



HAL
open science

U - Statistics and non parametric statistics for dependent data

Sinda Kharrat

► **To cite this version:**

Sinda Kharrat. U - Statistics and non parametric statistics for dependent data. Statistics [math.ST].
Université Paris Cité, 2023. English. NNT : 2023UNIP7042 . tel-04519696

HAL Id: tel-04519696

<https://theses.hal.science/tel-04519696>

Submitted on 25 Mar 2024

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 Paris Cité

École doctorale Sciences Mathématiques de Paris Centre (ED386)
Laboratoire Mathématiques Appliquées à Paris 5 (MAP5), CNRS UMR 8145

U - Statistiques et statistique non paramétrique pour données dépendantes

par Sinda AMMOUS

Thèse de doctorat de mathématiques appliquées
Dirigée par Jérôme DEDECKER
et Codirigée par Céline DUVAL

Présentée et soutenue publiquement le 17 mars 2023

Devant un jury composé de :

Jérôme DEDECKER	Professeur, Université Paris Cité	Directeur
Céline DUVAL	Professeure, Université de Lille	Co-directrice
Mohamed EL MACHKOURI	Maître de conférences et HDR, Université de Rouen	Rapporteur
Clémentine PRIEUR	Professeure, Université de Grenoble Alpes	Rapporteuse
Claire LACOUR	Professeure, Université Gustave Eiffel	Présidente du jury
Gabriel LANG	Ingénieur HDR, MIA Paris, AgroParisTech	Examinateur
Florence MERLEVÈDE	Professeure, Université Gustave Eiffel	Examinatrice

Remerciements

Tout d'abord, j'adresse toute ma gratitude à mes deux directeurs de thèse, Jérôme Dedecker et Céline Duval de m'avoir fait honneur d'encadrer cette thèse. Je souhaite également les remercier pour leur bienveillance, disponibilité, gentillesse. Merci aussi pour vos conseils, encouragements qui m'ont permis d'arriver à ce jour. Je vous suis très reconnaissant de m'avoir initié et fait découvrir le monde de la recherche, j'ai beaucoup appris à vos côtés.

Merci beaucoup à Clémentine Prieur et Mohamed El Machkouri d'avoir consacré leur temps et leur énergie à rapporter ce manuscrit. Je vous remercie pour votre réception si positive de mon travail. Merci également à Claire Lacour, Gabriel Lang et Florence Merlevède d'avoir accepté d'être membre de jury.

Je remercie bien sûr tous les membres du laboratoire MAP5, ainsi que le personnel administratif et informatique L'UFR Math-info et de l'IUT de l'Université de Paris Cité, pour leur accueil chaleureux et bienveillant. Merci encore à Fabienne Comte et à Anne Estrade (directrice du MAP5) de permettre aux doctorants d'avoir un cadre très agréables de travail. Marie-Hélène, merci pour ta gentillesse, ta bonne humeur et tout le travail que tu accomplis. J'en profite pour remercier Max Plaisley, Arnaud Meunier et Rémy Abregel pour leur aide et dépannages informatiques.

Je remercie Florence Muri, Mohamed Mellouk et Jérôme Dedecker de m'avoir fait confiance pour m'occuper de leurs groupes de TD. Je souhaite encore remercier Fabienne Comte et Clémentine Prieur d'avoir fait partie de mon comité du mi-parcours.

Un grand merci à toutes et à tous les anciens et anciennes doctorant(e)s, les doctorant(e)s, post-doctorant(e)s, ATER du MAP5: Andréa, Allesandro, Arthur, claire, Julie, Juliana, Laurent, Mehdi, Diala, Sonia, Alexandre, Antoine, Rémi B, Rémi L, Zoé, Fabien, Allan, Vivien, Florian, Antoine M, Osmane, Yen, Safa, Mariem, Warith, Vincent, Charlie, Chabane, Eloi, Ariane, Ivan, Martin. Je vous remercie de m'avoir si bien accueilli au MAP5 pour ces 5 belles années.

Je souhaite aussi remercier ma famille. Merci à mon Mari Mohamed Ali Kharrat qui m'a soutenu durant cette aventure. Merci à Mes enfants Assil et Nassim Kharrat qui m'ont données de l'énergie pour terminer ma thèse. Merci à Yosra et Amal qui m'ont aider pour préparer le pot de ma thèse.

Je tiens à remercier également mes parent Lassaad Ammous et Souhira Rekik de m'avoir autorisé pour terminer mon Master en France ainsi pour vos encouragements et votre soutien. Je termine par remercier mes frères et mes sœurs : Nassredine, Yassine, Yosra, Syrine, Adnen.

Résumé

Dans cette thèse intitulée "U - Statistiques et statistique non paramétrique pour données dépendantes", on s'intéresse à trois problèmes distincts issus de la statistique non paramétrique pour des données dépendantes: le test de corrélation de Kendall pour données dépendantes, l'estimation adaptative de densité pour des suites α -mélangeantes et la concentration de la mesure empirique pour la distance de Wasserstein. L'objectif a été d'étendre des résultats existants pour des données indépendantes dans un cadre dépendant. Chacun de ces problèmes a été étudié dans des chapitres distincts après le chapitre 1 introductif.

Chapitre 2 Le test non paramétrique de Kendall permet de tester l'existence d'une liaison monotone entre deux variables. Soit $(X_i, Y_i)_{i \in \mathbb{Z}}$ une suite strictement stationnaire de variables aléatoires à valeurs dans \mathbb{R}^2 , que nous ne supposons pas indépendantes. Pour tester si X_1 et Y_1 sont corrélées au sens de Kendall, nous proposons une correction robuste du test de Kendall habituel étudié pour une suite $(X_i, Y_i)_{i \in \mathbb{Z}}$ de variables aléatoires iid (indépendantes et identiquement distribuées). Nous commençons par montrer numériquement que ce test est mal calibré dans le cas de variables dépendantes. Pour résoudre ce problème, nous démontrons un Théorème Central Limite pour la U -statistique de Kendall valable pour une large classe de suites dépendantes en utilisant des coefficients de dépendance proches de ceux introduits par Dedecker et Prieur [28]. La loi limite obtenue fait intervenir une variance limite pour laquelle nous proposons un estimateur. Ces résultats permettent la construction d'une procédure de test asymptotiquement bien calibré. Cette correction est valable pour une grande classe de suites $\bar{\beta}_2$ dépendantes, sous des conditions quasi-optimales sur ces coefficients. Les résultats sont ensuite illustrés sur plusieurs jeux de simulation.

Les travaux de ce chapitre ont donné lieu à un article paru dans la revue *Statistics* en 2019. Ils sont également utilisés dans un preprint déposé sur HAL, qui décrit plusieurs procédures de tests corrigés du Package `robustTest`. Ce preprint a été rédigé par 7 auteurs et est présenté en annexe de ce manuscrit.

Chapitre 3 Dans ce second travail nous proposons un estimateur adaptatif de la densité invariante d'une suite $(X_i)_{i \in \mathbb{Z}}$ de variables aléatoires réelles α -mélangeantes au sens de Rosenblatt [65] à partir des observations $(X_i)_{1 \leq i \leq n}$ de cette suite strictement stationnaire. Pour cela, nous adaptons une procédure d'estimation adaptative de densité introduite récemment dans un cadre indépendant par Duval et Kappus (2019) dans [38]. Cet estimateur adaptatif est basé sur l'estimation de la fonction caractéristique des observations et ses performances sont étudiées pour des pertes L^2 .

Cette généralisation nécessite à la fois une étude précise du biais de l'estimateur non adaptatif dans le cas mélangeant, et l'utilisation d'une inégalité de Fuk-Nagaev due à Rio [64] pour démontrer la borne oracle pour l'estimateur adaptatif. Les résultats fournissent (presque) les mêmes vitesses que les vitesses minimax du cas iid, pourvu que les coefficients de mélange décroissent plus vite que k^{-a} pour $a > 5$, ce qui constitue une amélioration par rapport aux résultats existants dans ce cadre de dépendance. Les résultats sont illustrés via plusieurs jeux de simulation.

Chapitre 4 Dans ce chapitre, nous étudions le comportement de la distance de Wasserstein $W_p^p(\mu_n, \mu)$, pour $p > 1$, entre la distribution empirique $\mu_n = (\delta_{X_1} + \dots + \delta_{X_n})/n$ et la distribution marginale μ d'une suite strictement stationnaire de variables aléatoires $(X_i)_{i \in \mathbb{Z}}$ à valeurs dans \mathbb{R}^d . L'outil principal est une inégalité de Fournier et Guillin [43] qui permet de contrôler cette distance à partir d'un écart entre μ_n et μ de la forme $\sum_{A \in \mathcal{F}} |\mu_n(A) - \mu(A)|$, où \mathcal{F} est une collection d'hyper cubes qui forment une partition d'un domaine borné de \mathbb{R}^d . Les coefficients adaptés pour contrôler les normes de ces quantités sont des coefficients de dépendance introduits par Dedecker et Prieur [28], similaires à ceux utilisés dans le chapitre 2.

Pour contrôler les moments d'ordre $r \in (1, 2)$ on peut utiliser une inégalité de type von Bahr-Esseen; pour contrôler les moments d'ordre $r > 2$, on peut utiliser une inégalité de type Burkholder. En procédant ainsi, nous obtenons des inégalités de moment pour $W_p^p(\mu_n, \mu)$ qui dépendent des valeurs respectives de p, d ainsi que du moment des variables et du taux de décroissance des coefficients de dépendance. Dans le cas du moment d'ordre $r = 2$, on retrouve les conditions de dépendance obtenues par Dedecker et Merlevède [25] en dimension 1.

Mots-clés: Distance de Wasserstein, Estimateurs adaptatifs, Inégalités de concentration, Suites mélangeantes, Tests d'hypothèses, U -statistiques, Variables aléatoires dépendantes.

Abstract

In this thesis entitled "U - Statistics and Nonparametric Statistics for Dependent Data", we investigate three distinct problems of non parametric statistics for dependent data: the Kendall correlation test, an adaptive density estimator for α -dependent variables and the concentration of the empirical measure for the Wasserstein distance. The purpose is to generalize existing results of the independent setting to some dependent settings. Each of these problems is studied in distinct chapters after an introductory chapter 1.

Chapter 2 Kendall's test is a non parametric testing procedure enabling to detect the existence of a monotonic relation between two random variables. Let $(X_i, Y_i)_{i \in \mathbb{Z}}$ be a strictly stationary sequence of \mathbb{R}^2 -valued random variables. To test whether X_1 and Y_1 are correlated in Kendall's sense, we propose a robust correction of the usual Kendall test for iid (independent and identically distributed) random variables (X_i, Y_i) . We show numerically that the usual Kendall test does not have the desired level in case of dependent data. To solve this problem we prove a Central Limit Theorem for Kendall's U -statistic valid over a large class of dependent sequences, using dependence coefficients close to the ones introduced in Dedecker and Prieur [28]. The limit distribution depends on an unknown variance which is consistently estimated. These results enable to build an asymptotically well calibrated testing procedure. This correction of the usual Kendall procedure is valid for a large class of $\bar{\beta}_2$ dependent sequences, the conditions imposed on the coefficients are almost optimal. The results are then illustrated numerically on several simulated examples.

The results of this chapter are published in the journal *Statistics* in 2019. They are also used in a recent preprint available on HAL, which describes several corrected testing procedures of the Package `robustTest`. This preprint is a joint work with 7 authors and is presented in the Appendix of the present manuscript.

Chapter 3 In this second work we propose an adaptive density estimator of the common density of a strictly stationary sequence $(X_i)_{i \in \mathbb{Z}}$ of real-valued random variables. These variables are supposed to be α -mixing in the sense of Rosenblatt [65]. To that end we generalize an adaptive density estimation procedure recently introduced in an independent framework by Duval and Kappus [38]. This procedure relies on the estimation of the characteristic function of the observations and its performances are studied for the integrated L^2 loss. This generalization requires both a precise study of the bias of the non adaptive estimator in the case of α -mixing variables, and the use of a Fuk-Nagaev inequality proved by Rio [64] to establish the oracle inequality for the adaptive estimator. The obtained rates of convergence are (almost) the same as the minimax rates known in the iid case as long as the decay of the mixing coefficients is faster than k^{-a} for $a > 5$. This is an improvement of existing results in the α -mixing framework. The procedure is then illustrated on different sets of simulations.

Chapter 4 In this chapter we study the behavior of the Wasserstein distance $W_p^p(\mu_n, \mu)$, $p > 1$, between the empirical distribution $\mu_n = (\delta_{X_1} + \dots + \delta_{X_n})/n$ and the marginal distribution μ of a stationary sequence $(X_i)_{i \in \mathbb{Z}}$ of \mathbb{R}^d -valued random variables. The main tool is an inequality

due to Fournier and Guillin [43] which allows to control this distance by a quantity of the type $\sum_{A \in \mathcal{F}} |\mu_n(A) - \mu(A)|$, where \mathcal{F} is a collection of hyper cubes that forms a partition of a bounded domain of \mathbb{R}^d . The dependency coefficients introduced by Dedecker and Prieur in [28] are well adapted to control the norms of such quantities (these coefficients are similar to the ones used in chapter 2).

To get an upper bound for the moments of order $r \in (1, 2)$ we can apply an inequality of von Bahr-Esseen type; for the moments of order $r > 2$, we rely on an inequality of Burkholder type. Then, we derive moment inequalities for $W_p^p(\mu_n, \mu)$ depending on the quantities p, d , on the moments of the variables as well as on the decay rate of the dependency coefficients. In the case $r = 2$ we recover the same conditions on the dependency coefficients as the ones obtained by Dedecker and Merlevède [25] in dimension 1.

Keywords: Adaptive estimators, Concentration inequalities, Dependent random variables, Hypothesis testing, U -statistics, Wasserstein distance.

Sommaire

1	Introduction	1
1.1	Motivation	2
1.2	Coefficients de mélange considérés	2
1.3	Test de Kendall dans un cadre dépendant	4
1.3.1	Généralités sur le test de Kendall	4
1.3.2	Résultats	6
1.4	Estimation adaptative de densité	7
1.4.1	Motivations	7
1.4.2	Définitions et résultats	8
1.4.3	Estimation adaptative	9
1.4.4	Vitesse de convergence	10
1.4.5	Perspectives	11
1.5	Distance de Wasserstein	11
1.5.1	Notations	11
1.5.2	Généralités sur la distance de Wasserstein	12
1.5.3	Objectifs et outils utilisés	12
1.5.4	Les résultats obtenus	13
1.5.5	Perspectives	15
2	Testing Kendall's τ	17
2.1	Introduction and definitions	18
2.2	Main results	20
2.3	Quasi-optimality of the dependency conditions	23
2.4	Simulations	25
2.4.1	First example	25
2.4.2	Second example	27
2.4.3	Third example	28
2.5	Proofs	30
2.5.1	Proof of theorem 2.1	30
2.5.2	Proof of proposition 2.1	31
2.6	Appendix	34
2.6.1	Proof of lemma 2.1	34
2.6.2	Proof of proposition 2.3	36
2.6.3	Proof of lemma 2.2	42

3	Adaptive density estimation	45
3.1	Introduction	46
3.1.1	Motivations	46
3.1.2	α -mixing sequences	47
3.1.3	Fourier density estimator and the sinc kernel	48
3.2	Main results	48
3.2.1	An upper bound for the quadratic risk	49
3.2.2	Adaptive procedure	50
3.2.3	Comments and rates of convergence	51
3.3	Simulations	54
3.3.1	Methodology	54
3.3.2	Numerical results	56
3.3.3	Robustness to another dependent model	60
3.4	Proofs	61
3.4.1	Preliminary	61
3.4.2	Proof of theorem 3.1	63
4	The empirical Wasserstein distance	67
4.1	Introduction and notations	68
4.2	First upper bounds	70
4.3	Upper bounds under polynomial mixing conditions	75
4.3.1	Moment of order $r = 2$	75
4.3.2	Moment of order $r \in (1, 2)$	83
4.3.3	Moment of order $r > 2$	84
4.4	Proofs	85
4.4.1	Proof of proposition 4.5	86
4.4.2	Proof of proposition 4.6	88
4.4.3	Proof of proposition 4.7	92
4.4.4	Proof of proposition 4.8	95
4.5	Appendix	98
A	The <i>robustTest</i> package	107
A.1	Introduction	108
A.2	Robust tests for testing correlation and independence	109
A.2.1	Pearson's correlation test	109
A.2.2	Kendall's correlation test	110
A.2.3	Spearman's correlation test	111
A.2.4	Kolmogorov–Smirnov statistic for testing independence	112
A.2.5	Simulation study	112
A.3	Robust test for the equality of variances	115
A.3.1	Simulation study	117
A.4	Robust test for testing stochastic dominance	118
A.4.1	Simulations	119
A.5	Other tests of the <i>robustTest</i> package for paired two-sample	121
A.6	Testing correlation on a real data set	123

Chapter 1

Introduction:

Contents

1.1	Motivation	2
1.2	Coefficients de mélange considérés	2
1.3	Test de Kendall dans un cadre dépendant	4
1.3.1	Généralités sur le test de Kendall	4
1.3.2	Résultats	6
1.4	Estimation adaptative de densité	7
1.4.1	Motivations	7
1.4.2	Définitions et résultats	8
1.4.3	Estimation adaptative	9
1.4.4	Vitesse de convergence	10
1.4.5	Perspectives	11
1.5	Distance de Wasserstein	11
1.5.1	Notations	11
1.5.2	Généralités sur la distance de Wasserstein	12
1.5.3	Objectifs et outils utilisés	12
1.5.4	Les résultats obtenus	13
1.5.5	Perspectives	15

1.1 Motivation

L'objectif de cette thèse est d'étudier des problèmes issus de la statistique non paramétrique (i.e. où le modèle statistique n'est pas décrit par un nombre fini de paramètres) lorsque les observations sont issues de variables aléatoires dépendantes. On étudie trois problèmes distincts qui sont classiques en statistique non paramétrique : un test de corrélation construit à partir d'une U-statistique, la concentration de la mesure empirique et l'estimation de densité. Ces questions ont été largement étudiés pour des données indépendantes, et nous étendons certains résultats au cadre dépendant.

L'intérêt d'obtenir des résultats pour des données dépendantes est d'abord de montrer que certaines procédures sont robustes par rapport à l'hypothèse très contraignante de l'indépendance des variables, mais réside aussi dans le fait qu'en pratique il n'est souvent par raisonnable de supposer que les observations sont issues de variables indépendantes : beaucoup de données proviennent de séries temporelles pour lesquelles les variables sont naturellement corrélées; c'est le cas des données socio-économiques (données de production, de consommation, indices socio-économiques, données financières...), mais aussi de données météorologiques ou climatiques, des données de trafic passager (routiers, aérien, ferroviaire...), et la liste n'est bien sûr pas exhaustive.

Ce chapitre introduit les notions de dépendances considérées dans ce manuscrit ainsi que quelques outils probabilistes utilisés tout au long de la thèse. Nous détaillons également les trois problèmes étudiés et les résultats obtenus dans chacun des cas. Il s'agit d'adapter le test de corrélation de Kendall aux données $\bar{\beta}_2$ dépendantes (chapitre 2), de proposer une méthode adaptative d'estimation de la densité de variables α -mélangeantes (chapitre 3) et de donner des bornes pour les moments de la distance de Wasserstein entre la mesure empirique et la loi marginale de données $\tilde{\beta}_1$ -dépendantes à valeurs dans \mathbb{R}^d (chapitre 4).

1.2 Coefficients de mélange considérés

Dans l'ensemble de ce manuscrit, nous considérons des suites strictement stationnaires de variables aléatoires dépendantes à valeur dans \mathbb{R} ou dans \mathbb{R}^d . Rappelons qu'une suite $(X_i)_{i \in \mathbb{Z}}$ de variables aléatoires à valeur dans \mathbb{R}^d (ou dans un espace plus général) est strictement stationnaire lorsque : pour tous $k \in \mathbb{Z}$ et $\ell \in \mathbb{N}$, le $\ell + 1$ -uplet $(X_k, \dots, X_{k+\ell})$ a même loi que (X_0, \dots, X_ℓ) . La dépendance entre les variables peut être mesurée à l'aide de plusieurs coefficients de dépendance. Nous utilisons les coefficients $\tilde{\beta}_2$ et $\bar{\beta}_2$ dans le chapitre 2; le coefficient de mélange fort (α -mélange) dans le chapitre 3; le coefficient $\tilde{\beta}_1$ dans le chapitre 4.

Donnons à présent la définition de ces différents coefficients.

α -mélange Commençons par définir le coefficient de α -mélange, introduit par Rosenblatt en (1956) [65], qui sera utilisé dans le Chapitre 3.

Definition 1.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, et $(X_i)_{i \in \mathbb{Z}}$ une suite strictement stationnaire de variables aléatoires réelles. Soient $\mathcal{F}_0 = \sigma(X_i, i \leq 0)$ et $\mathcal{G}_k = \sigma(X_i, i \geq k)$. Le coefficient de mélange fort $(\alpha(n))_{n \geq 0}$ de $(X_i)_{i \in \mathbb{Z}}$ introduit par Rosenblatt (1956) [65], est défini par

$$\alpha(0) = \frac{1}{2} \text{ et } \alpha(n) = \alpha(\mathcal{F}_0, \mathcal{G}_n),$$

où, pour deux σ -algèbres \mathcal{F}, \mathcal{G} ,

$$\alpha(\mathcal{F}, \mathcal{G}) = 2 \sup_{A \in \mathcal{F}, B \in \mathcal{G}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| .$$

Remark 1.1. Nous utilisons ici la définition du livre de Rio [64], de sorte que la définition de $\alpha(\mathcal{F}, \mathcal{G})$ diffère de celle de Rosenblatt [65] d'un facteur 2.

Les liens entre les coefficients de mélange impliquent que toute suite β -mélangeante est aussi α mélangeante (avec une vitesse de décroissance au moins aussi rapide; voir Volkonskii et Rozanov [73] pour la définition du coefficient de β -mélange). Nous utiliserons des suites β -mélangeantes lors des simulations des chapitres 2 et 3, pour illustrer nos résultats théoriques.

Coefficients $\tilde{\beta}_1$ Définissons à présent les coefficient $\tilde{\beta}_1$ pour une suite de variables aléatoires strictement stationnaire et à valeurs dans \mathbb{R}^d . Ces coefficients ont été introduits par Dedecker et Prieur (2007) [28]. Ils seront utilisés dans le chapitre 2 (pour des variables à valeur dans \mathbb{R}^2) et dans le chapitre 4.

Definition 1.2. Soit $(X_i)_{i \in \mathbb{Z}}$ une suite strictement stationnaire de variables aléatoires réelles à valeurs dans \mathbb{R}^d avec μ comme loi de X_0 . Soit $\mathcal{F}_0 = \sigma(X_k; k \leq 0)$ et soit $\mathbb{P}_{X_k | \mathcal{F}_0}$ la distribution conditionnelle de X_k sachant \mathcal{F}_0 .

Définissons alors

$$b(k) = \sup_{(t_1, \dots, t_d) \in \mathbb{R}^d} |P_{X_k | \mathcal{F}_0}([-\infty, t_1] \times \dots \times]-\infty, t_d]) - \mu([-\infty, t_1] \times \dots \times]-\infty, t_d])|$$

et

$$\tilde{\beta}_1(k) = \mathbb{E}(b(k)) .$$

De même, ces coefficients sont moins restrictifs que les coefficients de β -mélange usuels. De nombreux exemples de suites non mélangeantes au sens de Rosenblatt [65] pour lesquelles les coefficients $\tilde{\beta}_1(k)$ peuvent être calculés sont présentés dans l'article [28].

Coefficients $\tilde{\beta}_2$ et $\bar{\beta}_2$ Dans le chapitre 2 et dans ce paragraphe, nous considérons $Z_i = (X_i, Y_i)_{i \in \mathbb{Z}}$ une suite strictement stationnaire de variables aléatoires à valeurs dans \mathbb{R}^2 . Le coefficient β_1 est donc défini pour cette suite comme dans le paragraphe précédent. Pour pouvoir contrôler les différents termes qui apparaissent lorsque l'on cherche à démontrer la normalité asymptotique de la U-statistique de Kendall dans un cadre dépendant, nous aurons aussi besoin des coefficients suivants.

Definition 1.3. Soit $Z_i = (X_i, Y_i)_{i \in \mathbb{Z}}$ une suite strictement stationnaire de variables aléatoires à valeurs dans \mathbb{R}^2 , et soit $\mathcal{F}_0 = \sigma(Z_k; k \leq 0)$. Soit $\mathbb{P}_{(Z_i, Z_j)}$ la loi du couple (Z_i, Z_j) , et $\mathbb{P}_{(Z_i, Z_j) | \mathcal{F}_0}$ la distribution conditionnelle du couple (Z_i, Z_j) sachant \mathcal{F}_0 .

Pour tous $s, t \in \mathbb{R}$ et $z = (x, y) \in \mathbb{R}^2$, on définit

$$f_z(s, t) = \mathbb{1}_{x \leq s} \mathbb{1}_{y \leq t} - \mathbb{P}(X_0 \leq s, Y_0 \leq t) .$$

On introduit de plus

$$b(i, j) = \sup_{(z_1, z_2) \in \mathbb{R}^2 \times \mathbb{R}^2} | \mathbb{P}_{(Z_i, Z_j) | \mathcal{F}_0}(f_{z_1} \otimes f_{z_2}) - \mathbb{E}(f_{z_1}(Z_i) f_{z_2}(Z_j)) |$$

où $f_{z_1} \otimes f_{z_2} = f_{z_1}(Z_i)f_{z_2}(Z_j)$. On définit enfin les coefficients

$$\begin{aligned}\tilde{\beta}_2(k) &= \max\{\tilde{\beta}_1(k), \sup_{i>j\geq k} \mathbb{E}[b(i,j)]\}, \\ \delta_2(k) &= \sup_{i>j\geq k} |\mathbb{E}(\mathbb{1}_{X_i \leq X_j} \mathbb{1}_{Y_i \leq Y_j} - \mathbb{P}(X_i \leq X_j, Y_i \leq Y_j) | \mathcal{F}_0)|, \\ \bar{\beta}_2(k) &= \max\{\tilde{\beta}_2(k), \delta_2(k)\}.\end{aligned}$$

Le coefficient $\bar{\beta}_2$ est moins restrictif que le coefficient de β -mélange usuel de la suite $(X_i, Y_i)_{i \in \mathbb{Z}}$. Par conséquent, tous les résultats de Chapitre 2 sont valables lorsque l'on remplace les coefficients $\tilde{\beta}_1, \tilde{\beta}_2, \bar{\beta}_2$ par le coefficient de β -mélange usuel.

Le coefficient $\bar{\beta}_2$ peut être calculé pour une grande classe de processus, y compris de nombreuses suites non-mélangeantes au sens de Rosenblatt [65]. Cela découle principalement de l'article de Dedecker et Prieur [28], qui fournit de nombreux exemples de processus (éventuellement non-mélangeants) pour lesquels le coefficient $\tilde{\beta}_2$ peut être facilement contrôlé. Le coefficient $\bar{\beta}_2$ est un peu plus restrictif que $\tilde{\beta}_2$, à cause du terme δ_2 , qui n'est pas si facile à manipuler. Toutefois, dans de nombreux cas, le coefficient δ_2 peut être traité comme $\tilde{\beta}_2$ en suivant les calculs de la Section 6 dans [28].

1.3 Test de Kendall dans un cadre dépendant

Le travail présenté au chapitre 2 a abouti à la rédaction de l'article [2] publié dans le journal *Statistics*. Une partie des résultats de cet article a aussi été repris dans le preprint [3] (soumis à *R journal*) qui illustre le package R `robustTest`. Il s'agit d'un travail en collaboration avec 6 autres auteurs, que nous présentons en annexe de cette thèse.

1.3.1 Généralités sur le test de Kendall

Il existe différents tests non paramétriques pour déterminer si deux variables aléatoires continues sont liées ou non : on peut citer le test de corrélation Pearson, ou les tests de Spearman ou Kendall qui sont bien adaptés pour mettre en évidence des liaisons monotones non linéaires. Ces deux derniers tests sont souvent utilisés par les bio-statisticiens, qui les préfèrent au test de Pearson pour sa robustesse et le fait qu'il ne requiert pas d'hypothèses de moment sur les variables.

Les résultats présentés ici concernent une correction robuste du test de Kendall usuel pour une grande classe de suites dépendantes, sous des conditions quasi-optimales sur les coefficients $\bar{\beta}_2$.

Commençons par définir le coefficient de corrélation de Kendall.

Definition 1.4. Soit (X, Y) un couple de variables aléatoires continues à valeurs réelles et (X^*, Y^*) une copie indépendante de (X, Y) . Le coefficient τ de corrélation de Kendall entre X et Y est défini par

$$\tau := 2(\mathbb{P}(\{(X^* - X)(Y^* - Y) > 0\}) - 0.5).$$

Si $\tau = 0$, X et Y ne sont pas corrélées au sens de Kendall, et si $\tau > 0$ (resp. $\tau < 0$), il y a une corrélation positive (resp. négative), ce qui signifie que X et Y ont tendance à varier dans

la même direction (resp. dans des directions opposées).

Nous disposons d'observations issues d'une suite stationnaire $Z_i = (X_i, Y_i)_{1 \leq i \leq n}$ de variables aléatoires à valeurs dans \mathbb{R}^2 , de même distribution marginale que (X, Y) . Pour estimer τ , l'idée de Kendall [54] est de compter le nombre de paires concordantes (les couples (X_i, Y_i) et (X_j, Y_j) qui vérifient: $(X_i - X_j)(Y_i - Y_j)$ est strictement positif).

Pour tester $H_0 : \tau = 0$ contre $H_1 : \tau \neq 0$, nous utilisons la U-statistique de Kendall

$$U_n = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{1}_{\{(X_j - X_i)(Y_j - Y_i) > 0\}},$$

qui a été étudié par Esscher [40], Lindeberg [58, 59], et Kendall [54] dans le cas où la suite $Z_i = (X_i, Y_i)$ est une suites de variables indépendantes et identiquement distribuées (iid).

Notons que, dans le cas iid, cette U-statistique est un estimateur sans biais de $(\tau + 1)/2$ (d'où la lettre U pour *unbiased*), mais que ce n'est plus le cas lorsque les variables sont dépendantes.

D'après la définition de U_n , il est clair que si la suite (Z_i) est une suite de variables iid et que de plus les variables X et Y sont indépendantes (ce qui implique que $\tau = 0$), alors la loi de U_n ne dépend pas de la loi de (X, Y) (on peut se ramener à des lois uniformes). Dans le cas contraire, ce n'est a priori plus le cas, et U_n n'est plus une statistique libre sous $H_0 : \tau = 0$.

