\hat{\Theta} - \Lambda_0^+ \cdot (\hat{\mathcal{B}} - \tilde{\mathcal{B}})) \leq \text{Var}(\hat{\Theta}).$$

Si la covariance entre la fonction d'intérêt et l'information auxiliaire est nulle, c'est-à-dire si $\text{Cov}(\hat{\Theta}, \hat{\mathcal{B}}) = 0$ alors il n'y a pas de réduction de variance. Si les sources sont indépendantes entre elles alors la matrice de covariance $\text{Var}(\tilde{\mathcal{B}})$ est diagonale par blocs.

Estimateur adaptatif. Sous réserve d'avoir la valeur Λ_0 le statisticien est en mesure de calculer la nouvelle statistique $\hat{\Theta}_0$ qui exploite l'information auxiliaire. Il est assuré que, pour toute taille d'échantillon n fixée, d'améliorer assurément et de manière optimale la statistique initiale $\hat{\Theta}$. Dans la majorité des cas le statisticien n'aura pas accès à la vraie valeur de Λ_0 puisqu'il n'aura pas à sa disposition la valeur exacte des matrices $\text{Cov}(\hat{\Theta}, \hat{\mathcal{B}})$ ou \mathbb{V} ou bien ces deux dernières valeurs. Pour compenser ce manque d'information, on peut substituer Λ_0 par un estimateur consistant $\hat{\Lambda}_0$ de celui-ci. On note $\hat{\Theta}^*$ l'estimateur adaptatif défini par

$$\hat{\Theta}^* = \hat{\Theta} - \hat{\Lambda}_0 \cdot (\hat{\mathcal{B}} - \tilde{\mathcal{B}}),$$

que le statisticien pourra utiliser s'il lui manque la valeur de Λ_0 .

Hypothèses et résultats. Le résultat de Tarima et Pavlov reste général mais suppose que certaines conditions de comportement asymptotique gaussien soient vérifiées. Nous listons ces conditions ci-après et supposons qu'elles sont vérifiées.

- $a_n(\hat{\Theta} - \Theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \xi$ avec $\xi \sim \mathcal{N}(0, \Sigma_{11})$, a_n une suite de réels positifs, telle que $a_n \xrightarrow[n \rightarrow +\infty]{} +\infty$ et $a_n^2 \text{Var}(\hat{\Theta}) \xrightarrow[n \rightarrow +\infty]{} \Sigma_{11}$;
- $\tau_n = a_n(\hat{\mathcal{B}} - \mathcal{B}) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \tau$ avec $\tau \sim \mathcal{N}(0, \Sigma'_{22})$ et $a_n^2 \text{Cov}(\hat{\mathcal{B}}, \hat{\mathcal{B}}) \xrightarrow[n \rightarrow +\infty]{} \Sigma'_{22}$;
- $\zeta_{in} = b_{in}(\tilde{\mathcal{B}}_i - \mathcal{B}_i) \rightarrow \zeta_i$ où b_{in} est une suite de réels positifs telle que $b_{in} \xrightarrow[n \rightarrow +\infty]{} +\infty$, $\zeta_i \sim \mathcal{N}(0, \Sigma''_{22i})$, $i = 1, \dots, I$ et $b_{in}^2 \text{Cov}(\tilde{\mathcal{B}}_i, \tilde{\mathcal{B}}_i) \xrightarrow[n \rightarrow +\infty]{} \Sigma''_{22i}$.

Notons $\Sigma_{12} = \text{Cov}(\xi, \tau)$. Sous ces hypothèses de convergence, Tarima et Pavlov ont démontré le résultat suivant (voir Proposition 1 et 2 de [79]).

Proposition 4.2.1. *Si $a_n b_{in}^{-1} \rightarrow w_i \in [0, +\infty[$ et $\Sigma_{22} = \Sigma'_{22} + \text{diag}(w_i^2 \Sigma''_{22i})$ est définie positive alors $a_n(\hat{\Theta}^0 - \Theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^t)$ et $a_n(\hat{\Theta}^* - \Theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^t)$.*

Ce résultat est général par rapport aux statistiques concernées mais n'apporte pas d'information concernant la vitesse de convergence et concerne qu'une seule statistique à la fois.

4.3 Information auxiliaire donnée par des espérances

4.3.1 Introduction

Motivation. Nous avons étudié lors des parties précédentes comment injecter dans notre étude statistique une information auxiliaire donnée par la connaissance d'appartenir à l'ensemble d'une ou de plusieurs partitions. Dans tous les cas, nous avons montré qu'on augmentait nécessairement la qualité de nos estimations en baissant la variance des estimateurs et la puissance de certains tests statistiques. Un problème reste à résoudre, celle de savoir comment intégrer dans nos études statistiques l'information auxiliaire apportée par la connaissance d'une fonction quelconque $P(g)$, voir la connaissance de plusieurs fonctions quelconques $P(g_1), \dots, P(g_m)$. La question est de savoir combiner cette information pour créer un processus empirique qui tient compte de toute cette information à la fois et qui nous assurerait de baisser de manière optimale la variance du processus empirique. Peut-on créer un processus qui gère une information auxiliaire partielle, par exemple qui gère le cas où le statisticien possède la connaissance des $P(g_1), \dots, P(g_m)$ sans connaître nécessairement les covariances $P(g_i g_j)$ ou $P(f g_i)$ où $f(X)$ est la variable d'intérêt. Il ne s'agit pas ici du cas le plus général que sera étudié dans la thèse mais il n'en reste pas moins intéressant.

Organisation. Nous introduisons dans cette sous-section le processus empirique indexé par une classe de fonctions mesurables à valeurs réelles permettant de répondre à ce problème particulier. Dans la sous-section 4.3.2 nous établissons des résultats élémentaires comme la baisse de variance ainsi que des résultats plus puissants tels que l'approximation forte et la convergence faible de ce nouveau processus vers un processus gaussien centré indexé ayant une variance plus faible que le P -pont brownien. Nous donnons des exemples d'applications statistiques dans la sous-section 4.3.3 qui rentre dans notre cadre d'étude. Nous finissons dans la sous-section 4.3.4 par étudier comment cette partie généralise certains des chapitres précédents.

Notation. Nous supposons que le statisticien a à sa disposition la connaissance de l'espérance de plusieurs fonctions, *i.e.* qu'il connaît $P(g_i), \forall 1 \leq i \leq m$ avec $(g_i)_{1 \leq i \leq m}$ des fonctions mesurables à valeurs réelles. Nous supposons que la classe des fonctions $\{g_1, \dots, g_m\}$ vérifient l'hypothèse $\mathcal{F}.i$. Nous devons donc modifier la mesure empirique \mathbb{P}_n en prenant en compte les mesures empiriques $\mathbb{P}_n(g_i)$ de telle sorte à garder un processus centré. Il est naturel donc de faire intervenir une combinaison linéaire des termes centrés $\mathbb{P}_n(g_i) - P(g_i)$ avec des coefficients à optimiser afin de réduire au mieux la variance du processus empirique. On pose le vecteur $\mathbf{g} = (g_1, \dots, g_m)^t \in \mathcal{F}^m$ contenant les fonctions dont on connaît l'espérance $P[\mathbf{g}] = (P(g_1), \dots, P(g_m))^t$ et le vecteur des coefficients $\Lambda = (\lambda_1, \dots, \lambda_m)^t$ associés aux termes $\mathbb{P}_n(g_i) - P(g_i)$. On définit la mesure et le processus avec l'information auxiliaire

généralisée $P[\mathbf{g}]$ par :

$$\begin{aligned}\mathbb{P}_n^{\mathbf{g},\Lambda}(f) &= \mathbb{P}_n(f) + \Lambda \cdot (\mathbb{P}_n[\mathbf{g}] - P[\mathbf{g}]), \\ \alpha_n^{\mathbf{g},\Lambda}(f) &= \sqrt{n}(\mathbb{P}_n^{\mathbf{g},\Lambda}(f) - P(f)) = \alpha_n(f) + \Lambda \cdot \alpha_n[\mathbf{g}],\end{aligned}$$

avec $\mathbb{P}_n[\mathbf{g}] = (\mathbb{P}_n(g_1), \dots, \mathbb{P}_n(g_m))^t$ et $\alpha_n[\mathbf{g}] = \sqrt{n}(\mathbb{P}_n[\mathbf{g}] - P[\mathbf{g}])$.

Estimateur optimal. Le processus empirique $\alpha_n^{\mathbf{g},\Lambda}(\mathcal{F})$ est un cas particulier de celui de Tarima et Pavlov. Il suffit de prendre $\Theta = P(f)$, $\hat{\Theta} = \mathbb{P}_n(f)$, $\mathcal{B} = \hat{\mathcal{B}} = P[\mathbf{g}]$, $\hat{\mathcal{B}} = \mathbb{P}_n[\mathbf{g}]$. On note $\text{Var}(\mathbf{g})$ la matrice de variance définie pour tout $1 \leq i, j \leq m$ par

$$(\text{Var}(\mathbf{g}))_{i,j} = \text{Cov}(g_i(X), g_j(X)).$$

Elle correspond à la matrice \mathbb{V} introduite précédemment. On note $\text{Cov}(\mathbf{g}, f)$ la matrice définie pour tout $1 \leq i \leq m$ par

$$(\text{Cov}(\mathbf{g}, f))_i = \text{Cov}(g_i(X), f(X)).$$

Elle correspond à la matrice $\text{Cov}(\hat{\Theta}, \hat{\mathcal{B}})$ introduite précédemment. La valeur Λ_0 minimisant la variance du processus $\alpha_n^{\mathbf{g},\Lambda}(f)$ dépend de la fonction f . Sous réserve que la matrice de variance $\text{Var}(\mathbf{g})$ soit inversible, sa valeur est donnée grâce à (4.1) par

$$\Lambda_0 = -\text{Cov}(\mathbf{g}, f) \cdot \text{Var}(\mathbf{g})^{-1}.$$

Si la matrice de covariance $\text{Var}(\mathbf{g})$ n'est pas inversible il suffit de considérer la valeur

$$\Lambda_0^+ = -\text{Cov}(\mathbf{g}, f) \cdot \text{Var}(\mathbf{g})^+.$$

La mesure $\mathbb{P}_n^{\mathbf{g}}(\mathcal{F})$ et le processus empirique $\alpha_n^{\mathbf{g}}(\mathcal{F})$ avec information auxiliaire donnée par des espérances de fonctions sont respectivement définis pour tout $f \in \mathcal{F}$ par

$$\begin{aligned}\mathbb{P}_n^{\mathbf{g}}(f) &= \mathbb{P}_n^{\mathbf{g},\Lambda_0}(f) = \mathbb{P}_n(f) - \Lambda_0 \cdot \alpha_n[\mathbf{g}] \\ &= \mathbb{P}_n(f) - \text{Cov}(\mathbf{g}, f) \cdot \text{Var}(\mathbf{g})^{-1} \cdot (\mathbb{P}_n[\mathbf{g}] - P[\mathbf{g}]), \\ \alpha_n^{\mathbf{g}}(f) &= \alpha_n^{\mathbf{g},\Lambda_0}(f) = \alpha_n(f) - \Lambda_0 \cdot \alpha_n[\mathbf{g}] \\ &= \alpha_n(f) - \text{Cov}(\mathbf{g}, f) \cdot \text{Var}(\mathbf{g})^{-1} \cdot \alpha_n[\mathbf{g}].\end{aligned}$$

Comme pour l'article de Tarima et Pavlov, nous étudions ci-après le cas où le statisticien n'aurait pas accès à la valeur de Λ_0 .

Estimateur adaptatif. Comme nous l'avons évoqué précédemment, si le statisticien n'a pas accès à la valeur de $\text{Cov}(\mathbf{g}, f)$ ou celle de $\text{Var}(\mathbf{g})$ alors il peut la substituer la valeur manquante par un estimateur consistant. On note $\text{Var}_n(\mathbf{g})$ et $\text{Cov}_n(\mathbf{g}, f)$ des estimateurs consistants respectifs de $\text{Var}(\mathbf{g})$ et $\text{Cov}(\mathbf{g}, f)$. Par exemple on peut prendre les estimateurs suivants définies pour tout $1 \leq i, j \leq m$ par

$$\begin{aligned}(\text{Var}(\mathbf{g}))_{i,j} &= \text{Cov}_n(g_i, g_j) \\ &= \frac{1}{n} \sum_{k=1}^n g_i(X_k)g_j(X_k) - \left(\frac{1}{n} \sum_{k=1}^n g_i(X_k) \right) \left(\frac{1}{n} \sum_{k=1}^n g_j(X_k) \right), \\ (\text{Cov}(\mathbf{g}, f))_i &= \text{Cov}_n(g_i, f) \\ &= \frac{1}{n} \sum_{k=1}^n g_i(X_k)f(X_k) - \left(\frac{1}{n} \sum_{k=1}^n g_i(X_k) \right) \left(\frac{1}{n} \sum_{k=1}^n f(X_k) \right).\end{aligned}$$

Si $\text{Var}_n(\mathbf{g})$ est inversible il peut par exemple utiliser l'estimateur naturel

$$\widehat{\Lambda}_0 = -\text{Cov}_n(\mathbf{g}, f) \cdot \text{Var}_n(\mathbf{g})^{-1}. \quad (4.4)$$

Si $\text{Var}_n(\mathbf{g})$ n'est pas inversible il peut par exemple utiliser l'estimateur

$$\widehat{\Lambda}_0 = -\text{Cov}_n(\mathbf{g}, f) \cdot \text{Var}_n(\mathbf{g})^+. \quad (4.5)$$

Dans un tel cas, le statisticien exploitera l'estimateur adaptatif suivant

$$\begin{aligned} \mathbb{P}_n^{\mathbf{g}*}(f) &= \mathbb{P}_n(f) - \widehat{\Lambda}_0 \cdot (\mathbb{P}_n[\mathbf{g}] - P[\mathbf{g}]), \\ \alpha_n^{\mathbf{g}*}(f) &= \alpha_n(f) - \widehat{\Lambda}_0 \cdot \alpha_n[\mathbf{g}], \end{aligned}$$

où $\widehat{\Lambda}_0$ est un estimateur consistant de Λ_0 . L'exemple qui suit reprend l'information auxiliaire apportée par la connaissance de la probabilité d'appartenir à l'ensemble d'une partition, cas étudié au chapitre 2.

Exemple 4.3.1. Prenons $g_1 = \mathbb{1}_{A_1}, \dots, g_m = \mathbb{1}_{A_m}$ avec $(A_i)_{1 \leq i \leq m}$ une partition de \mathcal{X} . On suppose que le statisticien a à sa disposition les valeurs $\mathbb{E}[g_i] = P(A_i)$. Les matrices $\text{Cov}(\mathbf{g}, f)$ et $\text{Var}(\mathbf{g})$ valent respectivement

$$\begin{aligned} (\text{Cov}(\mathbf{g}, f))_i &= P(A_i)(\mathbb{E}[f|A_i] - \mathbb{E}[f]), \\ \text{Var}(\mathbf{g}) &= \text{diag}(P[\mathcal{A}]) - P[\mathcal{A}] \cdot P[\mathcal{A}]^t, \end{aligned}$$

où $1 \leq i \leq m$ et $P[\mathcal{A}] = (P(A_1), \dots, P(A_m))$. La solution optimale est

$$\Lambda_0 = -(\mathbb{E}[f|A_1], \dots, \mathbb{E}[f|A_m]).$$

Si le statisticien connaît les valeurs $\mathbb{E}[f|A_i]$ il n'aurait pas besoin d'utiliser l'échantillon pour estimer $\mathbb{E}[f]$ puisqu'il pourrait effectuer le calcul direct $\mathbb{E}[f] = -\sum_{i=1}^m \mathbb{E}[f|A_i]\mathbb{E}[g_i]$.

Hypothèse supplémentaire. Nous établissons des résultats aussi bien pour le processus $\alpha_n^{\mathbf{g}}(\mathcal{F})$ que pour le processus adaptatif $\alpha_n^{\mathbf{g}*}(\mathcal{F})$. Nous devons donc supposer une vitesse de convergence supplémentaire et suffisante pour l'estimateur $\widehat{\Lambda}_0$ de Λ_0 . On pose $\|\widehat{\Lambda}_0 - \Lambda_0\|_{\mathcal{F}} = \max_{i=1, \dots, m} \|(\widehat{\Lambda}_0 - \Lambda_0)_i\|_{\mathcal{F}}$. Nous ferons régulièrement appel à l'hypothèse suivante.

Hypothèse (CV). Il existe $C_0 = C_0(\mathcal{F}) > 0$ tel que pour tout $t > 0$ et n suffisamment grand

$$\mathbb{P}\left(\|\widehat{\Lambda}_0 - \Lambda_0\|_{\mathcal{F}} > t\right) \leq \exp(-C_0 n t^2).$$

Cette hypothèse est naturellement vérifiée si $\widehat{\Lambda}_0$ est l'estimateur empirique de Λ_0 donnée par (4.4) ou (4.5). Elle est également vérifiée si on prend un estimateur au moins aussi efficace que ces derniers.

4.3.2 Résultats

Propriétés immédiates. La nouvelle mesure $\mathbb{P}_n^{\mathbf{g}}(\mathcal{F})$ est linéaire, par linéarité de $f \mapsto \mathbb{P}_n(f)$ et $f \mapsto \Lambda_0(f)$. Elle vérifie en outre l'information auxiliaire $P[\mathbf{g}]$ puisque

$$\begin{aligned} \mathbb{P}_n^{\mathbf{g}}[\mathbf{g}] &= (\mathbb{P}_n^{\mathbf{g}, \Lambda_0}(g_1), \dots, \mathbb{P}_n^{\mathbf{g}, \Lambda_0}(g_m))^t \\ &= \mathbb{P}_n(\mathbf{g}) - \text{Var}(\mathbf{g}) \cdot \text{Var}(\mathbf{g})^{-1} \cdot (\mathbb{P}_n(\mathbf{g}) - P(\mathbf{g})) = P[\mathbf{g}]. \end{aligned}$$

D'après (4.2) la variance du processus empirique avec information auxiliaire $\alpha_n^{\mathbf{g}}(\mathcal{F})$ est donnée par la relation suivante. Pour tout $f \in \mathcal{F}$,

$$\text{Var}(\alpha_n^{\mathbf{g}}(f)) = \text{Var}(f) - \text{Cov}(\mathbf{g}, f) \cdot \text{Var}(\mathbf{g})^{-1} \cdot \text{Cov}(\mathbf{g}, f)^t.$$

En particulier, comme (4.3), $\text{Var}(\alpha_n^{\mathbf{g}}(f)) \leq \text{Var}(\alpha_n(f))$ pour tout $f \in \mathcal{F}$. La matrice de covariance de $\alpha_n^{\mathbf{g}}(\mathcal{F})$ est donnée pour tout $f, g \in \mathcal{F}$ par

$$\text{Cov}(\alpha_n^{\mathbf{g}}(f), \alpha_n^{\mathbf{g}}(g)) = \text{Cov}(f, g) - \text{Cov}(\mathbf{g}, f) \cdot \text{Var}(\mathbf{g})^{-1} \cdot \text{Cov}(\mathbf{g}, g)^t.$$

Qualité de l'information. Une idée naturelle vérifiée par la généralisation du processus empirique avec information auxiliaire est que plus on a d'informations à disposition, plus la variance du processus diminue. Cette idée est confirmée par la proposition suivante.

Proposition 4.3.1. *Soient $m \leq m'$, $g_1, \dots, g_{m'}$ des fonctions mesurables de \mathcal{X} dans \mathbb{R} et $\mathbf{g} = (g_1, \dots, g_m)$, $\mathbf{g}' = (g_1, \dots, g_{m'})$. Alors, $\text{Var}(\alpha_n^{\mathbf{g}}(f)) \geq \text{Var}(\alpha_n^{\mathbf{g}'}(f))$.*

Démonstration. Il suffit de voir que pour $\Lambda = (\lambda_1, \dots, \lambda_m)$ on peut écrire que $\alpha_n^{\mathbf{g}, \Lambda} = \alpha_n^{\mathbf{g}', \Lambda'}$ avec $\Lambda' = (\lambda_1, \dots, \lambda_m, 0, \dots, 0)$ qui sera nécessairement de variance supérieure à $\alpha_n^{\mathbf{g}', \Lambda'}$. \square

Une autre question naturelle est de savoir si la qualité de l'information est améliorée si on décompose une information donnée par une fonction réelle en somme d'autres fonctions dont on connaît l'espérance. Par exemple dans le cas de partitions, nous nous demandons ici si le fait de connaître $P(A_1), \dots, P(A_m)$ où A_1, \dots, A_m est une partition de A est plus efficace que de simplement connaître $P(A)$. La proposition suivante montre qu'une information raffinée améliore l'efficacité du processus.

Proposition 4.3.2. *Soient $m' \geq m$ et $g_1, \dots, g_{m'}$ des fonctions mesurables de \mathcal{X} à valeurs réelles qui forment un raffinement de g_1, \dots, g_m dans le sens où les ensembles $(J_i)_{1 \leq i \leq m}$ définies par*

$$\forall i = 1, \dots, m, \quad g_i = \sum_{j \in J_i} g'_j$$

forment une partition de $\{1, \dots, m'\}$. Alors

$$\forall f \in \mathcal{F}, \quad \text{Var}(\alpha_n^{\mathbf{g}}(f)) \geq \text{Var}(\alpha_n^{\mathbf{g}'}(f))$$

où $\mathbf{g}' = (g'_1, \dots, g'_{m'})^t$.

Démonstration. Pour $j = 1, \dots, m'$ on note i_j l'indice vérifiant $j \in J_{i_j}$. Par hypothèse on peut écrire que

$$\begin{aligned} \mathbb{P}_n^{\mathbf{g}, \Lambda}(f) &= \mathbb{P}_n(f) + \sum_{i=1}^m \lambda_i (\mathbb{P}_n(g_i) - P(g_i)) = \mathbb{P}_n(f) + \sum_{i=1}^m \lambda_i \left(\sum_{j \in J_i} \mathbb{P}_n(g'_j) - P(g'_j) \right) \\ &= \mathbb{P}_n(f) + \sum_{j=1}^{m'} \lambda_{i_j} (\mathbb{P}_n(g'_{j'}) - P(g'_{j'})). \end{aligned}$$

Pour tout $f \in \mathcal{F}$ et $\Lambda \in \mathbb{R}^m$, $\mathbb{P}_n^{\mathbf{g}, \Lambda}(f)$ est nécessairement de variance supérieure à $\mathbb{P}_n^{\mathbf{g}'}(f)$. \square

Remarque 4.3.1. *La proposition précédente est une généralisation de la proposition 2.2.6.*

Approximation forte. Si la classe de fonctions \mathcal{F} vérifie les conditions d'entropie du théorème de Berthet-Mason alors nous sommes en mesure d'établir un résultat d'approximation forte du processus $\alpha_n^{\mathbf{g}}(\mathcal{F})$ avec information auxiliaire avec le processus gaussien centré $\mathbb{G}^{\mathbf{g}}(\mathcal{F})$ défini pour tout $f \in \mathcal{F}$ par

$$\mathbb{G}^{\mathbf{g}}(f) = \mathbb{G}(f) + \Lambda_0 \cdot \mathbb{G}[\mathbf{g}] = \mathbb{G}(f) - \text{Cov}(\mathbf{g}, f) \cdot \text{Var}(\mathbf{g})^{-1} \cdot \mathbb{G}[\mathbf{g}], \quad (4.6)$$

où $\mathbb{G}[\mathbf{g}] = (\mathbb{G}(g_1), \dots, \mathbb{G}(g_k))^t$. Ce processus gaussien possède la même variance que celui du processus $\alpha_n^{\mathbf{g}}(\mathcal{F})$, c'est-à-dire pour tout $f \in \mathcal{F}$,

$$\text{Var}(\mathbb{G}^{\mathbf{g}}(f)) = \text{Var}(f) - \text{Cov}(\mathbf{g}, f) \cdot \text{Var}(\mathbf{g})^{-1} \cdot \text{Cov}(\mathbf{g}, f)^t. \quad (4.7)$$

La fonction de covariance du processus $\mathbb{G}^{\mathbf{g}}(\mathcal{F})$ est donnée pour tout $f, g \in \mathcal{F}$ par

$$\text{Cov}(\mathbb{G}^{\mathbf{g}}(f), \mathbb{G}^{\mathbf{g}}(g)) = \text{Cov}(f, g) - \text{Cov}(\mathbf{g}, f) \cdot \text{Var}(\mathbf{g})^{-1} \cdot \text{Cov}(\mathbf{g}, g)^t.$$

Nous mentionnons ce dernier comme le P -pont brownien avec information auxiliaire donnée par des espérances de fonctions. Le théorème suivant conserve le même format que l'approximation forte de Berthet-Mason.

Théorème 4.3.3. *Si \mathcal{F} vérifie $\mathcal{F}.i$, $\mathcal{F}.ii$, BR ou VC alors pour tout $\theta > 0$ il existe $C_\theta > 0$, une suite v_n de la forme $v_n = n^{-\alpha}(\log n)^\beta$ si \mathcal{F} vérifie VC et $v_n = (\log n)^{-\beta}$ si \mathcal{F} vérifie BR (avec $\alpha, \beta > 0$), des variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi P et des processus $(\mathbb{G}_n)_n$ de P -pont brownien définis sur un même espace de probabilité vérifiant pour tout n ,*

$$\mathbb{P}(\|\alpha_n^{\mathbf{g}} - \mathbb{G}_n^{\mathbf{g}}\|_{\mathcal{F}} > C_\theta v_n) \leq \frac{1}{n^\theta}, \quad (4.8)$$

et presque sûrement

$$\|\alpha_n^{\mathbf{g}} - \mathbb{G}_n^{\mathbf{g}}\|_{\mathcal{F}} = O(v_n), \quad (4.9)$$

où $\mathbb{G}_n^{\mathbf{g}}$ est le P -pont brownien avec information auxiliaire donnée par des espérances de fonction défini par (4.6) via les \mathbb{G}_n . De plus si CV est vérifiée alors pour tout n suffisamment grand,

$$\mathbb{P}(\|\alpha_n^{\mathbf{g}^*} - \mathbb{G}_n^{\mathbf{g}}\|_{\mathcal{F}} > C_\theta v_n) \leq \frac{1}{n^\theta}, \quad (4.10)$$

et presque sûrement

$$\|\alpha_n^{\mathbf{g}^*} - \mathbb{G}_n^{\mathbf{g}}\|_{\mathcal{F}} = O(v_n). \quad (4.11)$$

Démonstration. D'après le théorème de Berthet-Mason, il existe $C_0 = C_0(\theta) > 0$, une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables i.i.d. de loi P , une suite $(\mathbb{G}_n)_{n \in \mathbb{N}^*}$ de P -pont browniens construits sur un même espace de probabilité et tels que pour tout n ,

$$\mathbb{P}(\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} > C_0 v_n) \leq \frac{1}{3n^\theta}. \quad (4.12)$$

Si $\Lambda_0 = (\lambda_{0,1}, \dots, \lambda_{0,m})$ alors $\|\alpha_n^{\mathbf{g}} - \mathbb{G}_n^{\mathbf{g}}\|_{\mathcal{F}} \leq (1 + \sum_{i=1}^m |\lambda_{0,i}|) \|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}}$ où $(\mathbb{G}_n^{\mathbf{g}}(\mathcal{F}))_n$ sont les P -ponts browniens avec information auxiliaire donnée par des espérances de fonctions \mathbf{g} construits à partir des $(\mathbb{G}_n(\mathcal{F}))_n$ via (4.6). Il suffit de prendre $C_1 = (1 + \sum_{i=1}^m \lambda_{0,i})C_0$ pour conclure grâce à (4.12) à l'inégalité

$$\mathbb{P}(\|\alpha_n^{\mathbf{g}} - \mathbb{G}_n^{\mathbf{g}}\|_{\mathcal{F}} > C_1 v_n) \leq \frac{1}{3n^\theta},$$

qui entraîne (4.8) avec $C_\theta = C_1$. Pour la seconde partie du théorème, il suffit d'établir une inégalité de concentration pour $\|\alpha_n^{\mathbf{g}^*} - \alpha_n^{\mathbf{g}}\|_{\mathcal{F}}$. Pour tout $n \in \mathbb{N}^*$,

$$\|\alpha_n^{\mathbf{g}^*} - \alpha_n^{\mathbf{g}}\|_{\mathcal{F}} = \|(\widehat{\Lambda}_0 - \Lambda_0) \cdot \alpha_n[\mathbf{g}]\|_{\mathcal{F}} \leq m \|\alpha_n\|_{\mathcal{F}} \|\widehat{\Lambda}_0 - \Lambda_0\|_{\mathcal{F}}.$$

D'après l'inégalité de Talagrand et l'hypothèse CV, il existe $D = D(\theta) > 0$ tel que pour tout n suffisamment grand,

$$\mathbb{P}(\|\alpha_n\|_{\mathcal{F}} > \sqrt{D \log n}) \leq \frac{1}{3n^\theta}, \quad \mathbb{P}\left(\|\widehat{\Lambda}_0 - \Lambda_0\|_{\mathcal{F}} > \sqrt{\frac{D \log n}{n}}\right) \leq \frac{1}{3n^\theta}.$$

On a ainsi montré que

$$\begin{aligned} \mathbb{P}\left(\|\alpha_n^{\mathbf{g}^*} - \mathbb{G}_n^{\mathbf{g}}\|_{\mathcal{F}} > C_1 v_n + \frac{Dm \log n}{\sqrt{n}}\right) &\leq \mathbb{P}(\|\alpha_n^{\mathbf{g}} - \mathbb{G}_n^{\mathbf{g}}\|_{\mathcal{F}} > C_1 v_n) \\ &+ \mathbb{P}(\|\alpha_n\|_{\mathcal{F}} > \sqrt{D \log n}) + \mathbb{P}\left(\|\widehat{\Lambda}_0 - \Lambda_0\|_{\mathcal{F}} > \sqrt{\frac{D \log n}{n}}\right), \end{aligned}$$

qui permet de conclure à (4.10) en prenant C_θ suffisamment grand pour avoir $C_\theta v_n > C_1 v_n + Dm \log n / \sqrt{n}$. Le lemme de Borel-Cantelli appliqué à (4.8), (4.10) et $\theta > 1$, nous permet de conclure à (4.9) et (4.11). \square

Convergence en loi. Le résultat de Tarima et Pavlov rappelé par proposition 4.2.1 montre que pour $f \in \mathcal{F}$ fixé, $\alpha_n^{\mathbf{g}}(f)$ converge en loi vers $\mathbb{G}^{\mathbf{g}}(f)$. La proposition suivante renforce ce résultat en montrant que la convergence reste vraie au niveau du processus $\alpha_n^{\mathbf{g}}(\mathcal{F})$ qui converge faiblement dans $\ell^\infty(\mathcal{F})$ vers $\mathbb{G}^{\mathbf{g}}(\mathcal{F})$ et donne même une vitesse de convergence.

Proposition 4.3.4. *On note d_{LP} la distance de Lévy-Prokhorov. Si \mathcal{F} vérifie $\mathcal{F}.i$ et $\mathcal{F}.ii$ alors il existe $C > 0$ tel que pour tout n , $d_{LP}(\alpha_n^{\mathbf{g}}, \mathbb{G}^{\mathbf{g}}) \leq C v_n$. Si l'hypothèse CV est vérifiée alors pour n suffisamment grand, $d_{LP}(\alpha_n^{\mathbf{g}^*}, \mathbb{G}^{\mathbf{g}}) \leq C v_n$.*

La preuve de cette proposition est en tout point similaire à celle de la proposition 2.2.3 en utilisant l'approximation forte du théorème 4.3.3 et ne sera donc pas présentée.

4.3.3 Exemples

Variable de contrôle. La variable de contrôle est une méthode statistique utilisée pour obtenir une réduction de la variance en utilisant la connaissance de la corrélation entre deux statistiques (voir par exemple [62]). Concrètement, si X, Y sont des variables corrélées et que l'on connaît $\mathbb{E}[Y]$ alors on peut utiliser cette information afin de réduire la variance de l'évaluation de X en considérant la variable X^* définie par

$$X^* = X - \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}(Y - \mathbb{E}[Y]).$$

La nouvelle variance est donnée par

$$\text{Var}(X^*) = \text{Var}(X) - \frac{\text{Cov}(X, Y)^2}{\text{Var}(Y)} = (1 - \text{Corr}(X, Y))\text{Var}(X),$$

où $\text{Corr}(X, Y) = \text{Cov}(X, Y) / \sigma_X \sigma_Y$ est le coefficient de corrélation entre X et Y . Si on travaille sur un échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$ il est possible d'améliorer de la même

façon l'estimation de la moyenne empirique de X en utilisant l'information auxiliaire $\mathbb{E}[Y]$. En effet, si on travaille avec la loi $P = \mathbb{P}^{(X,Y)}$, que notre fonction d'intérêt est $f(X, Y) = X$ et que l'information auxiliaire est donnée par $\mathbf{g} = (g)$ avec $g(X, Y) = Y$ c'est-à-dire que l'on a connaissance de $\mathbb{E}[g(X, Y)] = \mathbb{E}[Y]$ alors la mesure empirique avec information auxiliaire est donnée par

$$\mathbb{P}_n^{\mathbf{g}} = \mathbb{P}_n(X) - \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} (\mathbb{P}_n(Y) - \mathbb{E}[Y])$$

et sa variance asymptotique $\text{Var}(\widehat{\mathbb{G}}(f))$ est la même que $\text{Var}(X^*)$.

Fonction de répartition. Zhang [83] a étudié le comportement asymptotique de la fonction de répartition en présence d'informations auxiliaires donnée par des espérances de fonctions supposées nulles. Il a en particulier montré (voir théorème 3) que si X est une variable aléatoire réelle dont le support est contenu dans un intervalle $[a, b]$ et qui possède une fonction de répartition F continue alors le processus $\alpha_n^{\mathbf{g}}([a, b])$, où $\alpha_n^{\mathbf{g}}(t) = \alpha_n^{\mathbf{g}}(\mathbb{1}_{[a,t]})$ converge en loi sur $D([a, b])$ vers le processus gaussien centré $W([a, b])$ dont la fonction de covariance est donnée par

$$\text{Cov}(W(s), W(t)) = F(\min(s, t)) - F(s)F(t) - \mathbb{E}[\mathbf{g}[X]\mathbb{1}_{X \leq s}] \cdot \text{Var}(\mathbf{g})^{-1} \cdot \mathbb{E}[\mathbf{g}[X]\mathbb{1}_{X \leq t}]^t,$$

où $\mathbb{E}[\mathbf{g}[X]\mathbb{1}_{X \leq s}] = (\mathbb{E}[g_1(X)\mathbb{1}_{X \leq s}], \dots, \mathbb{E}[g_m(X)\mathbb{1}_{X \leq s}])$.

Vraisemblance empirique. La vraisemblance empirique est une méthode proche au niveau du concept de la méthode du maximum de vraisemblance. Alors que cette dernière est une méthode d'estimation nécessitant de travailler avec une famille paramétrique de lois, la méthode de vraisemblance empirique s'affranchit de cette contrainte. Cette méthode a été introduite par Owen [63] (voir également [64]).

Estimateur par le ratio. Le résumé de la méthode du redressement par ratio qui est faite dans ce paragraphe provient de l'article [24] de Bernard Bru présentant les estimations laplaciennes. L'estimateur par le ratio ou par le quotient est une méthode de redressement utilisé pour une variable d'intérêt qui suivrait une relation à peu près linéaire par rapport à une autre variable dont on connaîtrait l'espérance. On peut trouver une présentation détaillée de cette méthode dans le chapitre III.3 de [7]. C'est une méthode a été originellement introduite en France vers 1740 par le magistrat Jean-Baptiste François de La Michodière, a été validée par Laplace dans son mémoire intitulé « Sur les naissances, les mariages et les morts... » [56] et s'est répandue dans l'administration française. Elle avait pour but à l'époque de donner une approximation de la taille de la population française de l'époque en partant du principe selon Derham que « le nombre des mariages, des naissances et des morts est proportionné au nombre des personnes qui composent une nation entière, ou qui se trouvent dans toute une province, ou dans une paroisse » [31]. Ainsi les intendants obtenaient une approximation de la population d'une ville ou d'une province en multipliant le nombre de naissances exacte obtenu à partir des registres par un facteur longuement débattu au cours du XVIIIe siècle et qui variait selon la personne et le lieu d'intérêt (ville/campagne, localité, ...). Formellement on s'intéresse à la valeur $P(f) = \mathbb{E}[f(X)]$ et on suppose qu'il existe une relation presque linéaire de la forme $f(X) = Rg(X) + \epsilon$ où ϵ est une variable aléatoire centrée indépendante de X en ayant à disposition la valeur théorique $P(g) = \mathbb{E}[g(X)]$. La statistique proposée permettant ce redressement est donnée par

$$\widehat{\mathbb{P}}_n(f) = \widehat{R}_n P(g) = \frac{P(g)\mathbb{P}_n(f)}{\mathbb{P}_n(g)}. \quad (4.13)$$

On a estimé le coefficient de proportionnalité $R = P(f)/P(g)$ par le rapport $\hat{R}_n = \mathbb{P}_n(f)/\mathbb{P}_n(g)$. L'estimateur par le ratio est un cas particulier du processus empirique avec information auxiliaire que nous avons défini dans cette partie. En effet, l'information auxiliaire est donnée ici par la connaissance de $P(g)$ et la mesure empirique utilisant l'information est donnée par

$$\mathbb{P}_n^{\mathbf{g}}(f) = \mathbb{P}_n(f) - \frac{\text{Cov}(f, g)}{\text{Var}(g)}(\mathbb{P}_n(g) - P(g)).$$

En utilisant le fait que dans notre cas $\text{Cov}(f, g) = R\text{Var}(g)$ on obtient que

$$\mathbb{P}_n^{\mathbf{g}}(f) = \mathbb{P}_n(f) - R(\mathbb{P}_n(g) - P(g)).$$

L'approximation du ratio R par \hat{R}_n nous permet de retomber sur l'estimateur par le ratio donné par la formule (4.13). En utilisant (4.7) on obtient que la variance du processus empirique avec information auxiliaire tout comme la variance du processus empirique résultant de l'estimateur par le ratio est donnée par

$$\text{Var}(\mathbb{G}^{\mathbf{g}}(f)) = \text{Var}(f) - \frac{\text{Cov}(f, g)^2}{\text{Var}(g)} = \text{Var}(f) - R^2\text{Var}(g).$$

Jusqu'à présent nous avons supposé qu'il existait une relation de quasi-linéarité entre les variables $f(X)$ et $g(X)$. Si cette relation n'existait pas, rien ne garantit que l'estimateur par le ratio diminuerait la variance, pire on peut exhiber des cas où on empire les estimations en augmentant la variance de l'estimateur par le ratio. Avec les outils d'approximation forte on peut montrer que le processus $\sqrt{n}(P(g)\mathbb{P}_n(f)/\mathbb{P}_n(g) - P(f))$ converge en loi dans $\ell^\infty(\mathcal{F})$ vers le processus $\mathbb{G}(f) - P(f)\mathbb{G}(g)/P(g)$ de variance

$$\text{Var}(f) - 2\frac{P(f)}{P(g)}\text{Cov}(f, g) + \left(\frac{P(f)}{P(g)}\right)^2 \text{Var}(g).$$

Dans le cas trivial où la variable $g(X)$ est indépendante de $f(X)$, la variance serait de

$$\text{Var}(f) + \left(\frac{P(f)}{P(g)}\right)^2 \text{Var}(g) \geq \text{Var}(f)$$

ce qui augmente la variance, comparé au processus empirique usuel. Finalement, il serait préférable de privilégier la mesure empirique avec l'information auxiliaire $P(g)$ qui améliore les estimations que l'on soit dans un cas de quasi-linéarité ou non. Quand la covariance et la variance ne sont pas connues on peut utiliser

$$\mathbb{P}_n^{\mathbf{g}^*}(f) = \mathbb{P}_n(f) - \frac{\mathbb{P}_n(fg) - \mathbb{P}_n(f)P(g)}{\mathbb{P}_n(g^2) - P(g)^2}(\mathbb{P}_n(g) - P(g)).$$

La variance asymptotique est donnée par

$$\text{Var}(\mathbb{G}^{\mathbf{g}^*}(f)) = \text{Var}(f) - \frac{\text{Cov}(f, g)^2}{\text{Var}(g)} \leq \text{Var}(f).$$

Espérance du produit. Supposons que l'on travaille avec la loi $P = \mathbb{P}^{(X, Y)}$, que notre fonction d'intérêt est $f(X, Y) = X$ et que l'information auxiliaire est donnée par $\mathbf{g} = (g)$

avec $g(X, Y) = XY$, c'est-à-dire que l'on a connaissance de $\mathbb{E}[g(X, Y)] = \mathbb{E}[XY]$. La mesure empirique avec information auxiliaire est donnée par

$$\mathbb{P}_n^{\mathbf{g}}(f) = \mathbb{P}_n(f) - \frac{\text{Cov}(XY, X)}{\text{Var}(XY)} (\mathbb{P}_n(XY) - \mathbb{E}[XY]).$$

Si les valeurs $\text{Cov}(XY, X)$ et $\text{Var}(XY)$ ne sont pas connues il convient d'utiliser la mesure empirique

$$\mathbb{P}_n^{\mathbf{g}^*}(f) = \mathbb{P}_n(f) - \frac{\mathbb{P}_n(X^2Y) - \mathbb{E}[XY]\mathbb{P}_n(X)}{\text{Var}_n(XY)} (\mathbb{P}_n(XY) - \mathbb{E}[XY]),$$

où Var_n désigne la variance empirique. La variance du processus gaussien au point $f \in \mathcal{F}$ sans information auxiliaire est $\text{Var}(\mathbb{G}(f)) = \text{Var}(f(X, Y)) = \text{Var}(X)$ tandis que la variance du processus gaussien limite avec information auxiliaire est donnée par

$$\text{Var}(\mathbb{G}^{\mathbf{g}}(f)) = \text{Var}(X) - \frac{\text{Cov}(XY, X)^2}{\text{Var}(XY)}.$$

Le cas trivial où $Y = 1$ conduirait à connaître $\mathbb{E}[X]$ et on retombe $\text{Var}(\widehat{\mathbb{G}}(f)) = 0$. Le cas où $Y = X$ revient à connaître le moment d'ordre 2 de la variable X et dans ce cas la variance du processus gaussien limite avec information auxiliaire est donnée par

$$\text{Var}(\mathbb{G}^{\mathbf{g}}(f)) = \sigma^2 - \frac{(\mu_3 - \mu(\mu^2 + \sigma^2))^2}{\mu_4 - (\mu^2 + \sigma^2)^2},$$

où μ, σ^2, μ_k sont respectivement l'espérance, la variance et le moment d'ordre k de la variable X .