Ainsi, dans le cas de variables dépendantes, le test de Kendall usuel est mal calibré. Cette mauvaise calibration peut s'illustrer numériquement : simulons (X_i) et (Y_i) selon les mécanismes auto régressifs suivant :

$$\begin{aligned} & - \begin{cases} X_i = \frac{1}{2}(X_{i-1} + \varepsilon_i) & \text{avec } (\varepsilon_i)_{i \geq 1} \text{ iid, et } \varepsilon_i \sim \mathcal{B}\left(\frac{1}{2}\right) \\ X_0 \sim \mathcal{U}[0, 1] & \text{avec } X_0 \text{ indépendant de } (\varepsilon_i)_{i \geq 1} \end{cases} \\ & - \begin{cases} Y_i = \frac{1}{2}(Y_{i-1} + \varepsilon'_i) & \text{où } (\varepsilon'_i)_{i \geq 0} \text{ iid, et } \varepsilon'_i \sim \mathcal{B}\left(\frac{1}{2}\right) \\ Y_0 \sim \mathcal{U}[0, 1] & \text{avec } Y_0 \text{ indépendant de } (\varepsilon'_i)_{i \geq 1} \end{cases} \\ & - (X_0, (\varepsilon_i)) \text{ est indépendant de } (Y_0, (\varepsilon'_i)), \end{aligned}$$

Dans cet exemple, les deux suites (X_i) et (Y_i) sont indépendantes (de sorte que $\tau = 0$), mais composées de variables dépendantes (de loi marginales uniformes sur $[0, 1]$). On simule $N = 2000$ fois n observations $(x_k, y_k)_{1 \leq k \leq n}$ tirées selon le modèle décrit ci-dessus, pour différentes valeurs de n . On applique alors le test de Kendall usuel sur ces observations, et on calcule la fréquence de rejet de ces 2000 tests au niveau de risque 0.05. Puisque H_0 est vraie, si le test était bien calibré, la fréquence de rejet (c'est à dire le niveau estimé) devrait être proche de 0.05.

Nous présentons les résultats dans le tableau suivant:

n	150	200	250	300	350	400	500	600
Niveau estimé	0.132	0.127	0.1145	0.125	0.129	0.144	0.125	0.131

Le niveau estimé du test de Kendall usuel est de l'ordre de 0,13 au lieu de 0,05 ; ceci est dû au fait que le test de Kendall usuel ne prend pas en compte la dépendance des variables.

Pour résoudre ce problème, nous étendons l'étude de la distribution limite de $\sqrt{n}(2U_n - \tau - 1)$ donnée par Hoeffding [50] dans le cadre indépendant (voir aussi van der Vaart [71], Exemple 12.5) au cadre dépendant. Cette loi limite fera intervenir un terme de variance inconnu que nous estimons. A partir de ces résultats nous proposons une procédure asymptotiquement valide pour tester $H_0 : \tau = 0$ contre $H_1 : \tau \neq 0$.

1.3.2 Résultats

Introduisons d'abord quelques notations. Notons

$$\pi := \frac{\tau}{2} + 0.5 = \mathbb{P}(\{(X^* - X)(Y^* - Y) > 0\}),$$

$$F(x, y) = \mathbb{P}(X < x, Y < y), \quad H(x, y) = \mathbb{P}(X > x, Y > y),$$

et

$$F_X(x) = \mathbb{P}(X < x), \quad H_X(x) = \mathbb{P}(X > x).$$

Notre résultat principal est le théorème suivant. Les coefficients de dépendance qui y apparaissent ont été introduit en définition 1.3.

Theorem 1.1. *Soit $(X_i, Y_i)_{i \in \mathbb{Z}}$ une suite strictement stationnaire de variables aléatoires à valeurs dans \mathbb{R}^2 . Supposons que*

$$\sum_{k=1}^{\infty} \tilde{\beta}_1(k) < \infty \quad \text{et} \quad k \bar{\beta}_2(k) \xrightarrow[k \rightarrow +\infty]{} 0$$

alors,

$$\sqrt{n}(U_n - \pi) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, V),$$

avec

$$\begin{aligned} V = & 4\text{Var}(F(X_0, Y_0) + H(X_0, Y_0)) \\ & + 8 \sum_{k=1}^{\infty} \text{Cov}(F(X_0, Y_0) + H(X_0, Y_0), F(X_k, Y_k) + H(X_k, Y_k)). \end{aligned} \quad (1.1)$$

On constate que la loi limite de la statistique $\sqrt{n}(U_n - 0.5)$ dépend de la quantité inconnue V . Ainsi pour pouvoir calibrer un test de $H_0 : \tau = 0$ contre $H_1 : \tau \neq 0$ nous proposons un estimateur consistant de la variance V .

Proposition 1.1. *Soit $(X_i, Y_i)_{i \in \mathbb{Z}}$ une suite strictement stationnaire de variables aléatoires à valeurs dans \mathbb{R}^2 . Supposons que*

$$\sum_{k=1}^{\infty} \tilde{\beta}_2(k) < \infty,$$

et définissons

$$F_n(s, t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i < t} \mathbf{1}_{Y_i < t}, \quad H_n(s, t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i > s} \mathbf{1}_{Y_i > t}, \quad G_n(s, t) = 2(F_n(s, t) + H_n(s, t)),$$

et

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{i=1}^{n-k} (G_n(X_i, Y_i) - \bar{G}_n)(G_n(X_{i+k}, Y_{i+k}) - \bar{G}_n),$$

avec $\bar{G}_n = \frac{1}{n} \sum_{i=1}^n G_n(X_i, Y_i)$. Soit (a_n) une suite d'entiers positifs tendant vers l'infini lorsque n tend vers l'infini, de sorte que $a_n = o(\sqrt{n}/(\log n)^2)$. Alors,

$$V_n := \hat{\gamma}(0) + 2 \sum_{k=1}^{a_n} \hat{\gamma}(k) \xrightarrow[n \rightarrow +\infty]{\mathbb{L}^2} V.$$

L'estimateur V_n dépend de la suite (a_n) . Ce qui soulève la question du choix de la quantité (a_n) , qui est une question difficile. Nous proposons un moyen simple de choisir numériquement cette suite. On peut aussi envisager un estimateur data-driven de V , tel que celui proposé dans Caron, Dedecker, Michel [15], qui devrait fonctionner dans notre cas. Pour plus de détail voir la section 2.4 du chapitre 2.

En combinant ces deux résultats on obtient que, si $V > 0$,

$$T_n := \frac{\sqrt{n}(U_n - 1/2)}{\sqrt{|V_n|}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}, H_0} \mathcal{N}(0, 1).$$

La région de rejet du test de Kendall corrigé est de la forme $R_{n,\alpha} = \{|T_n| > q_\alpha\}$ où q_α , est le quantile d'ordre $1 - (\alpha/2)$ de la loi normale standard. Ce test est donc asymptotiquement bien calibré.

Comparaison avec un résultat existant: Une correction du test de Kendall dans un contexte dépendant à été récemment proposée par Dehling *et al.* [30]. Leur résultat est valable pour une grande classe de suites dépendantes, avec une intersection importante avec notre classe de suites $\bar{\beta}_2$ -dépendantes. Un avantage de notre approche est que nous sommes en mesure de prouver que notre condition sur le coefficient $\bar{\beta}_2$ est quasi optimale (voir la proposition 2.2 du chapitre 2). De plus, si nous ne considérons que la classe des suites β -mélangeantes, alors nos résultats sont valables sous la condition $\sum_{k>0} \beta(k) < \infty$, alors que la condition dans [30] ne peut pas être meilleure que $\sum_{k>0} k\beta(k) < \infty$.

1.4 Estimation adaptative de densité

Dans cette seconde partie, nous cherchons à mettre au point une procédure adaptative pour l'estimation de la densité marginale d'une suite stationnaire de variables α -mélangeantes. Cette procédure est basée sur l'estimation de la fonction caractéristique des variables.

1.4.1 Motivations

Il existe une vaste littérature sur les procédures adaptatives en statistique non paramétrique, elles consistent à construire un estimateur sans aucune connaissance préalable de la régularité de la fonction à estimer. Il existe plusieurs méthodes pour obtenir des estimateurs de densité adaptatifs; parmi elles nous mentionnons la méthode de Goldenshluger et Lepski (voir e.g. [45] ou [46]) ou sa récente amélioration PCO (Penalized Comparison to Overfitting) proposée par Lacour *et al.* [55], la sélection de modèles (voir e.g. Massart [60], Barron *et al.* [8] ou Birgé et Massart [12]), ou encore le seuillage (on peut se reporter à Donoho *et al.* [34, 35, 36] ou Härdle *et al.* [49] par exemple). Ces travaux étudient le plus souvent le cadre indépendant.

Dans un article récent, Duval et Kappus [38] ont proposé une procédure d'adaptation simple dans un cadre i.i.d. pour des estimateurs obtenus à partir de l'estimation de la fonction caractéristique. Cette méthode présente l'avantage suivant: la preuve de l'inégalité oracle est basée sur un simple découpage du risque quadratique, et l'outil principal pour contrôler les fluctuations de l'estimateur est une inégalité de Hoeffding, alors que, les outils habituellement utilisés pour établir des inégalités oracles reposent sur des résultats de concentration beaucoup plus exigeants comme les inégalités de concentration de type Talagrand. Cependant, avec cette procédure nous

perdrons une puissance d'un facteur logarithmique dans la vitesse. Cette méthode peut être adapté au cas α -dépendant en utilisant l'inégalité de Fuk-Nagaev suivante (voir Rio [64] inégalité (1) page 111).

Inégalité de Fuk-Nagaev : Soit $(X_i)_{i>0}$ une suite stationnaire de variables aléatoires à valeurs réelles telle que $\|X_i\|_\infty \leq 1/2$, et soit $(\alpha(n))_{n \geq 0}$ la suite de coefficients de mélange fort (Définition 1.1). Soit $S_k = \sum_{i=1}^k (X_i - \mathbb{E}(X_i))$ et

$$s_n^2 \geq \max \left(1, \sum_{i=1}^n \sum_{j=1}^n |\text{Cov}(X_i, X_j)| \right).$$

Alors pour tout $\lambda \in [s_n, s_n^2]$,

$$\mathbb{P} \left(\sup_{k \in [1, n]} |S_k| \geq 4\lambda \right) \leq 4 \exp \left(\frac{-\lambda^2}{4s_n^2} \right) + 4n\lambda^{-1} \alpha \left(\left\lceil \frac{s_n^2}{\lambda} \right\rceil \right). \quad (1.2)$$

1.4.2 Définitions et résultats

Soit $(X_i)_{i \in \mathbb{Z}}$ une suite stationnaire de variables aléatoires α -mélangeantes de densité f par rapport à la mesure Lebesgue. La notion de α -mélange que nous considérons est celle introduite par Rosenblatt [65], voir définition 1.1.

Nous introduisons quelques notations qui seront utilisées dans le reste de cette partie. Soit $\varphi_X(u) = \mathbb{E}[e^{iuX}]$ la fonction caractéristique d'une variable aléatoire X . On écrit $\|\cdot\|$ pour la norme L^2 d'une fonction c'est à dire $\|f\|^2 = \int |f(x)|^2 dx$.

Pour estimer la densité f de X , nous construisons un estimateur $\widehat{\varphi}_X$ de φ_X comme suit

$$\widehat{\varphi}_{X,n}(u) = \frac{1}{n} \sum_{j=1}^n e^{iuX_j}, u \in \mathbb{R}.$$

En appliquant la transformation de Fourier inverse et en tronquant le domaine spectral (en effet $\widehat{\varphi}_{X,n}(u)$ n'appartient pas toujours à $L^1(\mathbb{R})$), nous obtenons un estimateur de f de la forme

$$\widehat{f}_m(x) = \frac{1}{2\pi} \int_{-m}^m e^{-iux} \widehat{\varphi}_{X,n}(u) du, \quad m > 0, x \in \mathbb{R}.$$

Afin de mesurer la qualité de cet estimateur, nous étudions le risque quadratique associé. L'étude diffère du cadre i.i.d. classique pour le terme de variance; pour cela comme dans Comte, Dedecker et Taupin (2008)[16], nous utilisons l'inégalité de covariance d'Ibragimov (voir [64] theorem 1.1 page 5) qui fait apparaître les coefficients du mélange.

Proposition 1.2. Soit m une constante positive, alors l'estimateur \widehat{f}_m défini dans (3.2) satisfait l'inégalité suivante,

$$\mathbb{E} \left[\|\widehat{f}_m - f\|_2^2 \right] \leq \frac{1}{2\pi} \int_{[-m, m]^c} |\varphi_X(u)|^2 du + \frac{m}{n\pi} \left(1 + 8 \sum_{k=1}^{n-1} \alpha(k) \right).$$

Pour calculer des vitesses de convergence induites par cette borne de risque, nous considérons des classes de régularité de type Sobolev afin de contrôler l'ordre du terme de biais. Supposons que f appartienne à une boule de régularité Sobolev s , $s \geq 0$, définie comme suit

$$\mathcal{S}(s, L) := \left\{ f \in \mathbf{F}; \int_{\mathbb{R}} (1 + |u|)^{2s} |\mathcal{F}f(u)|^2 du \leq L \right\}$$

avec \mathbf{F} l'ensemble des densités par rapport à la mesure Lebesgue et $\mathcal{F}f(u) = \int e^{iux} f(x) dx$ est la transformée de Fourier de f . D'après la Proposition 1.2, pour $f \in \mathcal{S}(s, L)$ et C une constante positive indépendante de m et n , nous trouvons que

$$\mathbb{E} \left[\|\widehat{f}_m - f\|^2 \right] \leq C m^{-2s} + \frac{m}{n\pi} \left(1 + 8 \sum_{k=1}^{n-1} \alpha(k) \right) \leq C' \left(m^{-2s} + \frac{m}{n} \right)$$

sous l'hypothèse additionnelle suivante: il existe une constante positive A_α telle que,

$$\sum_{k=1}^{\infty} \alpha(k) \leq A_\alpha < \infty. \quad (\mathcal{A}(\alpha))$$

En sélectionnant $m = m^* \asymp n^{\frac{1}{2s+1}}$, qui réalise le meilleur compromis entre le terme de biais et le terme de variance, on constate que la vitesse de convergence de l'estimateur \widehat{f}_{m^*} est d'ordre $n^{\frac{-2s}{2s+1}}$, qui est la vitesse de convergence optimale pour des suites i.i.d. Cependant, m^* n'est pas calculable en pratique car il dépend de la régularité inconnue s de f , c'est pourquoi nous proposons une procédure adaptative pour le choix de m .

1.4.3 Estimation adaptative

Nous proposons une procédure basée sur les données pour sélectionner un seuil \widehat{m}_n en s'inspirant de Duval et Kappus [38] et telle que $\widehat{f}_{\widehat{m}_n}$ atteigne les mêmes vitesses (à une puissance de log près) que \widehat{f}_{m^*} . L'intuition derrière la procédure est la suivante. Si on cherche à minimiser la borne supérieure donnée par la Proposition 1.2, cela revient à sélectionner m telle que la dérivée en m de cette borne s'annule : m^* satisfait l'équation suivante, où on note $C_\alpha = 1 + 8 \sum_{k=1}^{n-1} \alpha(k)$,

$$|\varphi_X(m^*)|^2 = \frac{C_\alpha}{n} \iff |\varphi_X(m^*)| = \sqrt{C_\alpha n^{-1/2}}.$$

L'avantage de cette dernière équation est qu'elle a une contrepartie empirique. La constante C_α est inconnue, mais sous $(\mathcal{A}(\alpha))$ cette quantité est bornée. Comme nous souhaitons estimer l'ordre de grandeur m^* , à constante près, nous estimons donc m^* par \widehat{m}_n qui est défini comme suit

$$\widehat{m}_n = \max\{m > 0 : |\widehat{\varphi}_{X,n}(m)| = \kappa_n n^{-1/2}\} \wedge n.$$

où $\kappa_n = 1 + \kappa(\log n)^\gamma$ et κ et γ sont des constantes positives. Notons que l'équation $|\widehat{\varphi}_{X,n}(m)| = \kappa_n/\sqrt{n}$ peut avoir plusieurs solutions, nous considérons le maximum pour obtenir une définition unique de \widehat{m}_n . Nous ne considérons pas de seuils supérieurs à n car pour $m \geq n$ le majorant du terme de variance est borné loin de 0 et la borne de risque précédente ne permet plus de conclure à la consistance de l'estimateur. Les constantes κ et γ doivent être calibrées, nous discutons de

leur choix dans le chapitre 3.

Nous considérons l'estimateur adaptatif suivant,

$$\tilde{f}_{\hat{m}_n}(x) = \frac{1}{2\pi} \int_{-\hat{m}_n}^{\hat{m}_n} e^{-iux} \tilde{\varphi}_{X,n}(u) du, \quad \forall x \in \mathbb{R},$$

où pour des raisons techniques nous seuilons l'estimateur $\hat{\varphi}_{X,n}$ en le remplaçant par $\tilde{\varphi}_{X,n} := \hat{\varphi}_{X,n} \mathbb{1}_{|\hat{\varphi}_{X,n}| \geq \kappa_n n^{-1/2}}$ (voir le chapitre 3 pour plus de détails).

Nous énonçons ici un corollaire du théorème 3.1 du chapitre 3 qui permet d'étudier les vitesses de convergence de $\tilde{f}_{\hat{m}_n}$ sur la classe $\mathcal{S}(s, L)$.

Theorem 1.2. *Soit \hat{m}_n défini comme ci-dessus avec $\gamma > 1/2$. Si les coefficients de mélange fort $(\alpha(k))_{k \geq 0}$ satisfont $\alpha(k) = O(k^{-a})$ pour $a > 3$, alors*

$$\mathbb{E} \left[\|\tilde{f}_{\hat{m}_n} - f\|_2^2 \right] \leq C_1 \inf_{m \in [0, n]} \left\{ \int_{|u| > m} |\varphi_X(u)|^2 du + \frac{m \log(n)^{2\gamma}}{n} \right\} + C_2 \left(\frac{1}{n} + \frac{(\log n)^{\gamma(a-1)}}{n^{\frac{a-3}{2}}} \right),$$

pour des constantes $C_1, C_2 > 0$ qui dépendent de κ , γ et a .

Par conséquent, si le taux de mélange est en $\alpha(k) = O(k^{-a})$ pour $a > 3$, nous obtenons toujours un estimateur consistant de f . Notons aussi que dans le terme de variance apparaît une perte en puissance de \log qui dépend de γ , pour limiter l'ordre de ce terme il faut choisir des valeurs de γ pas trop grandes qui respectent la contrainte $\gamma > 1/2$.

1.4.4 Vitesse de convergence

En évaluant la borne du Théorème 1.2 pour des densité $f \in \mathcal{S}(s, L)$ on obtient

$$\|\tilde{f}_{\hat{m}_n} - f\|_2^2 \leq C \left\{ (n \log(n)^{-2\gamma})^{-\frac{2s}{2s+1}} + (\log n)^{\gamma(a-1)} n^{\frac{3-a}{2}} + \frac{1}{n} \right\}.$$

- Si $a > 5$, $(\log n)^{\gamma(a-1)} n^{\frac{3-a}{2}}$ est négligeable devant n^{-1} et on retrouve une vitesse minimax $n^{-\frac{2s}{2s+1}}$ du cadre indépendant à une puissance de \log près.
- Si $a \in]3, 5]$, nous obtenons un estimateur qui converge à la vitesse

$$(n \log(n)^{-2\gamma})^{-\frac{2s}{2s+1}} \vee (\log n)^{\gamma(a-1)} n^{(3-a)/2},$$

mais cette vitesse n'est a priori pas optimale. Nous voyons que (en négligeant la contribution des termes logarithmiques) une condition suffisante pour atteindre la vitesse minimax du cadre indépendant est d'imposer

$$n^{-\frac{2s}{2s+1}} < n^{\frac{3-a}{2}} \quad \Leftrightarrow \quad a > 3 + \frac{4s}{2s+1}.$$

Si la fonction f est très peu régulière (moralement s proche de 0), on retrouve la contrainte $a > 3$.

Beaucoup d'autres auteurs ont étudié le problème de l'estimation adaptative de densité dans un cadre dépendant. Par exemple, Tribouley et Viennet (1998) [70], Comte et Merlevède (2002) [17], Comte et al (2008) [16], Lerasle [56], [57] ainsi que Asin et Johannes (2018) [7] propose des estimateurs adaptatifs pour des suites β -mélangeantes. Notons que le cas des suites α mélangeantes est aussi étudié par Comte et Merlevède (2002) [17], et plus récemment par Bertin et Klutchnikoff (2017)[11].

En examinant la littérature citée ci-dessus, on constate que, que ce soit pour des suites β -mélangeantes ou α -mélangeantes, la contrainte $a > 3$ sur la vitesse de mélange est toujours requise pour l'estimation adaptative. Cela correspond bien à la contrainte minimale que nous avons trouvée, bien qu'il n'y ait pas d'explication heuristique à cela, puisque lorsque la classe de régularité est connue, les estimateurs non adaptatifs atteignent la vitesse minimax i.i.d. dès que $\sum \alpha(k) < \infty$. Dans le cas $a \in]3, 5]$ notre estimateur demeure consistant mais nous n'avons plus d'assurance que la vitesse soit quasi-optimale.

Dans le chapitre 3, nous illustrons numériquement les performances de cet estimateur sur différents exemples simulés (avec différentes vitesses de mélange). Nous avons aussi considéré le cas limite $a = 3$ ainsi que des suites dépendantes qui ne sont pas α -mélangeantes, afin de mettre en évidence la robustesse de la procédure.

1.4.5 Perspectives

Une extension naturelle de ce travail serait de considérer des problèmes inverses comme dans Duval et Kappus [38]. Par exemple le modèle de déconvolution où les observations X_i sont observées avec un bruit additif indépendant des X_i . Une autre piste serait de regarder le cas multidimensionnel où les X_i sont des variables aléatoires à valeurs dans \mathbb{R}^d , $d \geq 2$.

1.5 Distance de Wasserstein

1.5.1 Notations

Nous commençons par quelques notations qui seront utilisées tout au long de cette section et du chapitre 4. Soit $(X_i)_{i \in \mathbb{Z}}$ une suite stationnaire de variables aléatoires à valeurs dans \mathbb{R}^d , pour $d \geq 1$, de distribution marginale commune μ . Soit μ_n la mesure empirique de (X_1, \dots, X_n) définie par

$$\mu_n = \sum_{k=1}^n \delta_{X_k}.$$

Soit X une variable aléatoire de distribution μ . Pour tout $x \in \mathbb{R}^d$, notons $|x| = \max\{|x_1|, \dots, |x_d|\}$. Définissons à présent, pour $b \geq 1$, le moment faible d'ordre b de X :

$$\|X\|_{b,w}^b := \sup_{t>0} t^b \mathbb{P}(|X| > t).$$

Notons \mathcal{L} la distribution de la suite $(X_i)_{i \in \mathbb{Z}}$. Soient alors $a_n(\mathcal{L})$ et $b_n(\mathcal{L})$ deux suites de nombres réels dépendant de \mathcal{L} . La notation

$$a_n(\mathcal{L}) \ll b_n(\mathcal{L})$$

signifie qu'il existe une constant C ne dépendant pas de n ni de \mathcal{L} telle que $a_n(\mathcal{L}) \leq C b_n(\mathcal{L})$.

1.5.2 Généralités sur la distance de Wasserstein

Définition 1.5. Soit $p \in [1, \infty[$ et soit μ et ν deux mesures de probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. La distance de Wasserstein d'ordre p entre μ et ν est définie ainsi

$$W_p^p(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|_2^p \pi(dx, dy)$$

avec $\Pi(\mu, \nu)$ l'ensemble des mesures de probabilité sur l'espace produit $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d))$ dont les lois marginales sont μ et ν et $|\cdot|_2$ la norme euclidienne sur \mathbb{R}^d .

Si π^* est une solution du problème de minimisation ci-dessus, alors un couple (X^*, Y^*) de loi π^* fournit deux variables de distribution μ et ν aussi proches que possible au sens de la distance \mathbb{L}^p . On dit alors que (X^*, Y^*) est un couplage optimal de marginales (μ, ν) pour le coût \mathbb{L}^p , ou encore que le couple (X^*, Y^*) minimise le coût de transport \mathbb{L}^p entre μ et ν .

La première occurrence d'un problème de transport optimal clairement posé remonte à l'article de Monge daté de 1781 [61], sur le fameux problème des "déblais et remblais". L'intérêt de ces distances de couplage dans un cadre d'optimisation des coûts économiques a ensuite été clairement mis en évidence par Kantorovich dans son célèbre article "The Mathematical Method of Production Planning and Organization" (traduction du manuscrit original russe de 1939; voir aussi [53]). Les distances de Wasserstein sont parfois nommées distances de Kantorovich. L'importance des questions de coût de transport ou de couplage optimal n'a cessé de croître depuis, trouvant des applications dans de nombreux domaines (par exemple, cette distance est très utilisée pour la comparaison d'images et la reconnaissance de formes). Pour des références générales sur ce sujet, on peut citer par exemple les deux livres de Rachev et Ruschendorf [62] [63], ainsi que le livre de Villani [72].

Si l'on s'intéresse à présent au coût de transport \mathbb{L}^p entre la mesure empirique μ_n et la mesure marginale μ d'une suite iid, c'est à dire à la distance $W_p^p(\mu_n, \mu)$, des résultats presque sûr optimaux sont obtenus par Ajtai *et al.* (1984) [1] (case $p = 1, d = 2$), Talagrand (1992) [67] (case $p = 1, d > 2$), Barthe et Bordenave (2013) [9]. Pour les contrôles de $\mathbb{E}(W_p^p(\mu_n, \mu))$, citons les résultats marquants et relativement récents de Dereich *et al.* (2013) [33] et Fournier et Guillin (2015) [43]. Pour des bornes de déviations et des contrôles de moment, citons Dedecker et Merlevède (2019) [25].

1.5.3 Objectifs et outils utilisés

Dans le chapitre 4, nous généralisons certains résultats obtenus dans Dedecker et Merlevède (2019) [25] au cadre des suites $\tilde{\beta}$ -dépendantes. Comme nous n'utiliserons ici que les coefficients $\tilde{\beta}_1(k)$ (voir Définition 1.2), nous les noterons simplement $\tilde{\beta}(k)$.

Nous étudions le comportement de la quantité $W_p(\mu_n, \mu)$ pour $p > 1$. Plus précisément, nous donnons des inégalités de moment d'ordre $r > 1$ pour la quantité aléatoire $W_p^p(\mu_n, \mu)$ lorsque X a un moment faible d'ordre b . Les inégalités de moments que nous établissons sont analogues à celles obtenues par Dedecker et Merlevède (2019) (voir la Section 5 dans [25]) pour les suites de variables iid à valeurs dans \mathbb{R}^d . Ces bornes permettent en particulier d'évaluer la déviation de $W_p^p(\mu_n, \mu)$ grâce à l'inégalité de Markov :

$$\mathbb{P}(|W_p^p(\mu_n, \mu)| > t) \leq \frac{\|W_p^p(\mu_n, \mu)\|_r^r}{t^r}.$$

Dans [25], les inégalités de moment obtenues sont basées sur une inégalité de Fournier et Guillin (2015) [43] (voir (4.3) dans la section 4.2 du chapitre 4, et pour plus de détails voir [43]). Cette inégalité fait intervenir la quantité

$$\sum_{A \in \mathcal{F}} |\mu_n(A) - \mu(A)|,$$

où \mathcal{F} est une collection de cubes qui forment une partition d'un domaine borné de \mathbb{R}^d . Il se trouve que, pour contrôler la norme \mathbb{L}^r de telles quantités, le coefficient $\tilde{\beta}(k)$ est particulièrement bien adapté.

Les outils mathématiques utilisés pour déterminer une borne supérieure du moment $\|W_p(\mu_n, \mu)\|_r$ dépendent de l'ordre r , pour $r > 1$. Pour $r \in (1, 2)$ nous utilisons une inégalité de type Von Bahr-Esseen, et pour $r > 2$ nous utilisons l'inégalité de Burkholder démontrée par Dedecker et Doukhan (2003) [21] (voir [21] Proposition 4).

Nous énonçons à présent ces deux inégalités.

Proposition 1.3. *Soit $(Y_i)_{i \geq 0}$ une suite stationnaire de variables aléatoires à valeurs réelles adaptées à une filtration croissante $(\mathcal{G}_i)_{i \geq 0}$. Supposons que Y_0 ait un moment d'ordre $r \in (1, 2]$ et que $\mathbb{E}(Y_0) = 0$. Soit $T_n = Y_1 + \dots + Y_n$, et soit a un nombre réel quelconque. Alors*

$$\mathbb{E}(\|T_n\|_r^r) \leq 2^{2-r} n \left(\mathbb{E}(|Y_0|^r) + r \sum_{j=1}^{n-1} \mathbb{E}(|Y_0 - a|^{r-1} |\mathbb{E}(Y_j | \mathcal{G}_0)|) \right).$$

La preuve d'une version non stationnaire de cette inégalité est donnée en annexe du Chapitre 4. Cette inégalité est une généralisation d'une inégalité similaire énoncée par Cuny, Dedecker et Jan (2017) [18], avec $a = 0$ (voir la Proposition 10 dans [18]).

Proposition 1.4. *Supposons que $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires centrées et de carrés intégrables, et $\mathcal{M}_i = \sigma(X_j, 0 \leq j \leq i)$. Soit $S_n = X_1 + \dots + X_n$ et, pour $r \geq 2$,*

$$b_{i,n} = \max_{i \leq l \leq n} \left\| X_i \sum_{k=i}^l \mathbb{E}(X_k | \mathcal{M}_i) \right\|_{r/2}.$$

Alors

$$\|S_n\|_r \leq \left(2r \sum_{i=1}^n b_{i,n} \right)^{1/2}.$$

1.5.4 Les résultats obtenus

Par soucis de simplicité, nous n'énonçons dans cette introduction que les contrôles de moment d'ordre $r = 2$. Comme dans Fournier et Guillin [43] (qui donnent des bornes pour $\mathbb{E}(W_p^p(\mu_n, \mu))$) et dans Dedecker et Merlevède [25], les inégalités obtenues sont différentes selon que $p > d/2$ ou $p < d/2$.

Proposition 1.5. *Soit $\|X\|_{b,w} < +\infty$ pour $b > 2p$, et supposons que $\tilde{\beta}(k) = O(k^{-a})$ pour $a > 0$.
• Si $p > d/2$ et $a > 1$, alors*

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \begin{cases} \frac{\|X\|_{b,w}^p}{\sqrt{n}} & \text{if } b > \frac{2ap}{a-1}, \\ \frac{\|X\|_{b,w}^p \log n}{\sqrt{n}} & \text{if } b = \frac{2ap}{a-1}, \\ \frac{\|X\|_{b,w}^p}{n^{a(\frac{1}{2}-\frac{p}{b})}} & \text{if } b < \frac{2ap}{a-1}. \end{cases}$$

• Si $p > d/2$ et $a < 1$, alors

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \frac{\|X\|_{b,w}^p}{n^{a(\frac{1}{2}-\frac{p}{b})}}.$$

• Si $p > d/2$ et $a = 1$, alors

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \frac{\|X\|_{b,w}^p \sqrt{\log n}}{n^{(\frac{1}{2}-\frac{p}{b})}}.$$

Cette première série d'inégalités requière quelques commentaires. Le cas $a > 1$ correspond au cas de la courte mémoire, au sens où les sommes partielles de fonctions bornées de la suite $(X_i)_{i \in \mathbb{Z}}$ vérifient le théorème limite central avec la normalisation \sqrt{n} . Dans ce cas, on peut espérer obtenir des vitesses comparables à celle du cas iid, pourvu que les variables aient des moments d'ordre suffisamment élevés. C'est en effet le cas, puisque lorsque $b > 2ap/(a-1)$, on obtient une décroissance d'ordre $n^{-1/2}$ pour $\|W_p^p(\mu_n, \mu)\|_2$, comme dans le théorème 5.4 de Dedecker et Merlevède (2019) [25] (cas $p > d/2$). Nous ne sommes pas en mesure, pour l'instant, de montrer que la condition de moment que nous imposons pour obtenir la vitesse $n^{-1/2}$ est optimale. Cependant, dans le cas de la dimension $d = 1$, nous trouvons la même condition de moment que dans l'article [24] (en partant de l'inégalité (3.9) de cet article). En fait, on peut facilement remarquer que les vitesses que l'on obtient à partir de l'inégalité (3.9) dans [24], en supposant que $\|X\|_{b,w} < +\infty$ et que $\tilde{\beta}(k) = O(k^{-a})$, sont exactement les mêmes que les nôtres, quel que soit $a > 0$ et $b > 2p$ (notons cependant que les coefficients de dépendance utilisés dans [24] sont légèrement moins restrictifs que les coefficients $\tilde{\beta}(k)$).