Espérance tronquée. Supposons que le statisticien ait à sa disposition la connaissance de l'espérance de sa variable tronquée, c'est-à-dire qu'il ait connaissance de la valeur de l'espérance $\mathbb{E}[g(X)]$ avec $g(X) = X\mathbb{1}_{a \leq X \leq b} + a\mathbb{1}_{X < a} + b\mathbb{1}_{X > b}$. On peut imaginer par exemple qu'une étude préliminaire a été menée sur la variable d'intérêt avec des outils dont soit une capacité réduite, soit un mauvais paramétrage de la part de l'utilisateur ou bien encore un dysfonctionnement a tronqué les valeurs observées. Si une seconde étude permet l'observation non tronquée de ces valeurs, il serait alors dommage de ne pas utiliser les résultats établis lors de l'enquête préliminaire. On pourrait intégrer cette information auxiliaire pour améliorer l'estimation de l'espérance des X en utilisant la mesure empirique avec information auxiliaire

$$\mathbb{P}_n^{\mathbf{g}}(f) = \mathbb{P}_n(X) - \frac{\text{Cov}_n(X, g(X))}{\text{Var}_n(g(X))} (\mathbb{P}_n(g) - \mathbb{E}[g(X)]).$$

Par exemple si $X \sim \mathcal{N}(0, 1)$ et $b = -a = 1$ alors $\text{Cov}(X, g(X)) \simeq 0.683$, $\text{Var}(g(X)) \simeq 0.516$ et donc la variance baisse de $\text{Cov}(X, g(X))^2 / \text{Var}(g(X)) \simeq 0.903$. La figure 4.1 est une simulation de la loi du processus empirique avec et sans l'information auxiliaire $\mathbb{E}[X\mathbb{1}_{-1 \leq X \leq 1} + \mathbb{1}_{X > 1} - \mathbb{1}_{X < -1}] = 0$.

4.3.4 Généralisation des chapitres précédents

Information auxiliaire d'une partition. On travaille avec $P = \mathbb{P}^X$ et notre fonction d'intérêt est $f(X) = X$. Supposons que l'information auxiliaire est donnée par la connaissance d'appartenir à un ensemble d'une partition de \mathcal{X} , autrement dit supposons que l'on

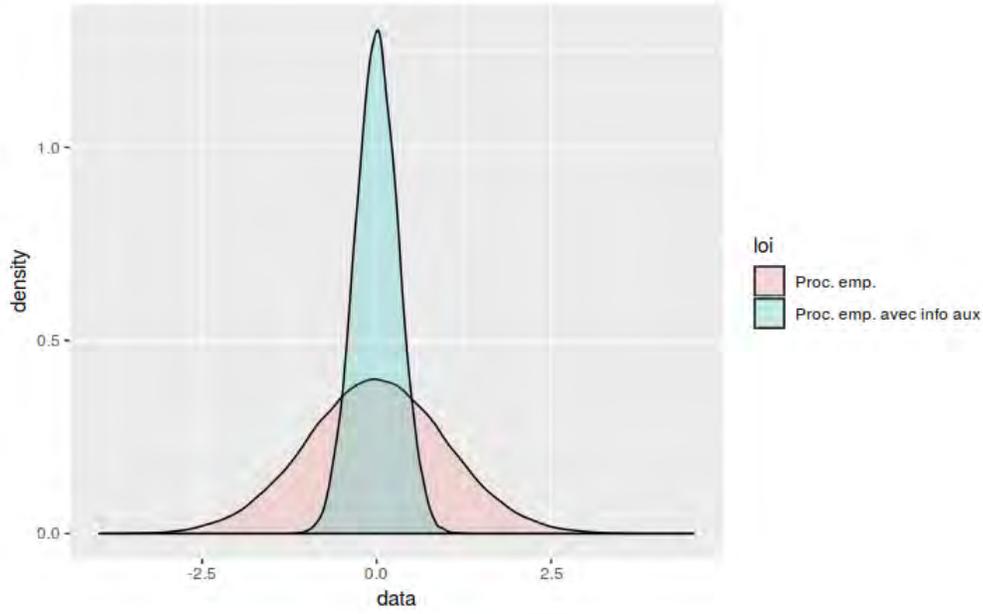


FIG. 4.1: Simulation numérique du processus avec/sans informations auxiliaires

ait à disposition $\mathbb{E}[g_i] = P(A_i)$ avec $g_i = \mathbb{1}_{A_i}$, $i = 1, \dots, m$ où $\mathcal{A} = \{A_1, \dots, A_m\}$ forme une partition de \mathcal{X} . La valeur optimale Λ_0 à estimer est $\Lambda_0 = (\mathbb{E}[f|A_1], \dots, \mathbb{E}[f|A_m])$. Si l'utilisateur ne connaît aucune valeur de ces espérances conditionnelles il peut les estimer grâce aux valeurs $-\mathbb{P}_n(f\mathbb{1}_{A_i})/\mathbb{P}_n(A_i)$ ou bien $-\mathbb{P}_n(f\mathbb{1}_{A_i})/P(A_i)$. Ces estimateurs permettent de retrouver le processus empirique avec information auxiliaire d'une partition. En effet d'après (2.1), on obtient dans les deux cas que $\mathbb{P}_n^{\mathbf{g}^*}(f) = \hat{\mathbb{P}}_n^A(f)$ où $\hat{\mathbb{P}}_n^A(f)$ est le processus défini et étudié dans la partie avec l'information auxiliaire d'une partition. Connaître toutes les valeurs b_i reviendrait à connaître $\mathbb{E}[f]$, et par conséquent que l'on aurait $\alpha_n^{\mathbf{g}}(f) = 0$ pour tout $f \in \mathcal{F}$.

Raking-Ratio. Supposons que l'information auxiliaire est donnée par la connaissance d'appartenir pour tout $N \in \mathbb{N}^*$ à un ensemble $(A_i^{(N)})_{i \leq m_N}$ d'une partition $A^{(N)} = \{A_1^{(N)}, \dots, A_{m_N}^{(N)}\}$ de \mathcal{X} . Le Raking-Ratio a pour objectif d'intégrer itérativement par rapport à N l'information auxiliaire donnée par $\mathcal{A}^{(N)}$ en utilisant la suite de mesure empirique modifiée $\mathbb{P}_n^{(N)}$ définie par $\mathbb{P}_n^{(0)} = \mathbb{P}_n$ et

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

Pour tout $f \in \mathcal{F}$ on note $\alpha_n^{(N)}(f) = \sqrt{n}(\mathbb{P}_n^{(N)}(f) - P(f))$ le processus empirique associé à la méthode du Raking-Ratio. Le résultat suivant montre que notre généralisation est plus optimale que le Raking-Ratio par rapport à l'intégration de cette information.

Proposition 4.3.5. *Pour tout $N \in \mathbb{N}^*$, il existe $(b_{i,j}^{(N)})_{i \leq N, j \leq m_N}$ tel que pour tout $f \in \mathcal{F}$,*

$$\mathbb{P}_n^{(N)}(f) = \mathbb{P}_n(f) + \sum_{i=1}^N \sum_{j=1}^{m_i} b_{i,j}^{(N)} (\mathbb{P}_n(A_j^{(i)}) - P(A_j^{(i)})).$$

Démonstration. La mesure empirique avec l'information auxiliaire d'une partition $\mathbb{P}_n^{(1)}(\mathcal{F}) = \widehat{\mathbb{P}}_n^{A^{(1)}}(\mathcal{F})$ étudié au deuxième chapitre vérifie cette propriété. Supposons que l'hypothèse soit vérifiée au rang N . Montrons qu'elle l'est également au rang $N + 1$.

$$\begin{aligned}
\mathbb{P}_n^{(N+1)}(f) &= \sum_{k=1}^{m_{N+1}} \frac{P(A_k^{(N+1)})}{\mathbb{P}_n^{(N)}(A_k^{(N+1)})} \mathbb{P}_n^{(N)}(f \mathbf{1}_{A_k^{(N+1)}}) \\
&= \sum_{k=1}^{m_{N+1}} \frac{P(A_k^{(N+1)})}{\mathbb{P}_n^{(N)}(A_k^{(N+1)})} \left[\mathbb{P}_n(f \mathbf{1}_{A_k^{(N+1)}}) + \sum_{i=1}^N \sum_{j=1}^{m_i} b_{i,j}^{(N)} (\mathbb{P}_n(A_j^{(i)}) - P(A_j^{(i)})) \right] \\
&= \mathbb{P}_n(f) + \sum_{i=1}^N \sum_{j=1}^{m_i} b_{i,j}^{(N)} \left(\sum_{k=1}^{m_{N+1}} \frac{P(A_k^{(N+1)})}{\mathbb{P}_n^{(N)}(A_k^{(N+1)})} \right) (\mathbb{P}_n(A_j^{(i)}) - P(A_j^{(i)})) \\
&\quad - \sum_{k=1}^{m_{N+1}} \frac{\mathbb{P}_n(f \mathbf{1}_{A_k^{(N+1)}})}{\mathbb{P}_n^{(N)}(A_k^{(N+1)})} (\mathbb{P}_n^{(N)}(A_k^{(N+1)}) - P(A_k^{(N+1)}))
\end{aligned}$$

Or,

$$\begin{aligned}
&\sum_{k=1}^{m_{N+1}} \frac{\mathbb{P}_n(f \mathbf{1}_{A_k^{(N+1)}})}{\mathbb{P}_n^{(N)}(A_k^{(N+1)})} (\mathbb{P}_n^{(N)}(A_k^{(N+1)}) - P(A_k^{(N+1)})) \\
&= \sum_{k=1}^{m_{N+1}} \frac{\mathbb{P}_n(f \mathbf{1}_{A_k^{(N+1)}})}{\mathbb{P}_n^{(N)}(A_k^{(N+1)})} \left[(\mathbb{P}_n(A_k^{(N+1)}) - P(A_k^{(N+1)})) + \sum_{i=1}^N \sum_{j=1}^{m_i} b_{i,j}^{(N)} (\mathbb{P}_n(A_j^{(i)}) - P(A_j^{(i)})) \right],
\end{aligned}$$

d'où $\mathbb{P}_n^{(N+1)} = \mathbb{P}_n(f) + \sum_{i=1}^{N+1} \sum_{j=1}^{m_i} b_{i,j}^{(N+1)} (\mathbb{P}_n(A_j^{(i)}) - P(A_j^{(i)}))$ avec

$$\begin{aligned}
\forall 1 \leq i \leq N, \quad \forall 1 \leq j \leq m_i, \quad b_{i,j}^{(N+1)} &= b_{i,j}^{(N)} \sum_{k=1}^{m_{N+1}} \frac{P(A_k^{(N+1)}) - \mathbb{P}_n(f \mathbf{1}_{A_k^{(N+1)}})}{\mathbb{P}_n^{(N)}(A_k^{(N+1)})} \\
\forall 1 \leq k \leq m_{N+1}, \quad b_{N+1,k}^{(N+1)} &= - \frac{\mathbb{P}_n(f \mathbf{1}_{A_k^{(N+1)}})}{\mathbb{P}_n^{(N)}(A_k^{(N+1)})}
\end{aligned}$$

□

Compte tenu du résultat précédent et de la proposition 4.3.1 on a en particulier pour tout $f \in \mathcal{F}$,

$$\text{Var}(\alpha_n^{(N)}(f)) \geq \text{Var}(\alpha_n^{\mathbf{g}_N}(f)) \geq \text{Var}(\alpha_n^{\mathbf{g}}(f))$$

où $\mathbf{g}_N = \{\mathbf{1}_{A_j^{(i)}} : i \leq N, j \leq m_i\}$, $\mathbf{g} = \{\mathbf{1}_{A_j^{(N)}} : N \in \mathbb{N}^*, j \leq m_N\}$.

4.4 Information auxiliaire générale

4.4.1 Introduction

Motivation. Nous avons traité le cas d'une information auxiliaire de la forme $P(g_k)$, c'est-à-dire une information qui est l'espérance d'une fonction mesurable. Mais comme évoqué dans l'introduction certains cas ne rentrent malheureusement pas dans cette définition. Par exemple l'information auxiliaire donnée par la connaissance de la variance d'une variable ou de la covariance de deux variables ne rentre pas dans le cadre précédent puisque la variance et covariance empirique ne peuvent s'exprimer comme la moyenne empirique d'une

seule fonction, c'est-à-dire qu'on ne peut pas les écrire sous la forme $\mathbb{P}_n(g)$. C'est également le cas simple de la moyenne empirique conditionnelle, de l'inverse d'une moyenne, etc... De manière générale, il faut considérer que l'information auxiliaire peut s'exprimer comme une fonction compliquée à valeur réelle de la loi de l'échantillon qui est naturellement approchée asymptotiquement par l'image de la mesure empirique par cette même fonction. Cette fonction doit vérifier au minimum des conditions de dérivabilité afin que l'on puisse théoriquement assurer la convergence de ce processus empirique avec informations auxiliaires générales.

Organisation. Nous introduisons dans cette sous-section le processus empirique avec information auxiliaire générale. Dans la sous-section 4.4.2 nous établissons des résultats d'approximation forte et de convergence faible sur ce processus. Nous donnons des exemples d'application aux sous-sections 4.4.3, 4.4.4 et 4.4.5.

Notation et hypothèses. On fixe $t_0 > 0$. Nous allons travailler sur une classe de fonctions \mathcal{H} de fonctions définies sur $\ell^\infty(\mathcal{F})$ à valeurs réelles vérifiant les hypothèses suivantes.

Hypothèse ($\mathcal{H}.i$). Pour tout $h \in \mathcal{H}$, $h(P)$ est défini et pour tout $Q \in \ell^\infty(\mathcal{F})$ vérifiant $\|Q\|_{\mathcal{F}} < t_0$,

$$h(P + Q) = h(P) + Q \circ \varphi_h(P) + R_h(Q),$$

où $\varphi_h : \ell^\infty(\mathcal{F}) \rightarrow \text{Vec}(\mathcal{F})$ et $R_h : \ell^\infty(\mathcal{F}) \rightarrow \mathbb{R}$ une application vérifiant sur une boule centrée en l'origine \mathcal{B} , $|R_h(\cdot)| \leq \|\cdot\|_{\mathcal{F}}^q$ pour un certain $q > 1$.

Hypothèse ($\mathcal{H}.ii$). Pour tout $h \in \mathcal{H}$, $\text{Var}(\varphi_h(P))$ existent.

Cette classe contient nos fonctions statistiques d'intérêt et les informations auxiliaires. L'hypothèse $\mathcal{H}.i$ signifie que les fonctions sont différentiables en la mesure P tandis que l'hypothèse $\mathcal{H}.ii$ s'assurera de la bonne définition de l'objet limite. L'information auxiliaire sera représentée par la connaissance de $g_1(P), \dots, g_m(P)$ où $g_1, \dots, g_m \in \mathcal{H}$. Cette façon de représenter l'information auxiliaire permet d'inclure les cas que nous avons cité et que ne pouvons pas traiter jusqu'à présent. Plus de détails sont donnés dans les exemples qui suivent. On définit le processus empirique avec information auxiliaire générale $\mathbf{a}_n^{\mathbf{g}}(\mathcal{H})$ pour tout $n \in \mathbb{N}^*$ et $h \in \mathcal{H}$ par

$$\begin{aligned} \mathbf{P}_n^{\mathbf{g}, \Lambda}(h) &= h(\mathbb{P}_n) + \Lambda \cdot (\mathbf{g}[\mathbb{P}_n] - \mathbf{g}[P]), \\ \mathbf{a}_n^{\mathbf{g}, \Lambda}(h) &= \sqrt{n}(\mathbf{P}_n^{\mathbf{g}, \Lambda}(h) - h(P)) \\ &= \sqrt{n}(h(\mathbb{P}_n) - h(P) + \Lambda \cdot (\mathbf{g}[\mathbb{P}_n] - \mathbf{g}[P])) \end{aligned}$$

où $\mathbf{g}[Q] = (g_1(Q), \dots, g_m(Q))^t$.

Exemples de classe \mathcal{H} . Le tableau suivant donne des exemples de classes de fonctions \mathcal{H} d'intérêt avec leur différentielle et reste respectives.

$h(P)$	\mathcal{F}	$\varphi_h(P)$
Espérance $P(f_0)$	$\{f_0\}$	f_0
Variance $P(f_0^2) - P^2(f_0)$	$\{f_0^2, f_0\}$	$f_0^2 - 2f_0P(f_0)$
Inverse de l'espérance $1/P(f_0)$	$\{f_0\}$	$-f_0/P^2(f_0)$
Espérance conditionnelle $P(f_0 \mathbf{1}_A)/P(A)$	$\{f_0 \mathbf{1}_A, \mathbf{1}_A\}$	$(f_0 - P(f_0 A))\mathbf{1}_A/P(A)$
Produit des espérances $P(f_0)P(g_0)$	$\{f_0, g_0, f_0g_0\}$	$f_0g_0 - f_0P(g_0) - P(f_0)g_0$

TAB. 4.1: **Tableau des différentielles usuelles**

Une définition générale de l'information auxiliaire requiert donc d'étudier en détail la statistique afin d'exprimer pour tout $h \in \mathcal{H}$ la fonction φ_h .

Estimateur optimal. Le processus $\mathbf{a}_n^{\mathbf{g},\Lambda}(\mathcal{H})$ est comme $\alpha_n^{\mathbf{g},\Lambda}(\mathcal{F})$ un cas particulier de l'étude de Tarima et Pavlov. Il suffit de prendre en effet $\Theta = h(P), \hat{\Theta} = h(\mathbb{P}_n), \mathcal{B} = \tilde{\mathcal{B}} = \mathbf{g}[P], \hat{\mathcal{B}} = \mathbf{g}[\mathbb{P}_n]$. Notons Σ_1, Σ_2 les matrices de covariance suivantes définies par

$$\begin{aligned}\Sigma_1(h) &= \text{Cov}(\varphi_h(P), \varphi_{\mathbf{g}}[P]) \in \mathcal{M}_{1,m}(\mathbb{R}), \\ \Sigma_2 &= \text{Var}(\varphi_{\mathbf{g}}[P]) \in \mathcal{M}_{m,m}(\mathbb{R}),\end{aligned}$$

c'est-à-dire les matrices définies pour tout $1 \leq i, j \leq m$ par

$$\begin{aligned}(\Sigma_1(h))_i &= \text{Cov}(\varphi_h(P), \varphi_{g_i}(P)), \\ (\Sigma_2)_{i,j} &= \text{Cov}(\varphi_{g_i}(P), \varphi_{g_j}(P)).\end{aligned}$$

Sous réserve que Σ_2 soit inversible, la valeur Λ_0 minimisant la variance du processus $\mathbf{a}_n^{\mathbf{g},\Lambda}(h)$ est donnée grâce à (4.1) par

$$\Lambda_0 = -\Sigma_1(h) \cdot \Sigma_2^{-1}.$$

Si la matrice Σ_2 n'est pas inversible, il suffit de considérer la valeur

$$\Lambda_0 = -\Sigma_1(h) \cdot \Sigma_2^+.$$

La mesure $\mathbf{P}_n^{\mathbf{g}}(\mathcal{H})$ et le processus empirique $\mathbf{a}_n^{\mathbf{g}}(\mathcal{H})$ avec information auxiliaire générale sont respectivement définis pour tout $f \in \mathcal{F}$ par

$$\begin{aligned}\mathbf{P}_n^{\mathbf{g}}(h) &= h(\mathbb{P}_n) - \Lambda_0 \cdot (\mathbf{g}[\mathbb{P}_n] - \mathbf{g}[P]) \\ &= h(\mathbb{P}_n) - \Sigma_1(h) \cdot \Sigma_2^{-1} \cdot (\mathbf{g}[\mathbb{P}_n] - \mathbf{g}[P]),\end{aligned}\tag{4.14}$$

$$\begin{aligned}\mathbf{a}_n^{\mathbf{g}}(h) &= \sqrt{n} (h(\mathbb{P}_n) - h(P) - \Lambda_0 \cdot (\mathbf{g}[\mathbb{P}_n] - \mathbf{g}[P])) \\ &= \sqrt{n} (h(\mathbb{P}_n) - h(P) - \Sigma_1(h) \cdot \Sigma_2^{-1} \cdot (\mathbf{g}[\mathbb{P}_n] - \mathbf{g}[P])).\end{aligned}\tag{4.15}$$

Comme dans le cas précédent, nous étudions ci-après le cas où la valeur Λ_0 n'est pas connue.

Estimateur adaptatif. Si la valeur Λ_0 n'est pas connue du statisticien, celui-ci peut utiliser un estimateur consistant $\hat{\Lambda}_0$ de Λ_0 . Notons les matrices $\Sigma_{1,n} \in \mathcal{M}_{1,m}(\mathbb{R}), \Sigma_{2,n} \in \mathcal{M}_{m,m}(\mathbb{R})$ définies pour tout $1 \leq i, j \leq m$ par

$$\begin{aligned}(\Sigma_{1,n}(h))_i &= \text{Cov}_n(\varphi_h(\mathbb{P}_n), \varphi_{g_i}(\mathbb{P}_n)), \\ (\Sigma_{2,n})_{i,j} &= \text{Cov}_n(\varphi_{g_i}(\mathbb{P}_n), \varphi_{g_j}(\mathbb{P}_n)),\end{aligned}$$

Si $\Sigma_{2,n}$ est inversible alors le statisticien peut utiliser l'estimateur consistant naturel

$$\hat{\Lambda}_0 = -\Sigma_{1,n}(h) \cdot \Sigma_{2,n}^{-1}.$$

Si la matrice $\Sigma_{2,n}$ n'est pas inversible il peut par exemple utiliser l'estimateur consistant

$$\hat{\Lambda}_0 = -\Sigma_{1,n}(h) \cdot \Sigma_{2,n}^+.$$

Dans tous les cas on emploie l'estimateur adaptatif suivant

$$\begin{aligned}\mathbf{P}_n^{\mathbf{g}^*}(h) &= h(\mathbb{P}_n) - \hat{\Lambda}_0 \cdot (\mathbf{g}[\mathbb{P}_n] - \mathbf{g}[P]), \\ \mathbf{a}_n^{\mathbf{g}^*}(h) &= \sqrt{n} \left(h(\mathbb{P}_n) - h(P) - \hat{\Lambda}_0 \cdot (\mathbf{g}[\mathbb{P}_n] - \mathbf{g}[P]) \right).\end{aligned}$$

L'exemple qui suit reprend à nouveau l'information auxiliaire apportée par la connaissance de la probabilité d'appartenir à un ensemble d'une partition donnée mais avec les nouvelles notations.

Exemple 4.4.1. On pose $g_1(P) = P(A_1), \dots, g_m = P(A_m)$ avec $(A_i)_{1 \leq i \leq m}$ une partition de \mathcal{X} . On étudie la statistique $h(P) = P(f)$. Les fonctions g_i, h vérifient l'hypothèse $\mathcal{H}.i$ avec $\varphi_h(P) = f, \varphi_{g_i}(P) = \mathbb{1}_{A_i}$. On suppose que le statisticien a à sa disposition les valeurs $\mathbb{E}[g_i] = P(A_i)$. Comme pour l'exemple 4.3.1, les matrices $\Sigma_1(h), \Sigma_2$ valent respectivement

$$\begin{aligned} (\Sigma_1(h))_i &= \text{Cov}(f(X), \mathbb{1}_{A_i}(X)) = P(A_i)(\mathbb{E}[f|A_i] - \mathbb{E}[f]), \\ \Sigma_2 &= \text{diag}(P[\mathcal{A}]) - P[\mathcal{A}] \cdot P[\mathcal{A}]^t, \end{aligned}$$

où $1 \leq i \leq m$ et $P[\mathcal{A}] = (P(A_1), \dots, P(A_m))$.

4.4.2 Résultats

Approximation forte. Pour tout $h \in \mathcal{H}$, on définit le processus gaussien avec information auxiliaire générale par

$$\mathbf{G}^{\mathbf{g}}(h) = \mathbb{G}(\varphi_h(P)) + \Lambda_0 \cdot \mathbb{G}[\varphi_{\mathbf{g}}[P]], \quad (4.16)$$

où $\varphi_{\mathbf{g}}[P] = (\varphi_{g_1}(P), \dots, \varphi_{g_m}(P))^t$. Nous mentionnons de dernier comme le P -pont brownien avec information auxiliaire générale. Si la classe de fonctions \mathcal{F} vérifie les conditions d'entropie de Berthet-Mason il est encore possible d'établir une approximation forte du processus $\mathbf{a}_n^{\mathbf{g}}(\mathcal{H})$ avec le processus gaussien $\mathbb{G}^{\mathbf{g}}(\mathcal{H})$. Le théorème suivant énonce l'approximation forte.

Théorème 4.4.1. Supposons que \mathcal{F} vérifie $\mathcal{F}.i, \mathcal{F}.ii, VC$ ou BR et que \mathcal{H} vérifie $\mathcal{H}.i, \mathcal{H}.ii$. Pour tout $\theta > 0$ il existe $C_\theta > 0$, une suite v_n de la forme $v_n = n^{-\alpha}(\log n)^\beta$ si \mathcal{F} vérifie VC et $v_n = (\log n)^{-\beta}$ si \mathcal{F} vérifie BR (avec $\alpha, \beta > 0$), des variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi P et des processus $(\mathbb{G}_n)_n$ de P -pont brownien définis sur un même espace de probabilité vérifiant pour tout n ,

$$\mathbb{P}(\|\mathbf{a}_n^{\mathbf{g}} - \mathbf{G}_n^{\mathbf{g}}\|_{\mathcal{H}} > C_\theta v_n) \leq \frac{1}{n^\theta}, \quad (4.17)$$

et presque sûrement

$$\|\mathbf{a}_n^{\mathbf{g}} - \mathbf{G}_n^{\mathbf{g}}\|_{\mathcal{H}} = O(v_n), \quad (4.18)$$

où $\mathbf{G}_n^{\mathbf{g}}$ est le P -pont brownien avec information auxiliaire générale défini par (4.16) via les \mathbb{G}_n . De plus si l'hypothèse CV est vérifiée alors pour tout n suffisamment grand,

$$\mathbb{P}(\|\mathbf{a}_n^{\mathbf{g}^*} - \mathbf{G}_n^{\mathbf{g}^*}\|_{\mathcal{H}} > C_\theta v_n) \leq \frac{1}{n^\theta}, \quad (4.19)$$

et presque sûrement

$$\|\mathbf{a}_n^{\mathbf{g}^*} - \mathbf{G}_n^{\mathbf{g}^*}\|_{\mathcal{H}} = O(v_n). \quad (4.20)$$

Démonstration. D'après l'approximation forte de Berthet-Mason, il existe $C_0 > 0$, des variables $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. de loi P , des P -ponts browniens $(\mathbb{G}_n)_{n \in \mathbb{N}^*}$ sur un même espace de probabilité tels que

$$\mathbb{P}(\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} > C_0 v_n) \leq \frac{1}{3n^\theta}.$$

On note $\Lambda_0 = (\lambda_{0,1}, \dots, \lambda_{0,m})$. En utilisant les définitions de $\mathbf{a}_n^{\mathbf{g}}(\mathcal{H})$ et $\mathbf{G}^{\mathbf{g}}(\mathcal{H})$ données par (4.15) et (4.16) on peut écrire que

$$\|\mathbf{a}_n^{\mathbf{g}} - \mathbf{G}_n^{\mathbf{g}}\|_{\mathcal{H}} \leq \left(1 + \sum_{i=1}^m |\lambda_{0,i}|\right) (\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} + \|\alpha_n/\sqrt{n}\|_{\mathcal{F}}^q),$$

où les processus $\mathbf{G}_n^{\mathbf{g}}(\mathcal{H})$ sont définis avec de (4.16) à partir des $\mathbb{G}_n(\mathcal{F})$. Il existe C_1 tel que

$$\mathbb{P}\left(\|\alpha_n\|_{\mathcal{F}}^q > C_1 v_n n^{q/2}\right) \leq \frac{1}{3n^\theta}.$$

L'inégalité (4.17) est vérifiée en prenant $C_\theta > 0$ suffisamment grand pour avoir $C_\theta v_n > (1 + \sum_{i=1}^m |\lambda_{0,i}|)(C_0 + C_1)v_n$. On note que $\|\mathbf{a}_n^{\mathbf{g}^*} - \mathbf{G}_n^{\mathbf{g}}\|_{\mathcal{H}} \leq \|\mathbf{a}_n^{\mathbf{g}^*} - \mathbf{a}_n^{\mathbf{g}}\|_{\mathcal{H}} + \|\mathbf{a}_n^{\mathbf{g}} - \mathbf{G}_n^{\mathbf{g}}\|_{\mathcal{H}}$ et que

$$\begin{aligned} \|\mathbf{a}_n^{\mathbf{g}^*} - \mathbf{a}_n^{\mathbf{g}}\|_{\mathcal{H}} &\leq \|(\widehat{\Lambda}_0 - \Lambda_0) \cdot \sqrt{n}(\mathbf{g}[\mathbb{P}_n] - \mathbf{g}[P])\|_{\mathcal{H}} \\ &\leq \|(\widehat{\Lambda}_0 - \Lambda_0)\|_{\mathcal{F}}(2M_{\mathcal{F}} + \|\alpha_n\|_{\mathcal{F}}^q/n^{(q-1)/2}). \end{aligned}$$

D'après l'hypothèse CV, il existe $D > 0$ tel que

$$\mathbb{P}\left(\|(\widehat{\Lambda}_0 - \Lambda_0)\|_{\mathcal{F}}(2M_{\mathcal{F}} + \|\alpha_n\|_{\mathcal{F}}^q/n^{(q-1)/2}) > \sqrt{\frac{D \log n}{n}}(2M_{\mathcal{F}} + C_1 v_n n^{(q-1)/2})\right) \leq \frac{1}{3n^\theta},$$

ce qui entraîne (4.19) avec C_θ suffisamment grand pour avoir

$$C_\theta v_n > \left(1 + \sum_{i=1}^m |\lambda_{0,i}|\right) (C_0 + C_1)v_n + \sqrt{\frac{D \log n}{n}}(2M_{\mathcal{F}} + C_1 v_n n^{(q-1)/2}).$$

Les inégalités (4.18) et (4.20) découlent du lemme de Borel-Cantelli et des inégalités (4.17) et (4.19) en prenant $\theta > 1$. \square

Convergence en loi. Comme pour le processus $\alpha_n^{\mathbf{g}}(\mathcal{F})$ dans le cas où l'information auxiliaire est donnée par des espérances de fonctions, le processus $\mathbf{a}_n^{\mathbf{g}}(\mathcal{H})$ converge en loi dans $\ell^\infty(\mathcal{H})$ vers $\mathbf{G}^{\mathbf{g}}(\mathcal{H})$ avec une vitesse de l'ordre de v_n qui dépend de la condition d'entropie vérifiée par \mathcal{F} . Le résultat suivant est une conséquence de l'approximation forte donnée par le théorème 4.4.1 et la preuve est similaire à celle de la proposition 2.2.3.

Proposition 4.4.2. *On note d_{LP} la distance de Lévy-Prokhorov. Si \mathcal{F} vérifie $\mathcal{F}.i$ et $\mathcal{F}.ii$ alors il existe $C > 0$ tel que pour tout n , $d_{LP}(\mathbf{a}_n^{\mathbf{g}}, \mathbf{G}^{\mathbf{g}}) \leq C v_n$. Si l'hypothèse CV est vérifiée alors pour n suffisamment grand, $d_{LP}(\mathbf{a}_n^{\mathbf{g}^*}, \mathbf{G}^{\mathbf{g}}) \leq C v_n$.*

4.4.3 Application : covariance connue

Motivation. Plaçons-nous dans le cas où le statisticien connaît la covariance entre une variable d'intérêt X et une autre variable aléatoire Y et que celui-ci souhaite estimer l'espérance de X . Autrement dit nous travaillons avec la loi $P = \mathbb{P}^{(X,Y)}$ et l'ensemble $\mathcal{H} = \{h, g\}$ avec la fonction $h(P) = P(f_0)$, $f_0 : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $f_0(x, y) = x$ et la fonction auxiliaire $g(P) = P(f_0 g_0) - P(f_0)P(g_0)$, $g_0 : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $g_0(x, y) = y$. L'hypothèse $\mathcal{H}.i$ est vérifiée avec les fonctions

$$\varphi_h(P) = f_0, \quad R_h(Q) = 0, \quad \varphi_g(P) = g_0, \quad R_g(Q) = 0,$$

d'après le tableau 4.1. L'hypothèse $\mathcal{H}.ii$ est vérifiée si X, Y et XY admettent des moments d'ordre deux. Par conséquence, les matrices $\Sigma_1(h), \Sigma_2$ sont respectivement données par

$$\begin{aligned}\Sigma_1(h) &= \text{Cov}(X, XY - \mathbb{E}[X]Y - \mathbb{E}[Y]X), \\ \Sigma_2 &= \text{Var}(XY - Y\mathbb{E}[X] - X\mathbb{E}[Y]).\end{aligned}$$

D'après (4.14), la mesure empirique avec l'information auxiliaire donnée par $\text{Cov}(X, Y)$ est donnée par

$$\mathbf{P}_n^{\mathbf{g}}(h) = \mathbb{P}_n(X) - \frac{\text{Cov}(X, XY - Y\mathbb{E}[X] - X\mathbb{E}[Y])}{\text{Var}(XY - Y\mathbb{E}[X] - X\mathbb{E}[Y])} (\text{Cov}_n(X, Y) - \text{Cov}(X, Y)), \quad (4.21)$$

avec Cov_n la covariance empirique. Si le statisticien n'a pas accès aux valeurs de $\Sigma_1(h)$ ou Σ_2 il pourra utiliser des estimations en utilisant la connaissance de la covariance. Par exemple la matrice $\Sigma_1(h)$ peut être estimée par la valeur suivante

$$\begin{aligned}\widehat{\Sigma}_1(h) &= \mathbb{P}_n(X^2Y) - \mathbb{P}_n(Y)(\mathbb{P}_n^2(X) + \text{Var}_n(X)) - 2\mathbb{P}_n(X)\text{Cov}(X, Y), \\ \widehat{\Sigma}_2 &= \text{Var}_n(XY) + \mathbb{P}_n^2(X)\text{Var}_n(Y) + \mathbb{P}_n^2(Y)\text{Var}_n(X) \\ &\quad - 2(\mathbb{P}_n(X)\text{Cov}_n(XY, Y) + \mathbb{P}_n(Y)\text{Cov}_n(XY, X)) + 2\mathbb{P}_n(X)\mathbb{P}_n(Y)\text{Cov}(X, Y),\end{aligned}$$

avec Var_n la variance empirique. Asymptotiquement la variance du processus gaussien limite sera donnée par

$$\text{Var}(\mathbf{G}^{\mathbf{g}}(h)) = \text{Var}(X) - \frac{\text{Cov}(X, XY - Y\mathbb{E}[X] - X\mathbb{E}[Y])^2}{\text{Var}(XY - Y\mathbb{E}[X] - X\mathbb{E}[Y])}. \quad (4.22)$$

Un autre exemple qui pourrait être intéressant est la connaissance du fait que les variables aléatoires X, Y sont indépendantes et donc que la covariance entre les variables est nulle. Malheureusement, si X, Y sont indépendantes alors $\Sigma_1(h) = 0$, ce qui n'apporte aucune baisse de variance de l'estimation de l'espérance de X . Si les variables X, Y sont des lois normales, on peut également montrer que $\Sigma_1(h) = 0$ et donc encore une fois qu'il n'y a pas de baisse de variance.

Information donnée par $\text{Var}(X)$. Si $Y = X$, c'est-à-dire si la variance de X est connue alors il est nécessaire que X admette un moment d'ordre 4 et la mesure empirique avec l'information auxiliaire et la variance du processus gaussien limite données par (4.21), (4.22) deviennent

$$\begin{aligned}\mathbf{P}_n^{\mathbf{g}}(h) &= \mathbb{P}_n(X) - \frac{\text{Cov}(X, X^2) - 2\mathbb{E}[X]\text{Var}(X)}{\text{Var}(X^2 - 2X\mathbb{E}[X])} (\text{Var}_n(X) - \text{Var}(X)), \\ \text{Var}(\mathbf{G}^{\mathbf{g}}(h)) &= \text{Var}(X) - \frac{[\text{Cov}(X, X^2) - 2\mathbb{E}[X]\text{Var}(X)]^2}{\text{Var}(X^2 - 2X\mathbb{E}[X])}.\end{aligned}$$

Cette information peut être supposée connue dans le cadre de certains tests statistiques, comme nous le préciserons après. Nous donnons ci-après quelques exemples avec des lois usuelles.

Loi multivariée de Poisson. Un vecteur aléatoire suit une loi multivariée de Poisson si ses marginales suivent des lois de Poisson unidimensionnelles. Cette distribution a été proposée par Campbell en 1934 [25]. En 1964, Holgate [48] permet d'exprimer une variable de Poisson bivariée à l'aide de trois variables univariées de Poisson indépendantes. Pour plus

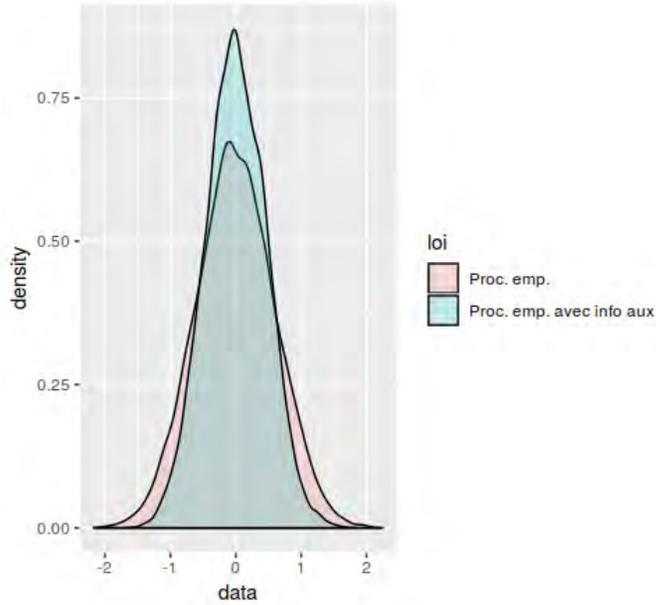


FIG. 4.2: Loi de $\alpha_n(X_1)$ et $\mathbf{a}_n^{\mathbf{g}}(X_1)$

de renseignements concernant cette loi, voir l'article de Morin [61]. Formellement, (X, Y) suit une loi multivariée de Poisson s'il existe X_1, X_2, X_3 tels que $X = X_1 + X_2, Y = X_1 + X_3$ avec $X_i \sim \mathcal{P}(\lambda_i)$ indépendantes deux à deux pour $i = 1, 2, 3$. Nous supposons que nous connaissons la covariance $\text{Cov}(X, Y) = \lambda_1$ et que $\lambda_1 \neq 0$ auquel cas nous retombons sur des variables indépendantes, cas évoqué précédemment. Si l'on souhaite estimer l'espérance de X avec l'information auxiliaire, les valeurs de Σ_1, Σ_2 sont données par

$$\Sigma_1(h) = \lambda_1, \quad \Sigma_2 = \lambda_1 + 2\lambda_1^2 + \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3.$$

La formule (4.22) de la variance asymptotique devient

$$\text{Var}(\mathbf{G}^{\mathbf{g}}(h)) = \lambda_1 + \lambda_2 - \frac{\lambda_1^2}{\lambda_1 + 2\lambda_1^2 + \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3}.$$

La figure 4.2 est une illustration numérique du processus empirique sans et avec l'information auxiliaire de la covariance. Les paramètres utilisés pour cette simulation sont $\lambda_1 = 0.2623, \lambda_2 = \lambda_3 = 0.1$. La courbe bleue représente la loi du processus empirique avec l'information auxiliaire et la rose celle sans. On utilise une méthode de Monte-Carlo avec des échantillons de taille $n = 1000$.

Généralisation à d'autres lois. On peut appliquer la même idée pour des variables X, Y définies par $X = X_1 + X_2, Y = X_1 + X_3$ avec $X_i \sim \mathcal{E}(\lambda_i)$ deux à deux indépendantes. Pour estimer l'espérance de X avec l'information auxiliaire, les valeurs Σ_1, Σ_2 sont données par

$$\Sigma_1(h) = \frac{2}{\lambda_1^3}, \quad \Sigma_2 = \frac{\lambda_1^4 + \lambda_1^2\lambda_2^2 + \lambda_1^2\lambda_3^2 + 8\lambda_2^2\lambda_3^2}{\lambda_1^4\lambda_2^2\lambda_3^2}.$$

```

MatrixForm[
Covariance[TransformedDistribution[
{X1 + X2, (X1 + X2) (X1 + X3) - (X1 + X3) * (1/lambda1 + 1/lambda2) - (X1 + X2) * (1/lambda1 + 1/lambda3)},
{X1 ~ ExponentialDistribution[lambda1]
, X2 ~ ExponentialDistribution[lambda2], X3 ~ ExponentialDistribution[lambda3]}]]]]

```

FIG. 4.3: Formule de Wolfram alpha

La variance asymptotique du processus avec informations auxiliaires est donnée par

$$\text{Var}(\mathbf{G}^{\mathbf{g}}(h)) = \frac{1}{\lambda_1^2} - \frac{4\lambda_2^2\lambda_3^2}{\lambda_1^2(\lambda_1^4 + \lambda_1^2\lambda_2^2 + \lambda_1^2\lambda_3^2 + 8\lambda_2^2\lambda_3^2)}.$$

On peut remarquer que plus la valeur λ_1 est faible, plus la réduction de variance sera importante. Si les X_i suivent une loi normale ou uniforme, $\Sigma_1(h) = 0$ et il n'y a donc pas réduction de variance. Les valeurs des variances et covariances ont pu être calculées grâce au logiciel Wolfram Alpha. La figure 4.3 est une capture d'écran de la formule permettant d'obtenir les valeurs dans le cas où les X_i suivent une loi exponentielle. Il suffit de changer la distribution pour obtenir les valeurs de $\Sigma_1(h), \Sigma_2$ dans le cas où les $X_i, i = 1, 2, 3$ suivent une autre loi.

4.4.4 Application : espérance conditionnelle connue

Motivation. Plaçons-nous cette fois-ci dans le cas où le statisticien connaît une espérance conditionnelle $\mathbb{E}[Y|A] = P(Y\mathbb{1}_A)/P(A)$. On peut imaginer par exemple qu'une étude préliminaire à moindre coût a été menée sur une variable spécifique auprès d'une certaine partie de la population et que l'on souhaite utiliser les résultats de cette étude afin de renforcer nos statistiques. Nous travaillons avec la loi $P = \mathbb{P}^{(X,Y)}$ et l'ensemble $\mathcal{H} = \{h, g\}$ avec la fonction $h(P) = P(f_0), f_0 : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $f_0(x, y) = x$ et la fonction auxiliaire $g(P) = P(g_0\mathbb{1}_A)/P(A), g_0 : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $g_0(x, y) = y$. L'hypothèse $\mathcal{H}.i$ est vérifiée avec les fonctions

$$\varphi_h(P) = f_0, R_h(Q) = 0, \quad \varphi_g(P) = (f_0 - P(f_0|A))\mathbb{1}_A, |R_g(Q)| \leq \|Q\|_{\mathcal{F}}.$$

L'hypothèse $\mathcal{H}.ii$ est satisfaite si les variables X et $Y\mathbb{1}_A$ admettent des moments d'ordre 2. Les valeurs $\Sigma_1(h), \Sigma_2$ sont données par

$$\begin{aligned} \Sigma_1(h) &= \frac{1}{P(A)} \text{Cov}(X, (Y - \mathbb{E}[Y|A])\mathbb{1}_A) \\ &= \mathbb{E}[XY|A] - \mathbb{E}[X|A]\mathbb{E}[Y|A] = \text{Cov}(X, Y|A), \\ \Sigma_2 &= \frac{1}{P(A)^2} \text{Var}((Y - \mathbb{E}[Y|A])\mathbb{1}_A) \\ &= \frac{1}{P(A)} (\mathbb{E}[Y^2|A] - \mathbb{E}[Y|A]^2) = \frac{\text{Var}(Y|A)}{P(A)}. \end{aligned}$$

Encore une fois, si Y est indépendante de X sur l'événement A alors $\Sigma_1(h) = 0$ et il n'y a pas de réduction de variance. Si le statisticien n'a pas accès aux valeurs de $\Sigma_1(h), \Sigma_2$ il pourra

les estimer respectivement par

$$\begin{aligned}\widehat{\Sigma}_1(h) &= \frac{\mathbb{P}_n(XY\mathbf{1}_A) - \mathbb{P}_n(X\mathbf{1}_A)\mathbb{E}[Y|A]}{\mathbb{P}_n(A)}, \\ \widehat{\Sigma}_2 &= \frac{1}{\mathbb{P}_n(A)} \left(\frac{\mathbb{P}_n(Y^2\mathbf{1}_A)}{\mathbb{P}_n(A)} - \mathbb{E}[Y|A]^2 \right).\end{aligned}$$

La mesure empirique et la variance asymptotique finale sont respectivement données par

$$\begin{aligned}\mathbf{P}_n^{\mathbf{g}} &= \mathbb{P}_n(X) - \frac{P(A)\text{Cov}(X, Y|A)}{\text{Var}(Y|A)} \left(\frac{\mathbb{P}_n(Y\mathbf{1}_A)}{\mathbb{P}_n(A)} - \mathbb{E}[Y|A] \right), \\ \text{Var}(\mathbf{G}^{\mathbf{g}}(h)) &= \text{Var}(X) - \frac{P(A)\text{Cov}(X, Y|A)^2}{\text{Var}(Y|A)}.\end{aligned}$$

Information donnée par $\mathbb{E}[X|A]$. Si $Y = X$ alors cela revient à dire que nous connaissons l'espérance de notre variable d'intérêt conditionnée à appartenir à l'ensemble A . Les valeurs de $\Sigma_1(h), \Sigma_2$ sont simplement exprimées par

$$\Sigma_1(h) = \text{Var}(X|A), \quad \Sigma_2 = \frac{\text{Var}(X|A)}{P(A)}.$$

La valeur $\Sigma_1(h)$ peut être estimée par $\widehat{\Sigma}_1(h) = \frac{\mathbb{P}_n(X^2\mathbf{1}_A) - \mathbb{E}[X|A]^2\mathbb{P}_n(A)}{\mathbb{P}_n(A)}$ et la variance asymptotique finale est donnée par

$$\text{Var}(\mathbf{G}^{\mathbf{g}}(h)) = \text{Var}(X) - P(A)\text{Var}(X|A).$$

La figure 4.4 est une simulation numérique de l'estimation de la moyenne empirique d'une variable X suivant une loi normale centrée réduite quand on injecte l'information $\mathbb{E}[X | -2 \leq X \leq 2] = 0$. Dans ce cas, les valeurs Σ_1, Σ_2 valent approximativement $\Sigma_1(h) \simeq 0.774, \Sigma_2 \simeq 0.811$, la réduction de variance est de 0.739.

Information donnée par $P(B|A)$. Supposons maintenant que la variable Y indique l'appartenance à une sous-population, *i.e.* $Y = \mathbf{1}_B$ et que l'on connaît $\mathbb{E}[Y|A] = P(B|A)$. Par exemple, on pourrait envisager que des experts ont mené une étude afin de connaître le taux d'infection ou d'une maladie dans une certaine partie de la population et que cette information puisse être utilisée dans une nouvelle étude. Les valeurs Σ_1, Σ_2 et la variance asymptotique sont données par

$$\begin{aligned}\Sigma_1 &= P(B|A)(\mathbb{E}[X|A \cap B] - \mathbb{E}[X|A]), \\ \Sigma_2 &= \frac{P(B|A)(1 - P(B|A))}{P(A)}, \\ \text{Var}(\mathbf{G}^{\mathbf{g}}(h)) &= \text{Var}(X) - \frac{P(A \cap B)(\mathbb{E}[X|A \cap B] - \mathbb{E}[X|A])^2}{1 - P(B|A)}.\end{aligned}$$

4.4.5 Application : coefficient de corrélation connu

Motivation. Soient X et Y des variables admettant des moments d'ordre 2. Supposons que nous sommes dans le cas où le statisticien connaît le coefficient de corrélation entre les variables aléatoires X et Y , c'est-à-dire la quantité $\rho(X, Y) = \text{Cov}(X, Y) / \sigma_X \sigma_Y$. Supposons

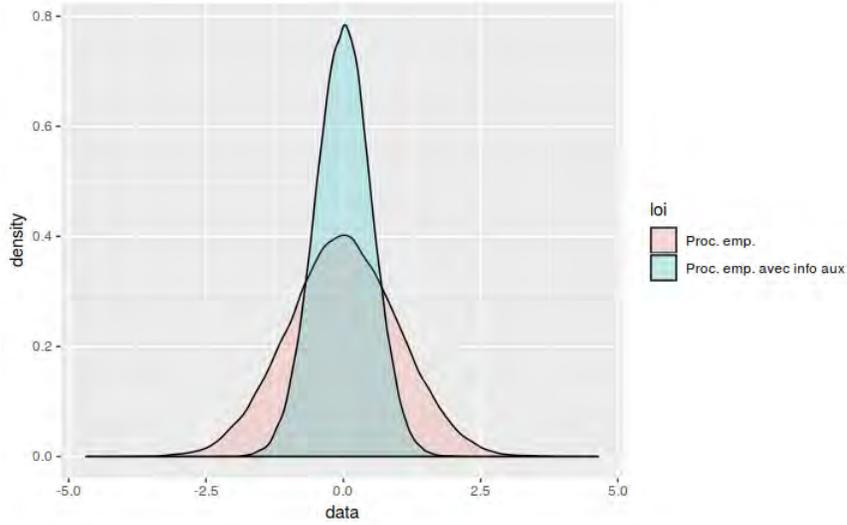


FIG. 4.4: Loi de $\alpha_n(X)$ et $\mathbf{a}_n^g(X)$

également que celui-ci souhaite estimer l'espérance de X , c'est-à-dire que $h(P) = P(X)$. La fonction d'information auxiliaire est donnée par

$$g(P) = \frac{P(XY) - P(X)P(Y)}{\sqrt{P(X^2) - P(X)^2} \sqrt{P(Y^2) - P(Y)^2}}.$$

Cette fonction est différentiable en la mesure P et admet pour différentielle appliquée en \mathbb{G} ,

$$d_P g(\mathbb{G}) = \frac{1}{\sigma_X \sigma_Y} \mathbb{G}(XY - P(Y)X - P(X)Y) - \frac{\rho(X, Y)}{2} \left(\frac{\mathbb{G}(X^2 - 2P(X)X)}{\text{Var}(X)} + \frac{\mathbb{G}(Y^2 - 2P(Y)Y)}{\text{Var}(Y)} \right)$$

L'hypothèse $\mathcal{H}.i$ est vérifiée et $\mathcal{H}.ii$ l'est si les variables X, Y et XY admettent des moments d'ordre deux. Les valeurs de Σ_1 et Σ_2 sont données par

$$\begin{aligned} \Sigma_1 &= \text{Cov}(d_P g(\mathbb{G}), \varphi_h(P)) \\ &= \frac{1}{\sigma_X \sigma_Y} (\text{Cov}(XY, X) - P(Y)\text{Var}(X)) - \frac{\rho(X, Y)}{2} \left(\frac{\text{Cov}(X^2, X)}{\text{Var}(X)} + \frac{\text{Cov}(Y^2 - 2P(Y)Y, X)}{\text{Var}(Y)} \right), \\ \Sigma_2 &= \text{Var}(d_P g(\mathbb{G})) \\ &= \text{Var} \left(\frac{XY - P(Y)X - P(X)Y}{\sigma_X \sigma_Y} - \frac{\rho(X, Y)}{2} \left(\frac{X^2 - 2P(X)X}{\text{Var}(X)} + \frac{Y^2 - P(Y)Y}{\text{Var}(Y)} \right) \right). \end{aligned}$$

Si le couple (X, Y) suit une loi normale multivariée, $\Sigma_1 = 0$ donc il n'y a pas réduction de variance.

4.4.6 Régression linéaire et information auxiliaire

Motivation. On se place dans le cadre d'un modèle de régression linéaire simple. On suppose que la relation suivante est vérifiée : $Y = \beta_0 + \beta_1 X + \varepsilon$ avec $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Nous ne faisons pas d'hypothèse sur la loi de X . La variable ε représente des erreurs potentielles, telles que des erreurs de prélèvement de données (liés à un matériel défectueux par exemple). On

suppose que les variables X et ε sont indépendantes. Les estimateurs classiques des moindres carrés sont données respectivement par

$$\hat{\beta}_0 = \mathbb{P}_n(Y) - \hat{\beta}_1 \mathbb{P}_n(X), \quad \hat{\beta}_1 = \frac{\text{Cov}_n(X, Y)}{\text{Var}_n(X)}.$$

Sous ce modèle ces estimateurs vérifient le théorème central limite suivante

$$\sqrt{n}(\hat{\beta}_0 - \beta_0) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \mathbb{E}[X^2]}{\text{Var}(X)}\right), \quad \sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2}{\text{Var}(X)}\right).$$

Régression avec information auxiliaire. On suppose qu'on travaille avec la loi $P = \mathbb{P}^{(X, Y)}$ et que nos fonctions d'intérêt sont données par

$$\begin{aligned} h_0(Q) &= Q(g_0) - h_1(Q)Q(f_0), \\ h_1(Q) &= \frac{Q(p_0) - Q(f_0)Q(g_0)}{Q(f_0^2) - Q(f_0)^2}, \end{aligned}$$

avec $f_0(x, y) = x, g_0(x, y) = y, p_0(x, y) = xy$. Avec ces notations, $\beta_0 = h_0(P), \beta_1 = h_1(P)$ et $\hat{\beta}_0 = h_0(\mathbb{P}_n), \hat{\beta}_1 = h_1(\mathbb{P}_n)$. Ces applications vérifient l'hypothèse $\mathcal{H}.i$ sous réserve que la variance de X soit non nulle. Les fonctions de cette hypothèse sont données par

$$\begin{aligned} \varphi_{h_0}(P) &= g_0 - h_1(P)f_0 - P(X)\varphi_{h_1}(P) = g_0 - \beta_1 f_0 - P(X)\varphi_{h_1}(P), \\ \varphi_{h_1}(P) &= \frac{1}{P(X^2) - P(X)^2} (p_0 - P(X)g_0 - P(Y)f_0) \\ &\quad - \frac{P(XY) - P(X)P(Y)}{(P(X^2) - P(X)^2)^2} (f_0^2 - 2P(X)f_0). \end{aligned}$$

Supposons que l'information auxiliaire est donnée par une statistique de la forme $g(P)$ avec $d_P g(\mathbb{G}) = \mathbb{G}(Z)$ où Z une variable aléatoire pouvant être corrélée aux paramètres de notre modèle. On suppose que Z admet un moment d'ordre 2. Alors $d_P g(Q) = Q(Z)$ et

$$\begin{aligned} \Sigma_1(h_1) &= \text{Cov}(d_P h_1(\mathbb{G}), d_P g(\mathbb{G})) = \frac{\text{Cov}(X, \varepsilon Z)}{\text{Var}(X)}, \\ \Sigma_1(h_0) &= \text{Cov}(d_P h_0(\mathbb{G}), d_P g(\mathbb{G})) = \text{Cov}(\varepsilon, Z) - \mathbb{E}[X] \text{Cov}(d_P h_1(\mathbb{G}), d_P g(\mathbb{G})) \\ &= \text{Cov}(\varepsilon, Z) - \frac{\mathbb{E}[X] \text{Cov}(X, \varepsilon Z)}{\text{Var}(X)}, \\ \Sigma_2 &= \text{Var}(d_P g(\mathbb{G})) = \text{Var}(\mathbb{G}(Z)) = \text{Var}(Z). \end{aligned}$$

Les covariances sont nulles si Z est indépendante de ε . Nous allons étudier plusieurs informations auxiliaires. Ainsi toute information auxiliaire donnée par une transformée de la variable X uniquement (l'espérance ou la variance de X par exemple) n'apporte aucune baisse de variance. Les nouveaux estimateurs de β_0 et β_1 sont dans ce cas respectivement donnés par

$$\begin{aligned} \hat{\beta}_0 &= \mathbf{P}_n^{\mathbf{g}}(h_0) = h_0(\mathbb{P}_n) - \frac{\Sigma_1(h_0)}{\Sigma_2} (g(\mathbb{P}_n) - g(P)) \\ &= \hat{\beta}_0 - \frac{\Sigma_1(h_0)}{\Sigma_2} (g(\mathbb{P}_n) - g(P)), \\ \hat{\beta}_1 &= \mathbf{P}_n^{\mathbf{g}}(h_1) = h_1(\mathbb{P}_n) - \frac{\Sigma_1(h_1)}{\Sigma_2} (g(\mathbb{P}_n) - g(P)) \\ &= \hat{\beta}_1 - \frac{\Sigma_1(h_1)}{\Sigma_2} (g(\mathbb{P}_n) - g(P)). \end{aligned}$$

Espérance des Y . Supposons que l'information auxiliaire soit donnée par la connaissance de $\mathbb{E}[Y] = \beta_0 + \beta_1\mathbb{E}[X]$. Cette information auxiliaire n'apporte aucune baisse de variance dans l'estimation de β_1 puisque $\Sigma_1(h_1) = \text{Cov}(d_P h_1(\mathbb{G}), \mathbb{G}(Y)) = 0$. Elle est néanmoins efficace dans l'estimation de β_0 . Les valeurs de Σ_1, Σ_2 sont données par

$$\begin{aligned}\Sigma_1(h_0) &= \text{Cov}(d_P h_0(\mathbb{G}), \mathbb{G}(Y)) = \text{Cov}(\varepsilon, Y) = \sigma^2, \\ \Sigma_2 &= \text{Var}(\mathbb{G}(Y)) = \text{Var}(Y),\end{aligned}$$

qui peuvent être estimés empiriquement et respectivement par

$$\begin{aligned}\widehat{\Sigma}_1(h_0) &= \text{Cov}_n(\widehat{\varepsilon}, Y) = \text{Cov}_n(Y - \widehat{\beta}_0 - \widehat{\beta}_1 X, Y) \\ &= \text{Var}_n(Y) - \frac{\text{Cov}_n(X, Y)^2}{\text{Var}_n(X)} = \text{Var}_n(Y) - \widehat{\beta}_1^2 \text{Var}_n(X), \\ \widehat{\Sigma}_2 &= \text{Var}_n(Y).\end{aligned}$$

Le nouvel estimateur de β_0 avec l'information auxiliaire est donnée par

$$\widehat{\widehat{\beta}}_0 = \widehat{\beta}_0 - \left(1 - \frac{\widehat{\beta}_1^2 \text{Var}_n(X)}{\text{Var}_n(Y)} (\mathbb{P}_n(Y) - \mathbb{E}[Y])\right).$$

La baisse de variance asymptotique de $\sqrt{n}(\widehat{\widehat{\beta}}_0 - \beta_0)$ par rapport à $\sqrt{n}(\widehat{\beta}_0 - \beta_0)$ est donc de $\Sigma_1^2/\Sigma_2 = \frac{\sigma^4}{\beta_1^2 \text{Var}(X) + \sigma^2}$. Cette réduction est donc d'autant plus importante que la variance de l'erreur est élevée et que la variance de X ou la valeur de $|\beta_1|$ est faible. La figure 4.5 représente en rouge $\sqrt{n}(\widehat{\widehat{\beta}}_0 - \beta_0)$ avec $X \sim \mathcal{N}(1, 0.8^2)$, $\sigma^2 = 4$, $\beta_0 = -2$, $\beta_1 = 0.5$. La diminution de variance asymptotique est donc approximativement de 3.85. Numériquement, on trouve par une méthode de Monte-Carlo pour $n = 2000$ que $\text{Var}(\sqrt{n}(\widehat{\beta}_0 - \beta_0)) \simeq 10.19$ et $\text{Var}(\sqrt{n}(\widehat{\widehat{\beta}}_0 - \beta_0)) \simeq 6.42$.

Espérance du produit XY . Supposons que l'information auxiliaire soit donnée par la connaissance de $\mathbb{E}[XY] = \mathbb{E}[X](\beta_0 + \beta_1\mathbb{E}[X]) + \beta_1 \text{Var}(X)$. L'estimation de β_0 n'est pas améliorée avec cette information auxiliaire puisque $\Sigma_1(h_0) = \text{Cov}(d_P h_0(\mathbb{G}), XY) = 0$. Les valeurs des matrices Σ_1, Σ_2 intervenant dans l'amélioration de $\widehat{\beta}_1$ sont données par

$$\begin{aligned}\Sigma_1(h_1) &= \frac{\text{Cov}(X, XY\varepsilon)}{\text{Var}(X)} = \sigma^2, \\ \Sigma_2 &= \text{Var}(XY) = \beta_0^2 \text{Var}(X) + \beta_1^2 \text{Var}(X^2) + \sigma^2 \mathbb{E}[X^2],\end{aligned}$$

qui peuvent être estimés empiriquement et respectivement par

$$\begin{aligned}\widehat{\Sigma}_1(h_1) &= \text{Var}_n(Y) - \widehat{\beta}_1^2 \text{Var}_n(X), \\ \widehat{\Sigma}_2 &= \widehat{\beta}_0^2 \text{Var}_n(X) + \widehat{\beta}_1^2 \text{Var}_n(X^2) + \text{Cov}_n(\widehat{\varepsilon}, Y) \mathbb{P}_n(X^2) \\ &= \widehat{\beta}_0^2 \text{Var}_n(X) + \widehat{\beta}_1^2 \text{Var}_n(X^2) + \mathbb{P}_n(X^2)(\text{Var}_n(Y) - \widehat{\beta}_1^2 \text{Var}_n(X)).\end{aligned}$$

La réduction de variance asymptotique de β_1 **Vraisemblance empirique.** La vraisemblance empirique est une méthode proche au niveau du concept de la méthode du maximum de vraisemblance. Alors que cette dernière est une méthode d'estimation nécessitant de travailler

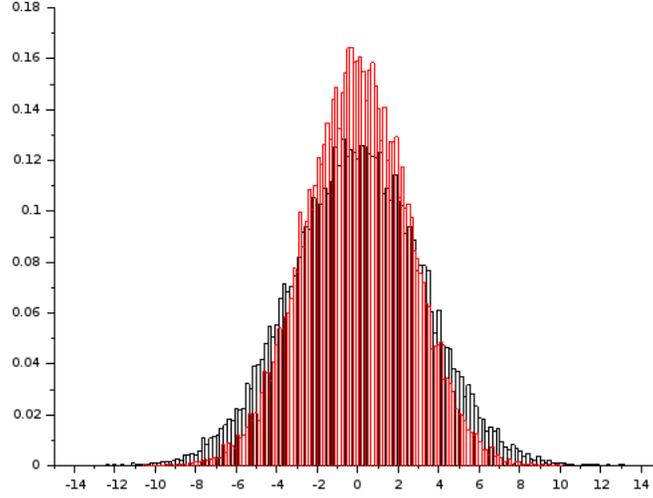


FIG. 4.5: Loi de $\sqrt{n}(\hat{\beta}_0 - \beta_0)$ et $\mathbf{a}_n^{\mathbf{g}}(h_0)$

avec une famille paramétrique de lois, la méthode de vraisemblance empirique s'affranchit de cette contrainte. Cette méthode a été introduite par Owen [63] (voir é sous information auxiliaire est de

$$\Sigma_1^2(h_1)/\Sigma_2 = \frac{\sigma^4}{\beta_0^2 \text{Var}(X) + \beta_1^2 \text{Var}(X^2) + \sigma^2 \mathbb{E}[X^2]}.$$

La figure 4.6 représente $\hat{\beta}_1$ avec $X \sim \mathcal{U}([0, 1])$, $\sigma^2 = 9$, $\beta_0 = -1$, $\beta_1 = 1$. La diminution de variance asymptotique de l'estimation de β_1 est approximativement de 25.43. Numériquement, on trouve par une méthode de Monte-Carlo pour $n = 2000$ que $\text{Var}(\sqrt{n}(\hat{\beta}_1 - \beta_1)) \simeq 107.47$, $\text{Var}(\sqrt{n}(\hat{\hat{\beta}}_1 - \beta_1)) \simeq 81.34$.

Covariance $\text{Cov}(X, Y)$. Supposons que l'information auxiliaire soit donnée par $g(P) = P(XY) - P(X)P(Y)$, c'est-à-dire que l'on connaît la valeur $\text{Cov}(X, Y) = \beta_1 \text{Var}(X)$. La différentielle de cette information auxiliaire est donnée par $d_P g(\mathbb{G}) = \mathbb{G}(XY - \mathbb{E}[X]Y - \mathbb{E}[Y]X)$. Les valeurs de Σ_1, Σ_2 sont données par

$$\begin{aligned} \Sigma_1(h_0) &= \text{Cov}(\varepsilon, XY - \mathbb{E}[X]Y - \mathbb{E}[Y]X) - \mathbb{E}[X]\Sigma_1(h_1) \\ &= -\mathbb{E}[X]\sigma^2, \\ \Sigma_1(h_1) &= \frac{\text{Cov}(X, \varepsilon(XY - \mathbb{E}[X]Y - \mathbb{E}[Y]X))}{\text{Var}(X)} = \sigma^2, \\ \Sigma_2 &= \text{Var}(XY - X\mathbb{E}[Y] - Y\mathbb{E}[X]) \\ &= \beta_1^2 \text{Var}(X^2) + \sigma^2 \text{Var}(X) - 4\mathbb{E}[X]\beta_1^2(\mathbb{E}[X^3] + \mathbb{E}[X]^3). \end{aligned}$$

La réduction de variance asymptotique de l'estimation de β_0 et β_1 est donnée respectivement

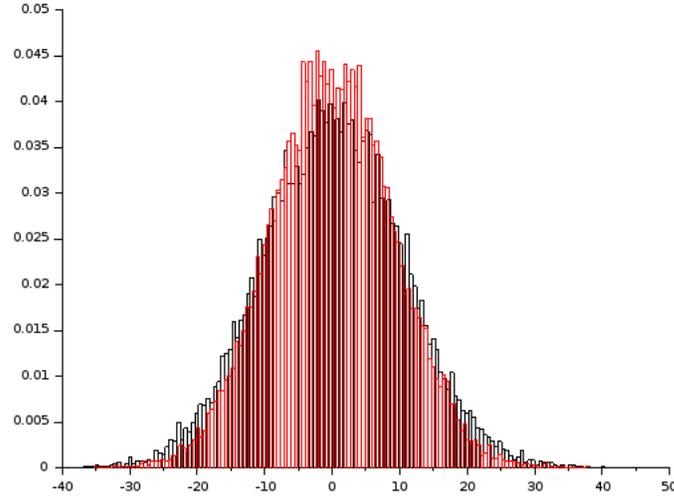


FIG. 4.6: Processus empirique de l'estimation de β_1 avec/sans l'information auxiliaire $\mathbb{E}[XY]$

par

$$\Sigma_1^2(h_0)/\Sigma_2 = \frac{\mathbb{E}[X]^2\sigma^4}{\text{Var}(XY - X\mathbb{E}[Y] - Y\mathbb{E}[X])},$$

$$\Sigma_1^2(h_1)/\Sigma_2 = \frac{\sigma^4}{\text{Var}(XY - X\mathbb{E}[Y] - Y\mathbb{E}[X])}.$$

La figure 4.7 et 4.8 illustrent respectivement $\widehat{\beta}_0$ et $\widehat{\beta}_1$ avec $X \sim \mathcal{N}(0.5, 0.5^2)$, $\sigma^2 = 16$, $\beta_0 = -1$, $\beta_1 = 3$. Les diminutions de variance asymptotique de l'estimation de β_0 et β_1 sont respectivement de 5.13 et de 20.5. Numériquement on trouve avec une méthode de Monte-Carlo pour $n = 2000$ que $\text{Var}(\sqrt{n}(\widehat{\beta}_0 - \beta_0)) \simeq 23.65$, $\text{Var}(\sqrt{n}(\widehat{\beta}_0 - \beta_0)) \simeq 18.84$, $\text{Var}(\sqrt{n}(\widehat{\beta}_1 - \beta_1)) \simeq 31.9$ et $\text{Var}(\sqrt{n}(\widehat{\beta}_1 - \beta_1)) \simeq 11.83$.

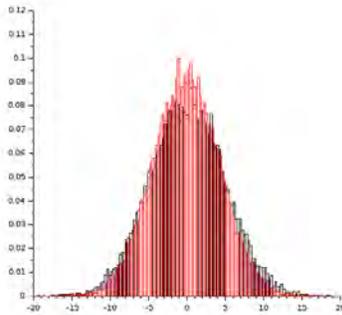


FIG. 4.7: Processus empirique de l'estimation de β_0 avec/sans l'information auxiliaire $\text{Cov}(X, Y)$

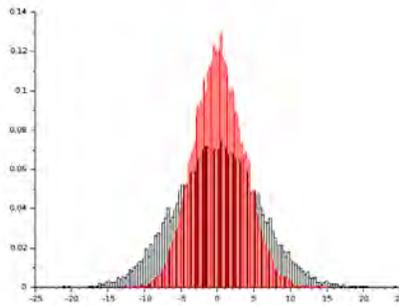


FIG. 4.8: Processus empirique de l'estimation de β_1 avec/sans l'information auxiliaire $\text{Cov}(X, Y)$

5.1 Introduction

5.1.1 The classical bootstrap

Presentation. The bootstrap is a very popular method of statistical inference introduced by Efron [37, 38, 40] that could be viewed as a generalization of the older jackknife method or leave k -out methods. Given $n \geq 1$ independent random variables X_1, \dots, X_n with common law P on a measurable space $(\mathcal{X}, \mathcal{A})$ let $\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ denote the associated empirical measure where δ_{X_i} is the Dirac measure at X_i . Any statistic of interest $S_n(X_1, \dots, X_n)$ is symmetric in its arguments, at least in distribution, and thus can be written $\varphi(\mathbb{P}_n)$. In practice one could face statistics that are not known to satisfy good estimation or test properties, or with unknown limiting variance thus preventing from computing confident bands. Each such S_n being unfortunately observed only once, moreover jointly to the same sample, it may seem irrelevant to infer their properties. The classical bootstrap aims to learn about unobserved properties of a statistic S_n by re-sampling X_1^*, \dots, X_n^* among X_1, \dots, X_n uniformly with replacement. The procedure consists in iterating to estimate by Monte Carlo methods the bias, variance or distribution of $S_n^* = S_n(X_1^*, \dots, X_n^*) = \varphi(\mathbb{P}_n^*)$ centered at the initially observed $S_n(X_1, \dots, X_n)$, where $\mathbb{P}_n^* = \frac{1}{n} \sum_{i=1}^n \delta_{X_i^*}$. The paradigm of Efron is that without any information on P the best way to mimic the unknown product measure $(P)^n = P \times \dots \times P$ of the original sample is to use the product empirical measure $(\mathbb{P}_n)^n = \mathbb{P}_n \times \dots \times \mathbb{P}_n$ and to center \mathbb{P}_n^* at \mathbb{P}_n instead of P . However, general assumptions ensuring non-asymptotic results are lacking to confirm such a belief, and to justify the Monte Carlo bootstrap approach by controlling the bias carried over each bootstrap experiment.

Motivation. From the mathematical statistics viewpoint a crucial question that has not been investigated in depth is to quantify the information one really gets about the distribution of S_n when bootstrapping b_n times, with $b_n \rightarrow +\infty$ as $n \rightarrow +\infty$. Towards this aim we address two unusual problems. The first problem is to choose b_n small enough to ensure the joint weak convergence of b_n Monte Carlo experiments and prove joint uniform Central Limit Theorems for an infinite collection of statistics, together with non asymptotic estimation properties of bootstrapped smooth statistics. The second problem is to incorporate an

auxiliary information on P while bootstrapping, and conversely to use the bootstrap method to simulate the behavior of an informed empirical process. The information we consider is the knowledge of the probability $P(A)$ of finitely many sets A or, equivalently, finitely many discrete marginals of P – in survey analysis, for instance. It is incorporated by the raking-ratio procedure we studied in [2], through an iterative modification of \mathbb{P}_n .

Monte Carlo bootstrap. A theoretical justification of the bootstrap method is not obvious even for a single explicit statistic $S_n = \varphi(\mathbb{P}_n)$ since it strongly depends on φ itself. One should evaluate how close the random experiments $\varphi(\mathbb{P}_n^*)$ and $\varphi(\mathbb{P}_n)$ are as $n \rightarrow +\infty$ to derive the consequences for the subsequent estimation procedure driven by b_n conditionally independent bootstraps. The known answers mainly concern empirical means and are asymptotic. Most often $b_n = 1$ when $n \rightarrow +\infty$ and, for fixed n , it is left implicit in practice that taking b_n large allows to numerically learn statistical properties of $\varphi(\mathbb{P}_n^*)$. Since no general study is available at fixed n , the balanced choice of b_n to guaranty the accuracy of Monte Carlo estimates can not be discussed. The probability distribution learned by bootstrapping being conditional to the initial sample, the difference with the unknown distribution of $\varphi(\mathbb{P}_n)$ could very well be misleading. Clearly, in the worse cases letting $b_n \rightarrow +\infty$ with n fixed is neither useful nor careful due to over-fitting and biasness. Our main contribution is to show that taking $c_1 \log n < b_n < c_2 n^{1/5}$ with c_1 large and c_2 small is a non-asymptotic rigorous compromise to control the bootstrap, in particular to estimate the variance and the distribution of a regular enough statistic – in a tractable setting.