Proposition 1.6. *Soit $\|X\|_{b,w} < +\infty$ pour $b > 2p$, et supposons que $\tilde{\beta}(k) = O(k^{-a})$ pour $a > 0$.*

• Si $p < d/2$ et $a > 1$, alors

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \begin{cases} \frac{\|X\|_{b,w}^p}{n^{p/d}} & \text{if } b > \frac{2adp}{ad-2p}, \\ \frac{\|X\|_{b,w}^p \log n}{n^{p/d}} & \text{if } b = \frac{2adp}{ad-2p}, \\ \frac{\|X\|_{b,w}^p}{n^{a(\frac{1}{2}-\frac{p}{b})}} & \text{if } b < \frac{2adp}{ad-2p}. \end{cases}$$

• Si $p < d/2$ et $a < 1$, alors

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \begin{cases} \frac{\|X\|_{b,w}^p}{n^{ap/d}} & \text{if } b > \frac{2dp}{d-2p}, \\ \frac{\|X\|_{b,w}^p \log n}{n^{ap/d}} & \text{if } b = \frac{2dp}{d-2p}, \\ \frac{\|X\|_{b,w}^p}{n^{a(\frac{1}{2}-\frac{p}{b})}} & \text{if } b < \frac{2dp}{d-2p}. \end{cases}$$

• Si $p < d/2$ et $a = 1$, alors

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \begin{cases} \frac{\|X\|_{b,w}^p \sqrt{\log n}}{n^{p/d}} & \text{if } b > \frac{2dp}{d-2p}, \\ \frac{\|X\|_{b,w}^p (\log n)^{3/2}}{n^{p/d}} & \text{if } b = \frac{2dp}{d-2p}, \\ \frac{\|X\|_{b,w}^p \sqrt{\log n}}{n^{(\frac{1}{2}-\frac{p}{b})}} & \text{if } b < \frac{2dp}{d-2p}. \end{cases}$$

Concernant cette seconde proposition, voici ce que l'on peut dire : lorsque $a > 1$ et $b > 2adp/(ad-2p)$, on obtient une vitesse d'ordre $n^{-p/d}$ pour $\|W_p^p(\mu_n, \mu)\|_2$, comme dans le Théorème 5.4 de Dedecker and Merlevède (2019) [25] (cas $p < d/2$). Nous ne sommes pas en mesure, pour l'instant, de montrer que la condition de moment que nous imposons pour obtenir la vitesse $n^{-p/d}$ est optimale.

1.5.5 Perspectives

Les bornes de moment que nous obtenons dans le cas $r > 2$ (que nous n'avons pas présentées dans cette introduction, mais qui sont énoncées dans le chapitre 4) semblent sous optimales dans les cas où $p < d/2$, au sens où nous ne retrouvons pas la vitesse $n^{-p/d}$ dans ce cas, même si l'on suppose des moments d'ordres élevés. Cela est peut-être dû au fait qu'on utilise une inégalité de type Burkholder plutôt qu'une inégalité de type Rosenthal. Une piste de travail serait donc d'essayer d'appliquer les inégalités de Rosenthal décrites dans l'article de Dedecker (2010) [20] afin de voir si une amélioration en découle (mais le coefficient de dépendance utilisé serait alors un coefficient $\tilde{\beta}_2$ similaire à celui introduit dans la définition 1.3).

Chapter 2

Testing Kendall's τ for a large class of dependent sequences

This chapter is a modified version of my article S. AMMOUS Testing Kendall's τ for a large class of dependent sequences. *Statistics* 54 (2020), no. 4, 686–713.

Contents

2.1	Introduction and definitions	18
2.2	Main results	20
2.3	Quasi-optimality of the dependency conditions	23
2.4	Simulations	25
2.4.1	First example	25
2.4.2	Second example	27
2.4.3	Third example	28
2.5	Proofs	30
2.5.1	Proof of theorem 2.1	30
2.5.2	Proof of proposition 2.1	31
2.6	Appendix	34
2.6.1	Proof of lemma 2.1	34
2.6.2	Proof of proposition 2.3	36
2.6.3	Proof of lemma 2.2	42

Abstract Let $(X_i, Y_i)_{i \in \mathbb{Z}}$ be a stationary sequence of \mathbb{R}^2 -valued random variables. To test if X_1 and Y_1 are correlated in the sense of Kendall, we propose a robust correction of the usual Kendall test, valid for a large class of dependent sequences. We also show that the condition on the dependency coefficients is quasi-optimal in a certain sense, and we illustrate our results through different sets of simulation.

2.1 Introduction and definitions

Let (X, Y) be a couple of real-valued and continuous random variables, and let (X^*, Y^*) be an independent copy of (X, Y) . The Kendall correlation coefficient τ between X and Y is then defined by

$$\tau := 2(\mathbb{P}(\{(X^* - X)(Y^* - Y) > 0\}) - 0.5). \quad (2.1)$$

By definition, $\tau \in [-1, 1]$. If $\tau = 0$, there is no correlation in the sense of Kendall. If $Y = f(X)$ for some increasing (resp. decreasing) function f , then $\tau = 1$ (resp. $\tau = -1$). If $\tau > 0$ (resp. $\tau < 0$), there is a positive correlation (resp. negative correlation), meaning that X and Y tend to vary in the same direction (resp. the opposite direction).

Let now $(X_i, Y_i)_{i \in \mathbb{Z}}$ be a stationary sequence of \mathbb{R}^2 -valued random variables, with the same marginal distribution as (X, Y) . To test $H_0 : \tau = 0$ against $H_1 : \tau \neq 0$ from the sequence $(X_i, Y_i)_{1 \leq i \leq n}$, one can use the U -statistic

$$U_n = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{1}_{\{(X_j - X_i)(Y_j - Y_i) > 0\}}, \quad (2.2)$$

which has been studied by Esscher [40], Lindeberg [58] [59], and Kendall [54] in the case where the random variables (X_i, Y_i) are independent and identically distributed (iid). In the general iid case (that is, without assuming that X and Y are independent), the asymptotic normal distribution of $\sqrt{n}(2U_n - \tau - 1)$ is given by Hoeffding [50] (see also van der Vaart [71], Example 12.5).

The Kendall correlation test is often preferred to the Pearson test for several reasons. Firstly, it is more suitable for detecting non-linear monotonic relations. Secondly, it does not assume that the variables are normally distributed (although a robust correction of the Pearson test can also be proposed for non-normal distributions). Finally it is valid without assumptions on the moments of the variables, and in particular when the variables have heavy tails, which is often the case for financial data (see for example [30] for more details on this subject).

The asymptotic behaviour, and in particular the asymptotic normality, of U -statistics in a dependent context has been studied by many authors since Sen's paper [66] (case of m -dependent sequences). Let us cite for instance the papers by Yoshihara [74] (β -mixing sequences) Denker and Keller [32] (β -mixing and φ -mixing sequences), Borovkova *et al.* [13] (function of β -mixing sequences), Hsing and Wu [51] (coupling conditions), Dehling and Wendler [31] (α -mixing sequences). The special case of Kendall U -statistics has been considered recently by Dehling *et al.* [30] for functions of β -mixing sequences. We shall say more about the two papers [32] and [30] at the end of this introduction.

In section 2.2 of the present paper, we extend Hoeffding's result on Kendall's U -statistic to a class of dependent sequences, and we propose an estimator of the limiting covariance of $\sqrt{n}(2U_n - \tau - 1)$. From these two results we derive an asymptotically valid procedure to test $H_0 : \tau = 0$ against $H_1 : \tau \neq 0$. Our results apply to a large class of dependent sequences, under a condition on the $\bar{\beta}$ -dependence coefficients of the sequence $(X_i, Y_i)_{i \in \mathbb{Z}}$, that are defined below. To be complete, we show in section 2.3 that the condition on the dependency coefficients is quasi-optimal in a certain sense, and we illustrate our results through different sets of simulation (see section 2.4).

Let us now introduce these dependence coefficients.

Definition 2.1. Let $(Z_i)_{i \in \mathbb{Z}} = (X_i, Y_i)_{i \in \mathbb{Z}}$ be a strictly stationary sequence of random variables with values in \mathbb{R}^2 . Let \mathbb{P} be the law of (X_0, Y_0) and $\mathbb{P}_{(Z_i, Z_j)}$ be the law of (Z_i, Z_j) . Define the

σ -algebra $\mathcal{F}_0 = \sigma(Z_k, k \leq 0)$, let $\mathbb{P}_{Z_k|\mathcal{F}_0}$ be the conditional distribution of Z_k given \mathcal{F}_0 , and let $\mathbb{P}_{(Z_i, Z_j)|\mathcal{F}_0}$ be the conditional distribution of (Z_i, Z_j) given \mathcal{F}_0 .

For any $s, t \in \mathbb{R}, z = (x, y) \in \mathbb{R}^2$, we define the function

$$f_z(s, t) := \mathbf{1}_{x \leq s} \mathbf{1}_{y \leq t} - \mathbb{P}(X_0 \leq s, Y_0 \leq t)$$

and the random variables

$$\begin{aligned} b(k) &= \sup_{z \in \mathbb{R}^2} | \mathbb{P}_{Z_k|\mathcal{F}_0}(f_z) |, \\ b(i, j) &= \sup_{(z_1, z_2) \in \mathbb{R}^2 \times \mathbb{R}^2} | \mathbb{P}_{(Z_i, Z_j)|\mathcal{F}_0}(f_{z_1} \otimes f_{z_2}) - \mathbb{E}(f_{z_1}(Z_i) f_{z_2}(Z_j)) | \end{aligned}$$

where as usual $f_{z_1} \otimes f_{z_2}(s, t) = f_{z_1}(t) f_{z_2}(s)$. Define now the coefficients

$$\begin{aligned} \tilde{\beta}_1(k) &= \mathbb{E}(b(k)), \\ \tilde{\beta}_2(k) &= \max\{\tilde{\beta}_1(k), \sup_{i>j \geq k} \mathbb{E}[b(i, j)]\}, \\ \delta_2(k) &= \sup_{i>j \geq k} \mathbb{E}(\mathbf{1}_{X_i \leq X_j} \mathbf{1}_{Y_i \leq Y_j} - \mathbb{P}(X_i \leq X_j, Y_i \leq Y_j) | \mathcal{F}_0), \\ \bar{\beta}_2(k) &= \max\{\tilde{\beta}_2(k), \delta_2(k)\}. \end{aligned}$$

Let us give some comments on these definitions. Our main result (theorem 2.1 below) is stated under a condition on the coefficient $\bar{\beta}_2$. The first important remark is that this coefficient is weaker than the usual β -mixing coefficient of the sequence $(X_i, Y_i)_{i \in \mathbb{Z}}$ (see Volkonskii and Rozanov [73] for the definition of the β -mixing coefficient). Hence all the results of the present paper hold true when replacing the coefficients $\tilde{\beta}_1, \tilde{\beta}_2, \bar{\beta}_2$ by the usual β -mixing coefficients.

The coefficient $\bar{\beta}_2$ can be computed for a large class of processes, including many non-mixing sequences in the sense of Rosenblatt [65]. This follows mostly from the paper by Dedecker and Prieur [28], which provides many examples of (possibly non-mixing) processes for which the coefficient β_2 can be easily controlled (see also the monograph [22] for more examples and a comparison with other notions of dependency). The coefficient $\bar{\beta}_2$ is a bit more restrictive than $\tilde{\beta}_2$, because of the term δ_2 , which is not so easy to handle. However, in many cases (if not all) the coefficient δ_2 may be handled as $\tilde{\beta}_2$ by following the thread of Section 6 in Dedecker-Prieur [28].

Let us give a simple example. Assume that $Z_i = (X_i, Y_i)^t$ is a \mathbb{R}^2 -valued linear process, defined by

$$Z_i = \sum_{k=0}^{\infty} A_k \varepsilon_{i-k},$$

where (ε_i) is a sequences of iid \mathbb{R}^2 -valued random variables with mean 0 and square integrable coordinates, and A_k is a deterministic sequence of 2×2 matrices such that $\sum_{k \geq 0} |A_k|^2 < \infty$ (here $|A_k|$ is the usual norm $|A_k| = \sup_{\|x\|=1} |A_k x|$, and $\|\cdot\|$ is the euclidean norm). Let $F_{X,0}$ and $F_{Y,0}$ be the distribution functions of X_0 and Y_0 , and, for $i > 0$, let

$$F_{X,i}(t) = \mathbb{P}(X_i - X_0 \leq t) \quad \text{and} \quad F_{Y,i}(t) = \mathbb{P}(Y_i - Y_0 \leq t).$$

If the functions $F_{X,i}, F_{Y,i}$ are uniformly Hölder of order $\gamma \in (0, 1]$, meaning that there exists a positive constant C such that

$$\sup_{i \geq 0} |F_{X,i}(s) - F_{X,i}(t)| \leq C|t - s|^\gamma,$$

then (following [28], Section 6),

$$\bar{\beta}_2(n) \leq C \left(\sum_{k \geq n} |A_k|^2 \right)^{\frac{\gamma}{\gamma+2}}.$$

In particular, if $|A_k|$ is geometrically decreasing, then so is $\bar{\beta}_2(k)$ (whatever the index γ).

Note that, without extra assumptions on the distribution of ε_0 , such linear processes have no reasons to be mixing in the sense of Rosenblatt. For instance, it is well known that the \mathbb{R} -valued linear process

$$X_i = \sum_{k \geq 0} \frac{\varepsilon_{i-k}}{2^{k+1}}, \quad \text{where } \mathbb{P}(\varepsilon_1 = -1/2) = \mathbb{P}(\varepsilon_1 = 1/2) = 1/2,$$

is not α -mixing (see for instance Bradley [14]). Hence, if $(Y_i)_{i \in \mathbb{Z}}$ is a sequence of iid random variables, independent of $(X_i)_{i \in \mathbb{Z}}$, then the sequence $(X_i, Y_i)_{i \in \mathbb{Z}}$ is not α -mixing. By contrast, for this particular example, one can check that $\bar{\beta}_2(k)$ is geometrically decreasing.

To conclude this section, let us quote that the asymptotic normality of Kendall's U -statistic for dependent sequences has been recently established by Dehling *et al.* [30] (note that these authors are able to deal with a large class of U -statistics, and that they also prove a functional central limit theorem for U -processes). Their result is valid for a large class of dependent sequences (including non-mixing sequences in the sense of Rosenblatt [65]), with a large intersection with our class of $\bar{\beta}_2$ -dependent sequences. Again, the advantage of our approach is that we are able to prove that our condition on the coefficient $\bar{\beta}_2$ is quasi-optimal in some sense (see section 2.3). Moreover, if we consider only the class of β -mixing sequences, then our results are valid under the condition $\sum_{k>0} \beta(k) < \infty$, while the condition in [30] cannot be better than $\sum_{k>0} k\beta(k) < \infty$.

Concerning the special case of β -mixing sequences, some general results are given in [32] for non-necessarily bounded U -statistics. For bounded U -statistics, the asymptotic normality is proved in [32] under the condition $\sum_{k>0} \beta(k) < \infty$. For Kendall's U -statistic, we shall also see in section 2.3 that this condition is close to the optimal condition (as a condition expressed in terms of β -mixing coefficients).

2.2 Main results

As in the introduction, $(X_i, Y_i)_{i \in \mathbb{Z}}$ is a stationary sequence of \mathbb{R}^2 -valued random variables, with the same marginal distribution as (X, Y) , and we denote by (X^*, Y^*) an independent copy of (X, Y) . Recall that Kendall's correlation coefficient τ is defined in (2.1). Let then

$$\pi := \frac{\tau}{2} + 0.5 = \mathbb{P}(\{(X^* - X)(Y^* - Y) > 0\}). \quad (2.3)$$

Define also

$$F(x, y) = \mathbb{P}(X < x, Y < y), \quad H(x, y) = \mathbb{P}(X > x, Y > y),$$

and

$$F_X(x) = \mathbb{P}(X < x), \quad H_X(x) = \mathbb{P}(X > x).$$

Our main result is the following theorem.

Theorem 2.1. *Let $(X_i, Y_i)_{i \in \mathbb{Z}}$ be a stationary sequence of \mathbb{R}^2 -valued random variables. Assume that*

$$\sum_{k=1}^{\infty} \tilde{\beta}_1(k) < \infty \quad \text{and} \quad k \bar{\beta}_2(k) \xrightarrow[k \rightarrow +\infty]{} 0 \quad (2.4)$$

then,

$$\sqrt{n}(U_n - \pi) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, V),$$

where

$$V = 4 \operatorname{Var}(F(X_0, Y_0) + H(X_0, Y_0)) + 8 \sum_{k=1}^{\infty} \operatorname{Cov}(F(X_0, Y_0) + H(X_0, Y_0), F(X_k, Y_k) + H(X_k, Y_k)). \quad (2.5)$$

Note that the statistic $\sqrt{n}(U_n - 0.5)$ cannot be used directly to test $H_0 : \tau = 0$ against $H_1 : \tau \neq 0$. Indeed, according to theorem 2.1, the asymptotic distribution of $\sqrt{n}(U_n - 0.5)$ under H_0 depends on the unknown quantity V . To resolve this problem, we propose in the next proposition a consistent estimator of V .

Proposition 2.1. *Let $(X_i, Y_i)_{i \in \mathbb{Z}}$ be a stationary sequence of \mathbb{R}^2 -valued random variables. Assume that*

$$\sum_{k=1}^{\infty} \tilde{\beta}_2(k) < \infty. \quad (2.6)$$

Let

$$F_n(s, t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i < t} \mathbb{1}_{Y_i < s}, \quad H_n(s, t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i > s} \mathbb{1}_{Y_i > t}, \quad G_n(s, t) = 2(F_n(s, t) + H_n(s, t)),$$

and

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{i=1}^{n-k} (G_n(X_i, Y_i) - \bar{G}_n)(G_n(X_{i+k}, Y_{i+k}) - \bar{G}_n),$$

where $\bar{G}_n = \frac{1}{n} \sum_{i=1}^n G_n(X_i, Y_i)$. Let (a_n) be a sequence of positive integers tending to infinity as n tends to infinity, such that $a_n = o(\sqrt{n}/(\log n)^2)$. Then,

$$V_n = \hat{\gamma}(0) + 2 \sum_{k=1}^{a_n} \hat{\gamma}(k)$$

converges in \mathbb{L}^2 to the quantity V defined in (2.5).

Note that the result of theorem 2.1 and the Proposition 2.1 are true without assuming that the random variables X and Y are continuous.

Combining theorem 2.1 and Proposition 2.1, we obtain that, under $H_0 : \tau = 0$ and if $V > 0$, the random variables

$$T_n := \frac{\sqrt{n}(U_n - 1/2)}{\sqrt{|V_n|}} \quad \text{converges in distribution to } \mathcal{N}(0, 1). \quad (2.7)$$

Therefore, for a significance level α , the rejection region of the corrected Kendall test is of the form $R_{n,\alpha} = \{|T_n| > q_\alpha\}$ where q_α is the quantile of order $1 - (\alpha/2)$ of the standard normal distribution. Moreover, the estimator V_n is always a consistent estimator of V (even if H_0 is not true). This ensures that the test is consistent under any point of the alternative hypothesis.

We now make several comments about the estimation of the quantity V defined in (2.5).

Remark 2.1. *The choice of the sequence (a_n) in Proposition 2.1 is a delicate matter. If the coefficients $\tilde{\beta}_2(k)$ decrease very quickly, then a_n should increase very slowly (it suffices to take $a_n \equiv 0$ in the iid setting). On the contrary, if $\tilde{\beta}_2(k) = O(k^{-1}(\log k)^{-a})$ for some $a > 1$, then the terms in the covariance series have no reason to be small, and one should take a_n close to \sqrt{n} to estimate many of these covariance terms. A data-driven criterion for choosing a_n is an interesting (but probably difficult) question, which is beyond the scope of the present paper.*

However, from a practical point of view, there is an easy way to proceed: one can plot the estimated covariances $\hat{\gamma}(k)$'s and choose a_n (not too large) in such a way that

$$\hat{\gamma}(0) + 2 \sum_{k=1}^{a_n} \hat{\gamma}(k)$$

should represent an important part of the unknown covariance series V defined in (2.5). As we shall see in the simulations (Section 2.4), if the decay of the covariances terms

$$\gamma(k) = 2\text{Cov}(F(X_0, Y_0) + H(X_0, Y_0), F(X_k, Y_k) + H(X_k, Y_k))$$

is not too slow, this provides an easy and reasonable choice for a_n .

Remark 2.2. *As for the residuals of the linear model, one can consider that the sequence $(G_n(X_i, Y_i))_{1 \leq i \leq n}$ is the sequence of "residuals". Then, our estimator belongs to the general class of kernel HAC estimators (cf. [6] and the references therein), based on the residuals and computed with the rectangular kernel. Other kernels are considered in [6], for example kernels with a positive Fourier transform which guarantee that the residual-based estimator of V is always positive. In [30], HAC estimators of V are proved to be consistent for a large class of kernels (see theorem 2.7 in [30]). However, in our context and without further assumptions than (2.6), the consistency of HAC estimators for a large class of kernels does not seem easy to obtain.*

Remark 2.3. *In [5] an automatic choice of the bandwidth h_n is given, for kernel HAC estimators computed from the residuals of a linear model. In [39] an automatic choice of a_n is given for the rectangular kernel, and minimax properties for the estimation of the spectral density are proved for ARMA processes. It would be interesting to see (for instance via a simulation study) if these procedures work well in our context. One can also refer to the recent paper [15] in which many automatic residual-based methods (including that given in [5] and [39]) for estimating a*

covariance series (or some related quantities) are studied through many simulation schemes. This paper deals with the linear model with stationary errors, and various estimators are computed from the residuals of the model. It appears that one of the best method (in terms of tests calibration) consists in fitting an AR model on the residuals, and estimating the unknown auto-covariances of the error process by the auto-covariances of the fitted AR process. Following [15] a strategy would be to fit an AR process on the "residuals" $(G_n(X_i, Y_i))_{1 \leq i \leq n}$, and to estimate the unknown quantity V by the same quantity computed with the auto-covariances of the fitted AR process. As in [15], the validity of this procedure should be checked through different sets of simulations (see remark 2.6 below for a first attempt in this direction).

2.3 Quasi-optimality of the dependency conditions

In this section, we prove that the dependency condition (2.4) is close to the optimal condition (as a condition expressed in terms of the $\bar{\beta}_2$ coefficients). More precisely, we shall exhibit a strictly stationary sequence $(X_i, Y_i)_{i \in \mathbb{Z}}$ for which $\bar{\beta}_2(n) = O(n^{-1})$ but (2.4) is not satisfied, and such that $\sqrt{n}(U_n - \pi)$ does not converge in distribution.

In fact, our example contains more information than the quasi-optimality of condition (2.4). Indeed, in our example, $Y_i = f(X_i)$, where $(X_i)_{i \in \mathbb{Z}}$ is a β -mixing Markov chain with β -mixing coefficients $\beta(n) \sim \frac{1}{n}$. Consequently, the β -mixing coefficients β_Z of $Z_i = (X_i, Y_i)$ are of order $\beta_Z(n) = O(n^{-1})$, proving that the condition $\sum_{k>0} \beta_Z(k) < \infty$ is also close to the optimal condition (as a condition expressed in terms of the β -mixing coefficients).

Proposition 2.2. *There exists a stationary Markov chain $(X_i)_{i \in \mathbb{Z}}$ with β -mixing coefficients $\beta(n) \sim \frac{1}{n}$ and invariant distribution $U[-\frac{1}{2}, \frac{1}{2}]$, such that*

$$\frac{\sqrt{n}}{\sqrt{\log n}}(U_n(X, X^2) - 0.5) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1), \quad (2.8)$$

where

$$U_n(X, X^2) = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ i < j}}^n \mathbb{1}_{\{(X_j - X_i)(X_j^2 - X_i^2) > 0\}}. \quad (2.9)$$

Remark 2.4. *Taking $Y_i = X_i^2$, we infer from proposition 2.2 that $\sqrt{n}(U_n - 0.5)$ does not converge in distribution, and is not even stochastically bounded. This implies in particular that the condition (2.4) cannot be satisfied. Since $\bar{\beta}_2(n)$ is dominated by $\beta_Z(n)$, we also infer that $\bar{\beta}_2(n) = O(n^{-1})$. This proves that the conclusion of theorem 1 cannot be true in general under the condition $\bar{\beta}_2(n) = O(n^{-1})$.*

Proof. We start from the Markov chain introduced by Doukhan, Massart and Rio [37].

Let λ be the uniform distribution on $[0, 1]$, and let ν be the probability with density $g(x) = 2x\mathbf{1}_{[0,1]}$. We define now a strictly stationary Markov chain by specifying its transition probabilities $K(x, A)$ as follows:

$$K(x, A) = (1 - x)\delta_x(A) + x\nu(A),$$

where δ_x denotes the Dirac measure. Then λ is the unique invariant probability measure of the chain with transition probabilities $K(x, \cdot)$. Let $(Z_i)_{i \in \mathbb{Z}}$ be the stationary Markov chain on $[0, 1]$

with transition probabilities $K(x, \cdot)$ and invariant distribution λ . From [37], we know that the β -mixing coefficients of this chain are such that $\beta(n) \sim \frac{1}{n}$.

We now define the random variables $X_i = Z_i - 0.5$. Hence $(X_i)_{i \in \mathbb{Z}}$ is a stationary Markov chain whose β -mixing coefficients are such that $\beta(n) \sim \frac{1}{n}$. Moreover the X_i 's are uniformly distributed over $[-0.5, 0.5]$.

Let $Y_i = X_i^2$. As quoted in Remark 2.4, the statistic $\sqrt{n/\log n}(U_n(X, X^2) - 0.5)$ in Proposition 2.2 is exactly $\sqrt{n/\log n}(U_n - 0.5)$. As in (2.11) (see the proof of theorem 2.1), we have the Hoeffding decomposition (used with $\pi = 1/2$):

$$\frac{\sqrt{n}}{\sqrt{\log n}}(U_n - 0.5) = T_n + R_n, \quad (2.10)$$

where

$$\begin{aligned} T_n &:= \frac{2}{\sqrt{n \log n}} \sum_{i=1}^n (F(X_i, Y_i) + H(X_i, Y_i) - 0.5), \\ R_n &:= \frac{2}{(n-1)\sqrt{n \log n}} \sum_{i=1}^n \sum_{j=i+1}^n (f(Z_i, Z_j) + f(Z_j, Z_i)), \\ f(Z_i, Z_j) &:= \mathbb{1}_{X_i < X_j} \mathbb{1}_{Y_i < Y_j} - F(X_j, Y_j) - H(X_i, Y_i) + 1/4. \end{aligned}$$

From the proof of proposition 2.3 (see subsection 2.6.2), we get the upper bound

$$\mathbb{E}(R_n^2) \leq \frac{C}{n \log n} \left(1 + \sum_{k=1}^n k \beta(k) \right),$$

for some positive constant C . Since $\beta(k) \sim \frac{1}{k}$, we easily infer that R_n converges to 0 in \mathbb{L}^2 . Hence, it remains to prove (2.8) with T_n instead of $\sqrt{n/\log n}(U_n(X, X^2) - 0.5)$.

Let us compute $F(X, X^2)$ and $H(X, X^2)$. For $x \in \mathbb{R}$ and $y > 0$,

$$F(x, y) = P(X < x, X^2 < y) = P(X < x, |X| < \sqrt{y}) = P(X < x, -\sqrt{y} < X < \sqrt{y})$$

Consequently

$$\begin{aligned} F(x, y) &= P(\min(x, -\sqrt{y}) < X < \min(x, \sqrt{y})) \\ &= F_X(\min(x, \sqrt{y})) - F_X(\min(x, -\sqrt{y})). \end{aligned}$$

Since $\min(X, |X|) = X$ and $\min(X, -|X|) = -|X|$, we infer that

$$F(X, X^2) = F_X(X) - F_X(-|X|) = F_X(X) - H_X(|X|).$$

In the same way, for $x \in \mathbb{R}$ and $y > 0$,

$$\begin{aligned} H(x, y) &= P(X > x, X^2 > y) = P(X > x, |X| > \sqrt{y}) \\ &= P(X > \max(x, \sqrt{y})) + P(x < X < -\sqrt{y}) \\ &= H_X(\max(x, \sqrt{y})) + [F_X(-\sqrt{y}) - F_X(x)] \mathbb{1}_{x < -\sqrt{y}} \end{aligned}$$

In our case, since $\mathbb{1}_{X < -|X|} = 0$, we infer that

$$H(X, X^2) = H_X(\max(X, |X|)) = H_X(|X|).$$

Altogether, this proves

$$F(X, X^2) + H(X, X^2) = F_X(X).$$

Note also that, since $X_i = Z_i - 0.5$, we have: $F_X(X_i) = Z_i$. Consequently

$$T_n = \frac{2}{\sqrt{n \log n}} \sum_{i=1}^n (Z_i - 0.5).$$

Proposition 2.2 then follows from lemma 2.1 below, whose proof will be done in Appendix.

Lemma 2.1. *Let $(Z_i)_{i \in \mathbb{Z}}$ be the stationary Markov chain on $[0, 1]$ with transition probabilities $K(x, \cdot)$ and invariant distribution λ . Then*

$$\frac{2}{\sqrt{n \log n}} \sum_{i=1}^n (Z_i - 0.5) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Remark 2.5. *This lemma can be proved by using Proposition 4 in [75]. However, checking the conditions given in that Proposition is quite long and not so easy. For the sake of clarity, we will provide a direct proof of lemma 2.1, going back to the initial result by Feller [42].*

2.4 Simulations

In this section, we compare the usual Kendall test with our corrected test in three different cases.

2.4.1 First example

In this first example, we consider two stationary sequences (X_i) and (Y_i) , with (X_i) independent of (Y_i) .

More precisely, we shall simulate X_i and Y_i according to the auto-regressive mechanisms:

$$\begin{aligned} - & \begin{cases} X_i = \frac{1}{2}(X_{i-1} + \varepsilon_i) & \text{with } (\varepsilon_i)_{i \geq 1} \text{ iid, and } \varepsilon_i \sim \mathcal{B}(\frac{1}{2}) \\ X_0 \sim \mathcal{U}[0, 1] & \text{with } X_0 \text{ independent of } (\varepsilon_i)_{i \geq 1} \end{cases} \\ - & \begin{cases} Y_i = \frac{1}{2}(Y_{i-1} + \varepsilon'_i) & \text{where } (\varepsilon'_i)_{i \geq 0} \text{ iid, and } \varepsilon'_i \sim \mathcal{B}(\frac{1}{2}) \\ Y_0 \sim \mathcal{U}[0, 1] & \text{with } Y_0 \text{ independent of } (\varepsilon'_i)_{i \geq 1} \end{cases} \end{aligned}$$

We assume moreover that $(X_0, (\varepsilon_i))$ is independent of $(Y_0, (\varepsilon'_i))$, so that the sequence (Y_i) is an independent copy of (X_i) . Moreover, it is well known that the uniform distribution $\mathcal{U}[0, 1]$ is the unique invariant distribution of the auto-regressive process (X_i) ; consequently, the two sequences (X_i) and (Y_i) are strictly stationary. Note also that the stationary process (X_i, Y_i) is not mixing in the sense of Rosenblatt (see for instance [65]). However, it follows from [28] that the coefficients $\bar{\beta}_2(k)$ converge to zero at an exponential rate.