Asymptotic justification. The statistics S_n we consider are of additive nature in the sense that they are sensitive to deviations between empirical and true expectations over a class of functions $\mathcal{F} \subset L_2(P)$. They are thus determined by $\mathbb{P}_n(f) = n^{-1} \sum_{i=1}^n f(X_i)$, $f \in \mathcal{F}$, and their behavior is led by $\alpha_n(f) = \sqrt{n}(\mathbb{P}_n(f) - P(f))$ where $P(f) = \mathbb{E}(f(X))$. The collection $\alpha_n(\mathcal{F}) = \{\alpha_n(f) : f \in \mathcal{F}\}$ is called the empirical process α_n indexed by \mathcal{F} . Its bootstrapped version is $\alpha_n^*(\mathcal{F}) = \{\alpha_n^*(f) : f \in \mathcal{F}\}$ where $\alpha_n^*(f) = \sqrt{n}(\mathbb{P}_n^*(f) - \mathbb{P}_n(f))$ and \mathbb{P}_n^* is the weighted bootstrap empirical measure introduced below. What has to be established is that conditionally to X_1, \dots, X_n the observed bootstrap process $\alpha_n^*(\mathcal{F})$ has the same behavior as the unobserved empirical process $\alpha_n(\mathcal{F})$ itself, for n large. Typically, with probability one α_n^* weakly converges to the weak limit of α_n as $n \rightarrow +\infty$, usually the P -Brownian bridge $\mathbb{G}(\mathcal{F})$ indexed by a P -Donsker class \mathcal{F} . The bootstrap method is therefore justified at the first order if $S_n = \varphi(\mathbb{P}_n) \in \mathbb{R}^d$, $S_n^* = \varphi(\mathbb{P}_n^*)$ and φ is Fréchet-differentiable at P with linear derivative operator $\varphi'(P)$ since then the distribution of $Y_n = \sqrt{n}(S_n - \varphi(P))$ can be estimated by the distribution of $Y_n^* = \sqrt{n}(S_n^* - S_n)$ which is in smooth cases asymptotically the same random vector $\varphi'(P) \cdot \mathbb{G}$ as the limit of Y_n provided that the differential distortion $\varphi'(\mathbb{P}_n) - \varphi'(P)$ vanishes. Whenever Y_n^* is simulated b_n times, the distortion generates a bias and b_n should be calibrated to avoid learning too much bias through $\varphi'(\mathbb{P}_n)$. In the general regular cases considered at Section 5.2.4 the weak distance between the distributions of Y_n , Y_n^* and the centered normal variable $\varphi'(P) \cdot \mathbb{G}$ or between the distributions of S_n^* and S_n are proved to vanish with explicit rates depending on b_n and entropy parameters.

Weak convergence approach. Giné and Zinn [44] proved that for any class of functions \mathcal{F} with envelope in $L_2(P)$ the weak convergence of $\alpha_n(\mathcal{F})$ to a – Gaussian or not – process \mathbb{G} indexed by \mathcal{F} is necessary and sufficient for the Efron’s bootstrap empirical process $\alpha_n^*(\mathcal{F})$ to almost surely converge weakly to $\mathbb{G}(\mathcal{F})$ also. This very nice statement is one of the most general results of the huge literature on the bootstrap methodology. We shall also work at this empirical process level. For a single real valued and regular statistic S_n a more common approach is through Edgeworth expansions, which exploits the cumulant expansion of the distribution function, see e.g. Hall [46] or Shao and Tu [71]. Other approaches rely on Berry-

Esseen bounds, like in Singh [73] or Mallows distances, as in Bickel and Freedman [13]. Under the name of Bayesian bootstrap, Rubin [69] defined an analogue of Efron’s bootstrap by resampling according to exchangeable weights that are independent of X_1, \dots, X_n rather than uniformly according to \mathbb{P}_n . In the case $\mathcal{X} = \mathbb{R}$, Mason and Newton [60] further generalized the bootstrap by independently assigning self-normalized random weights to the original data. If the weights are drawn independently from a multinomial distribution this reduces to the Efron’s bootstrap whereas if the weights come independently from a Dirichlet distribution this reduces to the Bayesian bootstrap. They established the weak convergence of this weighted real empirical process to a Brownian bridge provided the positive exchangeable weights satisfy a weak convergence condition. Unlike Giné and Zinn [44], their result does not handle the case indexed by \mathcal{F} . Præstgaard and Wellner [67] fills this gap, by still assuming that the exchangeable weights are independent of the data, with again a similar weak limit $\mathbb{G}(\mathcal{F})$ for $\alpha_n(\mathcal{F})$ and $\alpha_n^*(\mathcal{F})$, and $b_n = 1$.

About rates. The above weak convergence of Efron’s and weighted bootstrap processes are usually formulated in the sense of Hoffmann-Jørgensen to handle carefully measurability problems – see [6, 47]. The best known results – mainly [44] – are assembled in Chapter 3.6 of Van der Vaart and Wellner [80] – see also Kosorok [54]. The weak convergence is quantified in terms of the bounded Lipschitz norm between the processes $\alpha_n^*(\mathcal{F})$ and $\mathbb{G}(\mathcal{F})$, with no explicit rate. Obviously the rates could be arbitrarily slow for large \mathcal{F} or inadequate couples (P, \mathcal{F}) . One nice feature of our study is to provide general and explicit rates at the empirical process level for typical P -Donsker classes \mathcal{F} , with quantified statistical consequences on any bootstrapped regular statistic S_n^* . In this direction, a few authors considered the distance between the probability measures themselves, like \mathbb{P}_n^* , \mathbb{P}_n or P . For instance Shao [72] proved that the bounded Lipschitz distance between the uniform empirical measure \mathbb{P}_n on the d -dimensional unit cube and the Efron’s bootstrap empirical measure \mathbb{P}_n^* is bounded by $O(n^{-1/d})$ if $d > 2$ and $O(n^{-1/2}(\log n)^{(d-1)/2})$ if $d = 1, 2$. This improved Beran, Le Cam and Millar’s result [11] which only implies the convergence to zero, however in the more general indexed by sets setting. Other metrics have been studied in the indexed by \mathcal{F} setting. For instance, Barbe and Bertail [9] showed that various supremum type distances between the weighted bootstrap measure \mathbb{P}_n^* and P on \mathcal{F} are $O(n^{-1/2}(\log n)^{1/2})$ in probability where the extra $\log n$ term can be removed by following [67]. The forthcoming rates are of order $O(n^{-1/2}v_n)$ with $v_n \rightarrow 0$ at geometric or logarithmic decay.

Organization of the chapter. The weighted bootstrap we study is defined at Section 5.1.2. Our paradigm of auxiliary information from partitions is explained in Section 5.1.3 together with the two other bootstrap procedures to be considered. The main results are stated in Sections 5.2 and 5.3. More precisely, assumptions and notation are given at Section 5.2.1 then Section 5.2.2 and Section 5.2.3 provide the strong approximation of the weighted bootstrap empirical process iterated one and b_n times respectively. A few statistical consequences are derived in Sections 5.2.4 and 5.2.5 for statistics with Gaussian limits. Results of Section 5.3.1 show that we can apply the raking-ratio method after bootstrapping a sample in order to simulate the asymptotic law of the raking-ratio process. The results of Section 5.3.2 show how the performance of the basic bootstrap are improved by using a true information on P . The proof of all results are postponed until Section 5.4, focusing mainly on the classical case without information then avoiding straightforward but cumbersome details for the raked versions.

5.1.2 The weighted bootstrap

A strong approximation approach. By strong approximation we mean a coupling of a version of the process $\alpha_n^*(\mathcal{F})$ and a very close version $\mathbb{G}_n(\mathcal{F})$ of its Gaussian limiting process $\mathbb{G}(\mathcal{F})$ that is valid for all large n . This is easy to apply and provides a sharp control over the weak convergence metrics and rates. Such a Brownian coupling has been established in the very specific setting where KMT [51] can be applied, that is when P is the uniform law on $(0, 1)$. Alvarez-Andrade and Bouzebda [5] derived the usual almost sure rate $O(n^{-1/2} \log n)$ of strong approximation by a sequence of Brownian bridges of the weighted bootstrap process $\alpha_n^*(u) = \sqrt{n}(\mathbb{F}_n^*(u) - \mathbb{F}_n(u))$, where

$$\mathbb{F}_n(u) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq u\}}, \quad \mathbb{F}_n^*(u) = \sum_{i=1}^n W_{i,n} \mathbb{1}_{\{X_i \leq u\}}, \quad u \in (0, 1),$$

are the standard and weighted distribution functions respectively, with weights derived from independent and identically distributed (*i.i.d.*) positive random variables Z_1, \dots, Z_n also independent of X_1, \dots, X_n in the following way,

$$W_{i,n} = \frac{Z_i}{T_n}, \quad T_n = \sum_{i=1}^n Z_i, \quad (5.1)$$

provided $\mathbb{E}(Z) = \mathbb{V}(Z) = 1$ and Z_1 has a Laplace transform in a neighborhood of 0. By classical arguments this induces a distance $O(n^{-1/2} \log n)$ between \mathbb{F}_n and \mathbb{F}_n^* in various weak convergence metrics. We extend this indexed by $(0, 1)$ setting into the indexed by functions setting, and relax the usual assumption that the resampling weights Z_i/T_n are independent of the original sample by allowing (X_i, Z_i) to be *i.i.d.* with some distribution $P^{(X,Z)}$ while still controlling the marginal laws P of X and P^Z of Z . Notice that at that level of generality using KMT [51] is no more possible.

The weighted bootstrap empirical process. We revisit the results mentioned at Section 5.1.1 by analyzing the self-normalized weighted bootstrap empirical process

$$\alpha_n^*(f) = \sqrt{n}(\mathbb{P}_n^*(f) - \mathbb{P}_n(f)), \quad \mathbb{P}_n^*(f) = \sum_{i=1}^n W_{i,n} f(X_i), \quad f \in \mathcal{F}, \quad (5.2)$$

where the weights $W_{i,n}$ are defined at (5.1) and $(X_1, Z_1), \dots, (X_n, Z_n)$ are *i.i.d.* random vectors with distribution $P^{(X,Z)}$ on $(\mathcal{X}, \mathcal{A}) \times (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$. The conditional distributions $P^{(Z|X=x)}$ are assumed to exist and satisfy

$$\mathbb{E}(Z|X=x) = \text{Var}(Z|X=x) = 1, \quad x \in \mathcal{X}. \quad (5.3)$$

For sake of simplicity we shall assume

$$\mathbb{P}(0 < Z \leq M_Z) = 1, \quad 1 < M_Z < +\infty. \quad (5.4)$$

This is not restrictive when n is fixed since for $M_Z = F_Z^{-1}(1 - 1/n^{\theta+1} b_n)$ the random variable $\phi_{M_Z}(Z) = \mathbb{1}_{\{|Z| \leq M_Z\}} Z + \mathbb{1}_{\{|Z| > M_Z\}} M_Z$ behaves like Z over b_n bootstrapped samples with probability of the same order $1 - 1/n^\theta$ as our non asymptotic approximations. Hence the assumption (5.4) could be weakened by assuming light tails – like a finite Laplace transform or a finite high order moment – at the cost of additional technicalities as $n \rightarrow +\infty$ through truncation arguments. For $\mathcal{F} = \{\mathbb{1}_{\{\cdot \leq t\}} : t \in \mathbb{R}\}$ and the uniform distribution on $(0, 1)$ in (5.2) we recover the bootstrapped empirical process $\alpha_n^*(u)$ of [5] defined above, and the

assumptions stated at Section 5.2.1 allow any distribution P on \mathbb{R} for this class \mathcal{F} – not only compactly supported ones with positive continuous density away from zero as required by KMT approach. Moreover, we allow b_n bootstraps, which could not be handled by univariate KMT-type results.

New results in the purely nonparametric setting. Our main Theorem 5.2.5 is a non-asymptotic joint strong approximation of $b_n = o(n^{1/5})$ bootstrap iterations of $\alpha_n^*(\mathcal{F})$ by independent versions of the Gaussian limit process $\mathbb{G}(\mathcal{F})$ of $\alpha_n(\mathcal{F})$, jointly to the approximation of $\alpha_n(\mathcal{F})$ itself. This allows to turn the conditionally independent and orthogonal bootstrap empirical processes into *i.i.d.* Gaussian processes, which is a powerful transform. It ensues at Propositions 5.2.6 and 5.2.7 an uniform Central Limit Theorem for the bootstrap procedure (5.2) at first order rate $n^{-1/2}$ with second order rates in various weak convergence metrics, together with uniform Berry-Esseen type results at Corollary 5.2.10 and distances between distributions of S_n^* and S_n at Corollary 5.2.9, in particular an estimate of the variance at Corollary 5.2.8. The distance in distribution between S_n^* and S_n being controlled simultaneously for b_n bootstraps for statistics S_n built from $(\mathbb{P}_n, \mathcal{F})$ the Monte Carlo weighted bootstrap itself is justified, which answers to our first problem.

5.1.3 More weighted bootstraps

Bootstrapping in the auxiliary information nonparametric setting. Our second problem is to extend the bootstrap procedure to a less classical setting where an auxiliary information is known or learned about P . The motivation comes from the hasty development of distributed data. In this context it is realistic to consider a global statistical model where several sources learn deeply about one partial aspect of P – like a discrete marginal – and only communicate or sell their conclusions rather than their too large or confidential samples. In [2], by assuming this partial information to be exact we studied in details the raking-ratio empirical process obtained after N iterations of a reweighting procedure of \mathbb{P}_n that combines the auxiliary information given recursively by known partitions. In particular we showed that the limiting Gaussian process has a reduced variance. Suitable notation and a few results are recalled briefly below. In [1] this is extended to the actually learned information provided the auxiliary test samples are large enough. The underlying intuition connecting the bootstrap and the raking-ratio is twofold. On the one hand a better knowledge of P may help the bootstrap by either improving the initial \mathbb{P}_n and/or the redrawn \mathbb{P}_n^* . On the other hand bootstrapping the raked empirical measure may open access to the distribution of a raked statistic observed only once, to evaluate its lower variance, smaller bias and reduced risk.

The raking-ratio empirical process. The raking-ratio procedure consists in changing iteratively the weights of each X_i to match known probabilities of discrete marginals of P . This algorithm was introduced by Deming and Stephan [30] and rectified by Stephan [76] then justified by Lewis [57], Brown [23], Sinkhorn [74, 75] and finally Ireland and Kullback [50] who established the convergence to the Kullback projection. Special cases or closely related methods are stratification, calibration, fitting after sampling, iterative proportions, and some kind of nonparametric likelihood. For $N_0 \in \mathbb{N}_*$ and all $1 \leq N \leq N_0$ assume that $m_N \geq 1$ and $\mathcal{A}^{(N)} = (A_1^{(N)}, \dots, A_{m_N}^{(N)}) \in \mathcal{A}^{m_N}$ is a partition of \mathcal{X} such that the marginal $P(\mathcal{A}^{(N)}) = (P(A_1^{(1)}), \dots, P(A_{m_N}^{(N)}))$ is known and $p_N = \min_{1 \leq j \leq m_N} P(A_j^{(N)}) > 0$. Let $\mathbb{P}_n^{(N)}$ and $\alpha_n^{(N)}$ denote the empirical measure and process associated with the raking-ratio method. They are defined to be $\mathbb{P}_n^{(0)} = \mathbb{P}_n, \alpha_n^{(0)} = \alpha_n$ then recursively, on the event $\mathcal{B}_n^{N_0} = \bigcap_{1 \leq N \leq N_0} \mathcal{B}_n^{(N)}$ with

$\mathcal{B}_n^{(N)} = \{\min_{1 \leq j \leq m_N} \mathbb{P}_n^{(N-1)}(A_j^{(N)}) > 0\}$ for $1 \leq N \leq N_0$,

$$\begin{aligned} \mathbb{P}_n^{(N)}(f) &= \sum_{j=1}^{m_N} \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} \mathbb{P}_n^{(N-1)}(f 1_{A_j^{(N)}}), \\ \alpha_n^{(N)}(f) &= \sqrt{n}(\mathbb{P}_n^{(N)}(f) - P(f)), \quad f \in \mathcal{F}. \end{aligned} \quad (5.5)$$

In particular, $\mathbb{P}_n^{(N)}(A_j^{(N)}) = P(A_j^{(N)})$ and $\alpha_n^{(N)}(A_j^{(N)}) = 0$ for $j = 1, \dots, m_N$. By Proposition 2.3.3, $\mathbb{P}_n^{(N)}$ is the Kullback projection of $\mathbb{P}_n^{(N-1)}$ satisfying the N -th step constraint $\mathbb{P}_n^{(N)}(\mathcal{A}^{(N)}) = P(\mathcal{A}^{(N)})$. Under classical entropy conditions on \mathcal{F} , $\alpha_n^{(N)}(\mathcal{F})$ converges weakly as $n \rightarrow +\infty$ to a centered Gaussian process $\mathbb{G}^{(N)}$ defined in an iterative way similarly to $\alpha_n^{(N)}$ – see Proposition 3.3.8. More precisely, write $\mathbb{E}(f|A) = P(f 1_A)/P(A)$ and set $\mathbb{G}^{(0)} = \mathbb{G}$ to be the P -Brownian bridge indexed by \mathcal{F} then define the P -raked Brownian bridge to be, for $N \in \mathbb{N}_*$,

$$\mathbb{G}^{(N)}(f) = \mathbb{G}^{(N-1)}(f) - \sum_{j=1}^{m_N} \mathbb{E}(f|A_j^{(N)}) \mathbb{G}^{(N-1)}(A_j^{(N)}), \quad f \in \mathcal{F}. \quad (5.6)$$

The covariance of the limiting process $\mathbb{G}^{(N)}$ is

$$\text{Cov}(\mathbb{G}^{(N)}(f), \mathbb{G}^{(N)}(g)) = \text{Cov}(\mathbb{G}^{(0)}(f), \mathbb{G}^{(0)}(g)) - R_N(P, f, g) \quad (5.7)$$

where $R_N(P, f, g)$ has the closed form expression given at Proposition 3.3.3,

$$\begin{aligned} R_N(P, f, g) &= \sum_{k=1}^N \Phi_k^{(N)}(P, f) \cdot \text{Var}(\mathbb{G}[\mathcal{A}^{(k)}]) \cdot \Phi_k^{(N)}(P, g), \\ \Phi_k^{(N)}(P, f) &= \mathbb{E}[f|\mathcal{A}^{(k)}] + \sum_{l=k+1}^N \mathbf{P}(P, l) \cdot \mathbb{E}[f|\mathcal{A}^{(l)}], \end{aligned} \quad (5.8)$$

with $\mathbf{P}(P, l)$ a signed finite sum of $l \times l$ matrices depending on l and the values $P(A_{i_1}^{(j_1)}|A_{i_2}^{(j_2)})$ for $k < j_1 < j_2 \leq N$, $i_1 = 1, \dots, m_{j_1}$, $i_2 = 1, \dots, m_{j_2}$, whereas

$$\begin{aligned} \text{Var}(\mathbb{G}[\mathcal{A}^{(k)}]) &= \text{diag}(P(\mathcal{A}^{(k)})) - P(\mathcal{A}^{(k)})^t P(\mathcal{A}^{(k)}), \\ \mathbb{E}[f|\mathcal{A}^{(l)}] &= (\mathbb{E}(f|A_1^{(l)}), \dots, \mathbb{E}(f|A_{m_l}^{(l)}))^t. \end{aligned}$$

The asymptotic uniform variance reduction is induced by $R_N(P, f, f) \geq 0$ for all $N \geq 1$, $f \in \mathcal{F}$, and all finite dimensional covariance matrices of $\mathbb{G}^{(N)}(\mathcal{F})$ are decreasing compared to the initial one. The strong approximation of $\alpha_n^{(N)}$ by $\mathbb{G}^{(N)}$ established by Theorem 3.3.7 shows that the bias $\mathbb{E}(\mathbb{P}_n^{(N)}(f)) - P(f)$ is uniformly small and further provides rates of uniform quadratic risk reduction over \mathcal{F} . If recursive loops are performed among p partitions with known probabilities, for n sufficiently large the finite covariance matrices of $\alpha_n^{(kp)}$ decrease at each loop k . Theorem 3.5.6 provides a very simple expression for $\mathbb{G}^{(N)}$ as $N \rightarrow +\infty$ when raking with $p = 2$ partitions alternatively, which is the generic case of a two-way contingency table with known marginals enforced.

Raking the bootstrapped empirical process. Beyond the process α_n^* defined at (5.1) and (5.2) let us then introduce two other weighted bootstraps – a few variants satisfy similar results, unfortunately with heavier notation. According to the bootstrap paradigm, in

order to mimic the distribution of $\alpha_n^{(N)}$ one has to re-sample according to the weighted \mathbb{P}_n then apply the N -th order raking-ratio procedure to \mathbb{P}_n^* . This gives access by Monte Carlo approaches to the unknown distribution of $\alpha_n^{(N)}$ hence $\mathbb{G}^{(N)}$ for large n , which is useful since P is unknown. Define the raked bootstrapped empirical measure to be $\mathbb{P}_n^{*(0)} = \mathbb{P}_n^*$ then, recursively and conditionally to X_1, \dots, X_n on the event $\mathcal{B}_n^{*N_0} = \bigcap_{1 \leq N \leq N_0} \mathcal{B}_n^{*(N)}$ where $\mathcal{B}_n^{*(N)} = \{\min_{1 \leq j \leq m_N} \mathbb{P}_n^{*(N-1)}(A_j^{(N)}) > 0\}$,

$$\begin{aligned} \mathbb{P}_n^{*(N)}(f) &= \sum_{j=1}^{m_N} \frac{\mathbb{P}_n(A_j^{(N)})}{\mathbb{P}_n^{*(N-1)}(A_j^{(N)})} \mathbb{P}_n^{*(N-1)}(f 1_{A_j^{(N)}}), \\ \alpha_n^{*(N)}(f) &= \sqrt{n}(\mathbb{P}_n^{*(N)}(f) - \mathbb{P}_n(f)), \quad f \in \mathcal{F}. \end{aligned} \quad (5.9)$$

The centering with respect to \mathbb{P}_n in (5.9) should be discussed. In (5.2) the centering \mathbb{P}_n stands for the conditional expectation of \mathbb{P}_n^* and plays the role of the expectation P of \mathbb{P}_n in the centered process α_n . On the opposite, there is a bias inherent to the raking-ratio procedure so that P is no more the expectation of $\mathbb{P}_n^{(N)}$ in (5.5). This bias was established by Proposition 3.3.10 to be uniformly small and vanishing, hence P was confirmed as the targeted probability measure and the limiting process is centered. In order to simulate the influence of this bias we center the bootstrap (5.9) on \mathbb{P}_n . Therefore, the N -th order raked bootstrapped measure $\mathbb{P}_n^{*(N)}$ and process $\alpha_n^{*(N)}$ use the auxiliary information $\mathbb{P}_n(\mathcal{A}^{(N)}) = (\mathbb{P}_n(A_1^{(N)}), \dots, \mathbb{P}_n(A_{m_N}^{(N)}))$ instead of the original $P(\mathcal{A}^{(N)})$ thus

$$\mathbb{P}_n^{*(N)}(A_j^{(N)}) = \mathbb{P}_n(A_j^{(N)}), \quad \alpha_n^{*(N)}(A_j^{(N)}) = 0, \quad j = 1, \dots, m_N.$$

Bootstrapping the raked empirical process. A way to exploit directly the information of partitions is to bootstrap by using a probability that is possibly closer to P than \mathbb{P}_n actually is, namely $\mathbb{P}_n^{(N)}$. Let $T_n^{(N)}/n$ denote the mean of Z_1, \dots, Z_n under the discrete measure $\mathbb{P}_n^{(N)}$, that is $T_n^{(N)} = \sum_{i=1}^n n \mathbb{P}_n^{(N)}(\{X_i\}) Z_i$. In particular, $T_n^{(0)} = T_n$. Given $1 \leq N \leq N_0$, define the bootstrapped N -th order raked empirical measure and process to be, on the event $\mathcal{B}_n^{N_0}$,

$$\begin{aligned} \mathbb{P}_n^{(N)*}(f) &= \sum_{i=1}^n \frac{n \mathbb{P}_n^{(N)}(\{X_i\}) Z_i}{T_n^{(N)}} f(X_i), \\ \alpha_n^{(N)*}(f) &= \sqrt{n}(\mathbb{P}_n^{(N)*}(f) - \mathbb{P}_n^{(N)}(f)), \quad f \in \mathcal{F}. \end{aligned} \quad (5.10)$$

This reproducible imitation of α_n is a variant of (5.2).

5.2 Main results

5.2.1 The class \mathcal{F}

From now on, it is assumed that \mathcal{X} is measurable with σ -field \mathcal{A} . Notice that \mathcal{X} is not required to be metric separable nor \mathcal{A} to be Borel sets in order to establish approximation results, however this may help differentiability properties of the statistics to be bootstrapped. Let \mathcal{M} be the set of measurable real valued functions on $(\mathcal{X}, \mathcal{A})$ and $\tilde{\mathcal{F}} \subset \mathcal{F} \subset \mathcal{M}$ satisfying \mathcal{F} .i and \mathcal{F} .ii. Condition \mathcal{F} .ii ensures that the empirical process α_n and the variants defined in the previous section are point-wise separable and hence ball measurable, which allows to restrict their weak convergence to ball measurable test maps and avoid outer probabilities – see example 2.3.4 of [80]. Write $\sigma_f^2 = \text{Var}(f(X)) = P(f^2) - P(f)^2$ and $\sigma_{\mathcal{F}}^2 = \sup_{f \in \mathcal{F}} \sigma_f^2 \leq M_{\mathcal{F}}^2 < +\infty$.

5.2.2 Strong approximation for one bootstrap

For $n \in \mathbb{N}$, write $L(n) = \max(1, \log n)$ and $LL(n) = L(L(n))$.

Proposition 5.2.1. *Assume (5.3), (5.4) and either VC or BR. There exists a finite $K = K(\mathcal{F}, P^{(X,Z)}) > 0$ such that*

$$\limsup_{n \rightarrow +\infty} \sqrt{\frac{n}{LL(n)}} \|\mathbb{P}_n^* - P\|_{\mathcal{F}} \leq K \quad a.s.$$

Démonstration. This bounded LIL is established at Section 5.4.2. \square

In order to state the joint convergence of $(\alpha_n(\mathcal{F}), \alpha_n^*(\mathcal{F}))$ with rates, define

$$\begin{aligned} u_n &= n, v_n = n^{-\alpha_0} (\log n)^{\beta_0}, v_n^- = n^{-\alpha_1} (\log n)^{\beta_1}, \text{ if } \mathcal{F} \text{ satisfies VC,} \\ u_n &= \log n, v_n = (\log n)^{-\gamma_0}, v_n^- = (\log n)^{-\gamma_1}, \text{ if } (\mathcal{F}, P) \text{ satisfies BR.} \end{aligned}$$

Let formulate their joint approximation in the same way as in [12].

Théorème 5.2.2. *Assume (5.3), (5.4) and either VC or BR. For all $\theta > 0$ there exists $C_\theta > 0, n_\theta > 0$ and a probability space supporting a sequence $\{(X_n, Z_n)\}$ of i.i.d. random variables with distribution $P^{(X,Z)}$ and a sequence $\{(\mathbb{G}_n(\mathcal{F}), \mathbb{G}_n^*(\mathcal{F}))\}$ of pairs of independent P -Brownian bridges such that, for all $n > n_\theta$,*

$$\mathbb{P}(\{\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} > C_\theta v_n\} \cup \{\|\alpha_n^* - \mathbb{G}_n^*\|_{\mathcal{F}} > C_\theta v_n\}) \leq \frac{1}{n^\theta}. \quad (5.11)$$

Next assume either VC or BR with $r_0 > 1/2$. On the same probability space there exists two independent sequences $\{\mathbb{G}'_n(\mathcal{F})\}$ and $\{\mathbb{G}'_n^(\mathcal{F})\}$ of independent P -Brownian bridges such that, for all $n > n_\theta$,*

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \|\sqrt{k} \alpha_k - \sum_{i=1}^k \mathbb{G}'_i\|_{\mathcal{F}} > C_\theta v_n^-\right) \leq \frac{1}{u_n^\theta}, \quad (5.12)$$

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{k_{n,\theta} \leq k \leq n} \|\sqrt{k} \alpha_k^* - \sum_{i=1}^k \mathbb{G}'_i^*\|_{\mathcal{F}} > C_\theta v_n^-\right) \leq \frac{1}{u_n^\theta}. \quad (5.13)$$

Démonstration. See Steps 1 and 2 at Section 5.4.4. The constant C_θ, n_θ are different for each statement. The independence of \mathbb{G}_n and \mathbb{G}_n^* comes from Lemma 5.4.2. The processes \mathbb{G}_n^* (resp. \mathbb{G}_n) are not pair-wise independent, whereas by construction the \mathbb{G}'_n^* (resp. \mathbb{G}'_n) are mutually independent. \square

Notice that (5.11) applied with $\theta > 1$ not only implies that α_n^* almost surely weakly converges to \mathbb{G} but also provides a rate of weak convergence v_n . Let $d_{PL,1}(\mu, \nu)$ denote the Prokhorov-Lévy distance between two random elements of $\ell^\infty(\mathcal{F})$. The biased bootstrap process $\alpha_n^{*P}(\mathcal{F}) = \{\sqrt{n}(\mathbb{P}_n^*(f) - P(f)) : f \in \mathcal{F}\}$ is asymptotically unbiased, Gaussian and with doubled variance compared to the centered empirical process, that is $\text{Var}(\alpha_n^{*P}(f)) = 2\sigma_f^2$.

Corollaire 5.2.3. *Under the assumptions of Theorem 5.2.2, $\alpha_n^{*P}(\mathcal{F})$ has weak limit $2\mathbb{G}(\mathcal{F})$ on $\ell^\infty(\mathcal{F})$ and $d_{PL,1}(\alpha_n^{*P}(\mathcal{F}), 2\mathbb{G}(\mathcal{F})) \leq C v_n$ for some $C > 0$ and all n large enough.*

Démonstration. Since $\alpha_n^{*P} = \alpha_n^* + \alpha_n$ we get by (5.11) that for any $\theta > 0$, on the probability space of Theorem 5.2.2 it holds, for all $n > n_\theta$, $\|\alpha_n^{*P} - \mathbb{G}_n - \mathbb{G}_n^*\|_{\mathcal{F}} > 2C_\theta v_n$ with probability at most $2/n^\theta$. Now, $\mathbb{G}_n + \mathbb{G}_n^*$ is distributed as $2\mathbb{G}$ and $v_n > 1/n^\theta$ for n large. Therefore, by definition, $d_{PL,1}(\alpha_n^{*P}(\mathcal{F}), 2\mathbb{G}(\mathcal{F})) < 2C_\theta v_n$. \square

We also deduce the asymptotic independence of the coupled processes α_n and α_n^* , and their residual orthogonality is quantified by v_n .

Corollaire 5.2.4. *Under the assumptions of Theorem 5.2.2 it holds*

$$\limsup_{n \rightarrow +\infty} \frac{1}{v_n} \sup_{f, g \in \mathcal{F}} |\text{Cov}(\alpha_n(f), \alpha_n^*(g))| < +\infty.$$

Démonstration. Remind that $\text{Cov}(\alpha_n(f), \alpha_n^*(g))$ is deterministic and depends on the joint law of the processes (α_n, α_n^*) , not on the underlying probability space. Hence apply (5.11) and substitute \mathbb{G}_n and \mathbb{G}_n^* then follow the same path as the proof of Proposition 3.3.10. \square

5.2.3 Strong approximation for several bootstraps

Consider now the Monte Carlo experiment of b_n conditionally independent bootstraps. Evaluating the weak distance between α_n^* and its limit is crucial since each new bootstrap sample X_1^*, \dots, X_n^* is affected by it. In order to control the global distortion in play by not using P^n when bootstrapping a collection of estimators $\varphi(\mathbb{P}_n)$ we first approximate jointly the b_n bootstrapped empirical processes. The coupling error being quantified in the very strong sup-norm over \mathcal{F} , b_n has to be sufficiently small to guaranty that the confident bands for infinitely many estimated parameters are uniformly not over-biased.

Let $\mathbb{P}_{n,(j)}^*$ denote the j -th bootstrapped empirical measure built from the weights $(Z_{1,(j)}, \dots, Z_{n,(j)})$ drawn conditionally to X_1, \dots, X_n from the product distribution $P^{(Z|X=X_i)} \times \dots \times P^{(Z|X=X_n)}$. Write $\alpha_{n,(j)}^*(f) = \sqrt{n}(\mathbb{P}_{n,(j)}^*(f) - \mathbb{P}_n(f))$ the associated empirical process. For any fixed n and $1 \leq i \leq n$, the weights $(Z_{i,(1)}, \dots, Z_{i,(b_n)})$ thus correspond to X_i in each of the b_n bootstraps and they are not independent. Let denote $P^{(X_1, Z_{1,(1)}, \dots, Z_{1,(b_n)})}$ the ensuing joint distribution, so that the full experiment is distributed as the n -product of it.

For $b_n = 1$ the following result reduces to Theorem 5.2.2, otherwise the joint rate of approximation is slowed down. If \mathcal{F} satisfies (VC) then define

$$w_n = \left(\frac{b_n^5}{n} (\log n)^2 \right)^{\alpha_0} \left(\log \left(\frac{n}{b_n^5} \right) \right)^{5v/(4+10v)} \leq \left(\frac{b_n^5}{n} \right)^{\alpha_0} (\log n)^{\beta_0}. \quad (5.14)$$

If (\mathcal{F}, P) satisfy BR then define

$$w_n = \left(\frac{1}{\log(n/b_n^5)} \right)^{\gamma_0} \geq \left(\frac{1}{\log n} \right)^{\gamma_0}. \quad (5.15)$$

Théorème 5.2.5. *Assume (5.3), (5.4) and either VC or BR. Let $b_n \in \mathbb{N}_*$ be such that $b_n/n^{1/5} \rightarrow 0$. For all $\theta > 0$ there exists $C_\theta > 0$, $n_\theta > 0$ and a probability space supporting a triangular array $\{(X_n, Z_{n,(1)}, \dots, Z_{n,(b_n)})\}$ of i.i.d. random vectors distributed as $P^{(X_1, Z_{1,(1)}, \dots, Z_{1,(b_n)})}$ and a triangular array $\mathbb{G}_{n,b_n}(\mathcal{F}^{b_n+1}) = \{(\mathbb{G}_{n,(0)}^*(\mathcal{F}), \mathbb{G}_{n,(1)}^*(\mathcal{F}), \dots, \mathbb{G}_{n,(b_n)}^*(\mathcal{F}))\}$ of $(b_n + 1)$ -uplets of mutually independent P -Brownian bridges and a rectangular array $\{(X_n, Z_{n,(1)}, Z_{n,(2)}, \dots)\}$ of random variables such that, for all $n \geq n_\theta$, the n random vectors in \mathbb{R}^{b_n+1} ,*

$$(X_1, Z_{1,(1)}, \dots, Z_{1,(b_n)}), \dots, (X_n, Z_{n,(1)}, \dots, Z_{n,(b_n)})$$

are i.i.d. with distribution $P^{(X_1, Z_{1,(1)}, \dots, Z_{1,(b_n)})}$ and satisfy

$$\mathbb{P} \left(\left\{ \|\alpha_n - \mathbb{G}_{n,(0)}^*\|_{\mathcal{F}} > C_\theta w_n \right\} \cup \bigcup_{j=1}^{b_n} \left\{ \|\alpha_{n,(j)}^* - \mathbb{G}_{n,(j)}^*\|_{\mathcal{F}} > C_\theta w_n \right\} \right) \leq \frac{1}{n^\theta}. \quad (5.16)$$

Démonstration. See Step 3 at Section 5.4.4. The coupling performed there implies that the $(b_n + 1)$ -uplets $\mathbb{G}_{n,b_n}(\mathcal{F}^{b_n+1})$ are dependent, hence $\mathbb{G}_{m,(j)}^*$ and $\mathbb{G}_{n,(k)}^*$ are dependent if $m \neq n$ and independent if $m = n$ and $j \neq k$. \square

Remarque 5.2.1. *Theorem 5.2.5 allows to study the weak convergence of functionals estimated by b_n bootstrap experiments which are conditionally independent versions of α_n^* , by substituting the i.i.d. processes $\mathbb{G}_{n,(j)}^*$, $j \leq b_n$, to the $\alpha_{n,(j)}^*$. Moreover, the performance is good among a collection of statistics of interest.*

If $b_n = b$ is fixed, a multiple functional Central Limit Theorem immediately follows. Define the \mathbb{R}^{b+1} -valued empirical process indexed by \mathcal{F}^{b+1} to be

$$\Lambda_{n,b}(\mathcal{F}^{b+1}) = \{(\alpha_n(f_0), \alpha_{n,(1)}^*(f_1), \dots, \alpha_{n,(b)}^*(f_b)) : f = (f_0, f_1, \dots, f_b) \in \mathcal{F}^{b+1}\}.$$

Consider any norm $\|\cdot\|$ on \mathbb{R}^{b+1} then endow $\ell_{b+1}^\infty(\mathcal{F}) = \ell^\infty(\mathcal{F}^{b+1} \rightarrow \mathbb{R}^{b+1})$ with the distance associated to the sup-norm $\|\Lambda\|_{b+1} = \sup_{f \in \mathcal{F}^{b+1}} \|\Lambda(f)\|$.

Corollaire 5.2.6. *Under the assumptions of Theorem 5.2.5, for any fixed $b \in \mathbb{N}_*$ the sequence $\Lambda_{n,b}(\mathcal{F}^{b+1})$ converges weakly in $\ell_{b+1}^\infty(\mathcal{F})$ as $n \rightarrow +\infty$ to*

$$\mathbb{G}_b(\mathcal{F}^{b+1}) = \{(\mathbb{G}_{(0)}^*(f_0), \mathbb{G}_{(1)}^*(f_1), \dots, \mathbb{G}_{(b)}^*(f_b)) : f = (f_0, f_1, \dots, f_b) \in \mathcal{F}^{b+1}\}$$

where $\mathbb{G}_{(0)}^*, \mathbb{G}_{(1)}^*, \dots, \mathbb{G}_{(b)}^*$ are mutually independent P -Brownian bridges.

Démonstration. By Theorem 5.2.5, for any K -Lipschitz bounded function Ψ on $\ell_{b+1}^\infty(\mathcal{F})$,

$$\begin{aligned} & \mathbb{E}(\Psi(\Lambda_{n,b}(\mathcal{F}^{b+1}))) - \mathbb{E}(\Psi(\mathbb{G}_b(\mathcal{F}^{b+1}))) \\ & \leq K \mathbb{E}(1_{A_n} \|\Lambda_{n,b}(\mathcal{F}^{b+1}) - \mathbb{G}_{n,b}(\mathcal{F}^{b+1})\|_{b+1}) + \|\Psi\|_\infty \mathbb{P}(A_n) \\ & \leq C'_\theta K w_n + \frac{\|\Psi\|_\infty}{n^\theta} \end{aligned}$$

where A_n is the event in (5.16) and C'_θ depends on the norm chosen on \mathbb{R}^{b+1} . \square

Theorem 5.2.5 also implies that the distance in distribution between $\Lambda_{n,b_n}(\mathcal{F}^{b_n+1})$ and $\mathbb{G}_{n,b_n}(\mathcal{F}^{b_n+1})$ is at most $O(w_n)$, which is severely impacted by b_n and requires that $b_n/n^{1/5} \rightarrow 0$. Let $d_{PL,b_n}(\mu, \nu)$ denote the Prokhorov-Lévy distance between two random variables (μ, ν) with value in $\ell_{b+1}^\infty(\mathcal{F})$.