Since, for any positive integer i , the random variable X_i is independent of Y_i , it follows that $\pi = 1/2$. Hence, the statistic T_n defined in (2.7) converges in distribution to the $\mathcal{N}(0, 1)$ distribution as $n \rightarrow \infty$.

We now simulate the random variables (X_i, Y_i) for $i = 1, \dots, n$, and we study the behaviour of T_n for different choices of a_n (recall that a_n appears in the definition of the estimator V_n). As explained in Remark 2, the choice of a_n may be done by analysing the graph of the auto-covariances $\hat{\gamma}(k)$ defined in Proposition 1.

We compute T_n for different choices of n from 150 to 600. We estimate the quantities $\text{Var}(T_n)$ and $\mathbb{P}(|T_n| > 1.96)$ (the estimated level) via a classical Monte-Carlo procedure, by averaging over $N = 2000$ independent trials. This procedure will also be applied in the two following subsections 2.4.2 and 2.4.3.

If a_n is well chosen, the estimate value of $\text{Var}(T_n)$ should be close to 1 and the estimated level should be close to 0.05. The graph of the auto-covariances suggests a choice $a_n = 2$ (see Figure 2.1). For $k > 2$ the auto-covariances $\hat{\gamma}(k)$ seem to be close to 0.

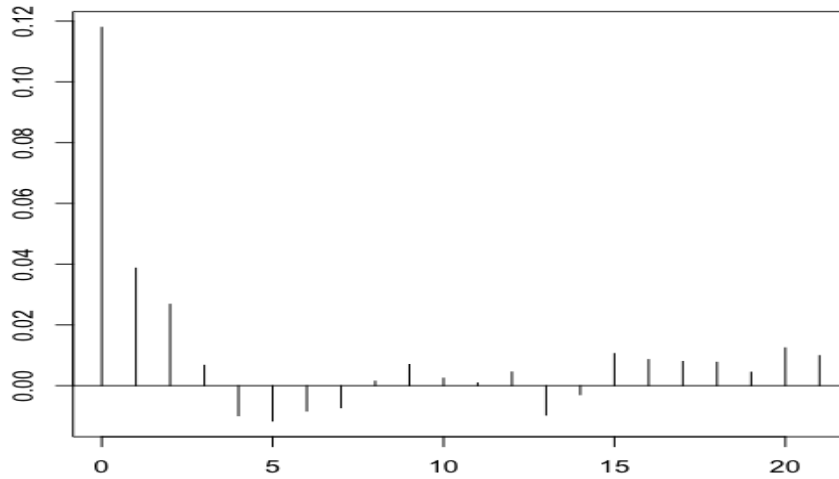


Figure 2.1: Graph of the auto-covariances $\hat{\gamma}(k)$ for example 1, with $n = 150$

The results for $a_n = 2$ are presented below. We also give the rejection frequency of the usual (non corrected) Kendall test.

- $a_n = 2$

n	150	200	250	300	350	400	500	600
Estimated variance	1.177	1.077	1.045	1.079	1.066	1.138	1.043	1.066
Est. corrected test level	0.066	0.067	0.061	0.057	0.0595	0.067	0.051	0.052
Est. Kendall test level	0.132	0.127	0.1145	0.125	0.129	0.144	0.125	0.131

As suggested by Figure 1, the choice $a_n = 2$ gives a reasonable estimated variance.

The estimated level lies always between 5% and 7% even for moderately large samples ($n = 150$); for $n \geq 500$ it is around 0.052, which is quite satisfactory.

The estimated level of the uncorrected Kendall test is around 0.13 instead of 0.05; this is due to the fact that the usual Kendall test does not take into account the dependency of the variables.

To be complete, we verify that the power of the corrected test tends to 1 as $n \rightarrow \infty$ for a model such that $\tau > 0$. This model is the following: (X_i) is simulated as before, (Z_i) is an independent copy of (X_i) , and

$$Y_i = Z_i + 0.2\sqrt{X_i}.$$

As before, we take $a_n = 2$ for the estimator V_n . The results are given below

n	150	200	250	300	350	400	500	600
Estimated variance	1.123	1.326	1.191	1.226	1.08	1.165	1.19	1.116
Estimated power	0.32	0.416	0.481	0.564	0.616	0.679	0.743	0.829

2.4.2 Second example

This is an example where $(X_i, Y_i)_{i \in \mathbb{Z}}$ is a sequence of iid random vectors, but the random variables X_i and Y_i are dependent.

More precisely, let $(X_i)_{i \in \mathbb{Z}}$ and $(\varepsilon_i)_{i \in \mathbb{Z}}$ be two independent sequences of iid random variables, with $X_i \sim \mathcal{U}[0, 1]$ and $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = \frac{1}{2}$. Define then $Y_i = X_i \varepsilon_i$.

First, we compute the value of π .

$$\begin{aligned} \pi &= \mathbb{P}((X_2 - X_1)(Y_2 - Y_1) > 0) \\ &= \mathbb{P}(\{(X_2 - X_1)^2 > 0\} \cap \{\varepsilon_1 = \varepsilon_2 = 1\}) + \mathbb{P}(\{-(X_2 - X_1)^2 > 0\} \cap \{\varepsilon_1 = \varepsilon_2 = -1\}) \\ &\quad + \mathbb{P}(\{(X_2 - X_1)(X_2 + X_1) > 0\} \cap \{\varepsilon_1 = -1, \varepsilon_2 = 1\}) \\ &\quad + \mathbb{P}(\{-(X_2 - X_1)(X_2 + X_1) > 0\} \cap \{\varepsilon_1 = 1, \varepsilon_2 = -1\}) \\ &= \frac{1}{4} + 0 + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}. \end{aligned}$$

For this example, it is clear that X_i and Y_i are not independent; however $\pi = 1/2$ (and hence $\tau = 0$) which means that there is no correlation in the sense of Kendall.

For the simulations, since we are in the usual situation where $(X_i, Y_i)_{i \in \mathbb{Z}}$ is a sequence of independent and identically distributed random vectors, we take $a_n = 0$, and we use the statistic T_n defined in (2.7) with $V_n = \hat{\gamma}(0)$ (the choice $a_n = 0$ is confirmed by the graph of the auto-covariances, see figure 2.2). The results are given below

n	150	200	250	300	350	400	500	600
Estimated variance	1.105	1.077	1.07	1.085	1.045	1.035	0.981	1.029
Est. corrected test level	0.059	0.059	0.058	0.055	0.056	0.056	0.051	0.051
Est. Kendall test level	0.248	0.255	0.253	0.254	0.26	0.262	0.242	0.265

As expected, the estimated variance is around 1 and the estimated level of the corrected test is around 0.05, even for moderately large samples.

It is important to notice that the usual Kendall test is not well calibrated in that case, with an estimated significance level around 0.25 instead of 0.05. The reason is in fact simple: the

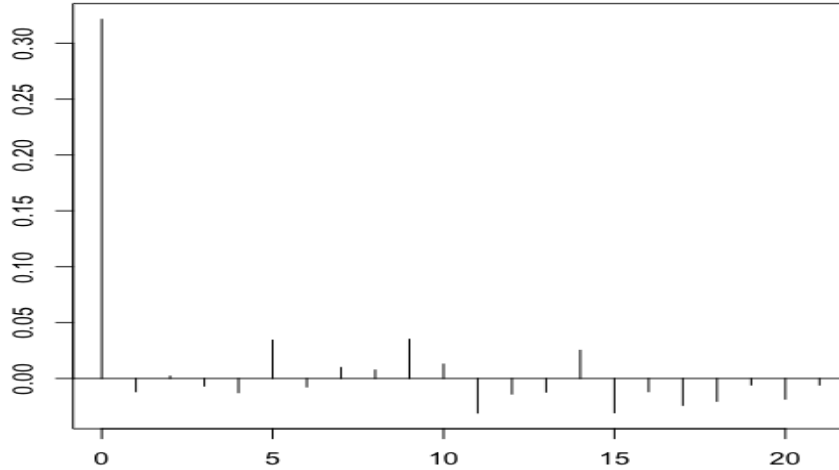


Figure 2.2: Graph of the auto-covariances $\hat{\gamma}(k)$ for example 2, with $n=150$

usual Kendall test is well calibrated if X_i and Y_i are independent (because in that case the term $\text{Var}(F(X_1, Y_1) + H(X_1, Y_1))$ can be explicitly computed), but it is not under the general hypothesis $H_0 : \pi = 1/2$. The conclusion is that: even in the usual case where $(X_i, Y_i)_{i \in \mathbb{Z}}$ is a sequence of iid random vectors, a correction should be made on the usual Kendall test. More precisely, to get an asymptotically well calibrated test procedure, the statistic T_n should be used with $V_n = \hat{\gamma}(0)$.

To be complete, we verify that the power of the corrected test tends to 1 as $n \rightarrow \infty$ for a model such that $\tau > 0$. This model is the following: (X_i) is simulated as before and $Y_i = X_i \varepsilon_i$, with $\mathbb{P}(\varepsilon_i = 1) = 0.58$ and $\mathbb{P}(\varepsilon_i = -1) = 0.42$.

As before, we take $V_n = \hat{\gamma}(0)$. The results are given below

n	150	200	250	300	350	400	500	600
Estimated variance	1.191	1.066	1.077	1.085	1.074	1.031	1.038	1.054
Estimated power	0.386	0.501	0.595	0.636	0.732	0.794	0.871	0.929

2.4.3 Third example

In this last example, the X_i 's are dependent random variables, and so are the Y_i 's. Moreover, the variables X_i and Y_i are also dependent.

Let first (Z_i) be generated according to the auto-regressive mechanism:

$$\begin{cases} Z_i = \frac{1}{2}(Z_{i-1} + \varepsilon_i) & \text{with } (\varepsilon_i)_{i \geq 1} \text{ iid, and } \varepsilon_i \sim \mathcal{B}\left(\frac{1}{2}\right) \\ Z_0 \sim \mathcal{U}[0, 1] & \text{with } Z_0 \text{ independent of } (\varepsilon_i)_{i \geq 1}. \end{cases}$$

Define then $X_i = Z_i - 0.5$ and $Y_i = X_i^2$. Once again, one can easily check that $\pi = 0.5$, meaning that there is no correlation in the sense of Kendall.

For the simulations, the graph of the auto-covariances suggests to take $a_n = 4$ (see Figure 2.3). For $k > 4$ the auto-covariances $\hat{\gamma}(k)$ seem to be close to 0. We shall also consider the choice $a_n = 5$ to see if it gives better results than $a_n = 4$ for large samples. The results for $a_n = 4$ and $a_n = 5$ are given below.

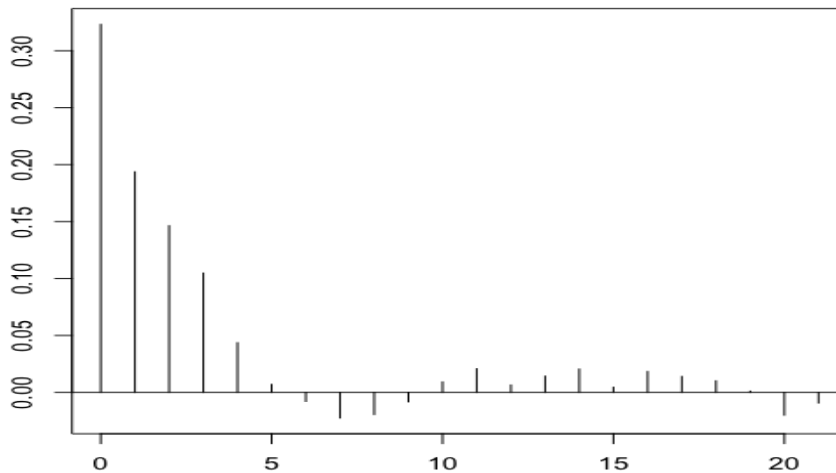


Figure 2.3: Graph of the auto-covariances $\hat{\gamma}(k)$ for example 3, with $n=150$

- $a_n = 4$

n	150	200	250	300	350	400	500	600
Estimated variance	1.395	1.288	1.181	1.242	1.138	1.124	1.087	1.093
Est. corrected test level	0.096	0.083	0.072	0.078	0.066	0.064	0.065	0.064
Est. Kendall test level	0.525	0.515	0.518	0.522	0.516	0.516	0.514	0.513

- $a_n = 5$

n	150	200	250	300	350	400	500	600
Estimated variance	1.383	1.341	1.178	1.284	1.155	1.113	1.051	1.044
Est. corrected test level	0.092	0.088	0.074	0.082	0.068	0.063	0.06	0.059
Est. Kendall test level	0.512	0.53	0.50	0.522	0.522	0.495	0.502	0.507

One can see that the choices $a_n = 4$ and $a_n = 5$ lead to similar results, except for large sample ($n \geq 500$) where the estimated level is slightly better for $a_n = 5$ (around 6%), in accordance with Proposition 2.1.

For this example, the usual Kendall test leads to a disastrous result, with an estimated level around 51%.

Remark 2.6. Following remark 2.3, we tried to implement the method described in [15], based on an AR fitting on the residuals. The results are presented below:

n	150	200	250	300	350	400	500	600
Estimated variance	1.260	1.231	1.197	1.100	1.121	1.106	1.096	1.011
Est. corrected test level	0.081	0.071	0.068	0.060	0.065	0.063	0.059	0.056
Est. Kendall test level	0.524	0.531	0.527	0.506	0.502	0.512	0.513	0.519

On this set of simulations, this method seems to work better than the "manual" choice of a_n . We can see that the estimated level is always between 0.05 and 0.08.

To be complete, we verify that the power of the corrected test tends to 1 as $n \rightarrow \infty$ for a model such that $\tau > 0$. This model is the following: (Z_i) is simulated as before, $X_i = Z_i - 0.44$ and $Y_i = X_i^2$.

We take $a_n = 4$ for the estimator V_n . The results are given below

n	150	200	250	300	350	400	500	600
Estimated variance	1.498	1.632	1.534	1.484	1.428	1.427	1.326	1.311
Estimated power	0.359	0.455	0.517	0.561	0.624	0.682	0.776	0.828

2.5 Proofs

2.5.1 Proof of theorem 2.1

Let us recall that $Z_i = (X_i, Y_i)$. We first recall the Hoeffding decomposition (see for instance Hoeffding [50]):

$$\sqrt{n}(U_n - \pi) = T_n + R_n, \quad (2.11)$$

$$\text{where } T_n := \frac{2}{\sqrt{n}} \sum_{i=1}^n (F(X_i, Y_i) + H(X_i, Y_i) - \pi),$$

$$R_n := \frac{2}{\sqrt{n}(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n (f(Z_i, Z_j) + f(Z_j, Z_i)),$$

$$\text{and } f(Z_i, Z_j) := \mathbb{1}_{X_i < X_j} \mathbb{1}_{Y_i < Y_j} - F(X_j, Y_j) - H(X_i, Y_i) + \pi/2.$$

The term T_n is the main term of this decomposition, and the term R_n is asymptotically negligible, as shown by the following proposition.

Proposition 2.3. *Let $(X_i, Y_i)_{i \in \mathbb{Z}}$ be a stationary sequence of \mathbb{R}^2 -valued random variables. If*

$$k \bar{\beta}_2(k) \xrightarrow[k \rightarrow +\infty]{} 0,$$

then

$$R_n \xrightarrow[n \rightarrow +\infty]{\mathbb{L}^2} 0.$$

Let us admit this proposition, which will be proved in appendix (subsection 2.6.2). It remains to show that, under H_0 and some condition on the $\tilde{\beta}$ -dependence coefficient,

$$T_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, V). \quad (2.12)$$

Let us recall the definition of the σ -algebra $\mathcal{F}_0 = \sigma(Z_k, k \leq 0)$, then according to Gordin [47] or Dedecker and Rio [29], if

$$\sum_{i=1}^{+\infty} \|\mathbb{E}(F(X_i, Y_i) + H(X_i, Y_i) - \pi | \mathcal{F}_0)\|_1 < +\infty, \quad (2.13)$$

then T_n converges in distribution to a mixture of Gaussian random variables. Now, if moreover $\tilde{\beta}_2(k) \rightarrow 0$ as $k \rightarrow \infty$ then the conditional variance in [29] is non random and (2.12) holds. Hence, it remains to check (2.13).

We first note that $\pi = 2\pi_1$ where $\pi_1 = \mathbb{P}(X^* < X, Y^* < Y)$. Indeed,

$$\begin{aligned} \pi &:= \mathbb{P}(\{(X^* - X)(Y^* - Y) > 0\}) \\ &= \mathbb{P}(\{\{X^* < X\} \cap \{Y^* < Y\}\} \cup \{\{X^* > X\} \cap \{Y^* > Y\}\}) \\ &= \mathbb{P}(\{\{X^* < X\} \cap \{Y^* < Y\}\}) + \mathbb{P}(\{\{X^* > X\} \cap \{Y^* > Y\}\}) = 2\pi_1. \end{aligned}$$

In order to verify (2.13), we control the conditional expectation of $F(X_i, Y_i) + H(X_i, Y_i) - \pi$ with respect to \mathcal{F}_0 .

Note first that,

$$\mathbb{E}(F(X_1, Y_1)) = \mathbb{E}[\mathbf{1}_{(X^* < X_1, Y^* < Y_1)}] = \mathbb{P}(X^* < X_1, Y^* < Y_1) = \pi_1. \quad (2.14)$$

Hence

$$\mathbb{E}[F(X_i, Y_i) - \pi_1 | \mathcal{F}_0] = \int [\mathbb{E}(\mathbf{1}_{x < X_i} \mathbf{1}_{y < Y_i} | \mathcal{F}_0) - \mathbb{E}(\mathbf{1}_{x < X_i} \mathbf{1}_{y < Y_i})] \mathbb{P}_{(X, Y)}(dx, dy),$$

and by definition of the coefficient $\tilde{\beta}_1$,

$$\|\mathbb{E}[F(X_i, Y_i) - \pi_1 | \mathcal{F}_0]\|_1 \leq \int \|\mathbb{E}(\mathbf{1}_{x < X_i} \mathbf{1}_{y < Y_i} | \mathcal{F}_0) - \mathbb{E}(\mathbf{1}_{x < X_i} \mathbf{1}_{y < Y_i})\|_1 \mathbb{P}_{(X, Y)}(dx, dy) \leq \tilde{\beta}_1(i). \quad (2.15)$$

It follows that

$$\begin{aligned} &\sum_{i=1}^n \|\mathbb{E}(F(X_i, Y_i) + H(X_i, Y_i) - \pi | \mathcal{F}_0)\|_1 \\ &\leq \sum_{i=1}^n \|\mathbb{E}(F(X_i, Y_i) - \pi_1 | \mathcal{F}_0)\|_1 + \sum_{i=1}^n \|\mathbb{E}(H(X_i, Y_i) - \pi_1 | \mathcal{F}_0)\|_1 \leq 2 \sum_{i=1}^n \tilde{\beta}_1(i). \end{aligned}$$

Hence, (2.13) is satisfied as soon as $\sum_{k=1}^{\infty} \tilde{\beta}_1(k) < \infty$, which concludes the proof of theorem 2.1.

2.5.2 Proof of proposition 2.1

Let us recall that $Z_i = (X_i, Y_i)$. The proof consists of three steps.

First step : We first introduce the function $G(s, t) = 2(F(s, t) + H(s, t))$. Let now

$$\gamma^*(k) = \frac{1}{n} \sum_{i=1}^{n-k} (G(Z_i) - 2\pi)(G(Z_{i+k}) - 2\pi) \quad \text{and} \quad \gamma(k) = \text{Cov}(G(Z_0), G(Z_k))$$

We shall prove that $V_n^* := \gamma^*(0) + 2 \sum_{k=1}^{a_n} \gamma^*(k)$ converges in \mathbb{L}^2 to V defined in theorem 1.

We first note that

$$\mathbb{E}(\gamma^*(k)) = \frac{1}{n} \sum_{i=1}^{n-k} \mathbb{E}((G(Z_i) - 2\pi)(G(Z_{i+k}) - 2\pi)) = \frac{n-k}{n} \gamma(k).$$

From (2.15) we get that $|\gamma(k)| \leq 4\tilde{\beta}_1(k)$. Since $\sum_{k>0} \tilde{\beta}_1(k) < \infty$ and since $a_n \rightarrow \infty$ as $n \rightarrow \infty$, we infer from the dominated convergence theorem that $\mathbb{E}(V_n^*)$ converges to V . Hence, it remains to show that

$$\text{Var}(V_n^*) \xrightarrow{n \rightarrow +\infty} 0. \quad (2.16)$$

By the ergodic theorem $\gamma^*(0)$ converge in \mathbb{L}^2 (and almost surely) to $\text{Var}(G(X_1, Y_1))$ (note that the sequence $(X_i, Y_i)_{i \in \mathbb{Z}}$ has no reason to be ergodic, but the fact that $\beta_1(n) \rightarrow 0$ as $n \rightarrow \infty$ implies that the limit is constant, and hence equal to $\text{Var}(G(X_1, Y_1))$). Hence, we only deal with the term

$$\text{Var} \left(\sum_{k=1}^{a_n} \gamma^*(k) \right) = \text{Var} \left(\frac{1}{n} \sum_{k=1}^{a_n} \sum_{i=1}^{n-k} (G(Z_i) - 2\pi)(G(Z_{i+k}) - 2\pi) \right).$$

For the sake of clarity, let $T_i = G(Z_i) - 2\pi$, and note that $\mathbb{E}(T_i T_{i+k}) = \gamma_k$. Then

$$\begin{aligned} \text{Var} \left(\sum_{k=1}^{a_n} \gamma^*(k) \right) &= \text{Var} \left(\frac{1}{n} \sum_{k=1}^{a_n} \sum_{i=1}^{n-k} T_i T_{i+k} \right) \\ &= \frac{1}{n^2} \sum_{k=1}^{a_n} \sum_{l=1}^{a_n} \sum_{i=1}^{n-k} \sum_{j=1}^{n-l} \text{Cov}(T_i T_{i+k}, T_j T_{j+l}) \\ &= \frac{1}{n^2} \sum_{k=1}^{a_n} \sum_{l=1}^{a_n} \sum_{i=1}^{n-k} \sum_{j=1}^{n-l} \mathbb{E}[(T_i T_{i+k} - \gamma_k)(T_j T_{j+l} - \gamma_l)]. \end{aligned}$$

We shall now control the terms $\mathbb{E}[(T_i T_{i+k} - \gamma_k)(T_j T_{j+l} - \gamma_l)]$ with the help of the coefficients $\tilde{\beta}$. As usual, this control depends on the gap between $i, i+k, j$ and $j+l$. Clearly, it suffices to deal with the sum over the set $\Gamma := \{i \leq j\}$. We then consider three distinct cases.

The sum over $\Gamma_1 = \{i+k \leq j\}$. In that case

$$|\mathbb{E}[(T_i T_{i+k} - \gamma_k)(T_j T_{j+l} - \gamma_l)]| \leq 2\tilde{\beta}_2(j - i - k).$$

Then for some positive constant C , we have (changing the indexes):

$$\frac{1}{n^2} \sum_{k=1}^{a_n} \sum_{l=1}^{a_n} \sum_{i=1}^{n-k} \sum_{j=1}^{n-l} |\mathbb{E}[(T_i T_{i+k} - \gamma_k)(T_j T_{j+l} - \gamma_l)]| \mathbf{1}_{\Gamma_1} \leq \frac{2}{n^2} \sum_{k=1}^{a_n} \sum_{l=1}^{a_n} \sum_{i=1}^n \sum_{j \geq 0} \tilde{\beta}_2(j) \leq \frac{C a_n^2}{n}.$$

The sum over $\Gamma_2 = \Gamma \cap \{j \leq i+k \leq j+l\}$

$$|\mathbb{E}[(T_i T_{i+k} - \gamma_k)(T_j T_{j+l} - \gamma_l)]| = |\mathbb{E}[(T_i T_{i+k} - \gamma_k) T_j T_{j+l}]| \leq 2\tilde{\beta}_1(j+l-i-k).$$

Then for some positive constant C , we have:

$$\frac{1}{n^2} \sum_{k=1}^{a_n} \sum_{l=1}^{a_n} \sum_{i=1}^{n-k} \sum_{j=1}^{n-l} |\mathbb{E} [(T_i T_{i+k} - \gamma_k) (T_j T_{j+l} - \gamma_l)]| \mathbb{1}_{\Gamma_2} \leq \frac{2}{n^2} \sum_{k=1}^{a_n} \sum_{l=1}^{a_n} \sum_{i=1}^n \sum_{j \geq 0} \tilde{\beta}_1(j) \leq \frac{C a_n^2}{n}.$$

The sum over $\Gamma_3 = \Gamma \cap \{i+k \geq j+l\}$

$$|\mathbb{E} [(T_i T_{i+k} - \gamma_k) (T_j T_{j+l} - \gamma_l)]| = |\mathbb{E} [(T_j T_{j+l} - \gamma_l) T_i T_{i+k}]| \leq 2 \tilde{\beta}_1(i+k-j-l).$$

Then for some positive constant C , we have:

$$\frac{1}{n^2} \sum_{k=1}^{a_n} \sum_{l=1}^{a_n} \sum_{i=1}^{n-k} \sum_{j=1}^{n-l} |\mathbb{E} [(T_i T_{i+k} - \gamma_k) (T_j T_{j+l} - \gamma_l)]| \mathbb{1}_{\Gamma_3} \leq \frac{2}{n^2} \sum_{k=1}^{a_n} \sum_{l=1}^{a_n} \sum_{j=1}^n \sum_{i \geq 0} \tilde{\beta}_1(i) \leq \frac{C a_n^2}{n}.$$

Consequently, from the last three upper bounds, we get

$$\frac{1}{n^2} \sum_{k=1}^{a_n} \sum_{l=1}^{a_n} \sum_{i=1}^{n-k} \sum_{j=1}^{n-l} |\mathbb{E} [(T_i T_{i+k} - \gamma_k) (T_j T_{j+l} - \gamma_l)]| \mathbb{1}_{\Gamma} \leq \frac{3C a_n^2}{n}.$$

Note that this result is still true on $\Gamma^c := \{i > j\}$ (interchanging i and j), in such a way that

$$\frac{1}{n^2} \sum_{k=1}^{a_n} \sum_{l=1}^{a_n} \sum_{i=1}^{n-k} \sum_{j=1}^{n-l} |\mathbb{E} [(T_i T_{i+k} - \gamma_k) (T_j T_{j+l} - \gamma_l)]| \leq \frac{6C a_n^2}{n}.$$

Finally, using the fact that $a_n = o(\sqrt{n})$, we conclude that $\sum_{k=1}^{a_n} \gamma^*(k)$ converges in \mathbb{L}^2 to $\sum_{k \geq 1} \gamma(k)$, and

$$V_n^* \xrightarrow[n \rightarrow +\infty]{\mathbb{L}^2} V. \quad (2.17)$$

Second step : In Appendix (Subsection 2.6.3), we shall prove the following lemma.

Lemma 2.2. *There exists a positive constant C such that*

$$\left\| \sup_{(s,t) \in \mathbb{R}^2} |G_n(s,t) - G(s,t)| \right\|_2^2 \leq C \frac{(\log(n))^4}{n} \sum_{k=0}^n \tilde{\beta}_1(k). \quad (2.18)$$

Last step :

Using (2.18), we see that we can replace the non-observable quantity $\gamma^*(k)$ by the estimator $\hat{\gamma}(k)$ in the expression of V_n^* , provided that $a_n(\log(n))^2 n^{-1/2}$ tends to zero as n tends to infinity. This completes the proof of Proposition 2.1.

2.6 Appendix

2.6.1 Proof of lemma 2.1

We first recall some elementary facts given in [23]. Let $(T_k)_{k \geq 0}$ be the sequence of stopping times defined by

$$T_0 = \inf\{i > 1 : Z_i \neq Z_{i-1}\} \quad \text{and} \quad T_k = \inf\{i > T_{k-1} : Z_i \neq Z_{i-1}\} \quad \text{for } k > 0.$$

Let $\tau_k = T_{k+1} - T_k$. The random variables $(Z_{T_k}, \tau_k)_{k \geq 0}$ are iid, Z_{T_k} has law ν , and the conditional distribution of τ_k given $Z_{T_k} = x$ is the geometric distribution $\mathcal{G}(x)$. Note that τ_0 has a weak moment of order 2: there exists $c > 0$ such that $\mathbb{P}(\tau_0 > x) \leq cx^{-2}$ for any $x > 0$.

Let now $N(n) = \inf\{T_k : T_k \leq n\}$. We can write

$$\sum_{i=1}^n (Z_i - 0.5) = (T_0 - 1)(Z_1 - 0.5) + \sum_{k=0}^{N(n)-1} \tau_k (Z_{T_k} - 0.5) + (n - T_{N(n)})(Z_{T_{N(n)}} - 0.5). \quad (2.19)$$

Clearly $(T_0 - 1)(Z_0 - 0.5)/\sqrt{n \log n}$ converges to 0 in probability, so this term is negligible for the convergence in distribution. Let us now consider the last term in (2.19), which is a bit more complicated. We first note that

$$|(n - T_{N(n)})(Z_{T_{N(n)}} - 0.5)| \leq n - T_{N(n)}. \quad (2.20)$$

Let t be any integer in $[0, n - 2]$, then

$$\begin{aligned} \mathbb{P}(n - T_{N(n)} > t) &= \sum_{k=1}^{n-t-1} \mathbb{P}(T_{N(n)} = k) = \sum_{k=1}^{n-t-1} \sum_{l=1}^k \mathbb{P}(T_l = k, N(n) = l) \\ &= \sum_{k=1}^{n-t-1} \sum_{l=1}^k \mathbb{P}(T_l = k, T_{l+1} > n) = \sum_{k=1}^{n-t-1} \sum_{l=1}^k \mathbb{P}(T_l = k, \tau_l > n - k). \end{aligned}$$

Since T_l is independent of τ_l , and since the τ_i 's are iid, we get

$$\mathbb{P}(n - T_{N(n)} > t) = \sum_{k=1}^{n-t-1} \sum_{l=1}^k \mathbb{P}(T_l = k) \mathbb{P}(\tau_0 > n - k) \leq \sum_{k=1}^{n-t-1} \mathbb{P}(\tau_0 > n - k).$$

Recall that τ_0 has a weak moment of order 2. Hence

$$\mathbb{P}(n - T_{N(n)} > t) \leq c \sum_{k=1}^{n-t-1} \frac{1}{(n - k)^2} \leq \frac{c'}{t} \quad (2.21)$$

for some $c' > 0$. From (2.20) and (2.21) we easily infer that $(n - T_{N(n)})(Z_{T_{N(n)}} - 0.5)/\sqrt{n \log n}$ converges to zero in probability.

From these considerations, we see that to prove Lemma 1, it is equivalent to prove that

$$\frac{2}{\sqrt{n \log n}} \sum_{k=0}^{N(n)-1} \tau_k (Z_{T_k} - 0.5) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1). \quad (2.22)$$

To do this, the main point is to prove that the random variable $\tau_0(Z_{T_0} - 0.5)$ has a weak moment of order 2, and to apply a result of Feller [42] on the domain of attraction of the normal distribution. So, let us compute the tails of $\tau_0(Z_{T_0} - 0.5)$. If $t > 0$,

$$\mathbb{P}(\tau_0(Z_{T_0} - 0.5) > t) = 2 \int_{1/2}^1 x(1-x)^{[t/(x-0.5)]} dx,$$

and one can easily see that

$$\lim_{t \rightarrow \infty} t^2 \mathbb{P}(\tau_0(Z_{T_0} - 0.5) > t) = 0.$$

Now, for $t > 0$,

$$\mathbb{P}(\tau_0(Z_{T_0} - 0.5) < -t) = \mathbb{P}(\tau_0(0.5 - Z_{T_0}) > t) = 2 \int_0^{1/2} x(1-x)^{[t/(0.5-x)]} dx,$$

Hence

$$\mathbb{P}(\tau_0(Z_{T_0} - 0.5) < -t) = \frac{2}{t^2} \int_0^{t/2} y(1-(y/t))^{[t/(0.5-(y/t))]} dy,$$

and by the dominated convergence theorem, we obtain that

$$\lim_{t \rightarrow \infty} t^2 \mathbb{P}(\tau_0(Z_{T_0} - 0.5) < -t) = 2 \int_0^\infty y \exp(-2y) dy = 0.5.$$

It follows that $\tau_0(Z_{T_0} - 0.5)$ is in the domain of attraction of the normal distribution (see Feller [42] and also Gouëzel [48], Section 1.2.2, for a short exposition), and that

$$\frac{\sqrt{2}}{\sqrt{n \log n}} \sum_{k=0}^{n-1} \tau_k(Z_{T_k} - 0.5) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1). \quad (2.23)$$

To conclude, it remains to replace n by $N(n)$ in (2.23), in order to get (2.22). More precisely, we shall prove that,

$$\frac{1}{\sqrt{n \log n}} \left| \sum_{k=0}^{[n/\mathbb{E}(\tau_0)]-1} \tau_k(Z_{T_k} - 0.5) - \sum_{k=0}^{N(n)-1} \tau_k(Z_{T_k} - 0.5) \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \quad (2.24)$$

Note that $\mathbb{E}(\tau_0) = 2$, and recall that $N(n)/n$ converges almost surely to $1/\mathbb{E}(\tau_0) = 1/2$.

Let us prove (2.24). Let $\epsilon > 0, \delta > 0$, and let $W_k = \tau_k(Z_{T_k} - 0.5)$. We have

$$\begin{aligned} \mathbb{P} \left(\left| \sum_{k=0}^{[n/2]-1} W_k - \sum_{k=0}^{N(n)-1} W_k \right| > \epsilon \sqrt{n \log n} \right) &\leq \mathbb{P}(|N(n) - [n/2]| > \delta n) \\ &+ \mathbb{P} \left(\left| \sum_{k=0}^{[n/2]-1} W_k - \sum_{k=0}^{N(n)-1} W_k \right| > \epsilon \sqrt{n \log n}, |N(n) - [n/2]| \leq \delta n \right). \end{aligned} \quad (2.25)$$

Now, the first term on right hand in (2.25) tends to zero as $n \rightarrow \infty$. For the second term, we easily see that it is smaller than

$$2\mathbb{P} \left(\max_{1 \leq k \leq \delta n} \left| \sum_{i=1}^k W_i \right| > \epsilon \sqrt{n \log n} \right).$$

Using Etemadi's inequality [41], we get that

$$\mathbb{P} \left(\max_{1 \leq k \leq \delta n} \left| \sum_{i=1}^k W_i \right| > \epsilon \sqrt{n \log n} \right) \leq 3 \max_{1 \leq k \leq \delta n} \mathbb{P} \left(3 \left| \sum_{i=1}^k W_i \right| > \epsilon \sqrt{n \log n} \right). \quad (2.26)$$

Since $(n \log n)^{-1/2} \sum_{i=1}^n W_k$ converges in distribution (see (2.23)), we easily see that

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \max_{1 \leq k \leq \delta n} \mathbb{P} \left(3 \left| \sum_{i=1}^k W_i \right| > \epsilon \sqrt{n \log n} \right) = 0.$$

Going back to (2.25), we obtain that, for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \sum_{k=0}^{\lfloor n/2 \rfloor - 1} W_k - \sum_{k=0}^{N(n)-1} W_k \right| > \epsilon \sqrt{n \log n} \right) = 0,$$

and (2.24) is proved.

Combining (2.23) and (2.24), and bearing in mind that $\mathbb{E}(\tau_0) = 2$, we infer that

$$\frac{2}{\sqrt{n \log n}} \sum_{k=0}^{N(n)-1} \tau_k (Z_{T_k} - 0.5) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1), \quad (2.27)$$

which is exactly (2.22). This completes the proof of Lemma 1.

2.6.2 Proof of proposition 2.3

Let us recall that $Z_i = (X_i, Y_i)$. We shall prove that:

$$\mathbb{E} \left(\left(\frac{2}{n\sqrt{n}} \sum_{i=1}^n \sum_{\substack{j=1 \\ i < j}}^n f(Z_i, Z_j) \right)^2 \right) = \frac{4}{n^3} \sum_{i=1}^n \sum_{\substack{j=1 \\ i < j}}^n \sum_{k=1}^n \sum_{\substack{l=1 \\ k < l}}^n \mathbb{E} [f(Z_i, Z_j) f(Z_k, Z_l)] \xrightarrow[n \rightarrow +\infty]{} 0. \quad (2.28)$$

There are many different cases, but it suffices to deal with the sum over the two sets

$$\Gamma_1 = \{(i, j, k, l) : i < j \leq k < l\} \quad \text{and} \quad \Gamma_2 = \{(i, j, k, l) : i < k < l < j\},$$

and the other cases can be handled in the same way (up to index permutations).

To prove the inequality (2.28), it remains to prove that

$$\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_1} |\mathbb{E}[f(Z_i, Z_j) f(Z_k, Z_l)]| \xrightarrow[n \rightarrow +\infty]{} 0 \quad \text{and} \quad \frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_2} |\mathbb{E}[f(Z_i, Z_j) f(Z_k, Z_l)]| \xrightarrow[n \rightarrow +\infty]{} 0.$$

For each case we proceed in two steps:

- First case.

Step 1. Let $(i, j, k, l) \in \Gamma_{1,1} = \{(i, j, k, l) : i < j \leq k < l, l - k > k - j\}$. We start from the inequality

$$\begin{aligned} |\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)]| &= \int |\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)|Z_i = z_i, Z_j = z_j, Z_k = z_k]| \mathbb{P}_{(Z_i, Z_j, Z_k)}(dz_i, dz_j, dz_k) \\ &\leq \int |f(z_i, z_j)| \cdot |\mathbb{E}[f(z_k, Z_l)|Z_i = z_i, Z_j = z_j, Z_k = z_k]| \mathbb{P}_{(Z_i, Z_j, Z_k)}(dz_i, dz_j, dz_k). \end{aligned}$$

Since $|f(x, y)| \leq 2$,

$$\begin{aligned} &|\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)]| \\ &\leq 2 \int |\mathbb{E}[\mathbb{1}_{x_k < X_l} \mathbb{1}_{y_k < Y_l} - F(X_l, Y_l) - H(x_k, y_k) + \pi_1 | Z_i = z_i, Z_j = z_j, Z_k = z_k]| \mathbb{P}_{(Z_i, Z_j, Z_k)}(dz_i, dz_j, dz_k) \\ &\leq 2 \int |\mathbb{E}[\mathbb{1}_{x_k < X_l} \mathbb{1}_{y_k < Y_l} - H(x_k, y_k) | Z_i = z_i, Z_j = z_j, Z_k = z_k]| \mathbb{P}_{(Z_i, Z_j, Z_k)}(dz_i, dz_j, dz_k) \\ &\quad + 2 \int |\mathbb{E}[F(X_l, Y_l) - \pi_1 | Z_i = z_i, Z_j = z_j, Z_k = z_k]| \mathbb{P}_{(Z_i, Z_j, Z_k)}(dz_i, dz_j, dz_k). \end{aligned}$$

For the first term on the right hand side, we use the fact that

$$\mathbb{E}[\mathbb{1}_{x_k < X_l} \mathbb{1}_{y_k < Y_l}] = \mathbb{P}(x_k < X_l, y_k < Y_l) = H(x_k, y_k),$$

and the definition of $\tilde{\beta}_1$. It follows that

$$\begin{aligned} &\int |\mathbb{E}[\mathbb{1}_{x_k < X_l} \mathbb{1}_{y_k < Y_l} - H(x_k, y_k) | Z_i = z_i, Z_j = z_j, Z_k = z_k]| \mathbb{P}_{(Z_i, Z_j, Z_k)}(dz_i, dz_j, dz_k) \\ &\leq \int \sup_{(x, y) \in \mathbb{R}^2} |\mathbb{E}[\mathbb{1}_{x < X_l} \mathbb{1}_{y < Y_l} - H(x, y) | Z_i = z_i, Z_j = z_j, Z_k = z_k]| \mathbb{P}_{(Z_i, Z_j, Z_k)}(dz_i, dz_j, dz_k) \\ &\leq \left\| \sup_{(x, y) \in \mathbb{R}^2} |\mathbb{E}[\mathbb{1}_{x < X_l} \mathbb{1}_{y < Y_l} - H(x, y) | Z_i, Z_j, Z_k]| \right\|_1 \leq \tilde{\beta}_1(l - k). \end{aligned}$$

Let us now control the second term. From (2.15), we have

$$\|\mathbb{E}[F(X_l, Y_l) - \pi_1 | Z_i, Z_j, Z_k]\|_1 \leq \tilde{\beta}_1(l - k).$$

Consequently,

$$\frac{1}{n^3} \sum_{(i, j, k, l) \in \Gamma_{1,1}} |\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)]| \leq \frac{1}{n^3} \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n \sum_{l=k+1}^n 4\beta_1(l - k) \mathbb{1}_{l-k > k-j}.$$

Setting $l - k = t$, we get that

$$\begin{aligned} \frac{1}{n^3} \sum_{(i, j, k, l) \in \Gamma_{1,1}} |\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)]| &\leq \frac{4}{n^3} \sum_{t=1}^n \beta_1(t) \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n \mathbb{1}_{k < t+j} \\ &\leq \frac{2}{n} \sum_{t=1}^n t \cdot \tilde{\beta}_1(t). \end{aligned}$$

Conclusion of Step 1: If $k\tilde{\beta}_1(k) \xrightarrow[k \rightarrow +\infty]{} 0$ then

$$\frac{4}{n^3} \sum_{(i,j,k,l) \in \Gamma_{1,1}} |\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)]| \xrightarrow[n \rightarrow +\infty]{} 0. \quad (2.29)$$

Step 2. Let $(i, j, k, l) \in \Gamma_{1,2} = \{(i, j, k, l) : i < j \leq k < l, k - j > l - k\}$. We start from the elementary decomposition

$$\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)] = \mathbb{E}[f(Z_i, Z_j)[f(Z_k, Z_l) - \mathbb{E}f(Z_k, Z_l)]] + \mathbb{E}f(Z_i, Z_j) \cdot \mathbb{E}f(Z_k, Z_l). \quad (2.30)$$

For the first term on the right hand side of (2.30), we work conditionally on Z_i, Z_j .

$$\begin{aligned} |\mathbb{E}[f(Z_i, Z_j)[f(Z_k, Z_l) - \mathbb{E}f(Z_k, Z_l)]]| &\leq \mathbb{E}[|f(Z_i, Z_j)\mathbb{E}[f(Z_k, Z_l) - \mathbb{E}f(Z_k, Z_l)|Z_i, Z_j]|] \\ &\leq 2 \|\mathbb{E}[f(Z_k, Z_l) - \mathbb{E}f(Z_k, Z_l)|Z_i, Z_j]\|_1. \end{aligned}$$

Recall that $f(Z_k, Z_l) = \mathbb{1}_{X_k < X_l} \mathbb{1}_{Y_k < Y_l} - F(X_l, Y_l) - H(X_k, Y_k) + \pi_1$. Hence

$$\begin{aligned} \|\mathbb{E}[f(Z_k, Z_l) - \mathbb{E}f(Z_k, Z_l)|Z_i, Z_j]\|_1 &= \|\mathbb{E}[\mathbb{1}_{X_k < X_l} \mathbb{1}_{Y_k < Y_l} - \mathbb{E}(\mathbb{1}_{X_k < X_l} \mathbb{1}_{Y_k < Y_l})|Z_i, Z_j]\|_1 \\ &\quad + \|\mathbb{E}[F(X_l, Y_l) - \mathbb{E}F(X_l, Y_l)|Z_i, Z_j]\|_1 \\ &\quad + \|\mathbb{E}[H(X_k, Y_k) - \mathbb{E}H(X_k, Y_k)|Z_i, Z_j]\|_1. \end{aligned}$$

We shall now give an upper bound for the last three terms by using the properties of the coefficient $\bar{\beta}_2$.

$$\|\mathbb{E}[\mathbb{1}_{X_k < X_l} \mathbb{1}_{Y_k < Y_l} - \mathbb{E}[\mathbb{1}_{X_k < X_l} \mathbb{1}_{Y_k < Y_l}]|Z_i, Z_j]\|_1 \leq \delta_2(k - j) \leq \bar{\beta}_2(k - j). \quad (2.31)$$

From (2.15), we infer that

$$\|\mathbb{E}[H(X_k, Y_k) - \mathbb{E}(H(X_k, Y_k))|Z_i, Z_j]\|_1 = \|\mathbb{E}[H(X_k, Y_k) - \pi_1|Z_i, Z_j]\|_1 \leq \tilde{\beta}_1(k - j) \leq \bar{\beta}_2(k - j). \quad (2.32)$$

For the last term, we use also the fact that $l - j > k - j$, and that the coefficient $\tilde{\beta}_1$ is decreasing. So

$$\|\mathbb{E}[F(X_l, Y_l) - \mathbb{E}(F(X_l, Y_l))|Z_i, Z_j]\|_1 \leq \tilde{\beta}_1(l - j) \leq \tilde{\beta}_1(k - j) \leq \bar{\beta}_2(k - j). \quad (2.33)$$

It follows from (2.31), (2.32) and (2.33) that

$$\|\mathbb{E}[f(Z_k, Z_l) - \mathbb{E}f(Z_k, Z_l)|Z_i, Z_j]\|_1 \leq 6\bar{\beta}_2(k - j).$$

Therefore,

$$\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{1,2}} |\mathbb{E}[f(Z_i, Z_j)[f(Z_k, Z_l) - \mathbb{E}f(Z_k, Z_l)]]| \leq \frac{6}{n^3} \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n \sum_{l=k+1}^n \bar{\beta}_2(k - j) \mathbb{1}_{k-j > l-k}.$$

Setting $k - j = s$ and $l - k = t$, we obtain

$$\begin{aligned} \frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{1,2}} |\mathbb{E}[f(Z_i, Z_j) [f(Z_k, Z_l) - \mathbb{E}f(Z_k, Z_l)]]| &\leq \frac{6}{n^3} \sum_{i=1}^n \sum_{j=i+1}^n \sum_{t=1}^n \sum_{s=1}^n \bar{\beta}_2(s) \mathbf{1}_{s>t} \\ &\leq \frac{3}{n} \sum_{s=1}^n s \bar{\beta}_2(s). \end{aligned}$$

Hence, we have proved that: if $k\bar{\beta}_2(k) \xrightarrow[k \rightarrow +\infty]{} 0$, then

$$\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{1,2}} |\mathbb{E}[f(Z_i, Z_j) [f(Z_k, Z_l) - \mathbb{E}f(Z_k, Z_l)]]| \xrightarrow[n \rightarrow +\infty]{} 0. \quad (2.34)$$

In the rest of the proof of Step 2, we control the second term on the right hand side of (2.30). We first note that

$$\begin{aligned} |\mathbb{E}f(Z_i, Z_j)| &\leq \int |\mathbb{E}[f(z_i, Z_j) | Z_i = z_i]| \mathbb{P}_{Z_i}(dz_i) \\ &\leq \int \sup_{z \in \mathbb{R}^2} |\mathbb{E}[f(z, Z_j) | Z_i = z_i]| \mathbb{P}_{Z_i}(dz_i) \\ &\leq \left\| \sup_{z \in \mathbb{R}^2} |\mathbb{E}[f(z, Z_j) | Z_i]| \right\|_1 \\ &\leq \left\| \sup_{z \in \mathbb{R}^2} |\mathbb{E}[\mathbf{1}_{x < X_j} \mathbf{1}_{y < Y_j} - H(x, y) | Z_i]| \right\|_1 + \left\| \sup_{z \in \mathbb{R}^2} |\mathbb{E}[F(X_j, Y_j) - \pi_1 | Z_i]| \right\|_1 \\ &\leq 2\tilde{\beta}_1(j - i). \end{aligned}$$

In the same way

$$|\mathbb{E}f(Z_k, Z_l)| \leq 2\tilde{\beta}_1(l - k).$$

Clearly

$$|\mathbb{E}f(Z_i, Z_j) \cdot \mathbb{E}f(Z_k, Z_l)| \leq 4\tilde{\beta}_1(j - i)\tilde{\beta}_1(l - k).$$

Hence

$$\begin{aligned} \frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{1,2}} |\mathbb{E}f(Z_i, Z_j) \cdot \mathbb{E}f(Z_k, Z_l)| &\leq \frac{4}{n^3} \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n \sum_{l=k+1}^n \tilde{\beta}_1(j - i)\tilde{\beta}_1(l - k) \\ &\leq \frac{4}{n} \sum_{p=1}^n \sum_{t=1}^n \tilde{\beta}_1(p)\tilde{\beta}_1(t) = \frac{4}{n} \left(\sum_{k=1}^n \tilde{\beta}_1(k) \right)^2. \quad (2.35) \end{aligned}$$

Since $k\tilde{\beta}_1(k) \xrightarrow[k \rightarrow +\infty]{} 0$, we get that

$$\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{1,2}} |\mathbb{E}f(Z_i, Z_j) \cdot \mathbb{E}f(Z_k, Z_l)| = O\left(\frac{(\log n)^2}{n}\right).$$

Hence, we have proved that: if $k\tilde{\beta}_1(k) \xrightarrow[k \rightarrow +\infty]{} 0$, then

$$\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{1,2}} |\mathbb{E}f(Z_i, Z_j) \cdot \mathbb{E}f(Z_k, Z_l)| \xrightarrow[n \rightarrow +\infty]{} 0. \quad (2.36)$$

Conclusion of Step 2: From (2.34) and (2.36), we infer that, if $k\bar{\beta}_2(k) \xrightarrow[k \rightarrow +\infty]{} 0$ then

$$\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{1,2}} |\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)]| \xrightarrow[n \rightarrow +\infty]{} 0. \quad (2.37)$$

Conclusion of the first case: From (2.29) and (2.37) we infer that, if $k\bar{\beta}_2(k) \xrightarrow[k \rightarrow +\infty]{} 0$, then

$$\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_1} |\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)]| \xrightarrow[n \rightarrow +\infty]{} 0. \quad (2.38)$$

• Second case.

Step 1. Let $(i, j, k, l) \in \Gamma_{2,1} = \{(i, j, k, l) : i < k < l < j, j - l > l - k\}$. Following the same strategy as to deal with $\Gamma_{1,1}$, we have that

$$|\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)]| \leq 4\tilde{\beta}_1(j - l).$$

Then, setting $j - l = t$, we get

$$\begin{aligned} \frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{2,1}} |\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)]| &\leq \frac{4}{n^3} \sum_{i=1}^n \sum_{k=i+1}^n \sum_{l=k+1}^n \sum_{j=l+1}^n \tilde{\beta}_1(j - l) \mathbf{1}_{j-l > l-k} \\ &\leq \frac{2}{n} \sum_{t=1}^n t \tilde{\beta}_1(t). \end{aligned}$$

Conclusion of Step 1: If $k\tilde{\beta}_1(k) \xrightarrow[k \rightarrow +\infty]{} 0$ then

$$\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{2,1}} |\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)]| \xrightarrow[n \rightarrow +\infty]{} 0. \quad (2.39)$$

Step 2. Let $(i, j, k, l) \in \Gamma_{2,2} = \{(i, j, k, l) : i < k < l < j, l - k > j - l\}$. Let also $(Z_i^*)_{i \in \mathbb{Z}}$ be an independent copy of $(Z_i)_{i \in \mathbb{Z}}$, and let us write

$$\begin{aligned} &\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)] \\ &= \mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)] - \mathbb{E}[f(Z_i, Z_j^*)f(Z_k, Z_l^*)] + \mathbb{E}[f(Z_i, Z_j^*)f(Z_k, Z_l^*)] \\ &= \mathbb{E} \left[(f(Z_i, Z_j)f(Z_k, Z_l)) - \int f(Z_i, z_j)f(Z_k, z_l) \mathbb{P}_{(Z_j, Z_l)}(dz_j, dz_l) \right] + \mathbb{E}[f(Z_i, Z_j^*)f(Z_k, Z_l^*)]. \end{aligned}$$

Let us first control the first term on the right hand side.

$$\begin{aligned}
& \left| \mathbb{E} \left[f(Z_i, Z_j) f(Z_k, Z_l) - \int f(Z_i, z_j) f(Z_k, z_l) \mathbb{P}_{(Z_j, Z_l)}(dz_j, dz_l) \right] \right| \\
&= \left| \int \mathbb{E} \left[\left[f(z_i, Z_j) f(z_k, Z_l) - \int f(z_i, z_j) f(z_k, z_l) \mathbb{P}_{(Z_j, Z_l)}(dz_j, dz_l) \right] \middle| Z_i = z_i, Z_k = z_k \right] \mathbb{P}_{(Z_i, Z_k)}(dz_i, dz_k) \right| \\
&\leq \left\| \sup_{(z_1, z_2) \in \mathbb{R}^2 \times \mathbb{R}^2} \left| \mathbb{E} [f(z_1, Z_j) f(z_2, Z_l) | Z_i, Z_k] - \mathbb{E} [f(z_1, Z_j) f(z_2, Z_l)] \right| \right\|_1.
\end{aligned}$$

Now, by definition of f and of the coefficients $\tilde{\beta}_2$, and since $l - k < j - k$, we easily infer that

$$\left| \mathbb{E} \left[f(Z_i, Z_j) f(Z_k, Z_l) - \int f(Z_i, z_j) f(Z_k, z_l) \mathbb{P}_{(Z_j, Z_l)}(dz_j, dz_l) \right] \right| \leq 9\tilde{\beta}_2(l - k) \quad (2.40)$$

Consequently, setting $l - k = t$ and $j - l = p$, we get

$$\begin{aligned}
\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{2,2}} \left| \mathbb{E} \left[f(Z_i, Z_j) f(Z_k, Z_l) - \int f(Z_i, z_j) f(Z_k, z_l) \mathbb{P}_{(Z_l, Z_j)}(dz_l, dz_j) \right] \right| \\
\leq \frac{9}{n^3} \sum_{i=1}^n \sum_{k=i+1}^n \sum_{t=1}^n \sum_{p=1}^n \tilde{\beta}_2(t) \mathbf{1}_{p < t} \leq \frac{9}{n} \sum_{t=1}^n t \tilde{\beta}_2(t).
\end{aligned}$$

Hence, we have proved that: if $k \tilde{\beta}_2(k) \xrightarrow[k \rightarrow +\infty]{} 0$, then

$$\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{2,2}} \left| \mathbb{E} \left[f(Z_i, Z_j) f(Z_k, Z_l) - \int f(Z_i, z_j) f(Z_k, z_l) \mathbb{P}_{(Z_l, Z_j)}(dz_l, dz_j) \right] \right| \xrightarrow[n \rightarrow +\infty]{} 0. \quad (2.41)$$

Now, it remains to control the term $\mathbb{E}[f(Z_i, Z_j^*) f(Z_k, Z_l^*)]$. Since Z and Z^* are independent, one can consider that either Z or Z^* are fixed. Hence, we obtain the upper bound

$$|\mathbb{E} [f(Z_i, Z_j^*) f(Z_k, Z_l^*)]| \leq 2 \cdot \min\{\tilde{\beta}_1(j - l), \tilde{\beta}_1(k - i)\}.$$

Therefore, setting $p = j - l$, $t = l - k$, $s = k - i$, we get

$$\begin{aligned}
\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{2,2}} |\mathbb{E} [f(Z_i, Z_j^*) f(Z_k, Z_l^*)]| &\leq \frac{2}{n^3} \sum_{i=1}^n \sum_{k=i+1}^n \sum_{l=k+1}^n \sum_{j=l+1}^n \min\{\tilde{\beta}_1(j - l), \tilde{\beta}_1(k - i)\} \\
&\leq \frac{2}{n} \sum_{p=1}^n \sum_{s=1}^n \min\{\tilde{\beta}_1(p), \tilde{\beta}_1(s)\} \\
&\leq \frac{4}{n} \sum_{p=1}^n p \tilde{\beta}_1(p)
\end{aligned}$$

It follows that: if $k \tilde{\beta}_1(k) \xrightarrow[k \rightarrow +\infty]{} 0$, then

$$\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{2,2}} |\mathbb{E} [f(Z_i, Z_j^*) f(Z_k, Z_l^*)]| \xrightarrow[n \rightarrow +\infty]{} 0. \quad (2.42)$$

Conclusion of Step 2: From (2.41) and (2.42), we infer that, if $k\tilde{\beta}_2(k) \xrightarrow[k \rightarrow +\infty]{} 0$ then

$$\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_{2,2}} |\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)]| \xrightarrow[n \rightarrow +\infty]{} 0. \quad (2.43)$$

Conclusion of the second case, and of the proof: From (2.39) and (2.43), we infer that

$$\frac{1}{n^3} \sum_{(i,j,k,l) \in \Gamma_2} |\mathbb{E}[f(Z_i, Z_j)f(Z_k, Z_l)]| \xrightarrow[n \rightarrow +\infty]{} 0. \quad (2.44)$$

Combining (2.38) and (2.44), the proof of Proposition 2.3 is complete.

2.6.3 Proof of lemma 2.2

In fact, it suffices to prove (2.18) for $F_n(s, t) - F(s, t)$, since the term involving $H_n(s, t) - H(s, t)$ can be handled similarly.

We define the empirical process by

$$\mu_n(s, t) = \sqrt{n}(F_n(s, t) - F(s, t)).$$

To simplify the rest of the proof, we reduce the interval of definition of (s, t) from \mathbb{R}^2 to $[0, 1]^2$. We define a random variable $(U_i, V_i)_{i \in \mathbb{Z}}$ with values in $[0, 1]^2$, such that $(U_i, V_i) = (F_X(X_i), F_Y(Y_i))$.

Without loss of generality, assume that the distribution functions F_X and F_Y are continuous, so that the random variables (U_i, V_i) have a uniform distribution (if it is not the case one can proceed as in the beginning of Section 4.1 in [26]). Then,

$$\left\| \sup_{(s,t) \in \mathbb{R}^2} |F_n(s, t) - F(s, t)| \right\|_2^2 = \left\| \sup_{(u,v) \in [0,1]^2} \left| \frac{1}{n} \sum_{i=1}^n (\mathbf{1}_{U_i \leq u, V_i \leq v} - \mathbb{P}(U_0 \leq u, V_0 \leq v)) \right| \right\|_2^2$$

In the rest of the proof, we apply a dyadic chaining (following [64], Chapter 7). Let K be some non-negative integer and, for any $z = (z_1, z_2)$ in the unit square $]0, 1]^2$, let

$$\Pi_K(z) = (\Pi_K(z_1), \Pi_K(z_2)), \quad \text{with} \quad \Pi_K(z_1) = 2^{-K} \lfloor 2^K z_1 \rfloor.$$

Let N be the unique integer such that $2^{N-1} < n \leq 2^N$. Clearly

$$\mu_n(z) = \mu_n(z) - \mu_n(\Pi_N(z)) + \mu_n(\Pi_N(z)).$$

Consequently,

$$\sup_{z \in [0,1]^2} |\mu_n(z)| \leq \underbrace{\sup_{z \in [0,1]^2} |\mu_n(z) - \mu_n(\Pi_N(z))|}_{R_N} + \underbrace{\sup_{z \in [0,1]^2} |\mu_n(\Pi_N(z))|}_{\Delta}.$$

Let us first control the main term Δ . For any $z = (z_1, z_2)$ in the unit square $]0, 1]^2$, let $]0, z] =]0, z_1] \times]0, z_2]$.

For any $j \in (1, 2)$ and any natural integer M ,

$$]0, \Pi_M(z_j)] = \bigcup_{L_j=0}^M]\Pi_{L_j-1}(z_j), \Pi_{L_j}(z_j)].$$

with the convention $\Pi_{-1}(z_j) = 0$. Then

$$]0, \Pi_M(z)] = \bigcup_{L \in [0, M]^2} \prod_{j=1}^2]\Pi_{L_j-1}(z_j), \Pi_{L_j}(z_j)].$$

Let \mathcal{D}_L be the class of dyadic boxes $\prod_{i=1}^2](k_i - 1)2^{-L_i}, k_i 2^{-L_i}]$ where $k = (k_1, k_2)$. Let $Z_n = \sqrt{n}(P_n - P)$ be the empirical and centered empirical measure where P denote the common marginal distribution of (U_i, V_i) and P_n the empirical measure. Define

$$\Delta_L := \sup_{S \in \mathcal{D}_L} |Z_n(S)|.$$

Then

$$\Delta \leq \sum_{L \in [0, N]^2} \Delta_L. \quad (2.45)$$

Moreover,

$$\begin{aligned} \|Z_n(S)\|_2^2 &= \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}_{(U_i, V_i) \in S} - \mathbb{P}((U_i, V_i) \in S) \right)^2 \right] \\ &\leq \text{Var}(\mathbf{1}_{(U_0, V_0) \in S}) + 2 \sum_{k=1}^{n-1} |\mathbb{E}((\mathbf{1}_{(U_0, V_0) \in S} - \mathbb{P}((U_i, V_i) \in S)) (\mathbf{1}_{(U_k, V_k) \in S} - \mathbb{P}((U_i, V_i) \in S)))| \\ &\leq \text{Var}(\mathbf{1}_{(U_0, V_0) \in S}) + 2 \sum_{k=1}^{n-1} |\mathbb{E}(\mathbf{1}_{(U_0, V_0) \in S} (\mathbf{1}_{(U_k, V_k) \in S} - \mathbb{P}((U_i, V_i) \in S)))|. \end{aligned}$$

Let $\mathcal{F}_0 = \sigma((U_i, V_i), i \leq 0)$, then

$$\begin{aligned} &\leq \text{Var}(\mathbf{1}_{(U_0, V_0) \in S}) + 2 \sum_{k=1}^n |\mathbb{E}(\mathbf{1}_{(U_0, V_0) \in S} \mathbb{E}(\mathbf{1}_{(U_k, V_k) \in S} - \mathbb{P}((U_i, V_i) \in S) | \mathcal{F}_0))| \\ &\leq \text{Var}(\mathbf{1}_{(U_0, V_0) \in S}) + 2 \sum_{k=1}^n |\mathbb{E}(\mathbf{1}_{(U_0, V_0) \in S} b(k))| \\ &\leq \mathbb{E} \left(\mathbf{1}_{(U_0, V_0) \in S} \left(1 + 2 \sum_{k=1}^n b(k) \right) \right). \end{aligned}$$

Hence,

$$\begin{aligned} \|\Delta_L\|_2^2 &\leq \sum_{S \in \mathcal{D}_L} \|Z_n(S)\|_2^2 \leq \sum_{S \in \mathcal{D}_L} \mathbb{E} \left(\mathbf{1}_{(U_0, V_0) \in S} \left(1 + 2 \sum_{k=1}^n b(k) \right) \right) \\ &\leq \mathbb{E} \left(\sum_{S \in \mathcal{D}_L} \mathbf{1}_{(U_0, V_0) \in S} \left(1 + 2 \sum_{k=1}^n b(k) \right) \right). \end{aligned}$$

Using the definition of $\tilde{\beta}_1$ and the fact that $\sum_{S \in \mathcal{D}_L} \mathbf{1}_{(U_0, V_0) \in S} = 1$, we infer that

$$\|\Delta_L\|_2^2 \leq \left(1 + 2 \sum_{k=1}^n \tilde{\beta}_1(k) \right). \quad (2.46)$$

Combining (2.45) and (2.46), we obtain that

$$\|\Delta\|_2 = \sum_{L \in [0, N]^2} \|\Delta_L\|_2 \leq N^2 \left(1 + 2 \sum_{k=1}^n \tilde{\beta}_1(k) \right)^{1/2}. \quad (2.47)$$

We shall now give an upper bound for the term $\mathbb{E}(R_N^2)$. Using the result of Rio [64] (Chapter 7, page 123),

$$R_N \leq 2\sqrt{n}2^{-N} + \sum_{j=1}^2 \sup_{z_j \in [0,1]} \sqrt{n} (F_{n,j}(\Pi_N(z_j) + 2^{-N}) - F_{n,j}(\Pi_N(z_j)))$$

where $F_{n,1}$ (resp. $F_{n,2}$) is the empirical distribution function of the variables U_1, \dots, U_n (resp. V_1, \dots, V_n). Let

$$\Delta_{N,j} = \sup_{z_j \in [0,1]} \sqrt{n} (F_{n,j}(\Pi_N(z_j) + 2^{-N}) - F_{n,j}(\Pi_N(z_j)))$$

In order to give an upper bound for $\Delta_{N,j}$, we use exactly the same strategy as for Δ_L (with dyadic intervals instead of dyadic boxes), which gives

$$\|\Delta_{N,j}\|_2^2 \leq 1 + 2 \sum_{k=1}^n \tilde{\beta}_1(k).$$

Hence

$$\|R_N\|_2 \leq 2\sqrt{n}2^{-N} + \left(1 + 2 \sum_{k=1}^n \tilde{\beta}_1(k) \right)^{1/2}. \quad (2.48)$$

Combining the inequalities (2.47) and (2.48), we obtain that there exists a positive constant K such that

$$\left\| \sup_{z \in [0,1]^2} |\mu_n(z)| \right\|_2^2 \leq Kn2^{-2N} + KN^4 \left(1 + 2 \sum_{k=1}^n \tilde{\beta}_1(k) \right).$$

Taking $N = \lceil (2 \log 2)^{-1} \log n \rceil$, we infer that there exists a positive constant C such that

$$\left\| \sup_{z \in [0,1]^2} |\mu_n(z)| \right\|_2^2 \leq C(\log(n))^4 \sum_{k=0}^n \tilde{\beta}_1(k),$$

which is exactly (2.18). The proof of Lemma 2.2 is complete.