Corollaire 5.2.7. *Under the assumptions of Theorem 5.2.5 we have*

$$d_{PL,b_n}(\Lambda_{n,b_n}(\mathcal{F}^{b_n+1}), \mathbb{G}_{n,b_n}(\mathcal{F}^{b_n+1})) = O(w_n).$$

Démonstration. By definition of d_{PL,b_n} , (5.16) and since $w_n < 1/n^\theta$ for $\theta > 1/2$. \square

Remarque 5.2.2. *Statisticians used to bootstrap far more than n times may find the restriction $b_n/n^{1/5} \rightarrow 0$ rather penalizing. However, letting $b_n \rightarrow +\infty$ with n fixed simply simulates the distribution of $S_n^* = \varphi(\mathbb{P}_n^*)$ conditioned to \mathbb{P}_n , which is not the desired target – especially for small n . On the one hand, if the variance of $S_n = \varphi(\mathbb{P}_n)$ is not small the bootstrap can not help much as S_n^* is severely biased with high probability, and with high variance itself also, thus b_n should be kept carefully small. On the other hand, if n is large and the distributions of S_n is rather concentrated – and typically nearly Gaussian, like a maximum likelihood estimator – then generating $b_n = o(n^{1/5})$ resamples produces confident bands of width $1/\sqrt{b_n}$ times the already small variance – say, σ^2/n – thus the Monte Carlo information is significantly refined and relevant. This is quantified more precisely below by applying Theorem 5.2.5 to Frechet differentiable statistics.*

The above coupling results are useful to analyze estimators S_n bootstrapped b_n times as soon as they are smooth transforms of $\mathbb{P}_n(\mathcal{F})$. For instance, in Section 5.2.4 we evaluate the variance and the distribution function of S_n by assuming a local expansion – linear with second order error – which establishes in a wide setting a rigorous Monte Carlo bootstrap size compromise. In Section 5.2.5 we derive Berry-Esseen type bounds that are valid uniformly over \mathcal{F} .

5.2.4 Bootstrap estimation of variance and distribution function

Motivation. Consider the bootstrap Monte Carlo procedure in which \mathbb{P}_n^* is reevaluated b_n times by redrawing each time the n weights Z_i according to the conditional distribution $P(Z|X=X_i)$. When $b_n = 1$ the bootstrapped moments of smooth transforms of a mean estimator $\mathbb{P}_n^*(f)$ are classically controlled by using Edgeworth expansions – even themselves estimated, see Beran [10]. When $b_n = b > 1$ estimators of the moments of $\mathbb{P}_n^*(f)$ have also been studied, but not jointly, and heuristically. For instance, in the spirit of Efron [39], by assuming that the statistic of interest is *approximately* normal Booth and Sarkar [17] argued that the distribution of the bootstrap estimator of its variance is *approximately* a chi-squared distribution with $b - 1$ degrees of freedom. They deduce the $b > 1$ necessary to get a relative error less than a fixed bound with some probability, assuming n large – for an error less than 10% with probability 0.95 about $b = 800$ are required. To consolidate these *approximative* statements one may adapt the convergence of transforms of the empirical mean to a centered chi-squared distribution obtained by Chandra and Ghosh [26] through Edgeworth expansions. In this vein, Babu [8] showed that the bootstrapped version of smooth transforms of a single empirical mean has a similar weak asymptotic behavior by assuming that the functional is three times continuously differentiable. We revisit and complement this kind of bootstrap estimation of moments and distribution functions in the general case of functionals of the empirical measure itself instead of one or several empirical means. This follows from the above Brownian approximation under a generic assumption of first order differentiability on the space of measures.

Some regularity. Consider a statistic $S_n = \varphi(\mathbb{P}_n)$ with φ differentiable at P in the following weak sense.

Hypothesis (FR). Let $\varphi : \ell^\infty(\mathcal{F}) \rightarrow \mathbb{R}$ be a real-valued function and $t_0 > 0$ such that for all $Q \in \ell^\infty(\mathcal{F})$ with $\|Q\|_{\mathcal{F}} < t_0$,

$$\varphi(P + Q) = \varphi(P) + \varphi'(P) \cdot Q + R(Q), \quad (5.17)$$

where $\varphi'(P) : \ell^\infty(\mathcal{F}) \rightarrow \mathbb{R}$ is a linear application satisfying $|\varphi'(P) \cdot Q| \leq M_{P,\mathcal{F}} \|Q\|_{\mathcal{F}}$ and $R : \ell^\infty(\mathcal{F}) \rightarrow \mathbb{R}$ is an application such that for a ball \mathcal{B} centered on the zero function, some $q > 1$ and $M_{P,\mathcal{F}} < +\infty$ it holds $|R(Q)| \leq M_{P,\mathcal{F}} \|Q\|_{\mathcal{F}}^q$ for all $Q \in \mathcal{B}$. Assume also that for some $\kappa > 0$, $C > 0$ and all n ,

$$S_n < Cn^\kappa, \quad (5.18)$$

and that $\sigma^2 = \text{Var}(\varphi'(P) \cdot \mathbb{G}(\mathcal{F})) < +\infty$ where \mathbb{G} is the P -Brownian bridge.

Remarque 5.2.3. *Typical examples for S_n are additive statistics, regular maximum likelihood estimators or a delta-method by-products. Since they converge almost surely by the strong law of large numbers, the truncated version $S_n \mathbf{1}_{|S_n| \leq n^\kappa} + \text{sign}(S_n) n^\kappa \mathbf{1}_{|S_n| > n^\kappa}$ satisfies (5.18) and eventually coincides with S_n . The local expansion (5.17) includes Frechet differentiability and examples pages 11-16 in Barbe and Bertail [9]. The main feature behind (5.17) is that it*

allows to evaluate the asymptotic variance of S_n or $\sqrt{n}(S_n - \mathbb{E}(S_n))$ by bootstrapping, and obviously the variance showing up when using $Q = \alpha_n/\sqrt{n}$ is σ^2/n .

Examples. In the following table we make (5.17) explicit for a few classical estimators and finite classes for which $q = 2$ and σ^2 and the differential $\varphi'(P) \cdot \mathbb{G}$ are easily computed. Let denote $S_{n,1} = \mathbb{P}_n(f_0)$, $S_{n,2} = \mathbb{P}_n(f_0^2) - \mathbb{P}_n(f_0)^2$, $S_{n,3} = \mathbb{P}_n(f_0 \mathbf{1}_A)/\mathbb{P}_n(A)$ and, for $\varepsilon_n > \exp(-n^\kappa)$, $0 < \kappa < 1$,

$$S_{n,4} = \frac{1}{\mathbb{P}_n(f_0)} \mathbf{1}_{|\mathbb{P}_n(f_0)| > \varepsilon_n} + \text{sign}(\mathbb{P}_n(f_0)) \frac{1}{\varepsilon_n} \mathbf{1}_{|\mathbb{P}_n(f_0)| \leq \varepsilon_n}.$$

$S_n = \varphi(\mathbb{P}_n)$	\mathcal{F}	$\varphi'(P) \cdot \mathbb{G}$	σ^2
$S_{n,1}$	$\{f_0\}$	$\mathbb{G}(f_0)$	$\text{Var}(f_0(X))$
$S_{n,2}$	$\{f_0^2, f_0\}$	$\mathbb{G}(f_0^2) - 2P(f_0)\mathbb{G}(f_0)$	$\text{Var}(f_0^2(X))$
$S_{n,3}$	$\{f_0 \mathbf{1}_A, \mathbf{1}_A\}$	$\frac{\mathbb{G}(f_0 \mathbf{1}_A) - P(f_0 A)\mathbb{G}(A)}{P(A)}$	$\frac{\text{Var}((f_0(X) - \mathbb{E}(f_0(X) A))\mathbf{1}_{X \in A})}{P(A)^2}$
$S_{n,4}$	$\{f_0\}$	$-\frac{\mathbb{G}(f_0)}{P^2(f_0)}$	$\frac{\text{Var}(f_0(X))}{P^4(f_0)}$

Variance estimation by bootstrap. Write $S_{n,(j)}^* = \varphi(\mathbb{P}_{n,(j)}^*)$, for $1 \leq j \leq b_n$ and φ satisfying (5.17). The variance of S_n is estimated by

$$\widehat{\text{Var}}(S_n^*) = \frac{1}{b_n} \sum_{j=1}^{b_n} (S_{n,(j)}^* - S_n)^2 = \frac{1}{b_n} \sum_{j=1}^{b_n} \left(\varphi(\mathbb{P}_{n,(j)}^*) - \varphi(\mathbb{P}_n) \right)^2.$$

The next statement provides a confidence band for $\text{Var}(S_n)$ and establishes that $(nb_n/\sigma^2)\widehat{\text{Var}}(S_n^*)$ is close to the chi square distribution $\chi^2(b_n)$ with b_n degrees of freedom for large value of n .

Proposition 5.2.8. Assume (5.3), (5.4), (FR) and either (VC) or (BR) with $r_0 < 1/2$. Let $b_n \in \mathbb{N}_*$ be such that $b_n/n^{1/5} \rightarrow 0$. For $a_n \in (0, 1)$ define $\delta_n = \delta(a_n, b_n) > 0$ to be such that

$$\mathbb{P} \left(\left| \frac{\chi^2(b_n)}{b_n} - 1 \right| \geq \delta_n \right) \leq a_n.$$

Write $\xi_n^* = \max(w_n, (\log n)^{q/2}/n^{(q-1)/2})$. Then, for all $\theta > 0$ there exists $C_0 > 0$, $C_\theta > 0$ and $n_\theta > 0$ such that for all $n > n_\theta$,

$$\mathbb{P} \left(\frac{n}{\sigma^2} |\widehat{\text{Var}}(S_n^*) - \text{Var}(S_n)| \geq \delta_n + C_\theta \xi_n^* \sqrt{\log n} \right) \leq a_n + \frac{1}{n^\theta}, \quad (5.19)$$

$$|\text{Var}(S_n) - \sigma^2| \leq \frac{C_0 \xi_n^* \sqrt{\log n}}{n}. \quad (5.20)$$

Remarque 5.2.4. Observe that if $a_n = \alpha \in (0, 1)$ and $b_n \rightarrow +\infty$ then $\delta_n = \delta(\alpha, b_n)$ is of order $\sqrt{2}\Phi^{-1}(1 - \alpha/2)/\sqrt{b_n}$ whereas $\xi_n^* \rightarrow 0$ is most often of order w_n . Here Φ is the distribution function of the standard normal. Assuming (FR) and (BR) with $r_0 < 1/2$ entails that $\gamma_0 > 1/2$ in (5.15) thus $\xi_n^* \sqrt{\log n} \rightarrow 0$ with logarithmic decay rate and there is no guaranteed improvement by taking b_n larger than $(1/\xi_n^*)^2$ for a fixed or very slow a_n since then the main term in the deviation bound (5.19) remains $\xi_n^* \sqrt{\log n}$. On the opposite, assuming (FR) and (VC) together with $a_n = \alpha$ allows to choose $b_n = n^\beta$ with polynomial decay rate with power $\beta < \min(1/5, 2\alpha_0/(10\alpha_0 + 1))$ and keep the main order δ_n in (5.19).

Remarque 5.2.5. Precise approximations of the chi-square distribution by the standard normal could make the bivariate quantile function $\delta(a, b)$ tractable at appropriate joint scales $(a, b) \rightarrow (0, +\infty)$ and the relationship between a_n , b_n and δ_n more explicit. For instance, if $a_n > 1/n^\theta$ is slow enough and $b_n = n^\beta$ with $\beta < 1/5$ then $\delta_n^- = \delta_n \sqrt{b_n} / \sqrt{\log n}$ is slowly varying with respect to n . Assuming (FR) and (VC), ξ_n^* has first order rate $(b_n^5/n)^{\alpha_0}$ thus for β small enough we have $b_n < (\delta_n^-/\xi_n^*)^2$ which implies, by (5.19) and (5.20) that for some $C'_\theta > 0$ and all n sufficiently large,

$$\mathbb{P} \left(\left| \frac{\widehat{\text{Var}}(S_n^*)}{\sigma^2} - 1 \right| \geq \frac{C'_\theta \delta_n^- \sqrt{\log n}}{n \sqrt{b_n}} \right) \leq 2a_n.$$

Remarque 5.2.6. If $b_n = b$ is fixed large, and the class \mathcal{F} is finite we get the same approximation for b as suggested by Booth and Sarkar [17] to have a relative error for $\widehat{\text{Var}}(S_n^*)$ less than δ with a probability greater than $1 - \alpha$, namely $b \simeq 2\Phi^{-1}(1 - \alpha/2)^2/\delta^2$. Moreover, Proposition 5.2.8 provides a second order control as $\alpha \rightarrow 0$ of both the width and probability of the confident interval for the variance, uniformly in n and for infinite classes.

Remarque 5.2.7. Combining Theorem 5.2.8 with a crude moment bound to control a_n yields that there almost surely exists $C'_0 > 0$, $n'_0 = n'_0(\omega) > 0$ such that for all $n > n'_0$, $|\widehat{\text{Var}}(S_n^*) - \text{Var}(S_n)| \leq C'_0 (\log n)^\iota / \sqrt{nb_n}$ for $\iota > 1/2$ which is not efficient in view of (5.20) and far weaker than the rate $\delta_n^- \sqrt{\log n} / (n \sqrt{b_n})$ of Remark 5.2.5.

Distribution function estimation by bootstrap. The following result is a DKW-type inequality for bootstrapped statistics. It evaluates the uniform deviation between the unknown distribution function F_{S_n} of S_n and the estimated empirical distribution function

$$\widehat{F}_{S_n^*}(x) = \frac{1}{b_n} \sum_{j=1}^{b_n} \mathbb{1}_{\{S_{n,(j)}^* \leq x\}}.$$

The shift due to the unavoidable bias $B_n = S_n - \varphi(P)$ makes $\widehat{F}_{S_n^*}$ less accurate on the middle of the sample.

Proposition 5.2.9. Assume (5.3), (5.4), (FR) and either (BR) or (VC). Let $b_n \in \mathbb{N}_*$ be such that $b_n/n^{1/5} \rightarrow 0$. For any $\theta > 0$ there exists $C_\theta > 0$ and $n_\theta > 0$ such that, for all $n > n_\theta$,

$$\mathbb{P} \left(\sup_{x \in \mathcal{I}_{0,n}} |\widehat{F}_{S_n^*}(x) - F_{S_n}(x)| > C_\theta w_n \right) < \frac{1}{n^\theta} \quad (5.21)$$

$$\mathbb{P} \left(\sup_{x \notin \mathcal{I}_{0,n}} |\widehat{F}_{S_n^*}(x) - F_{S_n}(x)| > C_\theta \sqrt{\frac{\log n}{b_n}} \right) < \frac{1}{n^\theta} \quad (5.22)$$

where $\mathcal{I}_{0,n} = (\varphi(P) - z_n, \varphi(P) + z_n)$ and $z_n = \sqrt{\sigma^2/n} \sqrt{\log(b_n w_n^2 / \log n)}$ if $b_n w_n^2 > \log n$, otherwise $\mathcal{I}_{0,n} = \emptyset$.

Remarque 5.2.8. If b_n is slow then $\sup_{x \in \mathbb{R}} |\widehat{F}_{S_n^*}(x) - F_{S_n}(x)| = O(\sqrt{(\log n)/b_n})$ almost surely. If b_n is fast then $\sup_{x \notin \mathcal{I}_{0,n}} |\widehat{F}_{S_n^*}(x) - F_{S_n}(x)| = O(\sqrt{(\log n)/b_n})$ and $\sup_{x \in \mathcal{I}_{0,n}} |\widehat{F}_{S_n^*}(x) - F_{S_n}(x)| = O(w_n)$ almost surely, with $z_n = \sqrt{2\sigma^2(\log n)/n}$. In the latter case one can state (5.21) and (5.22) differently. Let Λ denote the class of strictly increasing continuous

mappings of \mathbb{R} onto itself. Consider the Skorokhod metric on the space of distribution functions – see Billingsley [14] –

$$d_1(F, G) = \inf_{\lambda \in \Lambda} \max \left(\sup_{x \in \mathbb{R}} |\lambda(x) - x|, \sup_{x \in \mathbb{R}} |G \circ \lambda(x) - F(x)| \right).$$

Since $\sqrt{n}B_n$ converges weakly to $\phi'(P) \cdot \mathbb{G}$ then $B_n = O(\sqrt{\log n/n})$ almost surely and it follows from the proof of Proposition 5.2.9 that for all $n > n_\theta$,

$$\mathbb{P} \left(d_1(F_{S_n}, \hat{F}_{S_n^*}) > C_\theta \sqrt{\frac{\log n}{b_n}} \right) \leq \frac{1}{n^\theta}.$$

5.2.5 Rates of weak convergence

Local Berry-Esseen bounds. Singh [73] – see also Section 3.1.3. of [71] – obtained Berry-Esseen type inequalities for the distribution of a bootstrapped mean, when $b = 1$. Under certain conditions he established that the uniform deviation between these distributions and Φ is almost-surely at most $O(n^{-1/2})$. With Edgeworth expansion techniques, Hall [45, 46] studied the leading term of the expansion of the uniform deviation conditionally to the sample and found the same rate as the statistic $n^{-3/2} |\sum_{j=1}^n X_j^3| + n^{-2} \sum_{j=1}^n X_j^4$ converges to 0 – that is also $O(n^{-1/2})$ if the fourth moment exists. In this section, we derive a Berry-Esseen inequality for the bootstrapped empirical process indexed by functions, *i.e.* uniform results among larger classes of statistics. The rates are slower than the $O(n^{-1/2})$ for a single smooth statistic.

Uniform Berry-Esseen bounds. Let \mathcal{L} be a set of Lipschitz functions defined on $\ell^\infty(\mathcal{F})$ such that all $\phi \in \mathcal{L}$ has a Lipschitz constant bounded by $C_1 < +\infty$ and the density of $\phi(\mathbb{G}(f))$ is bounded by $C_2 < +\infty$ for all $f \in \mathcal{F}$.

Proposition 5.2.10. *Assume (5.3), (5.4) and either (VC) or (BR). There exists $C_0 > 0, n_0 > 0$ such that for all $n > n_0$,*

$$\sup_{\phi \in \mathcal{L}} \sup_{f \in \mathcal{F}} \sup_{x \in \mathbb{R}} |\mathbb{P}(\phi(\alpha_n^*(f)) \leq x) - \mathbb{P}(\phi(\mathbb{G}(f)) \leq x)| \leq C_0 C_1 C_2 v_n, \quad (5.23)$$

$$\sup_{\phi \in \mathcal{L}} \sup_{f \in \mathcal{F}} \sup_{x \in \mathbb{R}} |\mathbb{P}(\phi(\alpha_n^*(f)) \leq x) - \mathbb{P}(\phi(\alpha_n(f)) \leq x)| \leq C_0 C_1 C_2 v_n. \quad (5.24)$$

In particular, if $\tilde{\sigma}_{\mathcal{F}}^2 = \inf_{f \in \mathcal{F}} \text{Var}(f(X)) > 0$ then

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \sup_{f \in \mathcal{F}} \left| \mathbb{P} \left(\frac{\alpha_n^*(f)}{\sqrt{\text{Var}(f(X))}} \leq x \right) - \Phi(x) \right| \leq \frac{C_0}{\sqrt{2\pi\tilde{\sigma}_{\mathcal{F}}}} v_n, \\ & \sup_{x \in \mathbb{R}} \sup_{f \in \mathcal{F}} \left| \mathbb{P} \left(\frac{\alpha_n^*(f)}{\sqrt{\text{Var}(f(X))}} \leq x \right) - \mathbb{P} \left(\frac{\alpha_n(f)}{\sqrt{\text{Var}(f(X))}} \leq x \right) \right| \leq \frac{C_0}{\sqrt{2\pi\tilde{\sigma}_{\mathcal{F}}}} v_n. \end{aligned}$$

5.3 Raking-ratio results

5.3.1 Strong approximation of $\alpha_n^{*(N)}$

Fix $N_0 \in \mathbb{N}^*$ and denote $P_{N_0} = \prod_{N=1}^{N_0} p_N, M_{N_0} = \prod_{N=1}^{N_0} m_N$. The bootstrapped empirical measure associated with the raking-ratio method obeys the bounded law of iterated logarithm.

Proposition 5.3.1. *If \mathcal{F} satisfies VC or BR then there exists $K = K(\mathcal{F}, Z) > 0$ such that,*

$$\limsup_{n \rightarrow +\infty} \max_{1 \leq N \leq N_0} \sqrt{\frac{n}{LL(n)}} \|\mathbb{P}_n^{*(N)} - P\|_{\mathcal{F}} \leq K b_{N_0} \quad a.s.,$$

where $b_0 = 1$ and $b_{N_0} = \prod_{N=1}^{N_0} (1 + M_{\mathcal{F}}/p_N)$.

The bootstrapped empirical process $\alpha_n^{*(N)}$ satisfies concentration inequalities with explicit but large constants.

Proposition 5.3.2. *Let $K_{\mathcal{F}} = \max(1, M_{\mathcal{F}})$ and assume (5.3), (5.4) and that $\alpha_n^{(N)}$ and $\alpha_n^{*(N)}$ are both defined for every $0 \leq N \leq N_0$. For all $n > 0, \lambda > 0$ we have*

$$\begin{aligned} \mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\alpha_n^{*(N)}\|_{\mathcal{F}} > \lambda \right) &\leq 4^{N_0+1} N_0^2 M_{N_0} \mathbb{P} \left(\|\alpha_n\|_{\mathcal{F}} > \frac{\lambda P_{N_0}^2}{3M_{N_0} (4K_{\mathcal{F}})^{N_0} (1 + \lambda/\sqrt{n})^{2N_0}} \right) \\ &\quad + 2^{N_0+2} N_0 \exp \left(-\frac{\lambda^2 P_{N_0}^4}{18M_{N_0}^2 M_Z^4 M_{\mathcal{F}}^2 (4K_{\mathcal{F}})^{2N_0} (1 + \lambda/\sqrt{n})^{4N_0}} \right). \end{aligned} \quad (5.25)$$

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

$$\mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\alpha_n^{*(N)}\|_{\mathcal{F}} > \lambda \right) \leq D_1 \exp(-D_2 \lambda^2), \quad (5.26)$$

for some positive values λ_0, D_0, D_1, D_2 . Under VC it holds, for $n > n_0$ and $\lambda_0 < \lambda < 2M_{\mathcal{F}} \sqrt{n}$,

$$\mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\alpha_n^{*(N)}\|_{\mathcal{F}} > \lambda \right) \leq D_3 \lambda_0^{\nu_0} \exp(-D_2 \lambda^2), \quad (5.27)$$

for some positive values λ_0, D_3, D_4 .

The approximation of $\alpha_n^{*(N)}$ by a version $\mathbb{G}^{(N)}$ of (5.6) holds with larger constants C_{θ} .

Théorème 5.3.3. *Assume (5.3), (5.4) and either VC or BR. Let $\theta > 0$. There exists $C_{\theta}, n_{\theta} > 0$ and a probability space supporting a sequence $\{(X_n, Z_n)\}$ of i.i.d. random variables with distribution $P^{(X, Z)}$ and sequences $\{(\mathbb{G}_n^{(N)}(\mathcal{F}), \mathbb{G}_n^{*(N)}(\mathcal{F}))\}$ of pairs of independent versions of $\mathbb{G}^{(N)}(\mathcal{F})$ for all $0 \leq N \leq N_0$ such that for all $n > n_{\theta}$,*

$$\mathbb{P} \left(\max_{1 \leq N \leq N_0} \|\alpha_n^{*(N)} - \mathbb{G}_n^{*(N)}\|_{\mathcal{F}} > C_{\theta} v_n \right) < \frac{1}{n^{\theta}}, \quad (5.28)$$

$$\mathbb{P} \left(\max_{1 \leq N \leq N_0} \|\alpha_n^{(N)} - \mathbb{G}_n^{(N)}\|_{\mathcal{F}} > C_{\theta} v_n \right) < \frac{1}{n^{\theta}}. \quad (5.29)$$

In the same spirit as Theorem 5.2.5 we also get the following.

Théorème 5.3.4. *Assume (5.3), (5.4) and either VC or BR. Let $\theta > 0$ and $b_n \in \mathbb{N}_*$ be such that $b_n/n^{1/5} \rightarrow 0$. There exists $C_{\theta} > 0, n_{\theta} > 0$ and a probability space supporting a triangular array $\{(X_n, Z_{n,(1)}, \dots, Z_{n,(b_n)})\}$ of i.i.d. random vectors distributed as $P^{(X_1, Z_{1,(1)}, \dots, Z_{1,(b_1)})}$ and a triangular array $\{(\mathbb{G}_{n,(0)}^*(\mathcal{F}), \mathbb{G}_{n,(1)}^*(\mathcal{F}), \dots, \mathbb{G}_{n,(b_n)}^*(\mathcal{F}))\}$ of $(b_n + 1)$ -uplets of mutually independent P -Brownian bridges such that, for $n \geq n_{\theta}$,*

$$\mathbb{P} \left(\max_{j=1, \dots, b_n} \max_{1 \leq N \leq N_0} \|\alpha_{n,(j)}^{*(N)} - \mathbb{G}_{n,(j)}^{*(N)}\|_{\mathcal{F}} > C_{\theta} w_n \right) < \frac{1}{n^{\theta}},$$

$$\mathbb{P} \left(\max_{j=1, \dots, b_n} \max_{1 \leq N \leq N_0} \|\alpha_{n,(j)}^{(N)} - \mathbb{G}_{n,(j)}^{(N)}\|_{\mathcal{F}} > C_{\theta} w_n \right) < \frac{1}{n^{\theta}}.$$

In particular, Theorem 5.3.3 implies that the bootstrapped empirical process associated with this method $\alpha_n^{*(N)}$ converges weakly in $\ell^\infty(\mathcal{F})$ to the Gaussian process $\mathbb{G}^{(N)}$ as the same way as $\alpha_n^{(N)}$. A simple way to simulate the law of the raking-ratio process is to bootstrap the initial available sample and apply the raking-ratio method. We can therefore estimate the covariance or variance of the raking-ratio empirical process by Monte-Carlo method without applying (5.7). The next result is a corollary of Theorem 5.3.4 and gives details about the precision of these estimators :

Proposition 5.3.5. *Assume (5.3), (5.4) and (VC). Let $b_n \in \mathbb{N}_*$ be such that $b_n/n^{1/5} \rightarrow 0$. There exists $C_0 > 0, n_0 = n_0(\omega) > 0$ such that for all $n > n_0$ it holds*

$$\max_{0 \leq N \leq N_0} \sup_{f, g \in \mathcal{F}} \left| \mathbb{E}[\alpha_n^{*(N)}(f)\alpha_n^{*(N)}(g)] - \text{Cov}(\mathbb{G}^{(N)}(f), \mathbb{G}^{(N)}(g)) \right| \leq C_0 v_n.$$

Démonstration. We adapt the proof of Proposition 3.3.10 by replacing $\mathbb{P}_n^{(N)}$ by $\mathbb{P}_n^{*(N)}$ and $\mathbb{G}_n^{(N)}$ by $\mathbb{G}_n^{*(N)}$ from Theorem 5.3.3. \square

Remarque 5.3.1. *Proposition 5.3.5 is an interesting result since we could estimate by Monte Carlo methods the variance and covariance of the raking-ratio process, and so the efficiency of an auxiliary information, without knowing probabilities $P(\mathcal{A}^{(N)})$ for $N \in \mathbb{N}^*$. If the auxiliary information has a cost – for example if this information is provided by experts, statistical learning or by data purchasing – the statistician could test the efficiency of the information before paying it.*

5.3.2 Strong approximation of $\alpha_n^{(N)*}$

We defined our raked bootstrapped empirical process $\alpha_n^{*(N)}$ by bootstrapping α_n in a first time and raking among known partitions after. We have shown that this process has the same asymptotic behavior as the raking-ratio Gaussian process $\mathbb{G}^{(N)}$. So a natural question is whether the behavior is the same for $\alpha_n^{(N)*}$, if we apply the raking-ratio at first and bootstrapped in a second time by adding random weights. The answer to this question is negative, main reason being that our process $\alpha_n^{*(N)}$ no longer satisfies the constraints given by the auxiliary information $\mathbb{P}_n(\mathcal{A}^{(N)})$ and has no reason to have the same asymptotic limit as that the non-bootstrapped process $\alpha_n^{(N)}$. However, we can prove that $\alpha_n^{(N)*}$ has the same asymptotic behavior as α_n^* .

Proposition 5.3.6. *Assume (5.3), (5.4) and either VC or BR. For all $\theta > 0$ there exists positive constants C_θ, n_θ and a probability space supporting a sequence $\{(X_n, Z_n)\}$ of i.i.d. random variables with distribution $P^{(X, Z)}$ and a sequence $\{(\mathbb{G}_n(\mathcal{F}), \mathbb{G}_n^{(0)*}(\mathcal{F}), \dots, \mathbb{G}_n^{(N_0)*}(\mathcal{F}))\}$ of P -Brownian bridges such that for all $n \geq n_\theta$,*

$$\mathbb{P} \left(\left\{ \|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} > C_\theta v_n \right\} \cup \bigcup_{N=0}^{N_0} \left\{ \|\alpha_n^{(N)*} - \mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > C_\theta v_n \right\} \right) \leq \frac{1}{n^\theta}. \quad (5.30)$$

5.4 Proofs

5.4.1 Decomposition of α_n^*

Let express the bootstrapped empirical process through the classical empirical process indexed by an appropriate extension of \mathcal{F} .

Step 1. In order to study $\alpha_n^*(\mathcal{F})$ from (5.2) it is more convenient to work with the conditionally centered version $\tilde{\alpha}_n^*(\mathcal{F})$ that we define to be

$$\tilde{\alpha}_n^*(f) = \sqrt{n}(\tilde{\mathbb{P}}_n^*(f) - \bar{Z}_n \mathbb{P}_n(f)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_i - \bar{Z}_n) f(X_i), \quad f \in \mathcal{F}, \quad (5.31)$$

where, by recalling also (5.1) and (5.2),

$$\tilde{\mathbb{P}}_n^*(f) = \frac{1}{n} \sum_{i=1}^n Z_i f(X_i), \quad \bar{Z}_n = \frac{T_n}{n} = \frac{1}{n} \sum_{i=1}^n Z_i, \quad \alpha_n^*(f) = \frac{n}{T_n} \tilde{\alpha}_n^*(f). \quad (5.32)$$

Step 2. Define $\phi_M(z) = \max(-M, \min(M, z))$ for $z \in \mathbb{R}$. Let $M > 0$ and write $\mathbf{1}$ the application constant to 1 on \mathbb{R} or \mathcal{X} . Given $f \in \mathcal{F}$ write $h_f = h_f^{(1)} - P(f)h^{(2)} - h_f^{(3)}$ where, for $(x, z) \in \mathcal{X} \times \mathbb{R}$,

$$h_f^{(1)}(x, z) = \phi_{M_Z}(z) f(x), \quad h^{(2)}(x, z) = \phi_{M_Z}(z), \quad h_f^{(3)}(x, z) = f(x). \quad (5.33)$$

The functions of (5.33) belong to the following extension \mathcal{F}_1 of \mathcal{F} ,

$$\mathcal{F}_1 = \{h(x, z) = f(x)g(z) : f \in \mathcal{F} \cup \{\mathbf{1}\}, g \in \{\phi_{M_Z}, \mathbf{1}\}\}, \quad (5.34)$$

and \mathcal{F}_1 has the same entropy as \mathcal{F} .

Lemme 5.4.1. *The class \mathcal{F}_1 defined by (5.34) is uniformly bounded, point-wise measurable, $\sigma_{\mathcal{F}_1} = \sigma_{\mathcal{F}}$ and satisfies conditions VC or BR with respect to $P^{(X, Z)}$ with the same powers as \mathcal{F} with respect to P , that is c_0, ν_0, b_0, r_0 .*

Démonstration. Let $h = fg \in \mathcal{F}_1$ with $f \in \mathcal{F} \cup \{\mathbf{1}\}$ and $g \in \{\mathbf{1}, \phi_M\}$. Clearly h is bounded by $M_1 = M_Z \max(1, M_{\mathcal{F}}) < +\infty$. For $(x, z) \in \mathcal{X} \times \mathbb{R}$, $f(x)$ is the limit of a sequence $\{f_n(x)\}$ with $f_n \in \tilde{\mathcal{F}}$, so $h(x, z)$ is the limit of $f_n(x)g(z)$ where $(f_n, g) \in \tilde{\mathcal{F}} \times \{\mathbf{1}, \phi_M\}$ which is countable and independent of h . By (5.3) it holds $P^{(X, Z)}(h) = \mathbb{E}(\mathbb{E}(Z|X)(f(X) - P(f))) - P(f) = -P(f)$ thus $h - P^{(X, Z)}(h) = (Z - 1)(f(X) - P(f))$ and $\mathbb{V}(h) = \mathbb{E}(\mathbb{E}((Z - 1)^2|X)(f(X) - P(f))^2) = \mathbb{V}(f)$ hence $\sigma_{\mathcal{F}_1} = \sigma_{\mathcal{F}}$. If \mathcal{F} is covered by $N(\mathcal{F}, \varepsilon, d_P)$ balls then \mathcal{F}_1 can be covered by $2N(\mathcal{F}, \varepsilon/M_Z, d_{P(x, z)})$ because $d_Q^2(h, f_0g) \leq M_Z^2 d_Q^2(f, f_0)$. If \mathcal{F} is covered by $N_{[\]}(\mathcal{F}, \varepsilon, d_P)$ ε -brackets then \mathcal{F}_1 can be covered by $2N_{[\]}(\mathcal{F}, \varepsilon, d_{P(x, z)})$ ε -brackets since $f_-(x) < f(x) < f_+(x)$ implies $f_-(x)g(z) < h(x, z) < f_+(x)g(z)$ for g positive. Hence if \mathcal{F} satisfies VC or BR it is the same for \mathcal{F}_1 with larger constants c_0 or b_0 . \square

Step 3. By (5.3) we have $\mathbb{E}(Zf(X)) = P(f) = \mathbb{E}(ZP(f))$, thus by (5.31), $\tilde{\alpha}_n^*$ can be split

into

$$\begin{aligned}
\sqrt{n}\tilde{\alpha}_n^*(f) &= \sum_{i=1}^n f(X_i)Z_i - \bar{Z}_n \left(\sum_{i=1}^n f(X_i) \right) \\
&= \sum_{i=1}^n Z_i(f(X_i) - P(f)) - \bar{Z}_n \sqrt{n}\alpha_n(f) \\
&= \sum_{i=1}^n (Z_i f(X_i) - \mathbb{E}(Zf(X))) - \sum_{i=1}^n (Z_i P(f) - \mathbb{E}(ZP(f))) \\
&\quad - \sqrt{n}\alpha_n(f) (1 + (\bar{Z}_n - 1)).
\end{aligned}$$

The last expression is a linear functional of the empirical process $\alpha_n^{(X,Z)}$ built from the sample $(X_1, Z_1), \dots, (X_n, Z_n)$ and indexed by \mathcal{F}_1 from (5.34) since

$$\begin{aligned}
\tilde{\alpha}_n^*(f) &= \alpha_n^{(X,Z)}(h_f^{(1)}) - P(f) \cdot \alpha_n^{(X,Z)}(h_f^{(2)}) - \alpha_n^{(X,Z)}(h_f^{(3)}) (1 + (\bar{Z}_n - 1)) \\
&= \alpha_n^{(X,Z)}(h_f) - \alpha_n^{(X,Z)}(f) (\bar{Z}_n - 1),
\end{aligned} \tag{5.35}$$

where $h_f^{(1)}, h_f^{(2)}, h_f^{(3)} \in \mathcal{F}_1$ are defined by (5.33), $\alpha_n^{(X,Z)} = \sqrt{n}(\mathbb{P}_n^{(X,Z)} - P^{(X,Z)})$ and $\mathbb{P}_n^{(X,Z)} = \frac{1}{n} \sum_{i=1}^n \delta_{(X_i, Z_i)}$.

5.4.2 Proof of Propositions 5.2.1, 5.3.1 and 5.3.2

At Step 1 we prove a law of iterated logarithm for the simpler empirical bootstrapped measure $\tilde{\mathbb{P}}_n^*$ and Proposition 5.2.1 follows. Then it is used to prove Proposition 5.3.1 at Step 2. Finally, Proposition 5.3.2 is established at Step 3.

Step 1. Let $t_n = \sqrt{n/LL(n)}$ and fix $\varepsilon > 0$. By Lemma 5.4.1, \mathcal{F} and \mathcal{F}_1 satisfy both VC or BR. As a consequence, the law of the iterated logarithm holds – see Alexander [4] – and with probability one there exists $n_0 = n_0(\omega) > 0$ such that for $n > n_0$,

$$t_n \|\mathbb{P}_n - P\|_{\mathcal{F}} \leq \sqrt{2}\sigma_{\mathcal{F}}(1 + \varepsilon), \tag{5.36}$$

$$t_n \|\mathbb{P}_n^{(X,Z)} - P^{(X,Z)}\|_{\mathcal{F}_1} \leq \sqrt{2}\sigma_{\mathcal{F}_1}(1 + \varepsilon),$$

$$t_n |\bar{Z}_n - 1| = t_n |\mathbb{P}_n^{(X,Z)}(h^{(2)}) - P^{(X,Z)}(h^{(2)})| \leq \sqrt{2}(1 + \varepsilon). \tag{5.37}$$

Let us bound the two terms

$$t_n \|\tilde{\mathbb{P}}_n^* - P\|_{\mathcal{F}} \leq t_n \|\tilde{\mathbb{P}}_n^* - \bar{Z}_n \mathbb{P}_n\|_{\mathcal{F}} + t_n \|\bar{Z}_n \mathbb{P}_n - P\|_{\mathcal{F}}. \tag{5.38}$$

By (5.35) and (5.36), for all $n \geq n_0$ we have

$$\begin{aligned}
t_n \|\tilde{\mathbb{P}}_n^* - \bar{Z}_n \mathbb{P}_n\|_{\mathcal{F}} &\leq t_n \|\mathbb{P}_n^{(X,Z)} - P^{(X,Z)}\|_{\mathcal{F}_1} (2 + M_{\mathcal{F}} + \varepsilon) \\
&\leq \sqrt{2}M_Z \sigma_{\mathcal{F}} (2 + M_{\mathcal{F}} + \varepsilon)(1 + \varepsilon).
\end{aligned} \tag{5.39}$$

Remind that $\mathcal{F} \subset \mathcal{F}_1$. By (5.36) and (5.37) we can write for all $n \geq \max(n_0, n_1)$,

$$\begin{aligned}
t_n \|\bar{Z}_n \mathbb{P}_n - P\|_{\mathcal{F}} &= t_n \left\| \frac{\sum_{i=1}^n Z_i}{n} \frac{\sum_{i=1}^n f(X_i)}{n} - \mathbb{E}[f(X)] \right\|_{\mathcal{F}} \\
&\leq t_n \left\| \mathbb{P}_n^{(X,Z)}(h^{(2)}) (\mathbb{P}_n(f) - P(f)) \right\|_{\mathcal{F}_1} \\
&\quad + M_{\mathcal{F}} t_n |\mathbb{P}_n^{(X,Z)}(h^{(2)}) - P^{(X,Z)}(h^{(2)})| \\
&\leq \sqrt{2}(M_Z \sigma_{\mathcal{F}} + M_{\mathcal{F}})(1 + \varepsilon).
\end{aligned} \tag{5.40}$$

We have shown by (5.38), (5.39) and (5.40) that for all $n > \max(n_0, n_1)$, $t_n \|\tilde{\mathbb{P}}_n^* - P\|_{\mathcal{F}} < K_0(1 + \varepsilon)$ with $K_0 = K_0(\mathcal{F}, P^{(X,Z)}) = \sqrt{2}M_Z\sigma_{\mathcal{F}}(3 + 2M_{\mathcal{F}}) > 0$. By (5.32) we have for all $n > 0$,

$$t_n \|\mathbb{P}_n^* - P\|_{\mathcal{F}} = \frac{t_n \|\tilde{\mathbb{P}}_n^* - P\|_{\mathcal{F}}}{1 + (\bar{Z}_n - 1)}.$$

According to the conclusion of Step 1 and (5.37) for all $n > \max(n_0, n_1)$ and $\varepsilon \in (0, 1/2)$ it holds,

$$t_n \|\mathbb{P}_n^* - P\|_{\mathcal{F}} \leq \frac{K_0}{1 - \varepsilon}(1 + \varepsilon) \leq 2K_0(1 + \varepsilon).$$

Proposition 5.2.1 is proved with $K = K(\mathcal{F}, P^{(X,Z)}) = 2K_0(\mathcal{F}, P^{(X,Z)}) > 0$.

Step 2. Let $b = \sqrt{2}\sigma_{\mathcal{F}}$. A sufficient condition to prove Proposition 5.3.1 is to show that with probability one there exists $K = K(\mathcal{F}, P^{(X,Z)}) > 0, n_2 = n_2(\omega) > 0$ such that for all $n > n_2$,

$$t_n \|\mathbb{P}_n^{*(N)} - P\|_{\mathcal{F}} \leq (Kb_N - b)(1 + \varepsilon). \quad (5.41)$$

The case $N = 0$ is proved by Proposition 5.2.1. Now let assume that this condition is satisfied for some fixed $0 \leq N < N_0$. Then by (5.41) for all $n \geq n_2$,

$$\begin{aligned} t_n \|\mathbb{P}_n^{*(N+1)} - P\|_{\mathcal{F}} &\leq t_n \|\mathbb{P}_n^{*(N+1)} - \mathbb{P}_n^{*(N)}\|_{\mathcal{F}} + t_n \|\mathbb{P}_n^{*(N)} - P\|_{\mathcal{F}} \\ &\leq t_n \|\mathbb{P}_n^{*(N+1)} - \mathbb{P}_n^{*(N)}\|_{\mathcal{F}} + (Kb_N - b)(1 + \varepsilon) \end{aligned} \quad (5.42)$$

Notice that $m_{N+1} \leq p_{(N_0)}$ and $\|\mathbb{P}_n^{*(N)}(f1_A)\|_{\mathcal{F}} \leq M_{\mathcal{F}}\mathbb{P}_n^{*(N)}(A)$. By (5.36) and (5.41) we have for all $n > \max(n_0, n_2)$,

$$\begin{aligned} t_n \|\mathbb{P}_n^{*(N+1)} - \mathbb{P}_n^{*(N)}\|_{\mathcal{F}} &= t_n \left\| \sum_{j \leq m_{N+1}} \frac{\mathbb{P}_n^{*(N)}(f1_{A_j^{(N+1)}})}{\mathbb{P}_n^{*(N)}(A_j^{(N+1)})} (\mathbb{P}_n^{*(N)}(A_j^{(N+1)}) - \mathbb{P}_n(A_j^{(N+1)})) \right\|_{\mathcal{F}} \\ &\leq \frac{M_{\mathcal{F}}}{p_{N+1}} (t_n \|\mathbb{P}_n^{*(N)} - P\|_{\mathcal{F}} + t_n \|\mathbb{P}_n - P\|_{\mathcal{F}}) \\ &\leq \frac{M_{\mathcal{F}}}{p_{N+1}} Kb_N(1 + \varepsilon). \end{aligned} \quad (5.43)$$

Remind that $b_{N+1} = b_N(1 + M_{\mathcal{F}}/p_{N+1})$. By (5.42) and (5.43) we have for all $n > \max(n_0, n_2)$, $t_n \|\mathbb{P}_n^{*(N+1)} - P\|_{\mathcal{F}} \leq (Kb_{N+1} - b)(1 + \varepsilon)$ which proves Proposition 5.3.1 by induction.

Step 3. Denote $\beta_n^{(N)}(\mathcal{F})$ the process defined by $\beta_n^{(N)}(f) = \sqrt{n}(\mathbb{P}_n^{*(N)}(f) - \mathbb{P}_n^{(N)}(f))$. Notice that $\alpha_n^{*(N)} = \beta_n^{(N)} + \alpha_n^{(N)} - \alpha_n$. For all $\lambda > 0$,

$$\mathbb{P} \left(\|\alpha_n^{*(N)}\|_{\mathcal{F}} > 3\lambda \right) \leq \mathbb{P} \left(\|\alpha_n\|_{\mathcal{F}} > \lambda \right) + \mathbb{P} \left(\|\alpha_n^{(N)}\|_{\mathcal{F}} > \lambda \right) + \mathbb{P} \left(\|\beta_n^{(N)}\|_{\mathcal{F}} > \lambda \right). \quad (5.44)$$

Remind that $K_{\mathcal{F}} = \max(1, M_{\mathcal{F}})$. For $K \leq p_N$ and $K' \leq p_N - K$ we have

$$\begin{aligned}
& \mathbb{P} \left(\|\beta_n^{(N)}\|_{\mathcal{F}} > \lambda \right) \\
& \leq \mathbb{P} \left(\sqrt{n} \sum_{j=1}^{m_N} \left| \frac{\mathbb{P}_n(A_j^{(N)})}{\mathbb{P}_n^{*(N-1)}(A_j^{(N)})} \mathbb{P}_n^{*(N-1)}(f \mathbf{1}_{A_j^{(N)}}) - \frac{P(A_j^{(N)})}{\mathbb{P}_n^{(N-1)}(A_j^{(N)})} \mathbb{P}_n^{(N-1)}(f \mathbf{1}_{A_j^{(N)}}) \right| > \lambda \right) \\
& \leq \mathbb{P} \left(\sum_{j=1}^{m_N} \frac{\|\alpha_n\|_{\mathcal{F}} + \|\beta_n^{(N-1)}\|_{\mathcal{F}}}{\mathbb{P}_n^{*(N-1)}(A_j^{(N)}) \mathbb{P}_n^{(N-1)}(A_j^{(N)})} > \frac{\lambda}{2K_{\mathcal{F}}} \right) \\
& \leq \mathbb{P} \left(\|\alpha_n\|_{\mathcal{F}} > \frac{\lambda K^2}{4m_N K_{\mathcal{F}}} \right) + \mathbb{P} \left(\|\beta_n^{(N-1)}\|_{\mathcal{F}} > \frac{\lambda K^2}{4m_N K_{\mathcal{F}}} \right) \\
& \quad + \mathbb{P} \left(\min_{j=1, \dots, m_N} \mathbb{P}_n^{*(N-1)}(A_j^{(N)}) < K \right) + \mathbb{P} \left(\min_{j=1, \dots, m_N} \mathbb{P}_n^{(N-1)}(A_j^{(N)}) < K \right) \\
& \leq \mathbb{P} \left(\|\alpha_n\|_{\mathcal{F}} > \frac{\lambda K^2}{4m_N K_{\mathcal{F}}} \right) + \mathbb{P} \left(\|\beta_n^{(N-1)}\|_{\mathcal{F}} > \frac{\lambda K^2}{4m_N K_{\mathcal{F}}} \right) \\
& \quad + \mathbb{P} \left(\|\alpha_n^{(N-1)}\|_{\mathcal{F}} + \|\beta_n^{(N-1)}\|_{\mathcal{F}} > K' \sqrt{n} \right) + \mathbb{P} \left(\|\alpha_n^{(N-1)}\|_{\mathcal{F}} > K' \sqrt{n} \right) \\
& \leq \mathbb{P} \left(\|\alpha_n\|_{\mathcal{F}} > \frac{\lambda K^2}{4m_N K_{\mathcal{F}}} \right) + 2\mathbb{P} \left(\|\beta_n^{(N-1)}\|_{\mathcal{F}} > \frac{\lambda K^2}{4m_N K_{\mathcal{F}}} \right) + 2\mathbb{P} \left(\|\alpha_n^{(N-1)}\|_{\mathcal{F}} > \frac{\lambda K^2}{4m_N K_{\mathcal{F}}} \right),
\end{aligned}$$

where the last inequality holds provided that $K' \sqrt{n} \geq \lambda K^2 / 2m_N K_{\mathcal{F}}$. Define

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

Since $p_N \leq 1/2$ for any $N \geq 1$ it holds $K' > 0$ and $K' \sqrt{n} \geq \lambda K^2 / m_N(1 + K_{\mathcal{F}})$. By iteration, we obtain that

$$\begin{aligned}
\mathbb{P} \left(\|\beta_n^{(N)}\|_{\mathcal{F}} > \lambda \right) & \leq 2^{N-1} \mathbb{P} \left(\|\alpha_n\|_{\mathcal{F}} > \frac{\lambda P_N^2}{M_N(4K_{\mathcal{F}})^N(1 + \lambda/\sqrt{n})^{2N}} \right) \\
& \quad + 2^N \mathbb{P} \left(\|\beta_n^{(0)}\|_{\mathcal{F}} > \frac{\lambda P_N^2}{M_N(4K_{\mathcal{F}})^N(1 + \lambda/\sqrt{n})^{2N}} \right) \\
& \quad + 2^N \mathbb{P} \left(\max_{K=1, \dots, N} \|\alpha_n^{(K)}\|_{\mathcal{F}} > \frac{\lambda P_N^2}{M_N(4K_{\mathcal{F}})^N(1 + \lambda/\sqrt{n})^{2N}} \right). \tag{5.45}
\end{aligned}$$

Notice that $P_N^2 / M_N(4K_{\mathcal{F}})^N(1 + \lambda/\sqrt{n})^{2N} \leq 1$. By using (5.44), (5.45) and Proposition 3.2.4 we have, under the event that $\alpha_n^{(N)}$ are defined for every $0 \leq N \leq N_0$, that

$$\begin{aligned}
\mathbb{P} \left(\|\alpha_n^{*(N)}\|_{\mathcal{F}} > \lambda \right) & \leq 4^{N+1} N M_N \mathbb{P} \left(\|\alpha_n\|_{\mathcal{F}} > \frac{\lambda P_N^2}{3M_N(4K_{\mathcal{F}})^N(1 + \lambda/\sqrt{n})^{2N}} \right) \\
& \quad + 2^{N+1} \mathbb{P} \left(\|\beta_n^{(0)}\|_{\mathcal{F}} > \frac{\lambda P_N^2}{3M_N(4K_{\mathcal{F}})^N(1 + \lambda/\sqrt{n})^{2N}} \right). \tag{5.46}
\end{aligned}$$

Since $\beta_n^{(0)} = \alpha_n^*$ and according to Hoeffding inequality we have

$$\begin{aligned}
\mathbb{P}(\|\beta_n^{(0)}\|_{\mathcal{F}} > \lambda) & \leq \mathbb{P} \left(\left| \frac{n}{T_n} - 1 \right| > \frac{\lambda}{M_Z M_{\mathcal{F}} \sqrt{n}} \right) = \mathbb{P} \left(|T_n - n| > \frac{\lambda \sqrt{n}}{M_Z M_{\mathcal{F}}} \right) \\
& \leq 2 \exp \left(-\frac{\lambda^2}{2M_Z^4 M_{\mathcal{F}}^2} \right). \tag{5.47}
\end{aligned}$$

With (5.46) and (5.47) we have

$$\begin{aligned} \mathbb{P}\left(\|\alpha_n^{*(N)}\|_{\mathcal{F}} > \lambda\right) &\leq 4^{N+1}NM_N\mathbb{P}\left(\|\alpha_n\|_{\mathcal{F}} > \frac{\lambda P_N^2}{3M_N(4K_{\mathcal{F}})^N(1+\lambda/\sqrt{n})^{2N}}\right) \\ &\quad + 2^{N+2}\exp\left(-\frac{\lambda^2 P_N^4}{18M_N^2 M_Z^4 M_{\mathcal{F}}^2 (4K_{\mathcal{F}})^{2N}(1+\lambda/\sqrt{n})^{4N}}\right). \end{aligned}$$

The latter bound is increasing with N and $\mathbb{P}(\max_{0 \leq N \leq N_0} \|\alpha_n^{*(N)}\|_{\mathcal{F}} > \lambda) \leq \sum_{N=0}^{N_0} \mathbb{P}(\|\alpha_n^{*(N)}\|_{\mathcal{F}} > \lambda)$ which leads to (5.25). Inequality (5.26) is a consequence of Theorem 2.14.2 and 2.14.25 of van der Vaart and Wellner [80] whereas (5.27) is a consequence of Theorem 2.14.9 of [80].

5.4.3 Construction of the limit Gaussian processes

Bootstrapped Gaussian processes. Remind (5.33) and (5.34). The Gaussian processes leading the asymptotic behavior of the weighted bootstrap empirical processes are as follows. Let $\mathbb{G}^0(\mathcal{F}_1)$ be the $P^{(X,Z)}$ -Brownian bridge indexed by \mathcal{F}_1 . The bootstrapped Gaussian process $\mathbb{G}^*(\mathcal{F})$ and the standard P -Brownian bridge $\mathbb{G}(\mathcal{F})$ – which is actually in this section the margin process of $\mathbb{G}^0(\mathcal{F}_1)$ on $\mathcal{F} \subset \mathcal{F}_1$ – are defined to be

$$\mathbb{G}(f) = \mathbb{G}^0(h_f^{(3)}) = \mathbb{G}^0(f \times \mathbf{1}), \quad (5.48)$$

$$\mathbb{G}^*(f) = \mathbb{G}^0(h_f^{(1)}) - P(f)\mathbb{G}^0(h^{(2)}) - \mathbb{G}^0(h_f^{(3)}), \quad f \in \mathcal{F}. \quad (5.49)$$

By linearity one can extend \mathbb{G}^0 to $\{h_f : f \in \mathcal{F}\}$ by setting $\mathbb{G}^0(h_f) = \mathbb{G}^*(f)$. The following lemma gives the joint distribution of $\mathbb{G}^*(\mathcal{F})$ and $\mathbb{G}(\mathcal{F})$.

Lemme 5.4.2. *If (5.3) and (5.4) hold then $\mathbb{G}^*(\mathcal{F})$ and $\mathbb{G}(\mathcal{F})$ are independent P -Brownian bridges.*

Démonstration. By definition of $\mathbb{G}^0(\mathcal{F}_1)$, the two processes $\mathbb{G}^*(\mathcal{F})$ and $\mathbb{G}(\mathcal{F})$ are centered linear Gaussian, indexed by \mathcal{F} such that, for all $f, g \in \mathcal{F}$, $\text{Cov}(\mathbb{G}(f), \mathbb{G}(g)) = P^{(X,Z)}(h_f^{(3)} h_g^{(3)}) - P^{(X,Z)}(h_f^{(3)})P^{(X,Z)}(h_g^{(3)}) = P(fg) - P(f)P(g)$. By (5.4) we have $h_f(X, Z) = (Z-1)(f(X) - P(f)) - P(f)$ almost surely thus (5.3) implies $P^{(X,Z)}(h_f) = \mathbb{E}(Z(f(X) - P(f)) - f(X)) = \mathbb{E}(\mathbb{E}(Z|X)(f(X) - P(f))) - P(f) = -P(f)$. Since \mathbb{G}^0 is also linear and (5.3) implies $\mathbb{E}((Z-1)^2|X) = 1$ we get

$$\begin{aligned} \text{Cov}(\mathbb{G}^*(f), \mathbb{G}^*(g)) &= P^{(X,Z)}((h_f + P(f))(h_g + P(g))) \\ &= \mathbb{E}((Z-1)^2(f(X) - P(f))(g(X) - P(g))) \\ &= P(fg) - P(f)P(g). \end{aligned}$$

Moreover, for all $f, g \in \mathcal{F}$ we get, since (5.3) implies $\mathbb{E}(Z-1|X) = 0$,

$$\begin{aligned} \text{Cov}(\mathbb{G}^*(f), \mathbb{G}(g)) &= \text{Cov}(\mathbb{G}^0(h_f), \mathbb{G}^0(h_g^{(3)})) \\ &= P^{(X,Z)}((h_f + P(f))(h_g^{(3)} - P(g))) \\ &= \mathbb{E}((Z-1)(f(X) - P(f))(g(X) - P(g))) = 0, \end{aligned}$$

which proves the independence. □

Raked bootstrapped Gaussian process. Let $\mathbb{G}^{(N)*}(\mathcal{F})$ be the raked bootstrapped P -Brownian bridge defined recursively by :

$$\mathbb{G}^{*(0)}(f) = \mathbb{G}^*(f), \quad \mathbb{G}^{*(N+1)}(f) = \mathbb{G}^{*(N)}(f) - \sum_{j=1}^{m_{N+1}} \mathbb{E}[f|A_j^{(N+1)}] \mathbb{G}^{*(N)}(A_j^{(N+1)}), \quad (5.50)$$

where $\mathbb{G}^*(\mathcal{F})$ is defined by (5.49) through $\mathbb{G}^0(\mathcal{F}_1)$. The following lemma establishes the distribution of $\mathbb{G}^{*(N)}$ and the independence between this process and $\mathbb{G}^{(N)}(\mathcal{F})$.

Lemme 5.4.3. *If (5.3) and (5.4) hold then $\mathbb{G}^{*(N)}(\mathcal{F})$ and $\mathbb{G}^{(N)}(\mathcal{F})$ are independent Gaussian processes with same distribution.*

Démonstration. Let $f, g \in \mathcal{F}$. By Lemma 5.4.2 we have

$$\begin{aligned} \text{Cov}(\mathbb{G}^{*(0)}(f), \mathbb{G}^{(0)}(g)) &= 0, \\ \text{Cov}(\mathbb{G}^{*(0)}(f), \mathbb{G}^{*(0)}(g)) &= \text{Cov}(\mathbb{G}^{(0)}(f), \mathbb{G}^{(0)}(g)). \end{aligned}$$

If we assume that

$$\begin{aligned} \text{Cov}(\mathbb{G}^{*(N)}(f), \mathbb{G}^{(N)}(g)) &= 0, \\ \text{Cov}(\mathbb{G}^{*(N)}(f), \mathbb{G}^{*(N)}(g)) &= \text{Cov}(\mathbb{G}^{(N)}(f), \mathbb{G}^{(N)}(g)), \end{aligned}$$

for some $N > 0$ then by construction of $\mathbb{G}^{(N)}(\mathcal{F}), \mathbb{G}^{*(N)}(\mathcal{F})$ respectively defined by (5.6) and (5.50) we have necessary

$$\begin{aligned} \text{Cov}(\mathbb{G}^{*(N+1)}(f), \mathbb{G}^{(N+1)}(g)) &= 0, \\ \text{Cov}(\mathbb{G}^{*(N+1)}(f), \mathbb{G}^{*(N+1)}(g)) &= \text{Cov}(\mathbb{G}^{(N+1)}(f), \mathbb{G}^{(N+1)}(g)), \end{aligned}$$

which prove the lemma by induction the same distribution. \square

Bootstrapped raked Gaussian process. Let $\mathbb{G}^{0,(N)}(\mathcal{F}_1)$ be the $P^{(X,Z)}$ -raked Brownian bridge defined recursively as in (5.6) from $\mathbb{G}^{0,(0)} = \mathbb{G}^0$ and the following auxiliary information, for all $N > 0$ and $j = 1, \dots, m_N$,

$$\mathcal{A}^{0,(N)} = \{A_1^{0,(N)}, \dots, A_{m_N}^{0,(N)}\}, \quad A_j^{0,(N)} = A_j^{(N)} \times [0, M_Z]. \quad (5.51)$$

Notice that $P^{(X,Z)}(\mathcal{A}^{0,(N)}) = P(\mathcal{A}^{(N)})$. The bootstrapped raking-ratio Gaussian process $\mathbb{G}^{(N)*}(\mathcal{F})$ is defined through

$$\mathbb{G}^{(N)*}(f) = \mathbb{G}^{0,(N)}(h_f^{(1)}) - P(f) \mathbb{G}^{0,(N)}(h_f^{(2)}) - \mathbb{G}^{0,(N)}(h_f^{(3)}), \quad f \in \mathcal{F}. \quad (5.52)$$

By linearity, one can define $\mathbb{G}^{0,(N)}(h_f) = \mathbb{G}^{(N)*}(f)$. The following lemma gives the distribution of $\mathbb{G}^{(N)*}(\mathcal{F})$.

Lemme 5.4.4. *If (5.3) and (5.4) hold then for all $\mathbb{G}^{(N)*}(\mathcal{F})$ and $\mathbb{G}(\mathcal{F})$ are independent P -Brownian bridges.*

Démonstration. By (5.7), Lemma 5.4.2 and definition of $\mathbb{G}^{(N)*}(\mathcal{F})$ given by (5.52) and we have for all $f, g \in \mathcal{F}$,

$$\begin{aligned} \text{Cov}(\mathbb{G}^{(N)*}(f), \mathbb{G}^{(N)*}(g)) &= \text{Cov}(\mathbb{G}^{0,(N)}(h_f), \mathbb{G}^{0,(N)}(h_g)) \\ &= \text{Cov}(\mathbb{G}^0(h_f), \mathbb{G}^0(h_g)) - R_N(P^{(X,Z)}, h_f, h_g) \\ &= \text{Cov}(\mathbb{G}(f), \mathbb{G}(g)) - R_N(P^{(X,Z)}, h_f, h_g). \end{aligned}$$

Let $h_1, h_2 \in \mathcal{F}_1$ defined by $h_i(x, z) = f_i(x)g_i(z)$ for $i = 1, 2$. By the definition of R_N given by (5.8) we have $\Phi_k^{(N)}(P^{(X,Z)}, h) = \Phi_k^{(N)}(P, f_1)$ and $\text{Var}(\mathbb{G}[\mathcal{A}^{(N)}]) = \text{Var}(\mathbb{G}^0[\mathcal{A}^{0,(N)}])$ since for every $1 \leq j_1, j_2 \leq m_N$,

$$\begin{aligned} \text{Cov}(\mathbb{G}^0(A_{j_1}^{0,(N)}), \mathbb{G}^0(A_{j_2}^{0,(N)})) &= P^{(X,Z)}(A_{j_1}^{0,(N)} \cap A_{j_2}^{0,(N)}) - P^{(X,Z)}(A_{j_1}^{0,(N)})P^{(X,Z)}(A_{j_2}^{0,(N)}) \\ &= P(A_{j_1}^{(N)} \cap A_{j_2}^{(N)}) - P(A_{j_1}^{(N)})P(A_{j_2}^{(N)}) \\ &= \text{Cov}(\mathbb{G}(A_{j_1}^{(N)}), \mathbb{G}(A_{j_2}^{(N)})). \end{aligned}$$

These remarks lead to $R_N(P^{(X,Z)}, h_1, h_2) = R_N(P, f_1, f_2)$. Then

$$\begin{aligned} R_N(P^{(X,Z)}, h_f, h_g) &= R_N(P^{(X,Z)}, h_f^{(1)} - P(f)h^{(2)} - h_f^{(3)}, h_g^{(1)} - P(g)h^{(2)} - h_g^{(3)}) \\ &= -P(f)P(g)R_N(P, \mathbf{1}, \mathbf{1}). \end{aligned}$$

By using (5.7) and the fact that $\mathbb{G}(\mathbf{1}) = 0, \mathbb{G}^{(N)}(\mathbf{1}) = 0$ we have

$$R_N(P, \mathbf{1}, \mathbf{1}) = \text{Cov}(\mathbb{G}(\mathbf{1}), \mathbb{G}(\mathbf{1})) - \text{Cov}(\mathbb{G}^{(N)}(\mathbf{1}), \mathbb{G}^{(N)}(\mathbf{1})) = 0.$$

We have shown that $\text{Cov}(\mathbb{G}^{(N)*}(f), \mathbb{G}^{(N)*}(g)) = \text{Cov}(\mathbb{G}(f), \mathbb{G}(g))$. By Lemma 5.4.2 $\mathbb{G}^{0,(0)} = \mathbb{G}^*$ is independent of \mathbb{G} and if for all $f, g \in \mathcal{F}$,

$$\text{Cov}(\mathbb{G}^{(N)*}(f), \mathbb{G}(g)) = \text{Cov}(\mathbb{G}^{0,(N)}(h_f), \mathbb{G}(g)) = 0,$$

then

$$\text{Cov}(\mathbb{G}^{(N+1)*}(f), \mathbb{G}(g)) = \text{Cov}(\mathbb{G}^{0,(N+1)}(h_f), \mathbb{G}(g)) = 0,$$

since $\mathbb{G}^{0,(N+1)}(h_f)$ is a linear combination of elements of $\mathbb{G}^{0,(N)}(\mathcal{F}_1)$. By induction, we have proved the independence between $\mathbb{G}^{(N)*}(\mathcal{F})$ and $\mathbb{G}(\mathcal{F})$. \square

5.4.4 Proof of Theorem 5.2.2 and 5.2.5

At Step 1 we prove Proposition 5.4.5 below, by applying the strong approximation results of [12] to $P^{(X,Z)}$ and the relevant class of functions to approximate in turn the weighted bootstrap empirical process α_n^* . At Step 2 we deduce Theorem 5.2.2 from Proposition 5.4.5. At Step 3 the previous arguments are adapted for another appropriate empirical process and class, to derive Theorem 5.2.5.

Proposition 5.4.5. *Assume (5.3), (5.4) and either VC or BR. Fix any $\theta > 0$. There exists constants $C_\theta > 0, n_\theta > 0$ and sequences $\{(X_n, Z_n)\}$ of i.i.d. random variables with distribution $P^{(X,Z)}$ and a sequence $\{(\mathbb{G}_n(\mathcal{F}), \mathbb{G}_n^*(\mathcal{F}))\}$ of pairs of independent P-Brownian bridges, all of these sequences being defined on the same probability space, such that for all $n > n_\theta$,*

$$\mathbb{P}(\{\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} > C_\theta v_n\} \cup \{\|\tilde{\alpha}_n^* - \mathbb{G}_n^*\|_{\mathcal{F}} > C_\theta v_n\}) \leq \frac{1}{n^\theta}. \quad (5.53)$$

On the above probability space there also exists independent sequences $\{\mathbb{G}'_n(\mathcal{F})\}$ and $\{\mathbb{G}^{'}_n(\mathcal{F})\}$ of independent versions of $\mathbb{G}(\mathcal{F})$ such that, for all $n > n_\theta$,*

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \|\sqrt{k}\alpha_k - \sum_{i=1}^k \mathbb{G}'_i\|_{\mathcal{F}} > C_\theta v_n^-\right) \leq \frac{1}{u_n^\theta}, \quad (5.54)$$

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \|\sqrt{k}\tilde{\alpha}_k^* - \sum_{i=1}^k \mathbb{G}^{*'}_i\|_{\mathcal{F}} > C_\theta v_n^-\right) \leq \frac{1}{u_n^\theta}. \quad (5.55)$$

In the following steps we use the fact that by definition of v_n and v_n^- there exists $\varepsilon_0 \in (0, 1/2)$ such that $v_n^- n^{1/2-\varepsilon_0} \leq v_n n^{1/2-\varepsilon_0} < n^\alpha (\log n)^\beta$ for all n large enough, and some $\alpha \in (0, 1/2)$, $\beta \in \mathbb{R}$.

Step 1. Berthet and Mason [12] proved that one can construct a probability space on which the sequence of empirical processes $\alpha_n(\mathcal{F})$ can be defined together with a coupling sequence of P -Brownian bridges $\mathbb{G}_n(\mathcal{F})$ such that, almost surely, $\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} \leq C v_n$ for some $C > 0$ and all n large enough. By Lemma 5.4.1 one can apply Propositions 1,2 and Theorems 1,2 of [12] to $\alpha_n^{(X,Z)}(\mathcal{F}_1)$. There exists $C_0 = C_0(\theta) > 0$, $n_0 = n_0(\theta) > 0$ and a probability space where one can construct a sequence of independent random variables $\{(X_n, Z_n)\}$ distributed as $P^{(X,Z)}$ and two sequences $\{\mathbb{G}_n^0(\mathcal{F}_1)\}$, $\{\mathbb{G}_n^{0'}(\mathcal{F}_1)\}$ of independent $P^{(X,Z)}$ -Brownian bridge satisfying for all $n > n_0$,

$$\mathbb{P}\left(\|\alpha_n^{(X,Z)} - \mathbb{G}_n^0\|_{\mathcal{F}_1} > C_0 v_n\right) \leq \frac{1}{n^\theta} \quad (5.56)$$

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \|\sqrt{k} \alpha_k^{(X,Z)} - \sum_{i=1}^k \mathbb{G}_i^{0'}\|_{\mathcal{F}_1} > C_0 v_n^-\right) \leq \frac{1}{6u_n^\theta}. \quad (5.57)$$

We only prove (5.54) and (5.55) by using (5.57). We can adapt the following proof to show (5.53) by using (5.56). Let denote $\{\mathbb{G}_n(\mathcal{F})\}$ and $\{\mathbb{G}_n^*(\mathcal{F})\}$ the processes defined respectively by (5.48) and (5.49) both built through $\{\mathbb{G}_n^0(\mathcal{F}_1)\}$. According to Lemma 5.4.2, $\{\mathbb{G}_n(\mathcal{F})\}$, $\{\mathbb{G}_n^*(\mathcal{F})\}$ are independent P -Brownian bridges. Since $\alpha_k^{(X,Z)}(h_f^{(3)}) = \alpha_k(f)$, inequality (5.54) is satisfied for the first event as a special case of (5.56), and the second is controlled by the same argument as below. Write $C'_0 = C_0 \max(2, M_{\mathcal{F}})$. By (5.57) one can write that for all $n > n_0$,

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \|\sqrt{k} \tilde{\alpha}_k^* - \sum_{i=1}^k \mathbb{G}_i^*\|_{\mathcal{F}} > C'_0 v_n^-\right) \\ & \leq \mathbb{P}\left(\frac{1}{\sqrt{n}} \sup_{f \in \mathcal{F}} \max_{1 \leq k \leq n} \left| \sqrt{k} \alpha_k^{(X,Z)}(h_f^{(1)}) - \sum_{i=1}^k \mathbb{G}_i^{0'}(h_f^{(1)}) \right| > C'_0 v_n^-\right) \\ & \quad + \mathbb{P}\left(\frac{1}{\sqrt{n}} \sup_{f \in \mathcal{F}} \max_{1 \leq k \leq n} \left| P(f) \left(\sqrt{k} \alpha_k^{(X,Z)}(h_f^{(2)}) - \sum_{i=1}^k \mathbb{G}_i^{0'}(h_f^{(2)}) \right) \right| > C'_0 v_n^-\right) \\ & \quad + \mathbb{P}\left(\frac{1}{\sqrt{n}} \sup_{f \in \mathcal{F}} \max_{1 \leq k \leq n} \left| \sqrt{k} \alpha_k^{(X,Z)}(h_f^{(3)}) - \sum_{i=1}^k \mathbb{G}_i^{0'}(h_f^{(3)}) \right| > C'_0 v_n^-\right) \\ & \quad + \mathbb{P}\left(\frac{1}{\sqrt{n}} \sup_{f \in \mathcal{F}} \max_{1 \leq k \leq n} \left| \sqrt{k} \alpha_k^{(X,Z)}(h_f^{(3)}) \cdot \frac{1}{k} \sum_{i=1}^k (Z_i - 1) \right| > C'_0 v_n^-\right) \\ & \leq \frac{1}{2u_n^\theta} + \mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \left[\|\sqrt{k} \alpha_k^{(X,Z)}\|_{\mathcal{F}_1} \cdot \left| \frac{1}{k} \sum_{i=1}^k (Z_i - 1) \right| \right] > C'_0 v_n^-\right). \end{aligned} \quad (5.58)$$

The last right-hand side member of (5.58) can be bounded by

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \left[\|\alpha_k^{(X,Z)}\|_{\mathcal{F}_1} \cdot \left| \frac{1}{\sqrt{k}} \sum_{i=1}^k (Z_i - 1) \right| \right] > C'_0 v_n^-\right) \\ & \leq \mathbb{P}\left(\max_{1 \leq k \leq n} \|\alpha_k^{(X,Z)}\|_{\mathcal{F}_1} > 2C_0 v_n^- n^{1/2-\varepsilon_0}\right) + \mathbb{P}\left(\max_{1 \leq k \leq n} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^k (Z_i - 1) \right| > n^{\varepsilon_0}\right). \end{aligned} \quad (5.59)$$

By Hoeffding inequality and the fact that $\varepsilon_0 \in (0, 1/2)$ there exists $n_1 = n_1(\theta) > 0$ such that for all $n > n_1$,

$$\mathbb{P} \left(\max_{1 \leq k \leq n} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^k (Z_i - 1) \right| > n^{\varepsilon_0} \right) \leq 2n \exp \left(-\frac{2n^{2\varepsilon_0}}{M_Z^2} \right) \leq \frac{1}{6u_n^\theta}. \quad (5.60)$$

Moreover by (5.57) it holds, for all $n > n_0$,

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq k \leq n} \|\alpha_k^{(X,Z)}\|_{\mathcal{F}_1} > 2C_0 v_n^- n^{1/2-\varepsilon_0} \right) \\ & \leq \mathbb{P} \left(\max_{1 \leq k \leq n} \|\alpha_k^{(X,Z)} - \frac{1}{\sqrt{k}} \sum_{i=1}^k \mathbb{G}_i^{0'}\|_{\mathcal{F}_1} > C_0 v_n^- n^{1/2-\varepsilon_0} \right) \\ & \quad + \mathbb{P} \left(\max_{1 \leq k \leq n} \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^k \mathbb{G}_i^{0'} \right\|_{\mathcal{F}_1} > C_0 v_n^- n^{1/2-\varepsilon_0} \right) \\ & \leq \frac{1}{6u_n^\theta} + \mathbb{P} \left(\max_{1 \leq k \leq n} \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^k \mathbb{G}_i^{0'} \right\|_{\mathcal{F}_1} > C_0 v_n^- n^{1/2-\varepsilon_0} \right), \end{aligned} \quad (5.61)$$

Write $\mathbb{G}_{(k)}^{0'} = \frac{1}{\sqrt{k}} \sum_{i=1}^k \mathbb{G}_i^{0'}$ for short, so that $\mathbb{G}_{(k)}^{0'}(\mathcal{F}_1)$ is a $P^{(X,Z)}$ -Brownian bridge. By Lemma 5.4.1, \mathcal{F}_1 has small entropy, bounded variance so that we deduce that $\mathbb{E}(\|\mathbb{G}_{(k)}^{0'}\|_{\mathcal{F}_1}) < K\sqrt{\log n} < C_0 v_n^- n^{1/2-\varepsilon_0}$ for a finite $K > 0$ and all n large enough. Thus, by Borell-TIS inequality – see annex A.2.1 of [80] – there exists $n_2 = n_2(\theta) > 0$ such that for all $n > n_2$, some $C_0'' < C_0$,

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq k \leq n} \|\mathbb{G}_{(k)}^{0'}\|_{\mathcal{F}_1} > C_0 v_n^- n^{1/2-\varepsilon_0} \right) \\ & \leq \sum_{k=1}^n \mathbb{P} \left(\|\mathbb{G}_{(k)}^{0'}\|_{\mathcal{F}_1} - \mathbb{E}(\|\mathbb{G}_{(k)}^{0'}\|_{\mathcal{F}_1}) > C_0'' v_n^- n^{1/2-\varepsilon_0} \right) \\ & \leq 2n \exp \left(-\frac{(C_0'' v_n^- n^{1/2-\varepsilon_0})^2}{2M_Z^2 \sigma_{\mathcal{F}_1}^2} \right) \leq \frac{1}{6u_n^\theta}. \end{aligned} \quad (5.62)$$

The bounds (5.60), (5.61) and (5.62) imply that for all $n > \max(n_0, n_1)$,

$$\mathbb{P} \left(\frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \left[\|\sqrt{k} \alpha_k^{(X,Z)}\|_{\mathcal{F}_1} \cdot \left| \frac{1}{k} \sum_{i=1}^k (Z_i - 1) \right| \right] > C_\theta' v_n \right) \leq \frac{1}{2u_n^\theta}$$

then (5.55) follows from (5.58) with $n_\theta = \max(n_0, n_1, n_2)$ and $C_\theta = C_\theta'$.

Step 2. Now we prove (5.12) and (5.13) of Theorem 5.2.2. By adapting the following proof we can also show (5.11). According to Proposition 5.4.5, there exists a constant $C_0 = C_0(\theta)$, $n_0 = n_0(\theta) > 0$, a sequence $\{(X_n, Z_n)\}$ of i.i.d. random variables with distribution $P^{(X,Z)}$, a sequence $\{\mathbb{G}'_n(\mathcal{F})\}$, $\{\mathbb{G}^{*'}_n(\mathcal{F})\}$ of independent P -brownian bridge satisfying, for all $n > n_0$, (5.12) and

$$\mathbb{P} \left(\frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \|\sqrt{k} \hat{\alpha}_k^* - \sum_{i=1}^k \mathbb{G}_i^{*'}\|_{\mathcal{F}} > C_0 v_n \right) \leq \frac{1}{4n^\theta}. \quad (5.63)$$

By the definition of $\tilde{\alpha}_n^*$ given at (5.32),

$$\begin{aligned}
& \mathbb{P} \left(\frac{1}{\sqrt{n}} \max_{k_{n,\theta} \leq k \leq n} \|\sqrt{k}\alpha_k^* - \sum_{i=1}^k \mathbb{G}_i^{*'}\|_{\mathcal{F}} > 3C_0v_n \right) \\
&= \mathbb{P} \left(\frac{1}{\sqrt{n}} \max_{k_{n,\theta} \leq k \leq n} \left\| \frac{k}{T_k} \sqrt{k}\tilde{\alpha}_k^* - \sum_{i=1}^k \mathbb{G}_i^{*'} \right\|_{\mathcal{F}} > 3C_0v_n \right) \\
&\leq \mathbb{P} \left(\frac{1}{\sqrt{n}} \max_{k_{n,\theta} \leq k \leq n} \left\| \frac{k}{T_k} (\sqrt{k}\tilde{\alpha}_k^* - \sum_{i=1}^k \mathbb{G}_i^{*'}) \right\|_{\mathcal{F}} > 2C_0v_n \right) \\
&\quad + \mathbb{P} \left(\frac{1}{\sqrt{n}} \max_{k_{n,\theta} \leq k \leq n} \left\| \left(\frac{k}{T_k} - 1 \right) \sum_{i=1}^k \mathbb{G}_i^{*'} \right\|_{\mathcal{F}} > C_0v_n \right). \tag{5.64}
\end{aligned}$$

According to Hoeffding inequality there exists $n_1 = n_1(\theta) > 0$ such that for all $n > n_1$,

$$\begin{aligned}
\mathbb{P} \left(\max_{k_{n,\theta} \leq k \leq n} \frac{k}{T_k} \geq 2 \right) &\leq \sum_{k=k_{n,\theta}}^n \mathbb{P} \left(\frac{T_k}{k} \leq \frac{1}{2} \right) \leq \sum_{k=k_{n,\theta}}^n \mathbb{P} (|T_k - k| \geq k/2) \\
&\leq 2 \sum_{k=k_{n,\theta}}^n \exp \left(-\frac{2(k^2/4)}{kM_Z^2} \right) \leq \frac{1}{4n^\theta}. \tag{5.65}
\end{aligned}$$

Combining (5.63) and (5.65), we obtain that for all $n > \max(n_0, n_1)$,

$$\mathbb{P} \left(\frac{1}{\sqrt{n}} \max_{k_{n,\theta} \leq k \leq n} \left\| \frac{k}{T_k} (\sqrt{k}\tilde{\alpha}_k^* - \sum_{i=1}^k \mathbb{G}_i^{*'}) \right\|_{\mathcal{F}} > 2C_0v_n \right) \leq \frac{1}{2n^\theta}. \tag{5.66}$$

Using again the definition of ε_0 we see that

$$\begin{aligned}
& \mathbb{P} \left(\frac{1}{\sqrt{n}} \max_{k_{n,\theta} \leq k \leq n} \left\| \left(\frac{k}{T_k} - 1 \right) \sum_{i=1}^k \mathbb{G}_i^{*'} \right\|_{\mathcal{F}} > C_0v_n \right) \\
&\leq \mathbb{P} \left(n^{-1/2+\varepsilon_0} \max_{k_{n,\theta} \leq k \leq n} \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^k \mathbb{G}_i^{*'} \right\|_{\mathcal{F}} > C_0v_n \right) \\
&\quad + \mathbb{P} \left(\max_{k_{n,\theta} \leq k \leq n} \left| \sqrt{k} \left(\frac{k}{T_k} - 1 \right) \right| > n^{\varepsilon_0} \right). \tag{5.67}
\end{aligned}$$

With the same calculations as (5.62) – but on the class \mathcal{F} – we show that there exists $n_2 = n_2(\theta) > 0$ such that for all $n > n_2$:

$$\mathbb{P} \left(n^{-1/2+\varepsilon} \max_{k_{n,\theta} \leq k \leq n} \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^k \mathbb{G}_i^{*'} \right\|_{\mathcal{F}} > C_0v_n \right) \leq \frac{1}{4n^\theta}, \tag{5.68}$$

By (5.65) and Hoeffding inequality, there exists $n_3 = n_3(\theta) > 0$ such that for all $n > n_3$,

$$\begin{aligned}
& \mathbb{P} \left(\max_{k_{n,\theta} \leq k \leq n} \left| \sqrt{k} \left(\frac{k}{T_k} - 1 \right) \right| > n^{\varepsilon_0} \right) = \mathbb{P} \left(\max_{k_{n,\theta} \leq k \leq n} \left| \frac{k}{T_k} \cdot \frac{T_k - k}{\sqrt{k}} \right| > n^{\varepsilon_0} \right) \\
&\leq \mathbb{P} \left(\max_{1 \leq k \leq n} \left| \frac{T_k - k}{\sqrt{k}} \right| > \frac{n^{\varepsilon_0}}{2} \right) + \mathbb{P} \left(\max_{k_{n,\theta} \leq k \leq n} \frac{k}{T_k} > 2 \right) \leq \frac{1}{4n^\theta}. \tag{5.69}
\end{aligned}$$

By (5.64), (5.66), (5.67), (5.68) and (5.69), we have proved (5.13) with $n_\theta = \max(n_0, n_1, n_2, n_3)$ and $C_\theta = 3C_0$. Note that we needed that $n \geq \max(n_\theta, k_{n,\theta})$, where $k_{n,\theta} = \lceil 2M_Z^2(\ln 8 + (1 + \theta) \ln n) \rceil$ which is true for all $n > n_\theta$, n_θ large.