Chapter 3

Adaptive density estimation for α mixing sequences

Contents

3.1	Introduction	46
3.1.1	Motivations	46
3.1.2	α -mixing sequences	47
3.1.3	Fourier density estimator and the sinc kernel	48
3.2	Main results	48
3.2.1	An upper bound for the quadratic risk	49
3.2.2	Adaptive procedure	50
3.2.3	Comments and rates of convergence	51
3.3	Simulations	54
3.3.1	Methodology	54
3.3.2	Numerical results	56
3.3.3	Robustness to another dependent model	60
3.4	Proofs	61
3.4.1	Preliminary	61
3.4.2	Proof of theorem 3.1	63

Abstract We extend a novel procedure to estimate adaptively a probability density on the real line to the case of strictly stationary and α -mixing sequences. First, we build an estimator based on Fourier methods and determine an upper bound for its L^2 -risk. Then, we propose an adaptive procedure which is proved to be minimax optimal (up to a power of a logarithmic term) and is easy to implement. Finally, we illustrate the performances of our method through different sets of simulation. We also illustrate numerically, that it is robust to other dependency conditions.

3.1 Introduction

3.1.1 Motivations

The general purpose of this paper is to provide an adaptive density estimator for stationary α -mixing sequences of real-valued random variables. The method proposed below is inspired by the adaptive procedure introduced recently by Duval and Kappus [38]. These authors introduce a new procedure to select the optimal cutoff parameter for density estimators based on a Fourier method. This procedure is numerically easy to implement and the concentration tool used to establish the oracle inequality is a basic Hoeffding inequality for partial sum of bounded random variables, which makes it eligible for an extension to the case of sequences of dependent variables. The procedure described in [38] is optimal up to the power of a logarithmic factor.

The problem of adaptive density estimation was mainly studied in a context of independent and identically distributed (iid) random variables, but also for dependent data since the late 90's. Tribouley and Viennet (1998) [70] propose L^p adaptive estimators for the common density f of a stationary β -mixing sequence (see Volkonski and Rozanov(1959)[73] for the definition of β -mixing coefficients) using wavelets, that are minimax optimal (in the sense that the procedure leads to the same minimax rate as in the iid case). Considering discrete or continuous time stationary processes, Comte and Merlevède (2002) [17] study the adaptive density estimation of the common density f of a α -mixing or β -mixing processes (see the definition below for the α -mixing coefficients in the sense of Rosenblatt (1956) [65]). They use a penalized least square method to compute the adaptive estimators. Comte et al. (2008) [16] propose a model selection procedure for projection estimators on the Shannon basis (i.e. the orthonormal basis generated by the sinus-cardinal function), to estimate f on the whole real line for stationary β -mixing sequences. Lerasle [56], [57] proposes an adaptive estimator based on model selection for the common density of a stationary process which is either β -mixing or τ -mixing (see [27] for the definition of τ -mixing coefficients). Gannaz and Wintenberger (2010) [44] consider other type of dependence; they estimate the common marginal density f by a wavelet type estimator, under some conditions on the λ or $\tilde{\varphi}$ coefficients (see [22] for their definition). Asin et Johannes (2017) [7] give a data-driven non-parametric estimation procedure for a density and a regression function. They use an orthogonal series estimator for β -mixing sequences. Bertin and Klutchnikoff (2017) [11] estimate the common density f of a weakly dependent process. They give upper bounds for the pointwise risk, and propose a data-driven procedure using a Goldenshluger and Lepskii method for α -mixing and λ -dependent sequences. We shall say more on these papers in subsection 3.2.3. The reference closest to our context is the paper by Comte and Merlevède [17].

We see that, to obtain adaptive density estimators in a dependent context, many papers deal with β -mixing sequences. The reason is a technical one: for β -mixing sequences, one can use Berbee's coupling lemma [10], which enables to go back to the iid case and to apply the powerful concentration inequalities of Talagrand [69]. This approach is much trickier in the α -mixing case, because the coupling tools for α -mixing sequences are much less efficient (see for instance [64], Chapter 5). Following the approach of Duval and Kappus [38], the only inequality we need is a deviation inequality for partial sums of bounded random variables. We shall use a Fuk-Nagev type inequality, as described in [64] (we recall this inequality in subsection 3.4.1).

Let us now briefly recall why the α -mixing case is so interesting. It is well known that the notion of α -mixing is the weakest type of mixing among the usual mixing conditions (see e.g. Bradley [14]). In particular it contains two large classes of examples: irreducible, aperiodic

and positively recurrent Markov chains (for which β -mixing is equivalent to α -mixing, see [64], Chapter 9), as well as functions of Gaussian processes which are naturally ρ -mixing (see [14] for the definition of ρ -mixing). Of course, weaker notions of dependence can also be defined: we refer to [27], [22] for an overview.

The article is organized as follows. In the remaining of this section we define the α -mixing coefficients of Rosenblatt (1956) [65], we recall Ibragimov's covariance inequality [52], and we define our density estimator \widehat{f}_m . In section 3.2, we give the main results, we describe the idea of our adaptive procedure, we state the oracle inequality for the adaptive estimator, and we determine its rate of convergence over Sobolev balls. Section 3.3 is dedicated to simulations. We illustrate our theoretical results through different sets of simulation, considering different probability densities under different rates of mixing. We also consider another dependent model which is not α -mixing, to illustrate the robustness of our method. Finally, section 3.4 gives the proofs of the main results.

3.1.2 α -mixing sequences

Let X_1, \dots, X_n be a stationary sequence of dependent random variables with density f with respect to the Lebesgue measure. We aim at estimating f on the whole real line; we will evaluate the performance of our estimator using the quadratic loss on \mathbb{R} .

To measure the dependence between the variables X_i , we consider the usual α -mixing coefficients introduced by Rosenblatt (1956) [65]. Let us recall the definition of these coefficients.

Definition 3.1. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and $(X_i)_{i \in \mathbb{Z}}$ be a sequence of real-valued random variables. Set $\mathcal{F}_k = \sigma(X_i : i \leq k)$ and $\mathcal{G}_k = \sigma(X_i : i \geq k)$. The strong mixing coefficients $(\alpha(n))_{n \geq 0}$ of $(X_i)_{i \in \mathbb{Z}}$ of Rosenblatt (1956) [65], are defined by*

$$\alpha(0) = \frac{1}{2} \quad \text{and} \quad \alpha(n) = \sup_{k \in \mathbb{Z}} \alpha(\mathcal{F}_k, \mathcal{G}_{k+n}) \quad \text{for any } n > 0,$$

where, for two σ -algebra \mathcal{F}, \mathcal{G} ,

$$\alpha(\mathcal{F}, \mathcal{G}) = 2 \sup_{A \in \mathcal{F}, B \in \mathcal{G}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

Note that we use here the convention of the book by Rio [64], so that the definition of $\alpha(\mathcal{F}, \mathcal{G})$ differs from that of Rosenblatt [65] from a factor 2.

Among the mathematical tools that will be used in the rest of this article, we introduce the following key inequality.

Ibragimov's covariance inequality. Let X, Y be two real-valued random variables that are almost surely bounded, and denote by $\|\cdot\|_\infty$ the essential supremum norm. Then

$$|\text{Cov}(X, Y)| \leq 2\alpha(\sigma(X), \sigma(Y))\|X\|_\infty\|Y\|_\infty.$$

This inequality is due to Ibragimov (1962) [52]. Now, for complex-valued random variables, whose modulus are almost surely bounded, one can easily prove that

$$|\text{Cov}(X, Y)| \leq 4\alpha(\sigma(X), \sigma(Y))\|X\|_\infty\|Y\|_\infty. \tag{3.1}$$

where $\|X\|_\infty$ denotes the essential supremum norm of $|X|$. See for instance [64], Chapter 1, Exercise 7.

3.1.3 Fourier density estimator and the sinc kernel

Let X_1, \dots, X_n be a stationary sequence of strongly mixing random variables with density f with respect to the Lebesgue measure. Let also X be a random variable with density f .

The characteristic function of X is denoted by $\varphi_X(u) = \mathbb{E}[e^{iuX}]$ and the Fourier transform of f , for $f \in \mathbb{L}^1(\mathbb{R})$, by $\mathcal{F}f(u) = \int e^{iux} f(x) dx$. We write $\|\cdot\|$ for the \mathbb{L}^2 -norm on the real line, that is $\|f\|^2 = \int |f(x)|^2 dx$. Finally, for $m > 0$, let f_m be such that its Fourier transform coincides with the Fourier transform of f on $[-m, m]$ and is zero outside this interval, i.e. such that $\mathcal{F}f_m = (\mathcal{F}f)\mathbb{1}_{[-m, m]}$.

To estimate the density f from the random variables X_1, \dots, X_n , we consider the following Fourier estimator \hat{f}_m (here m is implicitly tending to ∞ as $n \rightarrow \infty$). First, we denote by

$$\hat{\varphi}_{X,n}(u) = \frac{1}{n} \sum_{j=1}^n e^{iuX_j}, u \in \mathbb{R}$$

the estimated characteristic function of X . Applying the Fourier inversion formula and truncating the spectral domain (because $\hat{\varphi}_{X,n}(u)$ does not belong to $\mathbb{L}^1(\mathbb{R})$), the estimator of f is then

$$\hat{f}_m(x) = \frac{1}{2\pi} \int_{-m}^m e^{-iux} \hat{\varphi}_{X,n}(u) du, \quad x \in \mathbb{R}. \quad (3.2)$$

One can easily see that $\hat{f}_m(x)$ is in fact a kernel estimator, where the kernel is the sinus-cardinal (sinc) function. More precisely, letting

$$K(x) = \text{sinc}(x) = \frac{\sin(x)}{\pi x},$$

and choosing $m = 1/h$, we have

$$\hat{f}_m(x) = \frac{1}{nh} \sum_{k=1}^n K\left(\frac{x - X_k}{h}\right).$$

This simply follows from the fact that the Fourier transform of K is exactly the indicator function of the interval $[-1, 1]$. We refer to Davis (1977) [19] for the properties of this estimator in the iid case. The sinc kernel is very useful for deconvolution problems (see for instance Comte et al. (2008) [16] in a dependent context). The main drawback of this estimator is that it is oscillating even for relatively large values of x . In subsection 3.2.2, we will propose a slight modification of \hat{f}_m (denoted by \tilde{f}_m , see (3.8)) to deal with this issue, and to make our adaptive procedure feasible.

In the next section, we give a non asymptotic \mathbb{L}^2 risk bound for \hat{f}_m , and we propose an adaptive procedure to choose the cutoff parameter m for the modified estimator \tilde{f}_m .

3.2 Main results

In this section we present our main results. First, we compute an upper bound for the quadratic risk of \hat{f}_m , and we determine its rate of convergence knowing that f belongs to a given Sobolev ball. Next we explain our adaptive procedure and we present the resulting oracle inequality.

3.2.1 An upper bound for the quadratic risk

Proposition 3.1. *Let \widehat{f}_m be the estimator defined in (3.2). The following upper bound holds*

$$\mathbb{E} \left[\|\widehat{f}_m - f\|_2^2 \right] \leq \frac{1}{2\pi} \int_{[-m, m]^c} |\varphi_X(u)|^2 du + \frac{m}{n\pi} \left(1 + 8 \sum_{k=1}^{n-1} \alpha(k) \right). \quad (3.3)$$

Proof. We first note that $\mathbb{E}(\widehat{f}_m) = f_m$. Consequently, the quadratic risk admits the following decomposition

$$\mathbb{E} \left[\|\widehat{f}_m - f\|_2^2 \right] = \|f_m - f\|_2^2 + \mathbb{E} \left[\|\widehat{f}_m - f_m\|_2^2 \right]$$

where the first term at the right hand side is a bias term and the second is a variance term. First, by the Plancherel equality, we obtain

$$\|f_m - f\|_2^2 = \frac{1}{2\pi} \|\mathcal{F}f_m - \mathcal{F}f\|_2^2 = \frac{1}{2\pi} \|\mathcal{F}f \mathbf{1}_{[-m, m]^c}\|_2^2 = \frac{1}{2\pi} \int_{[-m, m]^c} |\varphi_X(u)|^2 du.$$

Now, we control the variance term.

$$\mathbb{E} \left[\|\widehat{f}_m - f_m\|_2^2 \right] = \frac{1}{2\pi} \int_{-m}^m \mathbb{E} \left[|\widehat{\varphi}_{X, n}(u) - \varphi_X(u)|^2 \right] du.$$

Since

$$\text{Var}(e^{iuX_1}) = \mathbb{E}(e^{iuX_1} e^{-iuX_1}) - \mathbb{E}(e^{iuX_1}) \mathbb{E}(e^{-iuX_1}) = 1 - |\varphi_X(u)|^2,$$

and using the stationarity of the sequence (X_i) , we get

$$\mathbb{E} \left[|\widehat{\varphi}_{X, n}(u) - \varphi_X(u)|^2 \right] = \text{Var} \left(\frac{1}{n} \sum_{j=1}^n e^{iuX_j} \right) \leq \frac{1 - |\varphi_X(u)|^2}{n} + \frac{2}{n} \sum_{k=2}^n |\text{Cov}(e^{iuX_1}, e^{iuX_k})|. \quad (3.4)$$

We shall now control the term of covariance in (3.4). Using the inequality (3.1) and the fact that $|e^{iuX_k}| < 1$, we get

$$\sum_{k=2}^n |\text{Cov}(e^{iuX_1}, e^{iuX_k})| \leq 4 \sum_{k=2}^n \alpha(k-1).$$

It follows that,

$$\mathbb{E} \left[|\widehat{\varphi}_{X, n}(u) - \varphi_{X, n}(u)|^2 \right] \leq \frac{1}{n} + \frac{8}{n} \sum_{k=1}^{n-1} \alpha(k), \quad (3.5)$$

and

$$\mathbb{E} \left[\|\widehat{f}_m - f_m\|_2^2 \right] \leq \frac{m}{\pi n} + \frac{8m}{n\pi} \sum_{k=1}^{n-1} \alpha(k).$$

This completes the proof of Proposition 3.1 □

To provide explicit rates of convergence, we consider Sobolev classes of regularity in order to control the bias term. Suppose that f is in a Sobolev ball of regularity s , $s \geq 0$, defined as follows

$$\mathcal{S}(s, L) := \left\{ f \in \mathbf{F}; \int_{\mathbb{R}} (1 + |u|)^{2s} |\mathcal{F}f(u)|^2 du \leq L \right\} \quad (3.6)$$

where \mathbf{F} is the set of densities with respect to the Lebesgue measure. From Proposition 3.1, for $f \in \mathcal{S}(s, L)$ it holds that

$$\mathbb{E} \left[\|\widehat{f}_m - f\|^2 \right] \leq C m^{-2s} + \frac{m}{n\pi} \left(1 + 8 \sum_{k=1}^{n-1} \alpha(k) \right)$$

where C is a positive constant independent of m and n . Consider additionally the following assumption,

$$\sum_{k=1}^{\infty} \alpha(k) \leq A_{\alpha} < \infty, \quad (\mathcal{A}(\alpha))$$

for some positive constant A_{α} . Under $(\mathcal{A}(\alpha))$ and selecting $m = m^* \asymp n^{\frac{1}{2s+1}}$, which performs the tradeoff between the bias term and the variance term, we find that the rate of convergence of the estimator \widehat{f}_{m^*} is of order $n^{\frac{-2s}{2s+1}}$, which corresponds to the optimal rate of convergence for iid sequences.

However, m^* is not feasible in practice since it depends on the unknown regularity s of the density f . In the following Section, we propose a data driven procedure to select a cutoff \widehat{m}_n adapted from the recent paper [38], and such that $\widehat{f}_{\widehat{m}_n}$ nearly achieves the optimal rate of convergence.

Remark 3.1. *If the condition $(\mathcal{A}(\alpha))$ holds one says that the sequence (X_i) is short-range dependent. In that case the estimator \widehat{f}_m behaves as in the iid case. If the condition $(\mathcal{A}(\alpha))$ is not satisfied, then the sequence (X_i) can be long-range dependent (meaning that the partial sum of bounded functions of the X_i 's may not satisfies the central limit theorem). Note, however, that if $\alpha(k) \rightarrow 0$ and if f belongs to $\mathcal{S}(s, L)$ for some $s > 0$, there is always a sequence $m_n \rightarrow \infty$ such that \widehat{f}_{m_n} is a consistent estimator of f . For instance, if $\alpha(k) = O(k^{-a})$ for $a \in (0, 1)$, using that $\sum_{1 \leq k \leq n-1} \alpha(k) = O(n^{1-a})$, one can take $m_n \asymp n^{\frac{a}{2s+1}}$, which gives the rate of convergence $n^{\frac{-2sa}{2s+1}}$.*

3.2.2 Adaptive procedure

The idea behind the adaptive procedure is that the optimal cutoff m^* that minimises the upper bound of Proposition 3.1 can be obtained by differentiating in m this upper bound and finding its zeroes. More precisely m^* satisfies the following equation: denote by $C_{\alpha} = 1 + 8 \sum_{k=1}^{n-1} \alpha(k)$, then m^* is such that

$$|\varphi_X(m^*)|^2 = \frac{C_{\alpha}}{n} \iff |\varphi_X(m^*)| = \sqrt{C_{\alpha}} n^{-1/2}.$$

The quantity C_{α} is unknown, but under $(\mathcal{A}(\alpha))$, this quantity is bounded. Since we are only interested in estimating the order of m^* and not in constants, we select the cutoff as described below.

The advantage of the latter equation is that it has an empirical counterpart, therefore we estimate m^* by \widehat{m}_n which is defined as follow

$$\widehat{m}_n = \max\{m > 0 : |\widehat{\varphi}_{X,n}(m)| = \kappa_n n^{-1/2}\} \wedge n. \quad (3.7)$$

where $\kappa_n = 1 + \kappa(\log n)^\gamma$, κ and γ are positive constants (a lower bound on their values can be derived from the following oracle inequality). Since the equation $|\widehat{\varphi}_{X,n}(m)| = \kappa_n/\sqrt{n}$ may have several solutions, we consider its largest solution to get a unique definition of \widehat{m}_n . Similarly, we do not consider cutoffs larger than n , since for $m \geq n$ the variance term is bounded away from 0 and the resulting estimator is not consistent.

For the same reason as in Duval and Kappus (2019) [38], we need to re-define our estimator. For u such that $|\varphi_X(u)| \leq Cn^{-1/2}$, namely below the variations of the empirical version of the characteristic function, it is not possible to distinguish a small signal $|\varphi_X(u)| > 0$, from a noise where $|\varphi_X(u)| = 0$. Therefore, the estimator can be reduced to zero. Then, we change the definition of our estimator and we introduce a threshold. We consider the estimator \widetilde{f}_m defined as in (3.2) but where $\widehat{\varphi}_{X,n}$ is replaced by

$$\widetilde{\varphi}_{X,n} := \widehat{\varphi}_{X,n} \mathbb{1}_{|\widehat{\varphi}_{X,n}| \geq \kappa_n n^{-1/2}}.$$

We obtain the estimator

$$\widetilde{f}_m(x) = \frac{1}{2\pi} \int_{-m}^m e^{-iux} \widetilde{\varphi}_{X,n}(u) du, \quad x \in \mathbb{R}. \quad (3.8)$$

We can show the following oracle inequality for the resulting estimator $\widetilde{f}_{\widehat{m}_n}$. The proof of this inequality is given in section 3.4.

Theorem 3.1. *Let \widehat{m}_n be defined as in (3.7). Let $(\alpha(n))_{n \geq 0}$ be the strong mixing coefficient defined as in 3.1 and let κ and γ be positive constants. Then, for $C_\alpha = 1 + 8 \sum_{k=1}^{n-1} \alpha(k)$, it holds that*

$$\begin{aligned} & \mathbb{E} \left[\|\widetilde{f}_{\widehat{m}_n} - f\|_2^2 \right] \\ & \leq \inf_{m \in [0, n]} \left\{ \frac{4 + C_\alpha}{2\pi} \int_{u \in [-m, m]^c} |\varphi_X(u)|^2 du + \frac{4m}{n\pi} (C_\alpha + (1 + 2\kappa(\log n)^\gamma)^2) \right\} \\ & \quad + \frac{(4 + 2^4)n}{\pi} \exp\left(-\frac{\kappa^2(\log n)^{2\gamma}}{64C_\alpha}\right) + \frac{2^6 + 16}{\kappa\pi} \frac{n^{3/2}}{(\log n)^\gamma} \alpha\left(\left\lceil \frac{4}{\kappa} \frac{\sqrt{n}C_\alpha}{(\log n)^\gamma} \right\rceil\right), \end{aligned}$$

where $[x]$ means the integer part of x .

3.2.3 Comments and rates of convergence

First, we propose a simplified writing of the latter upper bound under an additional decay assumption for the α -mixing sequence. If $\gamma > 1/2$ and if $\alpha(k) = O(k^{-a})$ with $a > 5$ and for a positive constant C , we get that

$$\begin{aligned} \mathbb{E} \left[\|\widetilde{f}_{\widehat{m}_n} - f\|_2^2 \right] & \leq \inf_{m \in [0, n]} \left\{ \frac{4 + C_\alpha}{2\pi} \int_{u \in [-m, m]^c} |\varphi_X(u)|^2 du + \frac{4m}{n\pi} (C_\alpha + (1 + 2\kappa(\log n)^\gamma)^2) \right\} \\ & \quad + \frac{C}{n}. \end{aligned}$$

This upper bound is classical, up to the power of the logarithmic factor $(\log n)^{2\gamma}$, when working with iid random variables. Indeed, as we observe below, when $a > 5$ we recover, up to the power of the logarithmic term, the optimal rate of convergence of a density estimation problem from independent and identically distributed random variables.

We remark that the upper bound in theorem 3.1 depends on two different constants that are chosen by the practitioner in (3.7): γ and κ . The choice of the parameters κ and γ depends on different conditions. As noted above, the logarithmic loss factor is an increasing function of γ . Therefore, it would be better to take the smallest possible value for γ such that $\gamma > \frac{1}{2}$. Then, for n and $\gamma > 1/2$ fixed, the parameter κ is a constant that needs to be calibrated. As explained in Duval and Kappus (2019) [38], subsection 2.3, the procedure seems robust to a wide range of possible values for κ , in the sense that different values of κ lead to similar performances. However, it should not be too large as $(1 + \kappa(\log n)^\gamma)n^{-1/2}$ is compared to the modulus of an empirical characteristic function whose value is bounded by 1. In the next section, we give an heuristic to choose κ . In the simulation section 3.3, we will use the values $\gamma = 2.05$ which is larger than $1/2$, and $\kappa = 0.05$.

To provide explicit rates of convergence, we consider Sobolev classes of regularity defined as in (3.6) in order to control the order of the bias term and we optimise the bound of theorem 3.1. Let $f \in \mathcal{S}(s, L)$, then the first bias term can be controlled as follows

$$\int_{|u|>m} |\varphi_X(u)|^2 du \leq Cm^{-2s}.$$

In the sequel, C denotes a generic constant whose value may change from line to line and does not depend on m or n . The second term which is a variance term is such that

$$\frac{4m}{n\pi} (C_\alpha + (1 + 2\kappa(\log n)^\gamma)^2) \leq C \frac{m}{n} (\log n)^{2\gamma}.$$

Since $\gamma > \frac{1}{2}$, the third exponential term is of order $o(\frac{1}{n})$ so it is a remainder term that will not intervene in the rate. Finally for the last term, where the mixing coefficients play a central role, we consider the decay assumption $\alpha(k) = O(k^{-a})$ with $a > 0$ introduced earlier. Under this assumption the last term can be controlled by

$$\frac{n^{3/2}}{(\log n)^\gamma} \alpha \left(\left[\frac{4}{\kappa} \frac{\sqrt{n}C_\alpha}{(\log n)^\gamma} \right] \right) \leq C(\log n)^{\gamma(a-1)} n^{\frac{3-a}{2}}.$$

To conclude, we obtain that for f in $\mathcal{S}(s, L)$

$$\|\tilde{f}_{\hat{m}_n} - f\|^2 \leq C \left(\inf_{m \in [0, n]} \left\{ m^{-2s} + \frac{m}{n\pi} (\log n)^{2\gamma} \right\} + (\log n)^{\gamma(a-1)} n^{\frac{3-a}{2}} + \frac{1}{n} \right).$$

Optimizing in m this bound we get that for every $f \in \mathcal{S}(s, L)$ it holds

$$\|\tilde{f}_{\hat{m}_n} - f\|^2 \leq C \left\{ (n \log(n)^{-2\gamma})^{-\frac{2s}{2s+1}} + (\log n)^{\gamma(a-1)} n^{\frac{3-a}{2}} + \frac{1}{n} \right\}.$$

Indeed, straightforward computations enable to derive that the order of the optimal cutoff m_{opt} is of the order of $(n \log(n)^{-2\gamma})^{\frac{1}{2s+1}}$. On another hand, a necessary condition for $(\log n)^{\gamma(a-1)} n^{\frac{3-a}{2}}$

to go to 0 is that $a > 3$. Below we distinguish two cases, delineated by the value of a , depending on whether this strategy attains the optimal rate of convergence of the independent case (up to a power of a logarithmic factor) or not.

- We easily see well that if $a > 5$, $(\log n)^{\gamma(a-1)}n^{\frac{3-a}{2}}$ is negligible in front of n^{-1} so we find that our procedure attains the minimax rate $n^{-\frac{2s}{2s+1}}$ up to a power of a logarithmic term.
- Now in the case where $a \in]3, 5]$, we see that (neglecting the contribution of the logarithmic terms) a sufficient condition to attain the classical rate is to ensure that

$$n^{-\frac{2s}{2s+1}} < n^{\frac{3-a}{2}} \quad \Leftrightarrow \quad a > 3 + \frac{4s}{2s+1}.$$

This constraint cannot be verified in practice, since s is unknown. However, if the function f is poorly regular, informally letting s tend to 0, we recover the constraint $a > 3$. Without any assumption on s , all one can say is that if $a \in]3, 5]$, we have an estimator that converges at the rate

$$(n \log(n)^{-2\gamma})^{-\frac{2s}{2s+1}} \vee (\log n)^{\gamma(a-1)}n^{(3-a)/2},$$

but this rate does not coincide with the minimax rate of the iid case.

To conclude this section, let us take a look at some articles that consider the problem of adaptive density estimation in a dependent context. We are going to treat separately the articles which deal with β -mixing sequences, from those which deal with α -mixing sequences, because the case of β -mixing is both simpler from a technical point of view, and more restrictive in terms of examples. Note that in some articles (such as [56], [57], [11]) other notions of dependencies are also considered.

In the β -mixing framework, Tribouley and Viennet (1998) [70] proposed a wavelet method to estimate the density on a compact support, under the condition $\beta(k) = O(k^{-a})$ for $a > 4$. The resulting adaptive estimator reaches the minimax rates of convergence over Besov classes. Comte and Merlevède (2002) [17] proposed a general model selection procedure (valid for a large class of models) to estimate the density on a compact support, under the condition $\beta(k) = O(k^{-a})$ for $a > 4$. The resulting adaptive estimator reaches the minimax rates of convergence over Besov classes. Comte et al (2008) [16] proposed a model selection procedure for projection estimators on the Shannon basis (i.e. the orthonormal basis generated by the sinus-cardinal function), to estimate the density on the whole real line. Here again, the results are valid under the condition that $\beta(k) = O(k^{-a})$ for some $a > 4$. In the same context, Lerasle [56], [57] gave a general model selection result (valid for a large class of models) under the condition $\beta(k) = O(k^{-a})$ for an $a > 3$. Note that in [57], the delicate question of the “data-driven” penalty is discussed. Asin and Johannes (2018) [7] proposed a model selection procedure to estimate the density, when the regularity of f is given by the decrease of the coefficients of the decomposition of f on a fixed orthonormal basis. Here again, the results are valid if $\beta(k) = O(k^{-a})$ for some $a > 4$ (see the comments on condition (4.6), after their Corollary 4.7).

In the α -mixing framework, Comte and Merlevède (2002) [17] proposed a general model selection procedure (valid for a large class of models) to estimate the density on a compact support, under the condition that $\alpha(k)$ decreases geometrically (see their theorem 3.2). The resulting adaptive estimator reaches the minimal rates of convergence over Besov classes, up to a power of a logarithmic term. In the same paper, the authors consider the case where

$\alpha(k) = O(k^{-a})$ (see their Proposition 3.2). Their result are valid provided that $a > 6$, and under an additional regularity assumption on the joint densities g_{X_k, X_ℓ} , $k \neq \ell$. All the previously cited articles give results for the integrated quadratic risk (integrated \mathbb{L}_p risk are also considered in [70]), like the one we are considering here. Let us also mention the article by Bertin and Klutchnikoff (2017) [11] which proposes an automatic bandwidth selection for a kernel estimator, and for the pointwise risk. Their result applies to a large class of dependent sequences, and in particular to α -mixing sequences for which the mixing coefficient decreases geometrically.

We can see that, whether for β or α mixing sequences, the constraint $a > 3$ on the mixing speed is always required for the adaptive estimation. This corresponds well to the minimum constraint that we found, although there is no heuristic explanation for this, since when the class of regularity is known, non adaptive estimators attain the minimax iid rate as soon as $\sum \alpha(k) < \infty$.