Step 3. Let $P^{(X, Z_1, \dots, Z_{b_n})}$ be the joint probability distribution of (X, Z_1, \dots, Z_{b_n}) where X has marginal P then each Z_j has conditional marginal $P^{Z|X}$ and is independent of $Z_{j'}$ conditionally to X , for $0 < j' < j \leq b_n$. Consider the class defined by

$$\mathcal{F}_{b_n} = \{h_j(f, g)(x, z_1, \dots, z_{b_n}) = f(x)g(z_j) : 1 \leq j \leq b_n, f \in \mathcal{F} \cup \{\mathbf{1}\}, g \in \{\phi_{M_Z}, \mathbf{1}\}\}.$$

By (5.4) \mathcal{F}_{b_n} is uniformly bounded by $M_Z \max(1, M_{\mathcal{F}})$ and by (5.3) we have $\sigma_{\mathcal{F}_{b_n}} = \sigma_{\mathcal{F}}$. The pointwise measurable property is justified as follows. For $j \leq b_n$, rewrite $h_j(f, g) \in \mathcal{F}_{b_n}$ into $f(x)g_j(z)$ with $f \in \mathcal{F} \cup \{\mathbf{1}\}$, $g_j = g \circ \pi_j$, $g \in \{\mathbf{1}, \phi_M\}$ and $\pi_j(z) = z_j$ is the j -th unit projector, $z \in \mathbb{R}^{b_n}$. Given $(x, z) \in \mathcal{X} \times \mathbb{R}^{b_n}$, if $f(x)$ is the limit of $\{f_n(x)\}$ with $f_n \in \tilde{\mathcal{F}}$ then $h_j(f, g)$ is the limit of $h_j(f_n, g)$ where $(f_n, g_j) \in \tilde{\mathcal{F}} \times \bigcup_{j=1}^{b_n} \{\mathbf{1}, \phi_M \circ \pi_j\}$ which is a countable family. These properties are inherited by the recombined class

$$\mathcal{H}_{b_n} = \{h_j(f) = h_j^{(1)}(f) - P(f)h_j^{(2)}(f) - h_j^{(3)}(f), 1 \leq j \leq b_n, f \in \mathcal{F}\}$$

where $h_j^{(1)}(f) = h_j(f, \phi_{M_Z})$, $h_j^{(2)}(f) = h_j(\mathbf{1}, \phi_{M_Z})$, $h_j^{(3)}(f) = h_j(f, \mathbf{1})$, that is

$$\begin{aligned} h_j^{(1)}(f)(x, z_1, \dots, z_{b_n}) &= f(x)\phi_{M_Z}(z_j), & h_j^{(2)}(f)(x, z_1, \dots, z_{b_n}) &= \phi_{M_Z}(z_j), \\ h_j^{(3)}(f)(x, z_1, \dots, z_{b_n}) &= f(x). \end{aligned}$$

We adapt the proof of Propositions 1 and 2 in [12] according to the following strategy. Consider a sequence $\varepsilon_n \rightarrow 0$, to be chosen at the end, and write $N(\varepsilon_n)$ the covering number in $L_2(P)$ driven by (VC) or (BR). We control the relevant indexing class \mathcal{H}_{b_n} by a covering of \mathcal{F} used simultaneously for the b_n bootstrap samples. The centers of balls used to cover \mathcal{F} are denoted f_k , for $k = 1, \dots, N(\varepsilon_n)$. We associate to each center f_k the b_n functions $h_j(f_k)(x, z)$, for $j = 1, \dots, b_n$ and also $h_0(f_k)(x, z) = f_k(x)$. For any fixed $j = 0, \dots, b_n$ the centers $h_j(f_k)$, for $k = 1, \dots, N(\varepsilon_n)$ are then used to construct the P -Brownian bridge $\mathbb{G}_{n,(j)}^*$ indexed by the functions $h_j(f)$ and thus by \mathcal{F} . It suffices to determine $\mathbb{G}_{n,(j)}^*(h_j(f_k))$ close to $\alpha_{n,(j)}^*$ then to extend it by using a union bound for the modulus of continuity of $\mathbb{G}_{n,(0)}^*$ which is their common distribution. The extension is built independently as j varies since the processes $\mathbb{G}_{n,(j)}^*$ for $j = 0, 1, \dots, b_n$ are orthogonal. More precisely, consider the column vector Y defined by $Y_{j,k} = h_j(f_k)(X, Z_1, \dots, Z_{b_n}) + P(f_k)$ indexed by (j, k) with j fixed first then running increasingly among the indexes k . Thus $Y = Y(X, Z_1, \dots, Z_{b_n})$ is a centered random vector of size $(b_n + 1)N(\varepsilon_n)$. Clearly, thanks to (5.4), as already seen in the proof of Lemma 5.4.1, any function $h_j(f) \in \mathcal{F}_{b_n}$ is at $L_2(P^{(X, Z_1, \dots, Z_{b_n})})$ distance $C\varepsilon_n$ of one of the $N(\varepsilon_n)$ centers $h_j(f_k)$ at least, for some $C > 0$. The covariance matrix $\Sigma_n = (\text{Cov}(Y_{j,k}, Y_{j',k'})_{(j,k),(j',k')})$ of Y is very sparse since it is given by

$$\text{Cov}(h_j(f_k)(X, Z_1, \dots, Z_{b_n}), h_{j'}(f_{k'})(X, Z_1, \dots, Z_{b_n})) = 0$$

if $j \neq j'$ and, recalling Lemma 5.4.2,

$$\text{Cov}(h_j(f_k)(X, Z_1, \dots, Z_{b_n}), h_j(f_{k'})(X, Z_1, \dots, Z_{b_n})) = \text{Cov}(f_k(X), f_{k'}(X))$$

otherwise. Next, on the same probability space, we approximate the *i.i.d.* sum

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i, \quad Y_i = Y(X_i, Z_{i,(1)}, \dots, Z_{i,(b_n)})$$

by a coupled Gaussian vector U with the same covariance matrix Σ_n as Y , thanks to Zaitzev's construction. For any k the coordinates $U_{j,k}$ for $k = 1, \dots, N(\varepsilon_n)$ are used to jointly define pointwise the P -Brownian Bridges $\mathbb{G}_{n,(j)}^*(f_k) = U_{j,k}$. Now by replacing everywhere in [12] the vector size $N(\varepsilon)$ with the current one $(b_n + 1)N(\varepsilon_n)$ that we write $b_n N(\varepsilon_n)$ by changing constants, we readily obtain the probability bound

$$\begin{aligned} & \mathbb{P} \left(\bigcup_{j=0}^{b_n} \left\{ \|\alpha_{n,(j)}^* - \mathbb{G}_{n,(j)}^*\|_{\mathcal{F}} > e_{n,(j)}^* + g_{n,(j)}^* + \delta + t \right\} \right) \\ & \leq b_n^2 N(\varepsilon_n)^2 \exp \left(-\frac{c_1 \delta \sqrt{n}}{b_n^{5/2} N(\varepsilon_n)^{5/2}} \right) + b_n \exp(-c_2 t \sqrt{n}) + b_n \exp \left(-\frac{c_3 t^2}{\varepsilon_n^2} \right) \end{aligned}$$

where, for independent Rademacher random variables $\epsilon_{i,(j)}$,

$$e_{n,(j)}^* = \mathbb{E} \left(\sup_{f,g \in \mathcal{F}, d_P(f,g) < \varepsilon_n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_{i,(j)} (h_j(f) - h_j(g))(X_i, Z_{i,(1)}, \dots, Z_{i,(b_n)}) \right| \right)$$

and

$$g_{n,(j)}^* = \mathbb{E} \left(\sup_{f,g \in \mathcal{F}, d_P(f,g) < \varepsilon_n} \left| \mathbb{G}_{n,(j)}^*(h_j(f)) - \mathbb{G}_{n,(j)}^*(h_j(g)) \right| \right)$$

are the suprema of the increments of length ε_n at most, for the symmetrized empirical process $\alpha_{n,(j)}^*$ and the Gaussian process $\mathbb{G}_{n,(j)}^*$, respectively. Therefore the same concentration inequalities as in [12] under (VC) or (BR), moment bounds and discussions to optimize the parameters t and δ can be reproduced by turning n into n/b_n^5 in the most critical first exponential term above, and remarking that only constants change through, at times, the fact that $1 < \log(n/b_n^5) < \log n$. This is why the approximation rates (5.14) and (5.15) finally show up. By construction the initial and bootstrap Gaussian processes thus built have the desired covariance structure – they are independent P -Brownian Bridges indexed by \mathcal{F} .

5.4.5 Proof of Theorem 5.3.3 and 5.3.4

The process $\alpha_n^{*(N)}(\mathcal{F})$ could be seen as the process $\alpha_n^{*(N-1)}(\mathcal{F})$ corrected with partition $\mathcal{A}^{(N)}$ since

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

where $\mathbb{E}_n[f | A] = \frac{\mathbb{P}(f 1_A)}{\mathbb{P}_n(A)}$. We denote the following random variables :

$$p_{n,j}^{(N)} = \frac{\mathbb{P}_n(A_j^{(N)})}{\mathbb{P}_n^{*(N-1)}(A_j^{(N)})} - 1, \quad q_{n,j}^{(N)}(f) = \mathbb{E}_n[f | A_j^{(N)}] - \mathbb{E}[f | A_j^{(N)}].$$

Then,

$$\begin{aligned}\alpha_n^{*(N)}(f) &= \alpha_n^{*(N-1)} \circ \phi_{(N)}(f) + \Gamma_n^{(N)}(f), \\ \Gamma_n^{(N)}(f) &= \sum_{j=1}^{m_N} p_{n,j}^{(N)} \alpha_n^{*(N-1)} \circ \phi_{(N),j}(f) - (1 + p_{n,j}^{(N)}) q_{n,j}^{(N)}(f) \alpha_n^{*(N-1)}(A_j^{(N)}).\end{aligned}$$

By successive iterations we get

$$\begin{aligned}\alpha_n^{*(N)}(f) &= \alpha_n^{*(0)} \circ \phi_{(1)} \circ \cdots \circ \phi_{(N)}(f) + F_n^{(N)}(f), \\ F_n^{(N)}(f) &= \sum_{k=1}^N \Gamma_n^{(k)}(\phi_{(k+1)} \circ \cdots \circ \phi_{(N)}(f)).\end{aligned}\tag{5.70}$$

Let introduce the same notations than the Chapter 3. For all $f \in \mathcal{F}$, $N \leq N_0$ for some fixed N_0 and $j = 1, \dots, m_N$ we denote

$$\phi_{(N,j)}(f) = (f - \mathbb{E}[f | A_j^{(N)}]) 1_{A_j^{(N)}}, \quad \phi_{(N)}(f) = \sum_{j=1}^{m_N} \phi_{(N,j)}(f),$$

and function classes

$$\begin{aligned}\mathcal{F}_{(N)} &= \phi_{(1)} \circ \cdots \circ \phi_{(N)}(\mathcal{F}), \\ \mathcal{H}_{(N)} &= \bigcup_{1 \leq k \leq N} \bigcup_{1 \leq j \leq m_k} \phi_{(j,k)} \circ \phi_{(k+1)} \circ \cdots \circ \phi_{(N)}(\mathcal{F}), \\ F_0 &= \bigcup_{0 \leq N \leq N_0} F_{(N)}, \quad \mathcal{H}_0 = \bigcup_{0 \leq N \leq N_0} \mathcal{H}_{(N)}.\end{aligned}$$

The following lemma establishes that the bootstrapped process $\alpha_n^*(\mathcal{F}_{(N)})$ is the main process contributing for $\alpha_n^{*(N)}(\mathcal{F})$ and $F_n^{(N)}$ is an error process. It is the equivalent of Lemma 3.3.6.

Lemma 5.4.6. *Assume (5.3), (5.4) and either VC or BR. For any $\xi, \theta > 0$ there exists $n_{\xi, \theta} > 0$ such that, for all $n > n_{\xi, \theta}$,*

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

Fix $\theta, \xi > 0$. Now we prove Lemma 5.4.6 at Step 1 and apply it at Step 2 to prove Theorem 5.3.3. In Step 3, we describe how to prove Theorem 5.3.4.

Step 1. Let bound the terms of error :

$$\begin{aligned}|p_{n,j}^{(N)}| &= \left| \frac{\mathbb{P}_n(A_j^{(N)}) - \mathbb{P}_n^{*(N-1)}(A_j^{(N)})}{\mathbb{P}_n^{*(N-1)}(A_j^{(N)})} \right| \leq \frac{\|\alpha_n\|_{\mathcal{F}} + \|\alpha_n^{*(N-1)}\|_{\mathcal{F}}}{\sqrt{n} p_N - \|\alpha_n^{*(N-1)}\|_{\mathcal{F}}}, \\ \|q_{n,j}^{(N)}\|_{\mathcal{F}} &= \left\| \frac{\mathbb{P}_n(f 1_{A_j^{(N)}}) - \mathbb{E}[f | A_j^{(N)}] \mathbb{P}_n(A_j^{(N)})}{\mathbb{P}_n(A_j^{(N)})} \right\|_{\mathcal{F}} \leq \frac{2M_{\mathcal{F}} \|\alpha_n\|_{\mathcal{F}}}{\sqrt{n} p_N - \|\alpha_n\|_{\mathcal{F}}}, \\ \|F_n^{(N)}\|_{\mathcal{F}} &\leq \sum_{k=1}^N \max_{1 \leq j \leq m_k} |p_{n,j}^{(k)}| \sum_{j=1}^{m_k} |\alpha_n^{*(k-1)} \circ \phi_{k,j} \circ \phi_{(k+1)} \circ \cdots \circ \phi_{(N)} f| \\ &\quad + \max_{1 \leq j \leq m_k} \left(|1 + p_{n,j}^{(k)}| |q_{n,j}^{(k)}| |\alpha_n^{*(k-1)}(A_j)| \right) \\ &\leq S_N \max_{0 \leq k \leq N-1} \|\alpha_n^{*(k)}\|_{\mathcal{H}_0} \left(P_n^{(N)} + (1 + P_n^{(N)}) Q_n^{(N)} \right), \\ \max_{1 \leq N \leq N_0} \|F_n^{(N)}\|_{\mathcal{F}} &\leq S_{N_0} \max_{0 \leq N \leq N_0-1} \|\alpha_n^{*(N)}\|_{\mathcal{H}_0} \left(P_n^{(N_0)} + Q_n^{(N_0)} + P_n^{(N_0)} Q_n^{(N_0)} \right),\end{aligned}$$

where $P_n^{(N)} = \max_{0 \leq k \leq N} \max_{1 \leq j \leq m_k} |p_{n,j}^{(k)}|$, $Q_n^{(N)} = \max_{0 \leq k \leq N} \max_{1 \leq j \leq m_k} |q_{n,j}^{(k)}|$. For $\varepsilon \in (0, 1/2)$ there exists $n_\xi > 0$ such that for all $n > n_\xi$, $\xi v_n > 4S_{N_0}(1 + 2M_{\mathcal{F}})n^\varepsilon/(\sqrt{np_{N_0}})$. Then, for every $n > n_\xi$ we have

$$\begin{aligned} \mathbb{P} \left(\max_{0 \leq N \leq N_0} \|F_n^{(N)}\|_{\mathcal{F}} > \xi v_n \right) &\leq \mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\alpha_n^{*(N)}\|_{\mathcal{H}_0} > n^{\varepsilon/2} \right) \\ &+ 2\mathbb{P} \left(P_n^{(N_0)} > \frac{4n^{\varepsilon/4}}{\sqrt{np_{N_0}}} \right) + 2\mathbb{P} \left(Q_n^{(N_0)} > \frac{4M_{\mathcal{F}}n^{\varepsilon/4}}{\sqrt{np_{N_0}}} \right). \end{aligned} \quad (5.71)$$

According to Lemma 3.3.5 and Proposition 5.3.2 there exists $n_0 = n_0(\theta) > 0$ such that for all $n > n_0$ we have

$$\mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\alpha_n^{*(N)}\|_{\mathcal{H}_0} > n^{\varepsilon/2} \right) \leq \frac{1}{3n^\theta}. \quad (5.72)$$

Moreover there exists $n_1 = n_1(\theta) > 0$ such that for all $n > n_1$,

$$\begin{aligned} \mathbb{P} \left(P_n^{(N_0)} > \frac{4n^{\varepsilon/4}}{\sqrt{np_{N_0}}} \right) &\leq \mathbb{P} \left(\|\alpha_n\|_{\mathcal{F}} > n^{\varepsilon/4} \right) + \mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\alpha_n^{*(N)}\|_{\mathcal{F}} > n^{\varepsilon/4} \right) \\ &+ \mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\alpha_n^{*(N)}\|_{\mathcal{F}} > \frac{\sqrt{np_{N_0}}}{2} \right) \leq \frac{1}{6n^\theta}, \end{aligned} \quad (5.73)$$

$$\begin{aligned} \mathbb{P} \left(Q_n^{(N_0)} > \frac{4M_{\mathcal{F}}n^{\varepsilon/4}}{\sqrt{np_{N_0}}} \right) &\leq \mathbb{P} \left(\|\alpha_n\|_{\mathcal{F}} > n^{\varepsilon/4} \right) + \mathbb{P} \left(\|\alpha_n\|_{\mathcal{F}} > \frac{\sqrt{np_{N_0}}}{2} \right) \\ &\leq \frac{1}{6n^\theta} \end{aligned} \quad (5.74)$$

By (5.71), (5.72), (5.73) and (5.74), Lemma 5.4.6 is proved with $n_{\xi,\theta} = \max(n_\xi, n_0, n_1)$.

Step 2. By Lemma 3.3.5 we can apply Lemma 5.4.6, Theorem 5.2.2 and Theorem 3.3.7 to \mathcal{F}_0 . Since these two last theorems are based on the Berthet and Mason strong approximation of $\alpha_n(\mathcal{F})$ to $\mathbb{G}(\mathcal{F})$, the following Gaussian approximation is satisfied. For some constant $C_0 = C_0(\theta) > 0$ and $n_\theta > 0$ one can build on a probability space $(\Omega, \mathcal{T}, \mathbb{P})$ a version of the sequence $\{(X_n, Z_n)\}$ of i.i.d. random variables with distribution $P^{(X,Z)}$ and a sequence $\{\mathbb{G}_n^*(\mathcal{F})\}$ of independent P -Brownian bridges in such a way that, that for all $n > n_\theta$, (5.29) holds and

$$\mathbb{P} \left(\{ \|\alpha_n^* - \mathbb{G}_n^*\|_{\mathcal{F}_0} > C_0 v_n \} \cup \left\{ \max_{0 \leq N \leq N_0} \|F_n^{(N)}\|_{\mathcal{F}} > C_0 v_n \right\} \right) \leq \frac{1}{2n^\theta}. \quad (5.75)$$

Let denote $\{\mathbb{G}_n^{*(N)}(\mathcal{F})\}$ the N -th raked P -Brownian bridge defined by (5.50) built through $\{\mathbb{G}_n^*\}$. This construction implies $\mathbb{G}_n^{*(0)} = \mathbb{G}_n^*$. Successive iterations given by (5.70) give $\max_{1 \leq N \leq N_0} \|\alpha_n^{*(N)} - \mathbb{G}_n^{*(N)}\| \leq \|\alpha_n^* - \mathbb{G}_n^*\|_{\mathcal{F}_0} + \max_{0 \leq N \leq N_0} \|F_n^{(N)}\|_{\mathcal{F}}$. By (5.75) we have for all $n > n_\theta$,

$$\begin{aligned} &\mathbb{P} \left(\max_{1 \leq N \leq N_0} \|\alpha_n^{*(N)} - \mathbb{G}_n^{*(N)}\|_{\mathcal{F}} > 2C_0 v_n \right) \\ &\leq \mathbb{P} \left(\|\alpha_n^* - \mathbb{G}_n^*\|_{\mathcal{F}_0} > C_0 v_n \right) + \mathbb{P} \left(\max_{0 \leq N \leq N_0} \|F_n^{(N)}\|_{\mathcal{F}} > C_0 v_n \right) \leq \frac{1}{n^\theta}. \end{aligned}$$

We have proved (5.28) of Theorem 5.3.3 with $C_\theta = 2C_0 > 0$.

Step 3. To prove Theorem 5.3.4 we only need to extend the notations of previous steps and use Theorem 5.2.5 instead of Theorem 5.2.2 at Step 2.

5.4.6 Proof of Propositions 5.2.8, 5.2.9 and 5.2.10

Assume (FR) and either VC or BR. Introduce, on the probability space of Theorem 5.2.5, for $j = 1, \dots, b_n$,

$$R_{n,0} = \sqrt{n}R \left(\frac{\alpha_{n,(0)}^*}{\sqrt{n}} \right), \quad R_{n,j} = \sqrt{n}R \left(\frac{\alpha_{n,(j)}^* + \alpha_{n,(0)}^*}{\sqrt{n}} \right), \quad (5.76)$$

with the notational convention $\alpha_{n,(0)}^* = \alpha_n$ and $\mathbb{P}_{n,(0)}^* = \mathbb{P}_n$. Let define the following events, for large enough constants $C_1, C_2 > 0$,

$$A_n^* = \bigcap_{j=0}^{b_n} \left\{ \max \left(\|\alpha_{n,(j)}^*\|_{\mathcal{F}}, \|\mathbb{G}_{n,(j)}^*\|_{\mathcal{F}} \right) < C_1 \sqrt{\log n} \right\},$$

$$B_n^* = \bigcap_{j=0}^{b_n} \left\{ \|\alpha_{n,(j)}^* - \mathbb{G}_{n,(j)}^*\|_{\mathcal{F}} < C_2 w_n \right\}.$$

Likewise, on the probability space of Theorem 5.2.2, which coincides with the previous one for $b_n = 1$ and $w_n = v_n$, consider $R_n = \sqrt{n}R(\alpha_n/\sqrt{n})$ and

$$A_n = \left\{ \max \left(\|\alpha_n\|_{\mathcal{F}}, \|\mathbb{G}_n\|_{\mathcal{F}}, \|\alpha_n^*\|_{\mathcal{F}}, \|\mathbb{G}_n^*\|_{\mathcal{F}} \right) < C_1 \sqrt{\log n} \right\},$$

$$B_n = \left\{ \|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} < C_2 v_n \right\} \cap \left\{ \|\alpha_n^* - \mathbb{G}_n^*\|_{\mathcal{F}} < C_2 v_n \right\}.$$

For any fixed $\theta > 0$ we have, with respect to the two distinct probability distributions and their associated n_θ , $\mathbb{P}(B_n^C) < 1/3n^\theta$ and $\mathbb{P}(B_n^{*C}) < 1/3n^\theta$ for all $n > n_\theta$. Choose $n_1 > 0$ such that for all $n > n_1$ and the fixed ball $\mathcal{B} \subset \ell^\infty(\mathcal{F})$ from (FR), we have

$$A_n^* \subset \left\{ \frac{\alpha_{n,(0)}^*}{\sqrt{n}} \in \mathcal{B} \right\} \cap \bigcap_{j=1}^{b_n} \left\{ \frac{\alpha_{n,(j)}^* + \alpha_{n,(0)}^*}{\sqrt{n}} \in \mathcal{B} \right\},$$

$$A_n \subset \left\{ \frac{\alpha_n}{\sqrt{n}} \in \mathcal{B} \right\} \cap \left\{ \frac{\alpha_n^* + \alpha_n}{\sqrt{n}} \in \mathcal{B} \right\}.$$

By (5.17) for all $n > n_1$, on the event A_n^* it holds, for $j = 1, \dots, b_n$,

$$S_{n,(0)}^* = \varphi(\mathbb{P}_{n,(0)}^*) = \varphi(P) + \frac{1}{\sqrt{n}}\varphi'(P) \cdot \alpha_{n,(0)}^* + \frac{1}{\sqrt{n}}R_{n,0},$$

$$S_{n,(j)}^* = \varphi(\mathbb{P}_{n,(j)}^*) = \varphi(P) + \frac{1}{\sqrt{n}}\varphi'(P) \cdot (\alpha_{n,(j)}^* + \alpha_{n,(0)}^*) + \frac{1}{\sqrt{n}}R_{n,j},$$

and, for $j = 0, \dots, b_n$,

$$\max \left(|\varphi'(P) \cdot \alpha_{n,(j)}^*|, |\varphi'(P) \cdot \mathbb{G}_{n,(j)}^*| \right) < M_{P,\mathcal{F}} C_1 \sqrt{\log n},$$

$$R_{n,j} < M_{P,\mathcal{F}} C_1^q \frac{(\log n)^{q/2}}{n^{(q-1)/2}}.$$

On the event B_n^* it is true that $|\varphi'(P) \cdot (\alpha_{n,(j)}^* - \mathbb{G}_{n,(j)}^*)| < M_{P,\mathcal{F}} C_2 w_n$ for $j = 0, \dots, b_n$. By Proposition 5.3.2 – reduced for $N = 0$ to the usual concentration inequality – it holds $\mathbb{P}(\|\alpha_n\|_{\mathcal{F}} > \lambda\sqrt{n}) + b_n \mathbb{P}(\|\alpha_n^*\|_{\mathcal{F}} > \lambda\sqrt{n}) < 1/3n^\theta$ for all $n > n_{\theta,\lambda}$ and a small enough $\lambda > 0$ depending on \mathcal{B} . By the Borell-Sudakov deviation inequality similar bounds hold for

$\|\mathbb{G}_{n,(j)}^*\|_{\mathcal{F}}$ thus $\mathbb{P}(A_n^{*C}) < 2/3n^\theta$. Concerning the bootstrap resampling, if $n > n_1$, on the event A_n^* we then have the following equalities, for $j = 1, \dots, b_n$,

$$S_{n,(j)}^* - S_{n,(0)}^* = \varphi'(P) \cdot \frac{\alpha_{n,(j)}^*}{\sqrt{n}} + \frac{1}{\sqrt{n}}(R_{n,j} - R_{n,0}). \quad (5.77)$$

Finally, the real random variables $Y_{n,(j)}^* = \varphi'(P) \cdot \mathbb{G}_{n,(j)}^*$ for $j = 0, \dots, b_n$ are independent, centered and Gaussian with common variance $\sigma^2 > 0$. Similar relations as above hold in the probability space of Theorem 5.2.2 on the event A_n which is also such that $\mathbb{P}(A_n^C) < 1/3n^\theta$ for $n > n_{\theta,\lambda}$ and entails

$$\mathbb{1}_{A_n} S_n = \mathbb{1}_{A_n} \varphi \left(P + \frac{\alpha_n}{\sqrt{n}} \right) = \frac{1}{\sqrt{n}} \mathbb{1}_{A_n} (\varphi'(P) \cdot \alpha_n + R_n). \quad (5.78)$$

On the event B_n we have $|\varphi'(P) \cdot (\alpha_n - \mathbb{G}_n)| < M_{P,\mathcal{F}} C_2 v_n$ and $\mathbb{P}(B_n^C) < 1/3n^\theta$. Also, $Y_n = \varphi'(P) \cdot \mathbb{G}_n$ and $Y_n^* = \varphi'(P) \cdot \mathbb{G}_n^*$ are independent, centered Gaussian with variance $\sigma^2 > 0$. We are now ready to prove Proposition 5.2.8 at Step 1, Proposition 5.2.9 at Step 2 and Proposition 5.2.10 at Step 3.

Step 1. It is important to point out that $\varphi(\mathbb{P}_{n,(0)}^*)$ and $\varphi(\mathbb{P}_n)$ below have the same distribution as S_n but are not defined on the same probability space. Let us first work on the space of Theorem 5.2.2. Consider the event $C_n = A_n \cap B_n$. The normal random variable $\varphi'(P) \cdot \mathbb{G}$ being centered we have

$$\begin{aligned} \sigma^2 - \mathbb{E}(\mathbb{1}_{C_n} (\varphi'(P) \cdot \mathbb{G})^2) &= \mathbb{E}(\mathbb{1}_{C_n^C} (\varphi'(P) \cdot \mathbb{G})^2) \\ &\leq \sqrt{(1 - \mathbb{P}(C_n)) \mathbb{E}((\varphi'(P) \cdot \mathbb{G})^4)} \leq \frac{\sqrt{\mu_4}}{\sigma^2 n^{\theta/2}} \end{aligned}$$

where μ_4 is the fourth moment of the standard Gaussian distribution. Moreover,

$$\begin{aligned} \mathbb{E}(\mathbb{1}_{C_n} (\varphi'(P) \cdot \mathbb{G})^2) - \text{Var}(\mathbb{1}_{C_n} (\varphi'(P) \cdot \mathbb{G})) &= \mathbb{E}(\mathbb{1}_{C_n} (\varphi'(P) \cdot \mathbb{G})^2) \\ &= \mathbb{E}(\mathbb{1}_{C_n^C} (\varphi'(P) \cdot \mathbb{G})^2) \leq \frac{\sigma^2}{n^\theta}. \end{aligned}$$

By (5.18) we have

$$\begin{aligned} |\text{Var}(S_n) - \text{Var}(\mathbb{1}_{C_n} S_n)| &= |\text{Var}(\mathbb{1}_{C_n^C} S_n) - 2\mathbb{E}(\mathbb{1}_{C_n} S_n) \mathbb{E}(\mathbb{1}_{C_n^C} S_n)| \\ &\leq 3(1 - \mathbb{P}(C_n)) n^{2\kappa} \leq \frac{2}{n^{\theta-2\kappa}}. \end{aligned}$$

Now, by denoting $\mathbb{R}_n = \alpha_n - \mathbb{G}_n^*$ that is a centered process indexed by \mathcal{F} , and $\xi_n = \mathbb{E}(\mathbb{1}_{C_n} (\varphi'(P) \cdot \mathbb{R}_n + R_n)^2)$ it holds

$$\begin{aligned} &|\text{Var}(\mathbb{1}_{C_n} (\varphi'(P) \cdot \alpha_n + R_n)) - \text{Var}(\mathbb{1}_{C_n} \varphi'(P) \cdot \mathbb{G})| \\ &= |\text{Var}(\mathbb{1}_{C_n} (\varphi'(P) \cdot \mathbb{R}_n + R_n)) + 2\text{Cov}(\mathbb{1}_{C_n} \varphi'(P) \cdot \mathbb{G}_n, \mathbb{1}_{C_n} (\varphi'(P) \cdot \mathbb{R}_n + R_n))| \\ &\leq \xi_n + 2\sigma\sqrt{\xi_n}. \end{aligned} \quad (5.79)$$

By proceeding as in [2] – in particular the arguments establishing (3.27) of that paper – we readily get that for some $d_0 > 0$ and n large enough,

$$\sup_{f \in \mathcal{F}} \text{Var}(\mathbb{1}_{C_n} \mathbb{R}_n(f)) = \sup_{f \in \mathcal{F}} \mathbb{E}(\mathbb{1}_{C_n} \mathbb{R}_n(f)^2) \leq \mathbb{E}(\|\mathbb{R}_n\|_{\mathcal{F}}^2) < d_0 v_n^2. \quad (5.80)$$

According to Theorem 2.14.9 of [80], under (VC) or (BR) there exists $C_{\mathcal{F}} > 0$ such that for n large enough, $\mathbb{E}(\|\alpha_n\|_{\mathcal{F}}^{2q}) \leq C_{\mathcal{F}}$ which implies

$$\mathbb{E}(\mathbf{1}_{C_n} R_n^2) \leq \frac{1}{n^{q-1}} \mathbb{E}(\|\alpha_n\|_{\mathcal{F}}^{2q}) \leq \frac{C_{\mathcal{F}}}{n^{q-1}}. \quad (5.81)$$

Combining (5.80) and (5.81), for n large enough, it follows from (FR),

$$\begin{aligned} \xi_n &\leq M_{P,\mathcal{F}}^2 \mathbb{E}(\|\mathbb{R}_n\|_{\mathcal{F}}^2) + 2M_{P,\mathcal{F}} \mathbb{E}(\|\mathbb{R}_n\|_{\mathcal{F}}^2)^{1/2} \mathbb{E}(R_n^2)^{1/2} + \mathbb{E}(R_n^2) \\ &\leq d_0(M_{P,\mathcal{F}} v_n)^2 + 2M_{P,\mathcal{F}} v_n \left(\frac{d_0 C_{\mathcal{F}}}{n^{q-1}} \right)^{1/2} + \frac{C_{\mathcal{F}}}{n^{q-1}} < d_1 \xi_n^+, \end{aligned} \quad (5.82)$$

for some $d_1 > 0$ depending on (P, \mathcal{F}) and $\xi_n^+ = \max(v_n^2, 1/n^{q-1}) \rightarrow 0$. By (5.79) and (5.82) we can find $n_3 > 0, C_3 > 0$ such that for all $n > n_3$,

$$|\text{Var}(\varphi'(P) \cdot \alpha_n + R_n) - \sigma^2| \leq d_1 \xi_n^+ + 2\sigma \sqrt{d_1 \xi_n^+} \leq C_3 \sqrt{\xi_n^+}. \quad (5.83)$$

Next, in the space of Theorem 5.2.5 consider $C_n^* = A_n^* \cap B_n^*$. By (5.76), (5.77) and (5.78) one can write, on the event A_n^* ,

$$\begin{aligned} &n(\widehat{\text{Var}}(S_n^*) - \text{Var}(S_n)) \\ &= \frac{n}{b_n} \sum_{j=1}^{b_n} \left((\varphi(\mathbb{P}_{n,(j)}^*) - \varphi(\mathbb{P}_{n,(0)}^*))^2 - \text{Var}(\varphi(\mathbb{P}_n)) \right) \\ &= \frac{1}{b_n} \sum_{j=1}^{b_n} \left(\varphi'(P) \cdot \alpha_{n,(j)}^* + R_{n,j} - R_{n,0} \right)^2 - \text{Var}(\varphi'(P) \cdot \alpha_n + R_n). \end{aligned}$$

One can find $C_4 > 2M_{P,\mathcal{F}}^2 C_1 C_2$ and $n_4 > n_1$ such that for all $n > n_4$ and $1 \leq j \leq b_n$, on the event C_n^* we have

$$\begin{aligned} &\left| (\varphi'(P) \cdot \alpha_{n,(j)}^* + R_{n,j} + R_{n,0})^2 - Y_{n,(j)}^{*2} \right| \\ &= \left| (\varphi'(P) \cdot \alpha_{n,(j)}^* + R_{n,j} + R_{n,0})^2 - (\varphi'(P) \cdot \mathbb{G}_{n,(j)}^*)^2 \right| \\ &\leq \left| \varphi'(P) \cdot (\alpha_{n,(j)}^* - \mathbb{G}_{n,(j)}^*) \right| \left| \varphi'(P) \cdot (\alpha_{n,(j)}^* + \mathbb{G}_{n,(j)}^*) \right| \\ &\quad + (R_{n,j} + R_{n,0})^2 + 2|R_{n,j} + R_{n,0}| |\varphi'(P) \cdot \alpha_{n,(j)}^*| \\ &\leq C_4 \xi_n^* \sqrt{\log n} \end{aligned} \quad (5.84)$$

where $\xi_n^* = \max(w_n, (\log n)^{q/2}/n^{(q-1)/2})$. Since $\sqrt{\xi_n^+} < \xi_n^*$ for all n large we get by (5.83) and (5.84) that for some $C_5 > C_4$, $n_5 > \max(n_3, n_4)$ and any $n > n_5$ such that $A_n^* \cap B_n^*$ is satisfied, it holds

$$n|\widehat{\text{Var}}(S_n^*) - \text{Var}(S_n)| \leq \left| \frac{1}{b_n} \sum_{j=1}^{b_n} (\varphi'(P) \cdot \mathbb{G}_{n,(j)}^*)^2 - \sigma^2 \right| + C_5 \xi_n^* \sqrt{\log n}. \quad (5.85)$$

Notice that $\sum_{j=1}^{b_n} \varphi'^2(P) \cdot \mathbb{G}_{n,(j)}^*$ has the same distribution as $\sigma^2 \chi^2(b_n)$. By (5.85) we have,

for all $n > n_5$,

$$\begin{aligned}
& \mathbb{P} \left(\frac{n}{\sigma^2} \left| \widehat{\text{Var}}(S_n^*) - \text{Var}(S_n) \right| \geq \delta + \frac{C_5}{\sigma^2} \xi_n^* \sqrt{\log n} \right) \\
& \leq \mathbb{P} \left(\frac{1}{\sigma^2} \left| \frac{1}{b_n} \sum_{j=1}^{b_n} \varphi'^2(P) \cdot \mathbb{G}_{n,(j)}^* - \sigma^2 \right| \geq \delta \right) + \mathbb{P}(A_n^{*C}) + \mathbb{P}(B_n^{*C}) \\
& \leq \mathbb{P} \left(\left| \frac{\chi^2(b_n)}{b_n} - 1 \right| \geq \delta \right) + \frac{1}{n^\theta} \leq a_n + \frac{1}{n^\theta}.
\end{aligned}$$

We have shown that (5.19) is satisfied for $C_0 = C_5$ and $n_0 = n_6$. According to Tchebychev inequality, $\mathbb{P} \left(\left| \frac{\chi^2(b_n)}{b_n} - 1 \right| > \delta \right) \leq 2/b_n \delta^2$. Take $\delta^2 = 2/a_n b_n$ and $a_n = 1/n(\log n)^{2\iota}$ with $\iota > 1/2$. Since $(\log n)^\iota / \sqrt{n b_n} > w_n \sqrt{\log n} / n$ eventually in n , Remark 5.2.7 follows from (5.19) and Borel-Cantelli lemma since, provided that $n > n_6$,

$$\begin{aligned}
& \mathbb{P} \left(\left| \widehat{\text{Var}}(S_n^*) - \text{Var}(S_n) \right| > C'_0 / \sqrt{b_n} \right) \\
& \leq \mathbb{P} \left(\left| \widehat{\text{Var}}(S_n^*) - \text{Var}(S_n) \right| > \sigma^2 \left(\frac{1}{\sqrt{b_n}} + \frac{C_0 w_n \sqrt{\log n}}{n} \right) \right) \leq \frac{3}{n^{1+\iota}}.
\end{aligned}$$

Step 2. As a consequence, by the concentration inequality we get

$$|F_{S_n}(x) - \mathbb{P}(\{S_n \leq x\} \cap A_n \cap B_n)| \leq \mathbb{P}(A_n^C) + \mathbb{P}(B_n^C) \leq \frac{2}{n^\theta}.$$

Write

$$y_n = \max \left(\left(\frac{\log n}{n} \right)^{q/2}, v_n \right), \quad C_\theta^* > M_{P,\mathcal{F}} \frac{C_1^q + C_2}{\sigma},$$

and observe that $y_n \rightarrow 0$. We have, by (FR),

$$\begin{aligned}
& \mathbb{P}(\{S_n \leq x\} \cap A_n \cap B_n) \\
& = \mathbb{P} \left(\left\{ \frac{1}{\sqrt{n}} \varphi'(P) \cdot \mathbb{G}_n \leq x - \varphi(P) - \frac{1}{\sqrt{n}} \varphi'(P) \cdot (\alpha_n - \mathbb{G}_n) - \frac{1}{\sqrt{n}} R_{n,0} \right\} \cap A_n \cap B_n \right) \\
& \in \left[\Phi \left(\sqrt{n} \frac{x - \varphi(P)}{\sigma} - C_\theta^* y_n \right); \Phi \left(\sqrt{n} \frac{x - \varphi(P)}{\sigma} + C_\theta^* y_n \right) \right]. \tag{5.86}
\end{aligned}$$

The length of the latter interval is of order $2d_n(x)$ where, for some $C'_\theta > C_\theta^*$ and all n large enough,

$$d_n(x) \leq \frac{C'_\theta y_n}{\sqrt{2\pi}} \exp \left(-n \frac{(x - \varphi(P))^2}{2\sigma^2} \right), \quad x \in \mathbb{R}.$$

Therefore we conclude the deterministic fact that

$$\left| F_{S_n}(x) - \Phi \left(\sqrt{n} \frac{x - \varphi(P)}{\sigma} \right) \right| \leq d_n(x) + \frac{2}{n^\theta}, \quad x \in \mathbb{R}.$$

The empirical distribution function $\widehat{F}_{Y_n^*}$ of variables $Y_{n,(j)}^*$ can be compared to Φ by using the DKW inequality. Whatever the probability space, hence on the previous one, we have

$$\mathbb{P} \left(\sup_{x \in \mathbb{R}} \left| \widehat{F}_{Y_n^*}(x) - \Phi \left(\frac{x}{\sigma} \right) \right| > \delta_n \right) \leq \frac{2}{n^\theta}, \quad \delta_n = \sqrt{\frac{\theta \sigma^2 \log n}{2b_n}}. \tag{5.87}$$

By (5.86) and (5.87) it follows, on the probability space of Theorem 5.2.5, for any interval \mathcal{I} ,

$$\begin{aligned} & \mathbb{P} \left(\sup_{x \in \mathcal{I}} \left| \widehat{F}_{Y_n^*}(\sqrt{n}(x - \varphi(P))) - F_{S_n}(x) \right| > \delta_n + \sup_{x \in \mathcal{I}} d_n(x) + \frac{2}{n^\theta} \right) \\ & \leq \frac{2}{n^\theta} + \mathbb{P} \left(\sup_{x \in \mathcal{I}} \left| F_{S_n}(x) - \Phi \left(\frac{\sqrt{n}(x - \varphi(P))}{\sigma} \right) \right| > \sup_{x \in \mathcal{I}} d_n(x) + \frac{2}{n^\theta} \right) = \frac{2}{n^\theta}. \end{aligned}$$

Finally we study the bootstrap empirical distribution function $\widehat{F}_{S_n^*}$. For $y_n \leq y_n^* = \max(y_n, w_n) \rightarrow 0$ we have

$$A_n^* \cap B_n^* \subset \bigcap_{j=0}^{b_n} \left\{ \left| S_{n,(j)}^* - \varphi(P) - \frac{Y_{n,(j)}^*}{\sqrt{n}} \right| < \sigma C_\theta^* \frac{y_n^*}{\sqrt{n}} \right\} = \bigcap_{j=0}^{b_n} \left\{ \left| S_{n,(j)}^* - T_{n,(j)}^* \right| < \varepsilon_n \right\},$$

where we write $T_{n,(j)}^* = \varphi(P) + Y_{n,(j)}^*/\sqrt{n}$ and $\varepsilon_n = \sigma C_\theta^* y_n^*/\sqrt{n}$. On the event $A_n^* \cap B_n^*$ the empirical distribution function $\widehat{F}_{Y_n^*}$ can be compared to $\widehat{F}_{S_n^*}$ through $\widehat{F}_{T_{n,(j)}^*}(x - \varepsilon_n) \leq \widehat{F}_{S_n^*}(x) \leq \widehat{F}_{T_{n,(j)}^*}(x + \varepsilon_n)$ and hence

$$\widehat{F}_{Y_n^*}(\sqrt{n}(x - \varphi(P)) - \sigma C_\theta^* y_n^*) \leq \widehat{F}_{S_n^*}(x) \leq \widehat{F}_{Y_n^*}(\sqrt{n}(x - \varphi(P)) + \sigma C_\theta^* y_n^*).$$

Therefore we can attest that on the probability space of Theorem 5.2.5 it holds, for any Borel set $\mathcal{I} \subset \mathbb{R}$, any $\delta_n^* > 0$ and $d_n = \delta_n + \sup_{x \in \mathcal{I}} d_n(x) + 2/n^\theta$,

$$\begin{aligned} & \mathbb{P} \left(\sup_{x \in \mathcal{I}} \left| \widehat{F}_{S_n^*}(x) - F_{S_n}(x) \right| > \delta_n^* + \delta_n + d_n \right) \\ & \leq \frac{2}{n^\theta} + \mathbb{P} \left(A_n^* \cap B_n^* \cap \left\{ \sup_{x \in \mathcal{I}} \left| \widehat{F}_{S_n^*}(x) - F_{S_n}(x) \right| > \delta_n^* + \delta_n + d_n \right\} \right) \\ & \leq \frac{4}{n^\theta} + \mathbb{P} \left(A_n^* \cap B_n^* \cap \left\{ \sup_{x \in \mathcal{I}} \left| \widehat{F}_{S_n^*}(x) - \widehat{F}_{Y_n^*}(\sqrt{n}(x - \varphi(P))) \right| > \delta_n^* + \delta_n \right\} \right) \\ & \leq \frac{4}{n^\theta} + \mathbb{P} \left(\max \left(\sup_{x \in \mathcal{I}} |\widehat{s}_n^-(x)|, \sup_{x \in \mathcal{I}} |\widehat{s}_n^+(x)| \right) > \delta_n^* + \delta_n \right) \end{aligned}$$

where $\widehat{s}_n^\pm(x) = \widehat{F}_{Y_n^*}(\sqrt{n}(x - \varphi(P)) \pm \sigma C_\theta^* y_n^*) - \widehat{F}_{Y_n^*}(\sqrt{n}(x - \varphi(P)))$ are local empirical Gaussian increments. By applying (5.87) again, we conclude that

$$\begin{aligned} & \mathbb{P} \left(\sup_{x \in \mathcal{I}} \left| \widehat{F}_{S_n^*}(x) - F_{S_n}(x) \right| > \delta_n^* + \delta_n + d_n \right) \\ & \leq \frac{4}{n^\theta} + \mathbb{P} \left(\max \left(\sup_{x \in \mathcal{I}} |s_n^-(x)|, \sup_{x \in \mathcal{I}} |s_n^+(x)| \right) > \delta_n^* \right) = \frac{4}{n^\theta} \end{aligned} \quad (5.88)$$

where $s_n^\pm(x) = \Phi(\sqrt{n}(x - \varphi(P))/\sigma) \pm C_\theta^* y_n^* - \Phi(\sqrt{n}(x - \varphi(P))/\sigma)$ are local deterministic Gaussian increments and

$$\delta_n^* \geq \max \left(\sup_{x \in \mathcal{I}} |s_n^-(x)|, \sup_{x \in \mathcal{I}} |s_n^+(x)| \right).$$

It remains to evaluate δ_n^* and $\sup_{x \in \mathcal{I}} d_n(x)$ for \mathcal{I} including $\varphi(P)$ and \mathcal{I} far from $\varphi(P)$ while comparing the sequences $\delta_n^*, \delta_n, d_n, y_n^*, w_n$. First assume that $w_n > \sqrt{(\log n)/b_n}$ and define $\mathcal{I} = \mathcal{I}_{0,n} = (\varphi(P) - z_n, \varphi(P) + z_n)$ where $z_n = \sqrt{\sigma^2/n} \sqrt{\log(b_n w_n^2/\log n)}$. We have $C_\theta y_n =$

$C_\theta v_n, C_\theta^* y_n^* = C_\theta^* w_n$ for all n large enough provided that $q > 1$ is large enough under (VC) and $q > 1$ under (BR), hence $\sup_{x \in \mathcal{I}_{0,n}} d_n(x) \leq D'_\theta v_n \leq D'_\theta w_n$ and in the same way we see that $\delta_n^* \leq D_\theta^* w_n$ for some constants $D'_\theta > 0, D_\theta^* > 0$ whereas $\delta_n = D_\theta \sqrt{(\log n)/b_n} < D_\theta w_n$ by the current assumption, so that we clearly have $\delta_n^* + \delta_n + d_n \leq D_\theta w_n$ and (5.88) implies (5.21) for an appropriate choice of constants $C_\theta > D_\theta$ and n_θ with respect to the previous ones. On the complement intervals $\mathcal{I} = \mathcal{I}_{0,n}^c$ the largest increment is achieved at the boundary – that is at $x = \varphi(P) \pm z_n$ and no more at $x = \varphi(P) \notin \mathcal{I}_{0,n}^c$ – and, by symmetry, $\sup_{x \in \mathcal{I}_{0,n}} d_n(x) \leq D'_\theta v_n \exp(-nz_n^2/2\sigma^2) = (D'_\theta v_n/w_n) \sqrt{(\log n)/b_n} \leq D'_\theta v_n$. We similarly get that $\delta_n^* \leq D_\theta^* w_n \exp(-nz_n^2/2\sigma^2) \leq D_\theta^* \sqrt{(\log n)/b_n}$ and this time d_n is of order δ_n thus $\delta_n^* + \delta_n + d_n \leq D_\theta \sqrt{(\log n)/b_n}$. Whenever we have everywhere a rate δ_n of order $\sqrt{(\log n)/b_n}$ or w_n faster than $\sqrt{(\log n)/b_n}$.

Step 3. According to Theorem 5.2.2, there exists $C_0 > 0, n_1 > 0$, a sequence $\{(X_n, Z_n)\}$ of i.i.d. random variables with distribution $P^{(X,Z)}$ and a sequence of $\{\mathbb{G}_n^*(\mathcal{F})\}$ of P -Brownian bridge such that for all $n > n_1, \mathbb{P}(A_n) \leq 1/n^2$ with

$$A_n = \{|\alpha_n - \mathbb{G}_n|_{\mathcal{F}} \geq C_0 v_n/4\} \cup \{|\alpha_n^* - \mathbb{G}_n^*|_{\mathcal{F}} \geq C_0 v_n/4\}.$$

Then we have for all $n > n_0$

$$\begin{aligned} \mathbb{P}(\phi(\alpha_n^*(f)) \leq x) &\leq \mathbb{P}\left(\phi(\mathbb{G}_n^*(f)) \leq x + \frac{1}{4}C_0 C_1 v_n\right) + P(A_n) \\ &\leq \mathbb{P}(\phi(\mathbb{G}_n^*(f)) \leq x) + \frac{1}{4}C_0 C_1 C_2 v_n + \frac{1}{n^2}. \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{P}(\phi(\mathbb{G}_n^*(f)) \leq x) - C_0 C_1 C_2 v_n &\leq \mathbb{P}\left(\phi(\mathbb{G}_n^*(f)) \leq x - \frac{1}{4}C_0 C_1 v_n\right) \\ &\leq \mathbb{P}(\phi(\alpha_n^*(f)) \leq x) + P(A_n) \\ &\leq \mathbb{P}(\phi(\alpha_n^*(f)) \leq x) + \frac{1}{n^2}. \end{aligned}$$

By definition of v_n , for some $n_2 > 0$ and all $n > n_2, 1/n^2 \leq C_0 C_1 C_2 v_n/4$. Thus (5.23) holds for $n_0 = \max(n_1, n_2)$. The same arguments yield

$$\sup_{\phi \in \mathcal{L}} \sup_{f \in \mathcal{F}} \sup_{x \in \mathbb{R}} |\mathbb{P}(\phi(\alpha_n(f)) \leq x) - \mathbb{P}(\phi(\mathbb{G}(f)) \leq x)| \leq \frac{1}{2}C_0 C_1 C_2 v_n.$$

thus inequality (5.24) follows.

5.4.7 Proof of Proposition 5.3.6

Let $\theta > 0$. Proposition 5.4.7 below is proved at Step 1 then Proposition 5.3.6 is proved at Step 2. Like $\tilde{\alpha}_n^*(\mathcal{F})$ given by (5.32) let us define the process $\tilde{\alpha}_n^{(N)*}(\mathcal{F})$ by $\tilde{\alpha}_n^{(N)*}(f) = (T_n^{(N)}/n)\alpha_n^{(N)*}(f)$. The following proposition establishes the strong approximation of $\tilde{\alpha}_n^{(N)*}(\mathcal{F})$ to $\mathbb{G}^{(N)*}(\mathcal{F})$.

Proposition 5.4.7. *Assume (5.3), (5.4) and either (VC) or (BR). There exists $C_\theta, n_\theta > 0$, a sequence $\{(X_n, Z_n)\}$ of i.i.d. random variables with distribution $P^{(X,Z)}$ and a sequence $\{(\mathbb{G}_n(\mathcal{F}), \mathbb{G}_n^{(0)*}(\mathcal{F})), \dots, \mathbb{G}_n^{(N_0)*}(\mathcal{F})\}$ of P -Brownian bridges, all of these sequences being defined on the same probability space, such that for all $n > n_\theta$,*

$$\mathbb{P} \left(\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} > C_\theta v_n \cup \bigcup_{N=0}^{N_0} \{\|\tilde{\alpha}_n^{(N)*} - \mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > C_\theta v_n\} \right) \leq \frac{1}{n^\theta}. \quad (5.89)$$

Step 1. By Lemma 5.4.1 one can apply Theorem 2.1 of [2] to $(\alpha_n^{(X,Z)})^{(N)}(\mathcal{F}_1)$ the N -th raked empirical process of $\alpha_n^{(X,Z)}$ defined by (5.5), the law $P^{(X,Z)}$ and the auxiliary information $\mathcal{A}^{0,(N)}$ defined by (5.51). There exists $C_0 = C_0(\theta), n_0 = n_0(\theta) > 0$ and a probability space which supports a sequence of independent random variables $\{(X_n, Z_n)\}$ distributed as $P^{(X,Z)}$ and a sequence $\{\mathbb{G}_n^{0,(0)}(\mathcal{F}_1), \dots, \mathbb{G}_n^{0,(N_0)}(\mathcal{F}_1)\}$ of process of independent $P^{(X,Z)}$ -raked Brownian bridge satisfying for all $n > n_0$,

$$\mathbb{P} \left(\max_{0 \leq N \leq N_0} \|(\alpha_n^{(X,Z)})^{(N)} - \mathbb{G}_n^{0,(N)}\|_{\mathcal{F}_1} > C_0 v_n \right) \leq \frac{1}{7n^\theta}. \quad (5.90)$$

Since $\alpha_n(f) = (\alpha_n^{(X,Z)})^{(0)}(h_f^{(3)})$ then (5.90) implies immediately

$$\mathbb{P}(\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} > C_\theta v_n) \leq \frac{1}{7n^\theta}. \quad (5.91)$$

Let decompose $\tilde{\alpha}_n^{(N)*}(f)$ as the same way as (5.35):

$$\begin{aligned} \tilde{\alpha}_n^{(N)*}(f) &= (\alpha_n^{(X,Z)})^{(N)}(h_f^{(1)}) - \mathbb{E}[f] (\alpha_n^{(X,Z)})^{(N)}(h_f^{(2)}) \\ &\quad - (\alpha_n^{(X,Z)})^{(N)}(h_f^{(3)}) \left(1 + \left(T_n^{(N)}/n - 1\right)\right), \end{aligned} \quad (5.92)$$

where $h_f^{(1)}, h_f^{(2)}, h_f^{(3)}$ are defined by (5.33). Let $C'_0 = C'_0(\theta) = C_0 \max(1, M_{\mathcal{F}})$. By (5.90), (5.92) and definition of $\mathbb{G}^{(N)*}(\mathcal{F})$ given by (5.52) it holds for all $n > n_0$,

$$\begin{aligned} &\mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\tilde{\alpha}_n^{(N)*} - \mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > C'_0 v_n \right) \\ &\leq \frac{3}{7n^\theta} + \mathbb{P} \left(\max_{0 \leq N \leq N_0} \left| \frac{T_n^{(N)}}{n} - 1 \right| \|(\alpha_n^{(X,Z)})^{(N)}\|_{\mathcal{F}_1} > C_0 v_n \right). \end{aligned} \quad (5.93)$$

Last term of (5.93) can be bounded as the same way as (5.59). By (5.90) and definition of ε we have for all $n > n_0$,

$$\begin{aligned} &\mathbb{P} \left(\max_{0 \leq N \leq N_0} \left| \frac{T_n^{(N)}}{n} - 1 \right| \|(\alpha_n^{(X,Z)})^{(N)}\|_{\mathcal{F}_1} > C_0 v_n \right) \\ &\leq \mathbb{P} \left(\|(\alpha_n^{(X,Z)})^{(N)}\|_{\mathcal{F}_1} > C_0 v_n n^{1/2-\varepsilon} \right) + \mathbb{P} \left(\max_{0 \leq N \leq N_0} \sqrt{n} \left| \frac{T_n^{(N)}}{n} - 1 \right| > n^\varepsilon \right) \\ &\leq \frac{1}{7n^\theta} + \mathbb{P} \left(\|\mathbb{G}_n^{0,(N)}\|_{\mathcal{F}_1} > C'_0 v_n n^{1/2-\varepsilon} \right) + \mathbb{P} \left(\max_{0 \leq N \leq N_0} \sqrt{n} \left| \frac{T_n^{(N)}}{n} - 1 \right| > n^\varepsilon \right) .. \end{aligned} \quad (5.94)$$

We do the same calculations as (5.62). By Borell-Sudakov there exists $n_1 = n_1(\theta) > 0$ such that for all $n > n_1$,

$$\mathbb{P}(\|\mathbb{G}_n^{0,(N)}\|_{\mathcal{F}_1} > C_\theta v_n n^{1/2-\varepsilon}) \leq \frac{1}{7n^\theta}. \quad (5.95)$$

Since $\sqrt{n}(T_n^{(N)}/n - 1) = (\alpha_n^{(X,Z)})^{(N)}(\mathbf{1} \times \phi_{M_Z})$, Proposition 3 of [2] implies that there exists $n_2 = n_2(\theta) > 0$ such that for all $n > n_2$,

$$\mathbb{P}\left(\max_{0 \leq N \leq N_0} \sqrt{n} \left| \frac{T_n^{(N)}}{n} - 1 \right| > n^\varepsilon\right) \leq \mathbb{P}\left(\max_{0 \leq N \leq N_0} \|(\alpha_n^{(X,Z)})^{(N)}\|_{\mathcal{F}_1} > n^\varepsilon\right) \leq \frac{1}{7n^\theta}. \quad (5.96)$$

By (5.91), (5.93), (5.94), (5.95) and (5.96) we have shown that (5.89) is satisfied with $n_\theta = \max(n_0, n_1, n_2)$ and $C_\theta = C'_\theta$.

Step 2. According to Proposition 5.4.7 there exists $C_0 = C_0(\theta) > 0, n_0 = n_0(\theta) > 0$, a sequence $\{(X_n, Z_n)\}$ of i.i.d. random variables with distribution $P^{(X,Z)}$ and a sequence $\{(\mathbb{G}_n(\mathcal{F}), \mathbb{G}_n^{(0)*}(\mathcal{F})), \dots, \mathbb{G}_n^{(N_0)*}(\mathcal{F})\}$ of P -Brownian bridges, all of these sequences being defined on the same probability space, such that for all $n > n_0$,

$$\mathbb{P}\left(\{ \|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} > C_0 v_n \} \cup \bigcup_{N=0}^{N_0} \{ \|\tilde{\alpha}_n^{(N)*} - \mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > C_0 v_n \}\right) \leq \frac{1}{5n^\theta}. \quad (5.97)$$

In particular for all $n > n_0$,

$$\mathbb{P}(\|\alpha_n - \mathbb{G}_n\|_{\mathcal{F}} > C_0 v_n) \leq \frac{1}{5n^\theta}. \quad (5.98)$$

To prove (5.30) we use the same method as the proof of Theorem 5.2.2. We write

$$\begin{aligned} & \mathbb{P}\left(\max_{0 \leq N \leq N_0} \|\alpha_n^{(N)*} - \mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > 3C_0 v_n\right) \\ & \leq \mathbb{P}\left(\max_{0 \leq N \leq N_0} \frac{n}{T_n^{(N)}} \|\tilde{\alpha}_n^{(N)*} - \mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > 2C_0 v_n\right) \\ & \quad + \mathbb{P}\left(\max_{0 \leq N \leq N_0} \left| \frac{n}{T_n^{(N)}} - 1 \right| \|\mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > C_0 v_n\right). \end{aligned} \quad (5.99)$$

From one hand, using (5.97) one can say that for all $n > n_0$,

$$\begin{aligned} & \mathbb{P}\left(\max_{0 \leq N \leq N_0} \frac{n}{T_n^{(N)}} \|\tilde{\alpha}_n^{(N)*} - \mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > 2C_0 v_n\right) \\ & \leq \mathbb{P}\left(\max_{0 \leq N \leq N_0} \|\tilde{\alpha}_n^{(N)*} - \mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > C_0 v_n\right) + \mathbb{P}\left(\max_{0 \leq N \leq N_0} \left| \frac{T_n^{(N)}}{n} - 1 \right| > 1/2\right) \\ & \leq \frac{1}{5n^\theta} + \mathbb{P}\left(\max_{0 \leq N \leq N_0} \|(\alpha_n^{(X,Z)})^{(N)}\|_{\mathcal{F}_1} > \sqrt{n}/2\right). \end{aligned} \quad (5.100)$$

By Proposition 3 of [2] there exists $n_1 = n_1(\theta) > 0$ such that for all $n > n_1$,

$$\mathbb{P}\left(\max_{0 \leq N \leq N_0} \|(\alpha_n^{(X,Z)})^{(N)}\|_{\mathcal{F}_1} > \sqrt{n}/2\right) \leq \frac{1}{5n^\theta}. \quad (5.101)$$

By (5.100) and (5.101) we have shown that for any $n > \max(n_0, n_1)$,

$$\mathbb{P} \left(\max_{0 \leq N \leq N_0} \frac{n}{T_n^{(N)}} \|\tilde{\alpha}_n^{(N)*} - \mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > 2C_0 v_n \right) \leq \frac{2}{5n^\theta}. \quad (5.102)$$

From the other hand, we write

$$\begin{aligned} & \mathbb{P} \left(\max_{0 \leq N \leq N_0} \left| \frac{n}{T_n^{(N)}} - 1 \right| \|\mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > C_0 v_n \right) \\ & \leq \mathbb{P} \left(\max_{0 \leq N \leq N_0} \|\mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > C_0 v_n n^{1/2-\varepsilon} \right) + \mathbb{P} \left(\max_{0 \leq N \leq N_0} \sqrt{n} \left| \frac{n}{T_n^{(N)}} - 1 \right| > n^\varepsilon \right). \end{aligned} \quad (5.103)$$

By definition of ε and Borell-Sudakov inequality, there $n_2 = n_2(\theta) > 0$ such that for all $n > n_2$,

$$\mathbb{P} \left(\|\mathbb{G}_n^*\|_{\mathcal{F}} > C v_n n^{1/2-\varepsilon} \right) \leq \frac{1}{5n^\theta}. \quad (5.104)$$

There also exists $n_3 = n_3(\theta) > 0$ such that for all $n > n_3$,

$$\mathbb{P} \left(\max_{0 \leq N \leq N_0} \sqrt{n} \left| \frac{n}{T_n^{(N)}} - 1 \right| > n^\varepsilon \right) \leq \mathbb{P} \left(\max_{0 \leq N \leq N_0} \|(\alpha_n^{(X,Z)})^{(N)}\|_{\mathcal{F}_1} > 2n^\varepsilon \right) \leq \frac{1}{5n^\theta}. \quad (5.105)$$

By (5.103), (5.104) and (5.105) it holds that for any $n > \max(n_2, n_3)$,

$$\mathbb{P} \left(\max_{0 \leq N \leq N_0} \left| \frac{n}{T_n^{(N)}} - 1 \right| \|\mathbb{G}_n^{(N)*}\|_{\mathcal{F}} > C_0 v_n \right) \leq \frac{2}{5n^\theta}. \quad (5.106)$$

By (5.98) (5.99), (5.102) and (5.106) we have shown (5.30) with $n_\theta = \max(n_0, n_1, n_2, n_3)$ and $C_\theta = 3C_0 > 0$.

Bibliographie

- [1] Albertus, M. (2019). Raking-ratio empirical process with auxiliary information learning. *Preprint*.
- [2] Albertus, M. and Berthet, P. (2019a). Auxiliary information : the raking-ratio empirical process. *Electron. J. Stat.*, 13(1) :120–165.
- [3] Albertus, M. and Berthet, P. (2019b). Rates of convergence for the weighted bootstrap of empirical and raking-ratio processes.
- [4] Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.*, 12(4) :1041–1067.
- [5] Alvarez-Andrade, S. and Bouzebda, S. (2013). Strong approximations for weighted bootstrap of empirical and quantile processes with applications. *Stat. Methodol.*, 11 :36–52.
- [6] Andersen, N. T. (1985). The central limit theorem for nonseparable valued functions. *Z. Wahrsch. Verw. Gebiete*, 70(3) :445–455.
- [7] Ardilly, P. (2006). *Les techniques de sondage*. Editions Technip.
- [8] Babu, G. J. (1984). Bootstrapping statistics with linear combinations of chi-squares as weak limit. *Sankhyā Ser. A*, 46(1) :85–93.
- [9] Barbe, P. and Bertail, P. (1995). *The weighted bootstrap*, volume 98 of *Lecture Notes in Statistics*. Springer-Verlag, New York.
- [10] Beran, R. (1984). Jackknife approximations to bootstrap estimates. *Ann. Statist.*, 12(1) :101–118.
- [11] Beran, R., Le Cam, L., and Millar, P. (1987). Convergence of stochastic empirical measures. *Journal of multivariate analysis*, 23(1) :159–168.
- [12] Berthet, P. and Mason, D. M. (2006). Revisiting two strong approximation results of Dudley and Philipp. 51 :155–172.
- [13] Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9(6) :1196–1217.

- [14] Billingsley, P. (1999). *Convergence of probability measures*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons, Inc., New York, second edition. A Wiley-Interscience Publication.
- [15] Binder, D. A. and Théberge, A. (1988). Estimating the variance of raking-ratio estimators. *Canad. J. Statist.*, 16(suppl.) :47–55.
- [16] Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves : exponential bounds and rates of convergence. *Bernoulli*, 4(3) :329–375.
- [17] Booth, J. G. and Sarkar, S. (1998). Monte carlo approximation of bootstrap variances. *The American Statistician*, 52(4) :354–357.
- [18] Borell, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30(2) :207–216.
- [19] Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities*. Oxford University Press, Oxford. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.
- [20] Bousquet, O. (2002). A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris*, 334(6) :495–500.
- [21] Brackstone, G. J. and Rao, J. N. K. (1979). An investigation of raking ratio estimators. *The Indian journal of Statistics*, Vol. 41 :97–114.
- [22] Bretagnolle, J. and Massart, P. (1989). Hungarian constructions from the nonasymptotic viewpoint. *Ann. Probab.*, 17(1) :239–256.
- [23] Brown, D. T. (1959). A note on approximations to discrete probability distributions. *Information and Control*, 2 :386–392.
- [24] Bru, B. (1988). Estimations laplaciennes. un exemple : la recherche de la population d’un grand empire, 1785-1812. *Journal de la société française de statistique*, 129(1-2) :6–45.
- [25] Campbell, J. (1934). The poisson correlation function. *Proceedings of the Edinburgh Mathematical Society*, 4(1) :18–26.
- [26] Chandra, T. K. and Ghosh, J. K. (1979). Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables. *Sankhyā Ser. A*, 41(1-2) :22–47.
- [27] Choudhry, G. and Lee, H. (1987). Variance estimation for the canadian labour force survey. *Survey Methodology*, 13(2) :147–161.
- [28] Csörgő, M. and Révész, P. (1981). *Strong approximations in probability and statistics*. Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London.
- [29] Deheuvels, P. and Mason, D. M. (1990). Bahadur-Kiefer-type processes. *Ann. Probab.*, 18(2) :669–697.
- [30] 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. *Ann. Math. Statistics*, 11 :427–444.

- [31] Derham, W. et al. (1769). *Théologie physique ou démonstration de l'existence et des attributs de Dieu, tirée des oeuvres de la création, accompagnée d'un grand nombre de remarques et d'observations curieuses par Guillaume Derham [...] traduite de l'anglois par Jacques Lufneu.*
- [32] Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.*, 87(418) :376–382.
- [33] Deville, J.-C., Särndal, C.-E., and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American statistical Association*, 88(423) :1013–1020.
- [34] Dmitriev, Y. G. and Tarasenko, F. P. (1978). On the use of a priori information in estimated linear functionals of distribution. *Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform.*, 7(6) :459–469.
- [35] Dudley, R. M. (1989). *Real analysis and probability.* The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.
- [36] Dudley, R. M. and Philipp, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. Verw. Gebiete*, 62(4) :509–552.
- [37] Efron, B. (1979). Bootstrap methods : another look at the jackknife. *Ann. Statist.*, 7(1) :1–26.
- [38] Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*, volume 38 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa.
- [39] Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American statistical Association*, 82(397) :171–185.
- [40] Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap.* CRC press.
- [41] Erdős, P. and Rényi, A. (1959). On the central limit theorem for samples from a finite population. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 4 :49–61.
- [42] Finkelstein, H. et al. (1971). The law of the iterated logarithm for empirical distribution. *The Annals of Mathematical Statistics*, 42(2) :607–615.
- [43] Franklin, J. and Lorenz, J. (1989). On the scaling of multidimensional matrices. *Linear Algebra Appl.*, 114/115 :717–735.
- [44] Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.*, 18(2) :851–869.
- [45] Hall, P. (1988). Rate of convergence in bootstrap approximations. *Ann. Probab.*, 16(4) :1665–1684.
- [46] Hall, P. (1992). *The bootstrap and Edgeworth expansion.* Springer Series in Statistics. Springer-Verlag, New York.
- [47] Hoffmann-Jørgensen, J. (1991). *Stochastic processes on Polish spaces.* Number 39. Aarhus Universitet. Matematisk Institut.

- [48] Holgate, P. (1964). Estimation for the bivariate poisson distribution. *Biometrika*, 51(1-2) :241–287.
- [49] Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, 47 :663–685.
- [50] Ireland, C. T. and Kullback, S. (1968). Contingency tables with given marginals. *Biometrika*, 55 :179–188.
- [51] Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32 :111–131.
- [52] Komlós, J., Major, P., and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 34(1) :33–58.
- [53] Konijn, H. S. (1981). Biases, variances and covariances of raking ratio estimators for marginal and cell totals and averages of observed characteristics. *Metrika*, 28(2) :109–121.
- [54] Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York.
- [55] Lamperti, J. (1977). *Stochastic processes*. Springer-Verlag, New York-Heidelberg. A survey of the mathematical theory, Applied Mathematical Sciences, Vol. 23.
- [56] Laplace, P. S. (1783). Sur les naissances, les mariages et les morts. *Histoire de l'Académie Royale des Sciences*, page 693.
- [57] Lewis, II, P. M. (1959). Approximating probability distributions to reduce storage requirements. *Information and Control*, 2 :214–225.
- [58] Lewis, T. O. and Newman, T. G. (1968). Pseudoinverses of positive semidefinite matrices. *SIAM J. Appl. Math.*, 16 :701–703.
- [59] Major, P. (1979). An improvement of Strassen's invariance principle. *Ann. Probab.*, 7(1) :55–61.
- [60] Mason, D. M. and Newton, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *Ann. Statist.*, 20(3) :1611–1624.
- [61] Morin, A. (1993). A propos de la loi de poisson multivariée. *Journal de la société française de statistique*, 134(2) :3–13.
- [62] Nelson, B. L. (1987). On control variate estimators. *Computers & OR*, 14(3) :219–225.
- [63] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2) :237–249.
- [64] Owen, A. B. (2001). *Empirical likelihood*. Chapman and Hall/CRC.
- [65] Pollard, D. (1989). Asymptotics via empirical processes. *Statist. Sci.*, 4(4) :341–366. With comments and a rejoinder by the author.
- [66] Pollard, D. (2012). *Convergence of stochastic processes*. Springer Science & Business Media.

- [67] Præ stgaard, J. and Wellner, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.*, 21(4) :2053–2086.
- [68] Robinson, P. M. (1982). On the convergence of the Horvitz-Thompson estimator. *Austral. J. Statist.*, 24(2) :234–238.
- [69] Rubin, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.*, 9(1) :130–134.
- [70] Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *J. Indian Soc. Agric. Statist.*, 5 :119–127.
- [71] Shao, J. and Tu, D. S. (1995). *The jackknife and bootstrap*. Springer Series in Statistics. Springer-Verlag, New York.
- [72] Shao, Y. (2001). Rate of convergence of bootstrapped empirical measures. *Statist. Probab. Lett.*, 53(3) :293–298.
- [73] Singh, K. (1981). On the asymptotic accuracy of Efron’s bootstrap. *Ann. Statist.*, 9(6) :1187–1195.
- [74] Sinkhorn, R. (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.*, 35 :876–879.
- [75] Sinkhorn, R. (1967). Diagonal equivalence to matrices with prescribed row and column sums. *Amer. Math. Monthly*, 74 :402–405.
- [76] Stephan, F. F. (1942). An iterative method of adjusting sample frequency tables when expected marginal totals are known. *The Annals of Mathematical Statistics*, 13(2) :166–178.
- [77] Strassen, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 3 :211–226 (1964).
- [78] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, 22(1) :28–76.
- [79] Tarima, S. and Pavlov, D. (2006). Using auxiliary information in statistical function estimation. *ESAIM Probab. Stat.*, 10 :11–23.
- [80] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.
- [81] Yates, F. and Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society : Series B (Methodological)*, 15(2) :253–261.
- [82] Zhang, B. (1995). M-estimation and quantile estimation in the presence of auxiliary information. *Journal of Statistical Planning and Inference*, 44(1) :77–94.
- [83] Zhang, B. (1997a). Estimating a distribution function in the presence of auxiliary information. *Metrika*, 46(1) :221–244.
- [84] Zhang, B. (1997b). Quantile processes in the presence of auxiliary information. *Annals of the institute of Statistical Mathematics*, 49(1) :35–55.

Étude du processus empirique avec information auxiliaire

RÉSUMÉ

Cette thèse porte sur l'étude du processus empirique avec information auxiliaire, c'est-à-dire une information que l'on aurait a priori ou bien que l'on aurait obtenu grâce à une source d'information. Nous montrons dans cette thèse comment modifier le processus empirique pour prendre en compte une information auxiliaire. Nous démontrons aussi qu'apporter une information auxiliaire au niveau du processus empirique permet d'améliorer la qualité des estimations statistiques ainsi que la puissance de tests statistiques usuels. Le premier chapitre regroupe les principales définitions ainsi que les résultats importants utilisés dans cette thèse. Dans le second et troisième chapitre, nous étudions le cas particulier où l'information auxiliaire est respectivement donnée par la probabilité d'ensembles d'une ou de plusieurs partition(s) donnée(s). En particulier, le troisième chapitre se focalise sur la méthode du Raking-Ratio, méthode très utilisée en statistique permettant de combiner la connaissance de la probabilité d'ensembles de plusieurs partitions. Dans le quatrième chapitre, nous généralisons la définition d'information auxiliaire tout en conservant la possibilité d'établir des résultats d'approximation forte, au prix d'une perte de généralisation. Dans le dernier chapitre, nous établissons l'approximation forte du processus empirique dans le cas de la méthode bootstrap et nous combinons la méthode bootstrap avec celle du Raking-Ratio.

ABSTRACT

This thesis deals with the study of the empirical process with auxiliary information, that is to say information that one would have a priori or that one would have obtained with a source of information. We show in this thesis how to modify the empirical process to take into account auxiliary information. We also show that providing auxiliary information at the empirical process level improves the quality of statistical estimates as well as the power of standard statistical tests. The first chapter contains the main definitions as well as the important results used in this thesis. In the second and third chapter, we study the particular case where the auxiliary information is respectively given by the probability of sets of one or more given partition(s). In particular, the third chapter focuses on the method of Raking-Ratio, a method widely used in statistics to combine the knowledge of the probability of sets of several partitions. In the fourth chapter, we generalize the definition of auxiliary information while retaining the possibility of establishing strong approximation results, at the cost of a loss of generalization. In the last chapter, we establish the strong approximation of the empirical process in the case of the bootstrap method and we combine the bootstrap method with that of the Raking-Ratio.