The article closest to ours is that of Comte and Merlevède (2002) [17], although the context and the method of estimation are different. The mixing speed we get to approach the minimax rate is a bit better than theirs ($a > 5$ instead of $a > 6$), but we still have a loss in a power of $\log n$. Moreover, we do not need any condition on the joint densities g_{X_k, X_ℓ} , which is quite satisfactory.

3.3 Simulations

3.3.1 Methodology

Examples considered and choice of the parameters. In this section, we illustrate our procedure with some simulations. We consider different distributions as Gamma $\Gamma(3, \frac{1}{2})$, Gaussian $\mathcal{N}(4, 1)$, the mixture $0.7\mathcal{N}(4, 1) + 0.3\Gamma(3, \frac{1}{2})$ and the mixture $0.7\mathcal{N}(3, 4) + 0.3\mathcal{N}(13, 2)$.

To show the performance of our method, we proceed as follows. First, we will present a cluster of estimators in different contexts with 50 estimators and the estimated density function (see Figures 3.1, 3.2, 3.3 and 3.4). We also compute the estimated \mathbb{L}^2 risks, noted by R , and its estimated standard deviation via a classical Monte-Carlo procedure, by averaging over $N = 500$ iterations. We estimate the average value $\overline{\widehat{m}_n}$ of \widehat{m}_n over 500 simulations, and we compute the associated estimated standard deviation. The results are displayed on Tables 3.1, 3.2, 3.3 and 3.4.

We will study the behaviour of the estimator in different cases. We will use different values of n , such as $n = 500, 2000$ and 5000 , and various values (noted by a) for the arithmetic decay of the mixing coefficients. From theorem 3.1, the upper bound holds for a larger than 5, moreover the largest a is, the closest we are to the independent case. In practice, we will test $a \leq 5$ to see if it can lead to a relevant estimate, and try to see if the procedure is robust when a gets small.

The procedure described above depends on parameters to be chosen: κ and γ . We know that using this procedure we loose a power of $\log n$, this power depending on the parameter γ (cf. theorem 3.1). On another hand, as explained in the preceding subsection, we need to take $\gamma > 1/2$ to simplify the expression of the exponential term in theorem 3.1 (otherwise we would have to choose κ as a function of the unknown constant C_α). We tried different values of γ for different densities and different mixing rates, and it seems that a value close to 2 is a reasonable choice. For these simulations, we choose $\gamma = 2.05$.

For the choice of κ , we propose the following heuristic: in the definition of \widehat{m}_n (see (3.7)),

the quantities κ , n , and γ determine a value below which the estimated characteristic function φ is set to 0. Since $|\varphi| \leq 1$ it is reasonable, for a fixed n and $\gamma = 2.05$ to ensure that $(1 + \kappa(\log(n)^\gamma))n^{-1/2} \leq 1/10$, for instance. Let $\kappa_{max}(n, \gamma)$ such that $(1 + \kappa(\log(n)^\gamma))n^{-1/2} = 1/10$, we compute κ_{max} ; for $n = 500$ we find that $\kappa_{max} = 0.029$, for $n = 2000$ we get that $\kappa_{max} = 0.054$ and finally for $n = 5000$, $\kappa_{max} = 0.075$. To conclude, taking $\kappa = 0.05$ satisfies the condition for various values of n .

Simulation of mixing sequences. For the simulations, we shall use the Markov chain introduced by Doukhan et al. (1994) [37]. The interest of this chain is first that it is easy to simulate (as many Markov chains for which one can exhibit explicitly the iterated random system), and secondly we can compute exactly its rate of mixing as a function of the parameters of the transition kernel. Note that, since the chain is irreducible, positively recurrent and aperiodic, it is also β -mixing in the sense of Volkonskii and Rozanov (1959) [73]. Recall that, for such Markov chains, β -mixing is equivalent to α -mixing.

Let us now describe this Markov chain in more details. Let $a > 1$ (as before a will calibrate the rate of mixing), and let μ and ν be the two probability measures on $[0, 1]$ with respective densities

$$f_\mu(x) = ax^{a-1}\mathbf{1}_{0 \leq x \leq 1} \quad \text{and} \quad f_\nu(x) = (a+1)x^a\mathbf{1}_{0 \leq x \leq 1}.$$

Let F_ν be the cumulative distribution function of ν , and let Y_1 be a random variable with law μ . Let $(\varepsilon_i)_{i \geq 1} = ((U_i, V_i))_{i \geq 2}$ be a sequence of iid random variables with uniform law over $[0, 1]^2$ and independent of Y_1 . For $k \geq 1$ define then

$$Y_{k+1} = Y_k \mathbf{1}_{U_{k+1} \geq Y_k} + F_\nu^{-1}(V_{k+1}) \mathbf{1}_{U_{k+1} < Y_k}.$$

It is proved in [37] that this chain is strictly stationary, with invariant distribution μ . Moreover, the β -mixing coefficients of this chain are exactly of order n^{-a} . Note also that Y_i^a is uniformly distributed over $[0, 1]$.

To simulate a sequence with distribution function F , we take

$$X_i = Q(Y_i^a)$$

where $Q = F^{-1}$ is the inverse of the cumulative distribution function.

The sequence $(X_i)_{i \geq 1}$ is also a stationary Markov chain (as an invertible function of a stationary Markov chain), and its β -mixing coefficients are such that: there exist two positive constants $B > A > 0$ such that, for any $n \geq 1$,

$$\frac{A}{n^a} \leq \beta(n) \leq \frac{B}{n^a}$$

(and the same is true for the coefficients $\alpha(n)$ for different constants A, B). By construction, the cumulative distribution function of X_i is exactly F . For instance to simulate a chain whose invariant distribution is $\mathcal{N}(4, 1)$, we take for Q the quantile function of the $\mathcal{N}(4, 1)$ -distribution.

This procedure will also be applied in the following subsection 3.3.3, where we start from a non irreducible Markov chain.

3.3.2 Numerical results

In this subsection, we present, for each density, a cluster of 50 estimators and the true density distribution. Also, we give the estimated quantities \widehat{m}_n , $\sqrt{\mathbb{V}(\widehat{m}_n)}$, R and $\sqrt{\mathbb{V}(R)}$ computed from 500 Monte-Carlo iterations.

Gamma Distribution

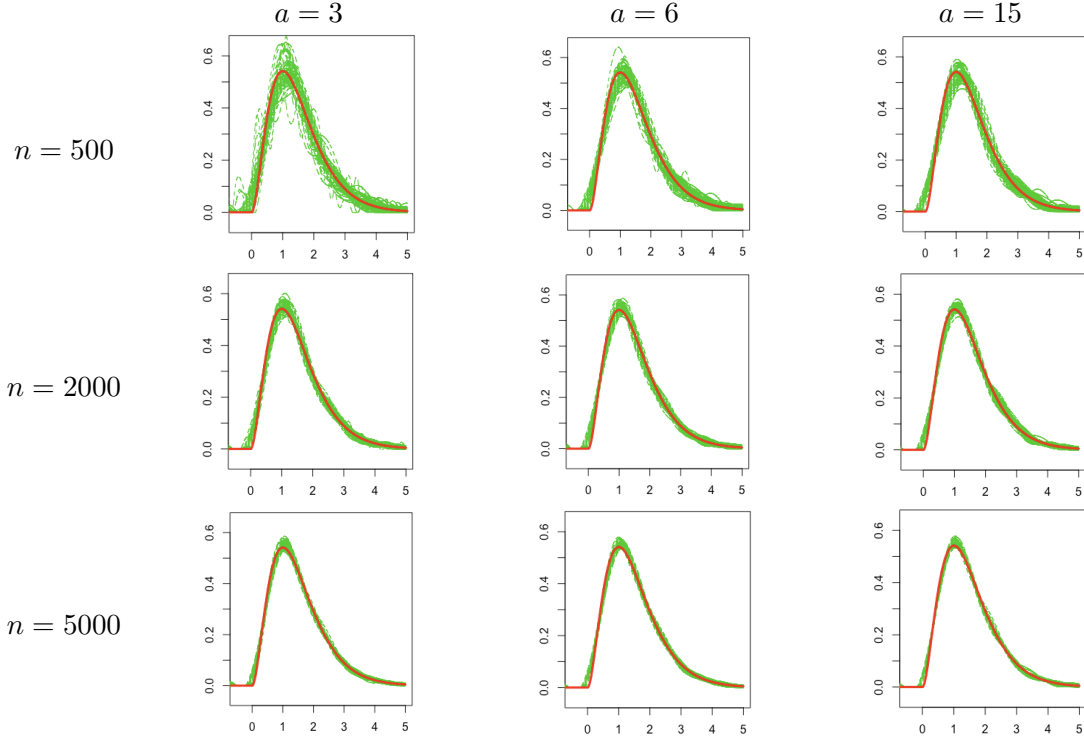


Figure 3.1: Estimation of $f \sim \Gamma(3, 1/2)$ (red) and the adaptive estimator (green) from $N = 50$ Monte Carlo iterations for $\kappa = 0.05$ and for different values of a and n .

		$a = 3$	$a = 6$	$a = 15$
$n = 500$	R	9.2×10^{-2}	5.43×10^{-2}	3.76×10^{-2}
	$\sqrt{\mathbb{V}(R)}$	8.15×10^{-2}	5.38×10^{-3}	3.29×10^{-2}
	\widehat{m}_n	11.78	8.69	6.38
	$\sqrt{\mathbb{V}(\widehat{m}_n)}$	5.97	5.60	4.29
$n = 2000$	R	1.59×10^{-2}	1.25×10^{-2}	1.14×10^{-2}
	$\sqrt{\mathbb{V}(R)}$	1.03×10^{-2}	4.22×10^{-3}	3.41×10^{-3}
	\widehat{m}_n	5.59	4.74	4.56
	$\sqrt{\mathbb{V}(\widehat{m}_n)}$	3.06	1.25	0.69
$n = 5000$	R	8.22×10^{-3}	7.51×10^{-3}	7.09×10^{-3}
	$\sqrt{\mathbb{V}(R)}$	3.32×10^{-3}	2.06×10^{-3}	1.65×10^{-3}
	\widehat{m}_n	4.98	4.93	4.89
	$\sqrt{\mathbb{V}(\widehat{m}_n)}$	0.85	0.46	0.4

Table 3.1: Computations via $N = 500$ Monte Carlo iterations of the \mathbb{L}^2 risks for the Gamma distribution $\Gamma(3, \frac{1}{2})$ and different values of $a = 3; 6; 15$.

Gaussian Distribution

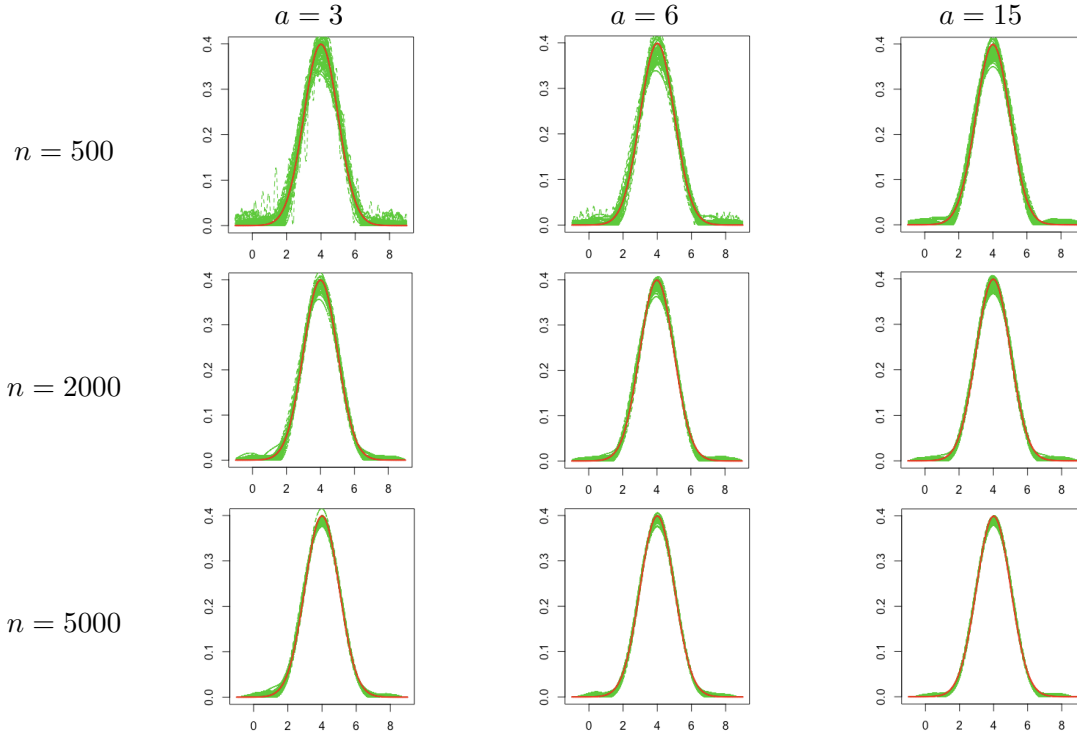


Figure 3.2: Estimation of $f \sim \mathcal{N}(4, 1)$ (red) and the adaptive estimator (green) from $N = 50$ Monte Carlo iterations for $\kappa = 0.05$ and for different values of a and n .

		$a = 3$	$a = 6$	$a = 15$
$n = 500$	R	5.68×10^{-2}	1.83×10^{-2}	1.02×10^{-2}
	$\sqrt{\mathbb{V}(R)}$	1.46×10^{-1}	2.29×10^{-2}	1.58×10^{-2}
	\widehat{m}_n	12.97	8.28	5.65
	$\sqrt{\mathbb{V}(\widehat{m}_n)}$	6.44	6.48	5.69
$n = 2000$	R	3.77×10^{-3}	1.85×10^{-3}	1.54×10^{-3}
	$\sqrt{\mathbb{V}(R)}$	5.79×10^{-3}	2.32×10^{-3}	7.98×10^{-4}
	\widehat{m}_n	4.10	2.56	2.35
	$\sqrt{\mathbb{V}(\widehat{m}_n)}$	4.41	1.69	0.58
$n = 5000$	R	1.03×10^{-3}	8.41×10^{-4}	7.57×10^{-4}
	$\sqrt{\mathbb{V}(R)}$	6.06×10^{-4}	4.04×10^{-4}	2.94×10^{-4}
	\widehat{m}_n	2.49	2.43	2.40
	$\sqrt{\mathbb{V}(\widehat{m}_n)}$	0.99	0.63	0.08

Table 3.2: Computations via $N = 500$ Monte Carlo iterations of the \mathbb{L}^2 risks for the Gaussian distribution $\mathcal{N}(4, 1)$ and different values of $a = 3; 6; 15$.

Mixture of Gamma and Gaussian distribution

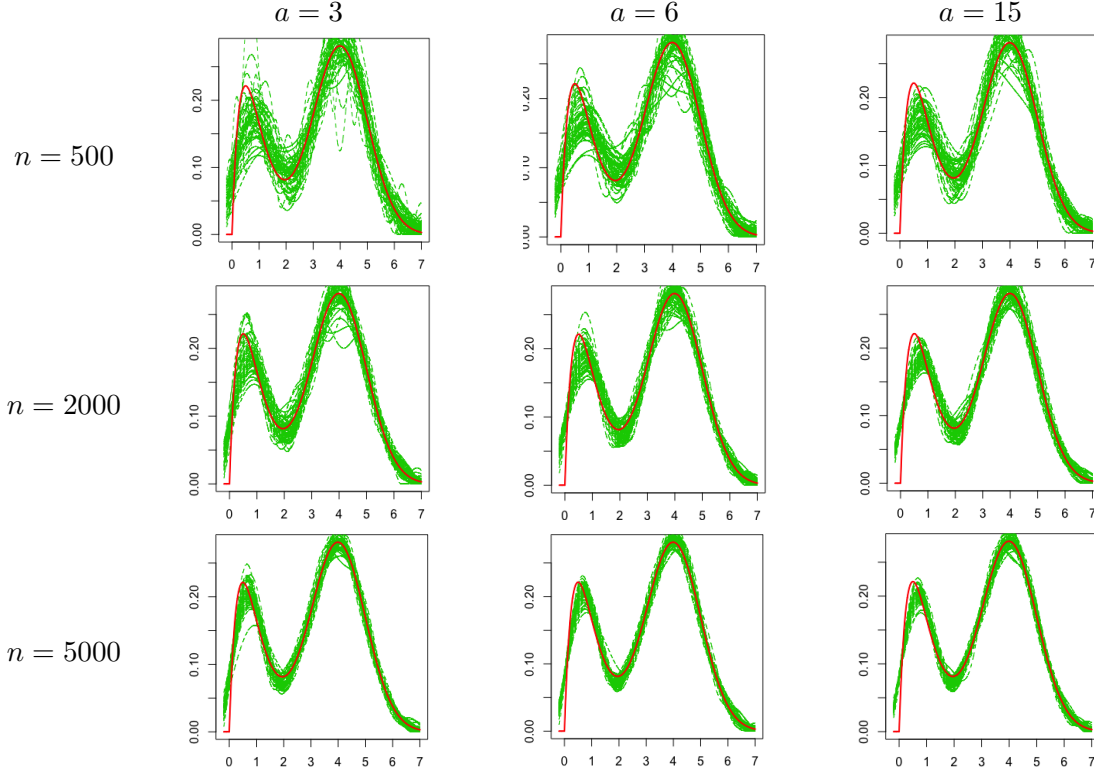


Figure 3.3: Estimation of $f \sim 0.7\mathcal{N}(4, 1) + 0.3\Gamma(2, \frac{1}{2})$ (red) and the adaptive estimator (green) from $N = 50$ Monte Carlo iterations for $\kappa = 0.05$ and for different values of a and n .

		$a = 3$	$a = 6$	$a = 15$
$n = 500$	R	10.53×10^{-2}	6.7×10^{-2}	4.6×10^{-2}
	$\sqrt{\mathbb{V}(R)}$	6.41×10^{-2}	4.08×10^{-2}	3.37×10^{-2}
	\widehat{m}_n	14.04	11.29	8.29
	$\sqrt{\mathbb{V}(\widehat{m}_n)}$	5.07	5.85	5.52
$n = 2000$	R	2.21×10^{-2}	1.61×10^{-2}	1.48×10^{-2}
	$\sqrt{\mathbb{V}(R)}$	1.56×10^{-2}	7.74×10^{-3}	5.32×10^{-3}
	\widehat{m}_n	5.85	4.55	4.20
	$\sqrt{\mathbb{V}(\widehat{m}_n)}$	3.73	2.01	1.35
$n = 5000$	R	11.18×10^{-3}	9.87×10^{-3}	9.52×10^{-3}
	$\sqrt{\mathbb{V}(R)}$	4.12×10^{-3}	2.45×10^{-3}	2.45×10^{-3}
	\widehat{m}_n	4.81	4.49	4.41
	$\sqrt{\mathbb{V}(\widehat{m}_n)}$	1.38	0.79	0.73

Table 3.3: Computations via $N = 500$ Monte Carlo iterations of the \mathbb{L}^2

risks for a mixture $0.7\mathcal{N}(4, 1) + 0.3\Gamma(2, \frac{1}{2})$ and different values of $a = 3; 6; 15$.

Mixture of Gaussian distribution

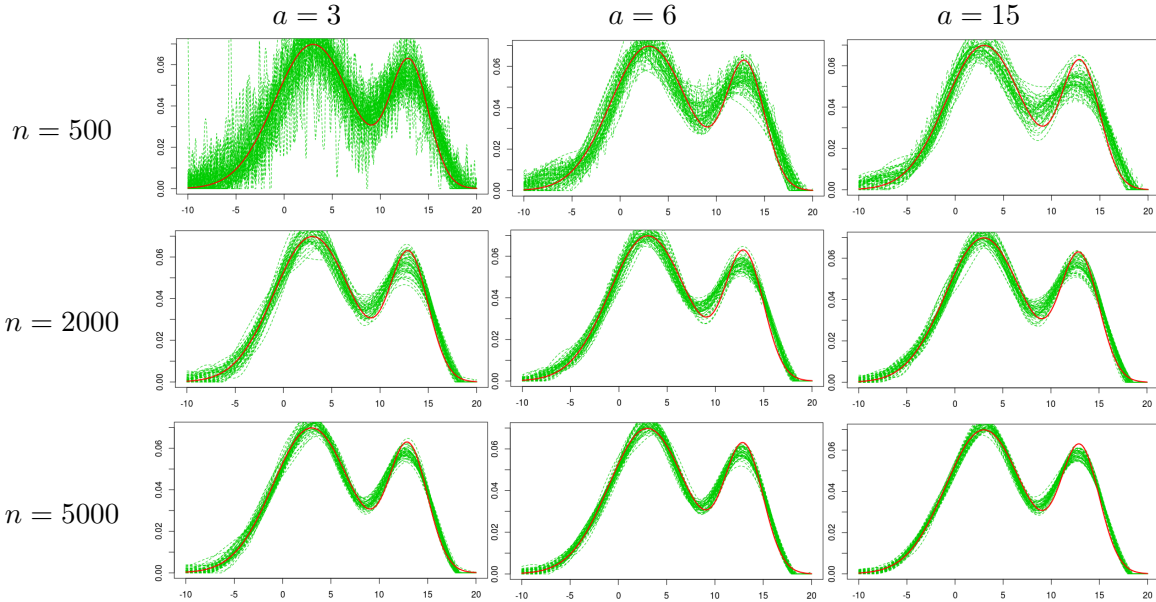


Figure 3.4: Estimation of $f \sim 0.7\mathcal{N}(3, 4) + 0.3\mathcal{N}(13, 2)$ (red) and the adaptive estimator (green) from $N = 50$ Monte Carlo iterations for $\kappa = 0.05$ and for different values of a and n .

		$a = 3$	$a = 6$	$a = 15$
$n = 500$	R	3.92×10^{-2}	1.88×10^{-2}	1.29×10^{-2}
	$\sqrt{\mathbb{V}(R)}$	3.89×10^{-2}	1.04×10^{-2}	8.62×10^{-3}
	\widehat{m}_n	18.44	16.19	12.96
	$\sqrt{\mathbb{V}(\widehat{m}_n)}$	1.86	3.75	6.15
$n = 2000$	R	3.65×10^{-3}	7.01×10^{-4}	4.19×10^{-4}
	$\sqrt{\mathbb{V}(R)}$	4.21×10^{-3}	1.28×10^{-3}	6.53×10^{-4}
	\widehat{m}_n	7.21	1.71	1.09
	$\sqrt{\mathbb{V}(\widehat{m}_n)}$	6.90	2.98	1.63
$n = 5000$	R	3.51×10^{-4}	2.01×10^{-4}	1.91×10^{-4}
	$\sqrt{\mathbb{V}(R)}$	6.41×10^{-4}	6.13×10^{-5}	5.23×10^{-5}
	\widehat{m}_n	1.43	0.92	0.91
	$\sqrt{\mathbb{V}(\widehat{m}_n)}$	2.56	6.25×10^{-2}	5.67×10^{-2}

Table 3.4: Computations via $N = 500$ Monte Carlo iterations of the \mathbb{L}^2 risks for a mixture $0.7\mathcal{N}(3, 4) + 0.3\mathcal{N}(13, 2)$ and different values of $a = 3; 6; 15$.

Comments on the figure and the numerical results On Figures 3.1, 3.2, 3.3 and 3.4, we observe the behaviour of 50 adaptive estimators around the true density distribution. On these four different cases, we see that the chosen values of κ and γ are meaningful.

For a fixed, as n increases, the adaptive estimators behave quite well. For n fixed, we see the difference according to the values of a that describes the speed of mixing. The bigger a , the better the graphs of the adaptive estimators.

For the case $a \leq 5$, we take $a = 3$ and we remark that, for a large n , the adaptive estimators give the same kind of performances, even though the results of theorem 3.1 do not apply. Taking the example of the mixture in table 3.4, comparing the numerical results for $n = 5000, a = 3$ and for $n = 2000, a = 15$, we find that they are of the same order. This comparison is also true for the other examples.

If we read the table diagonally, we remark that, as expected the larger a and n , the better the graphs of the adaptive estimators.

As expected, Tables 3.1, 3.2, 3.3 and 3.4 provide meaningful results. As we already explained, the larger a , the closer we are to the independent case.

The tables confirm that the \mathbb{L}^2 -risks decreases with n and a .

All these comments are true for the four different densities.

3.3.3 Robustness to another dependent model

To illustrate the robustness of our method, we consider another dependent model. We consider a stationary Markov chain (X_i) simulated according to the following auto-regressive mechanism:

$$\begin{cases} X_i = \frac{1}{2}(X_{i-1} + \varepsilon_i) & \text{for } i \geq 1, \text{ with } (\varepsilon_i)_{i \geq 1} \text{ iid, } \varepsilon_i \sim \mathcal{B}\left(\frac{1}{2}\right), \text{ and } (\varepsilon_i)_{i \geq 1} \text{ independent of } X_0 \\ X_0 \sim \mathcal{U}[0, 1] \end{cases}$$

Note that the X_i 's are uniformly distributed over $[0, 1]$. To generate another stationary Markov chain with invariant cumulative distribution function F , we proceed as in subsection 3.3.2 (using the transform $Q = F^{-1}$).

It is well known that the Markov chain $(X_i)_{i \geq 0}$ is not α -mixing in the sense of Rosenblatt [65] (see for instance [4]). However, one can prove that other dependence coefficients, such as the τ -dependence coefficients defined in [27], converge to 0 at exponential rate.

Since the chain is not α -mixing, theorem 3.1 does not apply. However, we will see that our procedure still gives interesting numerical results.

These results are presented through the following Figure and Table.

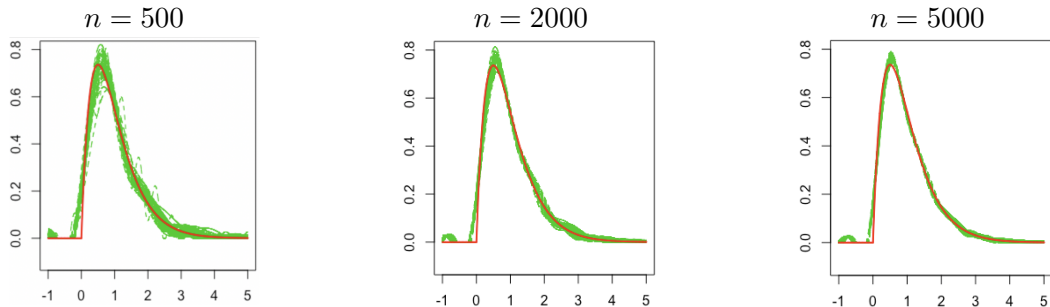


Figure 3.5: Estimation of autoregressive distribution with $f \sim \Gamma(2, \frac{1}{2})$ (red) and the adaptive estimator (green) from $N = 50$ Monte Carlo iterations and for different values of n and $\kappa = 0.05$.

	$n = 500$	$n = 2000$	$n = 5000$
R	17.26×10^{-2}	4.62×10^{-2}	3.05×10^{-2}
$\sqrt{\mathbb{V}(R)}$	17.02×10^{-2}	1.90×10^{-2}	4.97×10^{-3}
\hat{m}_n	14.36	7.85	8.44
$\sqrt{\mathbb{V}(\hat{m}_n)}$	12.97	2.45	0.93

Table 3.5: Computations via $N = 500$ Monte Carlo iterations of the \mathbb{L}^2 risks for a Markov chain generated from the non mixing auto-regressive model.

Comments on the numerical results On Figure 3.5 we observe the behaviour of 50 adaptive estimators of the $\Gamma(2, \frac{1}{2})$ -density distribution. We see that the chosen values of κ and γ are still meaningful.

As previously, we remark that the bigger n , the better the graphs of the adaptive estimators.

This graphical interpretation is confirmed by the numerical computations of the \mathbb{L}^2 risks given in table 3.5.

Hence, the procedure seems robust to other dependency structures. Note however that these simulations do not prove that we are close to the optimal choice of m . The \hat{m}_n chosen via the adaptive procedure can be a correct choice, ensuring the consistency of the estimator with a decent convergence rate, without even being of the right order.

3.4 Proofs

3.4.1 Preliminary

To prove theorem 3.1, we shall use a Fuk- Nagaev inequality stated in Rio [64]. More precisely, we need a version of this inequality which was proposed as an exercise in [64] (Exercise 1, Chapter 6). Let us recall and prove this inequality.

A Fuk-nagaev inequality :

Let $(X_i)_{i>0}$ be a sequence of real-valued random variables such that $\|X_i\|_\infty \leq 1/2$, and let $(\alpha(n))_{n \geq 0}$ be the sequence of strong mixing coefficients defined in (3.1). Let $S_k = \sum_{i=1}^k (X_i - \mathbb{E}(X_i))$ and

$$s_n^2 \geq \max \left(1, \sum_{i=1}^n \sum_{j=1}^n |\text{Cov}(X_i, X_j)| \right).$$

Then for any $\lambda \in [s_n, s_n^2]$,

$$\mathbb{P} \left(\sup_{k \in [1, n]} |S_k| \geq 4\lambda \right) \leq 4 \exp \left(\frac{-\lambda^2}{4s_n^2} \right) + 4n\lambda^{-1} \alpha \left(\left[\frac{s_n^2}{\lambda} \right] \right). \quad (3.9)$$

To prove (3.9), we start from theorem 6.2 Inequality (6.4) of [64], and we take $r = \frac{\lambda^2}{s_n^2}$ and $Q = 1$ (since $\|X_i - \mathbb{E}(X_i)\|_\infty \leq 1$). Then, with Rio's notations

$$H \left(\frac{\lambda}{r} \right) = \alpha \left(\left[\frac{\lambda}{r} \right] \right) = \alpha \left(\left[\frac{s_n^2}{\lambda} \right] \right).$$

We obtain that

$$\begin{aligned}
\mathbb{P}\left(\sup_{k \in [1, n]} |S_k| \geq 4\lambda\right) &\leq 4\left(1 + \frac{\lambda^2}{rs_n^2}\right)^{-\frac{r}{2}} + 4n\lambda^{-1} \int_0^{H(\lambda/r)} Q(u) du \\
&\leq 4\left(1 + \frac{\lambda^2}{\frac{\lambda^2}{s_n^2} s_n^2}\right)^{-\frac{\lambda^2}{2s_n^2}} + 4n\lambda^{-1} \int_0^{\alpha([s_n^2/\lambda])} 1 du \\
&\leq 4 \times 2^{-\frac{\lambda^2}{2s_n^2}} + 4n\lambda^{-1} \alpha\left(\left[\frac{s_n^2}{\lambda}\right]\right) \\
&\leq 4 \exp\left(-\frac{\lambda^2}{2s_n^2} \log 2\right) + 4n\lambda^{-1} \alpha\left(\left[\frac{s_n^2}{\lambda}\right]\right).
\end{aligned}$$

and (3.9) is proved.

Now, using Inequality (3.9) with

$$\sum_{k=1}^n \sum_{\ell=1}^n |\text{Cov}(e^{iuX_k}, e^{iuX_\ell})| \leq s_n^2 = n \left(1 + 8 \sum_{k=1}^n \alpha(k)\right) = nC_\alpha$$

and $\lambda = \kappa\sqrt{n}(\log n)^\gamma/4$, we get the following Lemma.

Lemma 3.1. *Let (X_i) be a stationary sequence of real-valued random variables, with strong mixing coefficients $(\alpha(n))_{n \geq 0}$. Then, for all $u \in \mathbb{R}$ it holds that*

$$\mathbb{P}\left(|\widehat{\varphi}_{X,n}(u) - \varphi_X(u)| > \kappa\sqrt{\frac{(\log n)^{2\gamma}}{n}}\right) \leq 4 \exp\left(-\frac{\kappa^2(\log n)^{2\gamma}}{64C_\alpha}\right) + \frac{16}{\kappa} \frac{\sqrt{n}}{(\log n)^\gamma} \alpha\left(\left[\frac{4}{\kappa} \frac{\sqrt{n}C_\alpha}{(\log n)^\gamma}\right]\right).$$

This result enables to determine an upper bound for the estimator \widetilde{f}_m similar to the one obtained in proposition 3.1. Following the lines of the proof of proposition 3.1, we only have to modify equation (3.5) as follows

$$\mathbb{E}[|\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2] \leq \mathbb{E}[|\widehat{\varphi}_{X,n}(u) - \varphi_X(u)|^2] + |\varphi_X(u)|^2 \mathbb{P}(|\widehat{\varphi}_{X,n}(u)| < \kappa_n/\sqrt{n}).$$

We decompose the second term on the set $A := \{u, |\varphi_X(u)| < (1 + 2\kappa(\log n)^\gamma)/\sqrt{n}\}$ and its complementary part as follows,

$$\begin{aligned}
&|\varphi_X(u)|^2 \mathbb{P}\left(|\widehat{\varphi}_{X,n}(u)| < \kappa_n/\sqrt{n}\right) \\
&= |\varphi_X(u)|^2 \mathbb{P}\left(|\widehat{\varphi}_{X,n}(u)| < \kappa_n/\sqrt{n}\right) \left(\mathbf{1}_{u \in A} + \mathbf{1}_{u \notin A}\right) \\
&\leq \frac{(1 + 2\kappa(\log n)^\gamma)^2}{n} \mathbf{1}_{u \in A} + \mathbf{1}_{u \notin A} \mathbb{P}\left(|\widehat{\varphi}_{X,n}(u)| < \frac{1 + \kappa(\log n)^\gamma}{\sqrt{n}} < \frac{1 + 2\kappa(\log n)^\gamma}{\sqrt{n}} \leq |\varphi_X(u)|\right) \\
&\leq \frac{(1 + 2\kappa(\log n)^\gamma)^2}{n} + \mathbb{P}\left(|\varphi_X(u)| - |\widehat{\varphi}_{X,n}(u)| > \kappa\sqrt{\frac{(\log n)^{2\gamma}}{n}}\right) \\
&\leq \frac{(1 + 2\kappa(\log n)^\gamma)^2}{n} + \mathbb{P}\left(|\widehat{\varphi}_{X,n}(u) - \varphi_X(u)| > \kappa\sqrt{\frac{(\log n)^{2\gamma}}{n}}\right),
\end{aligned}$$

where we used the triangle inequality. This together with (3.5) and lemma 3.1, leads to

$$\begin{aligned} \mathbb{E} [|\tilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2] &\leq \frac{C_\alpha + (1 + 2\kappa(\log n)^\gamma)^2}{n} + 4 \exp\left(-\frac{\kappa^2(\log n)^{2\gamma}}{64C_\alpha}\right) \\ &\quad + \frac{16}{\kappa} \frac{\sqrt{n}}{(\log n)^\gamma} \alpha\left(\left[\frac{4}{\kappa} \frac{\sqrt{n}C_\alpha}{(\log n)^\gamma}\right]\right). \end{aligned} \quad (3.10)$$

3.4.2 Proof of theorem 3.1

We decompose the \mathbb{L}_2 risk as follows

$$\mathbb{E} [\|\tilde{f}_{\hat{m}_n} - f\|^2] = \frac{1}{2\pi} \mathbb{E} \left[\int_{[-\hat{m}_n, \hat{m}_n]^c} |\varphi_X(u)|^2 du \right] + \frac{1}{2\pi} \mathbb{E} \left[\int_{-\hat{m}_n}^{\hat{m}_n} |\tilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 du \right]. \quad (3.11)$$

Let $0 \leq m < n$ be fixed and consider the event $\Lambda = \{\hat{m}_n < m\}$. We will control these terms on Λ and Λ^c . First, we focus on the event Λ , which is such that if $\hat{m}_n = n$ then $\Lambda = \emptyset$. We note that :

$$\frac{1}{2\pi} \mathbb{E} \left[\mathbf{1}_\Lambda \int_{-\hat{m}_n}^{\hat{m}_n} |\tilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 du \right] \leq \frac{1}{2\pi} \int_{-m}^m \mathbb{E} [|\tilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2] du,$$

and

$$\frac{1}{2\pi} \mathbf{1}_\Lambda \int_{u \in [-\hat{m}_n, \hat{m}_n]^c} |\varphi_X(u)|^2 du = \frac{1}{2\pi} \mathbf{1}_\Lambda \int_{u \in [-m, m]^c} |\varphi_X(u)|^2 du + \frac{1}{2\pi} \mathbf{1}_\Lambda \int_{|u| \in [\hat{m}_n, m]} |\varphi_X(u)|^2 du.$$

Then together with (3.10) we obtain that

$$\begin{aligned} \mathbb{E} [\|\tilde{f}_{\hat{m}_n} - f\|^2 \mathbf{1}_\Lambda] &\leq \frac{1}{2\pi} \int_{u \in [-m, m]^c} |\varphi_X(u)|^2 du + \frac{1}{2\pi} \mathbb{E} \left[\mathbf{1}_\Lambda \int_{|u| \in [\hat{m}_n, m]} |\varphi_X(u)|^2 du \right] \\ &\quad + m \frac{C_\alpha + (1 + 2\kappa(\log n)^\gamma)^2}{n\pi} + \frac{4m}{\pi} \exp\left(-\frac{\kappa^2(\log n)^{2\gamma}}{64C_\alpha}\right) + \frac{16m}{\kappa\pi} \frac{\sqrt{n}}{(\log n)^\gamma} \alpha\left(\left[\frac{4}{\kappa} \frac{\sqrt{n}C_\alpha}{(\log n)^\gamma}\right]\right). \end{aligned}$$

It suffices to control the second term on the right hand side. The triangle inequality gives

$$\mathbb{E} \left[\mathbf{1}_\Lambda \int_{|u| \in [\hat{m}_n, m]} |\varphi_X(u)|^2 du \right] \leq 2\mathbb{E} \left[\mathbf{1}_\Lambda \int_{|u| \in [\hat{m}_n, m]} |\hat{\varphi}_{X,n}(u)|^2 + |\varphi_X(u) - \hat{\varphi}_{X,n}(u)|^2 du \right].$$

Using the definition (3.7) of \hat{m}_n , we get that for any $|u| \in [\hat{m}_n, m]$, it holds $|\hat{\varphi}_{X,n}(u)| \leq \kappa_n/\sqrt{n}$. We then derive that

$$\mathbb{E} \left[\mathbf{1}_\Lambda \int_{|u| \in [\hat{m}_n, m]} |\varphi_X(u)|^2 du \right] \leq 2\mathbb{E} \left[\mathbf{1}_\Lambda \int_{|u| \in [\hat{m}_n, m]} \kappa_n^2 n^{-1} du \right] + 2 \int_{-m}^m \mathbb{E} [|\hat{\varphi}_{X,n}(u) - \varphi_X(u)|^2] du.$$

For the second term on the right hand side, we use Inequality (3.5). We derive that

$$\mathbb{E} \left[\mathbf{1}_\Lambda \int_{|u| \in [\hat{m}_n, m]} |\varphi_X(u)|^2 du \right] \leq \frac{4m\kappa_n^2}{n} + \frac{4m}{n} + \frac{32m}{n} \sum_{k=1}^{n-1} \alpha(k) = \frac{4m}{n} (\kappa_n^2 + C_\alpha).$$

Consequently, gathering all these inequalities and $m \leq n$ we have proved that

$$\begin{aligned} \mathbb{E} \left[\|\tilde{f}_{\hat{m}_n} - f\|^2 \mathbb{1}_\Lambda \right] &\leq \frac{1}{2\pi} \int_{u \in [-m, m]^c} |\varphi_X(u)|^2 du + \frac{m}{n\pi} (3C_\alpha + 3(1 + 2\kappa(\log n)^\gamma)^2) \\ &\quad + \frac{4n}{\pi} \exp\left(-\frac{\kappa^2(\log n)^{2\gamma}}{64C_\alpha}\right) + \frac{16}{\kappa\pi} \frac{n\sqrt{n}}{(\log n)^\gamma} \alpha\left(\left[\frac{4}{\kappa} \frac{\sqrt{n}C_\alpha}{(\log n)^\gamma}\right]\right). \end{aligned} \quad (3.12)$$

Secondly, we work on the event $\Lambda^c = \{\hat{m}_n \geq m\}$. Considering both terms on the right hand side of (3.11), we see that :

$$\frac{1}{2\pi} \mathbb{1}_{\Lambda^c} \int_{u \in [-\hat{m}_n, \hat{m}_n]^c} |\varphi_X(u)|^2 du \leq \frac{1}{2\pi} \int_{u \in [-m, m]^c} |\varphi_X(u)|^2 du,$$

and

$$\begin{aligned} \mathbb{E} \left[\mathbb{1}_{\Lambda^c} \int_{-\hat{m}_n}^{\hat{m}_n} |\tilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 du \right] &= \int_{-m}^m \mathbb{E} \left[\mathbb{1}_{\Lambda^c} |\tilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \right] du \\ &\quad + \mathbb{E} \left[\mathbb{1}_{\Lambda^c} \int_{|u| \in [m, \hat{m}_n]} |\tilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 du \right]. \end{aligned}$$

Then it follows that,

$$\begin{aligned} \mathbb{E} \left[\|\tilde{f}_{\hat{m}_n} - f\|^2 \mathbb{1}_{\Lambda^c} \right] &\leq \frac{1}{2\pi} \int_{u \in [-m, m]^c} |\varphi_X(u)|^2 du \\ &\quad + \frac{1}{2\pi} \int_{-m}^m \mathbb{E} \left[|\tilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \right] du + \frac{1}{2\pi} \mathbb{E} \left[\mathbb{1}_{\Lambda^c} \int_{|u| \in [m, \hat{m}_n]} |\tilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 du \right]. \end{aligned}$$

We will control the last term on the right hand side, which is a surplus in the variance of $\tilde{f}_{\hat{m}_n}$. We consider two distinct cases according to the value of $|\varphi_X|$. First, consider the set $\{|\varphi_X(u)| > n^{-1/2}\}$, using the definition of $\tilde{\varphi}_X$, Inequality (3.5) and that $|\varphi_X(u)| > n^{-1/2}$, we get

$$\begin{aligned} \mathbb{E}[|\tilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2] &\leq |\varphi_X(u)|^2 + \mathbb{E}[|\hat{\varphi}_{X,n}(u) - \varphi_X(u)|^2] \\ &\leq |\varphi_X(u)|^2 + \frac{1}{n} \left(1 + 8 \sum_{k=1}^{n-1} \alpha(k) \right) \leq |\varphi_X(u)|^2 (1 + C_\alpha). \end{aligned}$$

Hence, we derive

$$\mathbb{E} \left[\mathbb{1}_{\Lambda^c} \int_{|u| \in [m, \hat{m}_n]} |\tilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \mathbb{1}_{\{|\varphi_X(u)| > n^{-1/2}\}} du \right] \leq (1 + C_\alpha) \int_{[-m, m]^c} |\varphi_X(u)|^2 du,$$

which can be associated to the bias term. Secondly, we consider the set $\{|\varphi_X(u)| \leq n^{-1/2}\}$.

Noting that $|\widehat{\varphi}_{X,n}(u)| \leq 1$ and $\widehat{m}_n \leq n$, we obtain

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1}_{\Lambda^c} \int_{|u| \in [m, \widehat{m}_n]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \mathbf{1}_{\{|\varphi_X(u)| \leq n^{-1/2}\}} du \right] \\ = & \mathbb{E} \left[\mathbf{1}_{\Lambda^c} \int_{|u| \in [m, \widehat{m}_n]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \mathbf{1}_{\{|\varphi_X(u)| \leq n^{-1/2}\}} \left(\mathbf{1}_{\{|\widehat{\varphi}_{X,n}(u)| \leq \kappa_n n^{-1/2}\}} + \mathbf{1}_{\{|\widehat{\varphi}_{X,n}(u)| > \kappa_n n^{-1/2}\}} \right) du \right] \\ & \leq \int_{|u| \in [m, n]} |\varphi_X(u)|^2 du + 4 \int_{|u| \in [m, n]} \mathbb{P} \left(|\widehat{\varphi}_{X,n}(u) - \varphi_X(u)| > \kappa (\log n)^\gamma n^{-1/2} \right) du. \end{aligned}$$

Using Lemma 3.1 we derive that

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1}_{\Lambda^c} \int_{|u| \in [m, \widehat{m}_n]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \mathbf{1}_{\{|\varphi_X(u)| \leq n^{-1/2}\}} du \right] \\ & \leq \int_{|u| \in [m, n]} |\varphi_X(u)|^2 du + 2^5 n \exp \left(- \frac{\kappa^2 (\log n)^{2\gamma}}{64 C_\alpha} \right) + \frac{2^7}{\kappa} \frac{n \sqrt{n}}{(\log n)^\gamma} \alpha \left(\left[\frac{4}{\kappa} \frac{\sqrt{n} C_\alpha}{(\log n)^\gamma} \right] \right). \end{aligned}$$

It follows using (3.10) that

$$\begin{aligned} \mathbb{E} \left[\|\widetilde{f}_{\widehat{m}_n} - f\|^2 \mathbf{1}_{\Lambda^c} \right] & \leq \frac{3 + C_\alpha}{2\pi} \int_{u \in [-m, m]^c} |\varphi_X(u)|^2 du + \frac{m C_\alpha + (1 + 2\kappa (\log n)^\gamma)^2}{\pi n} \\ & \quad + \frac{2^4 n}{\pi} \exp \left(- \frac{\kappa^2 (\log n)^{2\gamma}}{64 C_\alpha} \right) + \frac{2^6}{\kappa \pi} \frac{n \sqrt{n}}{(\log n)^\gamma} \alpha \left(\left[\frac{4}{\kappa} \frac{\sqrt{n} C_\alpha}{(\log n)^\gamma} \right] \right). \quad (3.13) \end{aligned}$$

Combining (3.12) and (3.13), and taking the minimum over all m in $[0, n]$ leads to theorem 3.1.

Chapter 4

Moment bounds for the empirical Wasserstein distance of $\tilde{\beta}$ dependent sequences in \mathbb{R}^d

Contents

4.1	Introduction and notations	68
4.2	First upper bounds	70
4.3	Upper bounds under polynomial mixing conditions	75
4.3.1	Moment of order $r = 2$	75
4.3.1.1	Case $p > d/2$	75
4.3.1.2	Case $p < d/2$	79
4.3.2	Moment of order $r \in (1, 2)$	83
4.3.2.1	Case $p > d(r - 1)/r$	83
4.3.2.2	Case $p < d(r - 1)/r$	84
4.3.3	Moment of order $r > 2$	84
4.3.3.1	Case $p > d(r - 1)/r$	84
4.3.3.2	Case $p < d(r - 1)/r$	85
4.4	Proofs	85
4.4.1	Proof of proposition 4.5	86
4.4.2	Proof of proposition 4.6	88
4.4.3	Proof of proposition 4.7	92
4.4.4	Proof of proposition 4.8	95
4.5	Appendix	98

Abstract We study the behaviour of the Wasserstein distance between the empirical distribution μ_n and the marginal distribution μ of stationary sequences of \mathbb{R}^d -valued random variables. We give some moments inequalities of order r with $r \in (1, \infty)$.

4.1 Introduction and notations

The Wasserstein distance of order p ($p \geq 1$) between two probability measures μ and ν on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is defined by

$$W_p^p(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|_2^p \pi(dx, dy)$$

where $\Pi(\mu, \nu)$ is the set of probability measures on the product space $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d))$ with margins μ and ν , and $|\cdot|_2$ is the euclidean norm on \mathbb{R}^d .

If π^* is a solution of the minimization problem above, then a couple (X^*, Y^*) with law π^* provides two variables with marginal distributions μ and ν as close as possible in the sense of the \mathbb{L}^p -distance. One says that (X^*, Y^*) is an optimal coupling of marginals (μ, ν) for the \mathbb{L}^p -cost.

Let $(X_i)_{i \in \mathbb{Z}}$ be strictly stationary sequence of random variables with values in \mathbb{R}^d ($d \geq 1$) and with common distribution μ . Let μ_n be the empirical distribution of $(X_i)_{1 \leq i \leq n}$ such that

$$\mu_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}.$$

In this paper, we give moment inequalities of order $r > 1$ for $W_p^p(\mu_n, \mu)$, for a class of stationary sequences of \mathbb{R}^d -valued random variables. More precisely, our goal is to extend some moment inequalities obtained by Dedecker and Merlevède (2019) [25] for sequences of independent and identically distributed (iid) random variables. Note that these moment inequalities allow to evaluate the concentration of μ_n around μ with respect to the distance W_p . Indeed, by Markov's inequality

$$\mathbb{P}(W_p^p(\mu_n, \mu) > t) \leq \frac{\|W_p^p(\mu_n, \mu)\|_r^r}{t^r}.$$

The moment inequalities obtained in [25] are based on an inequality of Fournier and Guillin (2015) [43] (see inequality (4.3) in section 4.2). This inequality is written in terms of a sum

$$\sum_{A \in \mathcal{F}} |\mu_n(A) - \mu(A)|,$$

where \mathcal{F} is a collection of cubes which forms a partition of a bounded domain of \mathbb{R}^d . To control the \mathbb{L}^r norm of such quantities, a well-adapted dependence coefficient is the $\tilde{\beta}$ coefficient introduced by Dedecker and Prieur (2007) [28]. This coefficient is defined as follows:

Definition 4.1. *Let $(X_i)_{i \in \mathbb{Z}}$ be a strictly stationary sequence of random variables with values in \mathbb{R}^d . Let μ be the law of X_0 . Define the σ -algebra $\mathcal{F}_0 = \sigma(X_k, k \leq 0)$, and let $P_{X_k|\mathcal{F}_0}$ be the conditional distribution of X_k given \mathcal{F}_0 . Define then*

$$b(k) = \sup_{(t_1, \dots, t_d) \in \mathbb{R}^d} |P_{X_k|\mathcal{F}_0}([-\infty, t_1] \times \dots \times [-\infty, t_d]) - \mu([-\infty, t_1] \times \dots \times [-\infty, t_d])|$$

and

$$\tilde{\beta}(k) = \mathbb{E}(b(k)).$$

Note that the $\tilde{\beta}$ coefficients are less restrictive than the usual β -mixing coefficients defined by Volkonski and Rozanov [73]. Therefore, our results apply in particular to \mathbb{R}^d -valued functions of irreducible, aperiodic and positively recurrent Markov chains (we refer to Chapter 9 of Rio's book (2017) [64] for the mixing properties of these chains). Many other examples of $\tilde{\beta}$ dependent sequences are given in [28]. Let us give a simple example. Assume that $X_i = (X_{i,1}, \dots, X_{i,d})^t$ is a \mathbb{R}^d -valued linear process, defined by

$$X_i = \sum_{k=0}^{\infty} A_k \varepsilon_{i-k},$$

where (ε_i) is a sequences of iid \mathbb{R}^d -valued random variables with mean 0 and square integrable coordinates, and A_k is a deterministic sequence of $d \times d$ matrices such that $\sum_{k \geq 0} |A_k|^2 < \infty$ (here $|A_k|$ is the usual norm $|A_k| = \sup_{|x|_2=1} |A_k x|$, and $|\cdot|_2$ is the euclidean norm). Let F_j be the cumulative distribution functions of $X_{0,j}$. If there exist constants $C > 0$ and $\gamma \in (0, 1]$ such that

$$\sup_{1 \leq j \leq d} |F_j(s) - F_j(t)| \leq C|t - s|^\gamma,$$

then (following [28], Section 6),

$$\tilde{\beta}(n) \leq C \left(\sum_{k \geq n} |A_k|^2 \right)^{\frac{\gamma}{\gamma+2}}.$$

In particular, if $|A_k|$ is geometrically decreasing, then so is $\tilde{\beta}(k)$ (whatever the index γ).

Let us conclude this introduction with a few bibliographical notes (the list of articles quoted below being non-exhaustive). The control of the minimal distance $W_p(\mu_n, \mu)$, sometimes called "the optimal matching problem", has been studied for a long time in the iid case. In the case where μ is the uniform measure on the unit cube, precise results have been obtained by Ajtai and al. (1983) [1] (case $p = 1, d = 2$) and Talagrand (1994) [68] (case $p = 1, d > 2$). More recently, almost sure optimal results have been obtained by Barthe and Bordenave (2013) [9]. Regarding the control of $\mathbb{E}(W_p(\mu_n, \mu))$ recent progress has been made by Dereich et al (2013) [33] then by Fournier and Guillin (2015) [43]. As already mentioned, starting from an inequality given in [43], deviation bounds and moment inequalities are given in [25].

Concerning the dependent case, bounds for $\mathbb{E}(W_p(\mu_n, \mu))$ are given in [43] under ρ -mixing conditions (see Bradley (1986) [14] for the definition of ρ -mixing). In the case $d = 1$, moments bounds are established in [24], using slightly weaker coefficients than the coefficients $\tilde{\beta}(k)$.

The article is organised as follow. In Section 4.2, we give a first bound for $\|W_p^p(\mu_n, \mu)\|_r$, $r > 1$. Section 4.3 is devoted to the main results: we give some upper bounds for $\|W_p^p(\mu_n, \mu)\|_r$ under weak moment conditions on X_0 and under polynomial rates for $\tilde{\beta}(k)$. The results for $r = 2$ are presented in subsection 4.3.1 (which also contains the full proof in the case $r = 2$); the results for $r \in (1, 2)$ are presented in subsection 4.3.2; the results for $r > 2$ are presented in subsection 4.3.3. Finally, section 4.4 is devoted to the proofs of the results in the cases where $r \in (1, 2)$ and $r > 2$.

4.2 First upper bounds

Notation 4.1. We give some notations that will be used throughout the rest of paper. Let $(X_i)_{i \in \mathbb{Z}}$ be a strictly stationary sequence of random variables with values in \mathbb{R}^d . Denote by \mathcal{L} the distribution of the sequence $(X_i)_{i \in \mathbb{Z}}$. Let $a_n(\mathcal{L})$ and $b_n(\mathcal{L})$ be two sequences of real numbers depending on \mathcal{L} . The notation

$$a_n(\mathcal{L}) \ll b_n(\mathcal{L})$$

means that there exists a positive constant C not depending on n nor \mathcal{L} such that $a_n(\mathcal{L}) \leq Cb_n(\mathcal{L})$. Let $B_0 = (-1, 1]^d$, and for any integer $m \geq 1$, let $B_m = (-2^m, 2^m]^d \setminus (-2^{m-1}, 2^{m-1}]^d$.

Our first Proposition is the following:

Proposition 4.1. Let $r \in (1, 2)$. Then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \sum_{m \geq 0} 2^{pm} \sum_{l \geq 0} 2^{-pl} \min \left((\mu(B_m))^{1/r}, \frac{2^{\frac{ld(r-1)}{r}}}{n^{\frac{(r-1)}{r}}} \left(\sum_{k=0}^n \mu(B_m) \wedge \tilde{\beta}_k \right)^{1/r} \right). \quad (4.1)$$

Let $r > 2$. Then

$$\begin{aligned} & \|W_p^p(\mu_n, \mu)\|_r \\ & \ll \sum_{m \geq 0} 2^{pm} \sum_{l \geq 0} 2^{-pl} \min \left((\mu(B_m))^{1/r}, \frac{2^{\frac{ld(r-1)}{r}}}{\sqrt{n}} \left(\sum_{k=0}^n (k+1)^{(r-2)/2} (\mu(B_m) \wedge \tilde{\beta}_k) \right)^{1/r} \right). \end{aligned} \quad (4.2)$$

Proof. For any integer l , let \mathcal{P}_l be the natural partition of $(-1, 1]^d$ into 2^{dl} translations of $(-2^{-l}, 2^{-l}]^d$. We start from inequality (6.4) of Dedecker and Merlevède [25]

$$W_p^p(\mu_n, \mu) \leq C \Delta_p(\mu_n, \mu), \quad (4.3)$$

where C is a constant depending on p and d and

$$\Delta_p(\mu_n, \mu) = \sum_{m \geq 0} 2^{pm} \sum_{l \geq 0} 2^{-pl} \sum_{F \in \mathcal{P}_l} |\mu_n(2^m F \cap B_m) - \mu(2^m F \cap B_m)|.$$

Consequently,

$$\|\Delta_p(\mu_n, \mu)\|_r \leq \sum_{m \geq 0} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left\| \sum_{F \in \mathcal{P}_l} |\mu_n(2^m F \cap B_m) - \mu(2^m F \cap B_m)| \right\|_r \quad (4.4)$$

We shall give two upper bounds for the third sum on the right hand side. First, since $\{2^m F, F \in \mathcal{P}_l\} \cap B_m$ is a partition of B_m , we get that

$$\begin{aligned} \left\| \sum_{F \in \mathcal{P}_l} |\mu_n(2^m F \cap B_m) - \mu(2^m F \cap B_m)| \right\|_r & \leq \|\mu_n(B_m) + \mu(B_m)\|_r \\ & \leq \|\mu_n(B_m)\|_r + \mu(B_m) \\ & \leq \|\mathbf{1}_{X_0 \in B_m}\|_r + \mu(B_m) \\ & \leq (\mu(B_m))^{1/r} + \mu(B_m) \\ & \leq 2(\mu(B_m))^{1/r}. \end{aligned} \quad (4.5)$$

To conclude, gathering (4.3), (4.4) and (4.5), we get the first upper bound

$$\|W_p^p(\mu_n, \mu)\|_r \leq \|\Delta_p(\mu_n, \mu)\|_r \leq \sum_{m \geq 0} 2^{pm} \sum_{l \geq 0} 2^{-pl} \times 2(\mu(B_m))^{1/r}. \quad (4.6)$$

To obtain the second upper bound, we use Hölder inequality and the fact that $|\mathcal{P}_l| = 2^{ld}$. We obtain

$$\begin{aligned} & \sum_{F \in \mathcal{P}_l} \|(\mu_n(2^m F \cap B_m) - \mu(2^m F \cap B_m))\|_r \\ & \leq \left(\sum_{F \in \mathcal{P}_l} 1 \right)^{\frac{r-1}{r}} \left(\sum_{F \in \mathcal{P}_l} \mathbb{E} (|\mu_n(2^m F \cap B_m) - \mu(2^m F \cap B_m)|^r) \right)^{1/r} \\ & \leq 2^{\frac{ld(r-1)}{r}} \left(\sum_{F \in \mathcal{P}_l} \mathbb{E} (|\mu_n(2^m F \cap B_m) - \mu(2^m F \cap B_m)|^r) \right)^{1/r}. \quad (4.7) \end{aligned}$$

From (4.7), we will consider separately the case $r = 2$ and the case $r \in (1, 2)$.

Case $r = 2$. From (4.7), we obtain

$$\sum_{F \in \mathcal{P}_l} \|(\mu_n(2^m F \cap B_m) - \mu(2^m F \cap B_m))\|_2 \leq 2^{ld/2} \sqrt{\sum_{F \in \mathcal{P}_l} \frac{1}{n^2} \text{Var} \left(\sum_{k=1}^n \mathbb{1}_{X_k \in 2^m F \cap B_m} \right)}. \quad (4.8)$$

Using the stationarity of the sequence $(X_i)_{1 \leq i \leq n}$, we get

$$\frac{1}{n^2} \text{Var} \left(\sum_{k=0}^n \mathbb{1}_{X_k \in 2^m F \cap B_m} \right) \leq \frac{1}{n} \text{Var}(\mathbb{1}_{X_0 \in 2^m F \cap B_m}) + \frac{2}{n} \sum_{k=1}^n |\text{Cov}(\mathbb{1}_{X_0 \in 2^m F \cap B_m}, \mathbb{1}_{X_k \in 2^m F \cap B_m})|.$$

To control the covariance terms, we use some properties of the conditional expectation.

$$\begin{aligned} |\text{Cov}(\mathbb{1}_{X_0 \in 2^m F \cap B_m}, \mathbb{1}_{X_k \in 2^m F \cap B_m})| &= \mathbb{E}[\mathbb{1}_{X_0 \in 2^m F \cap B_m} (\mathbb{1}_{X_k \in 2^m F \cap B_m} - \mu(2^m F \cap B_m))] \\ &\leq \mathbb{E}[\mathbb{1}_{X_0 \in 2^m F \cap B_m} |\mathbb{E}(\mathbb{1}_{X_k \in 2^m F \cap B_m} | \mathcal{F}_0) - \mu(2^m F \cap B_m)|]. \end{aligned}$$

By a simple combinatorial argument and the definition of $b(k)$, we see that

$$|\mathbb{E}(\mathbb{1}_{X_k \in 2^m F \cap B_m} | \mathcal{F}_0) - \mu(2^m F \cap B_m)| \leq 2^d b(k). \quad (4.9)$$

Consequently

$$\frac{1}{n^2} \text{Var} \left(\sum_{k=0}^m \mathbb{1}_{X_k \in 2^m F \cap B_m} \right) \leq \frac{2}{n} \sum_{k=0}^m \mathbb{E}(\mathbb{1}_{X_0 \in 2^m F \cap B_m} 2^d b(k)). \quad (4.10)$$

Combining (4.8) and (4.10), we infer that

$$\begin{aligned}
\sum_{F \in \mathcal{P}_l} \|(\mu_n(2^m F \cap B_m) - \mu(2^m F \cap B_m))\|_2 &\leq \frac{2^{ld/2}}{\sqrt{n}} \sqrt{2^{d+1} \sum_{k=0}^n \mathbb{E} \left(\sum_{F \in \mathcal{P}_l} (\mathbf{1}_{X_0 \in 2^m F \cap B_m}) b(k) \right)} \\
&\leq \frac{2^{ld/2}}{\sqrt{n}} \sqrt{2^{d+1} \sum_{k=0}^n \mathbb{E}(\mathbf{1}_{X_0 \in B_m}) \wedge \tilde{\beta}(k)} \\
&\leq \frac{2^{ld/2}}{\sqrt{n}} \sqrt{2^{d+1} \sum_{k=0}^n \mu(B_m) \wedge \tilde{\beta}(k)} \tag{4.11}
\end{aligned}$$

Hence, in the case $r = 2$, Proposition 4.1 follows from (4.3), (4.4), (4.5) and (4.11).

Case $r \in (1, 2)$. For these values of r , we will need the following von Bahr-Esseen type inequality.

Proposition 4.2. *Let $r \in (1, 2]$. Let $(Y_i)_{i \geq 0}$ be a stationary sequence of real-valued random variables adapted to a non decreasing filtration $(\mathcal{G}_i)_{i \geq 0}$. Assume that Y_0 has a moment order r and that $\mathbb{E}(Y_0) = 0$. Let $T_n = Y_1 + \dots + Y_n$, and let a be any real number. Then*

$$\mathbb{E}(\|T_n\|_r^r) \leq 2^{2-r} n \left(\mathbb{E}(|Y_0|^r) + r \sum_{j=1}^{n-1} \mathbb{E}(|Y_0 - a|^{r-1} |\mathbb{E}(Y_j | \mathcal{G}_0)|) \right).$$

The proof of a non stationary version of this inequality is given in Appendix. This inequality is a generalisation of a similar inequality in Proposition 10 of Cuny, Dedecker and Jan (2017) [18], which was stated with $a = 0$.

Let us continue the proof. From (4.7), we obtain

$$\begin{aligned}
&\sum_{F \in \mathcal{P}_l} \|(\mu_n(2^m F \cap B_m) - \mu(2^m F \cap B_m))\|_r \\
&\leq 2^{\frac{ld(r-1)}{r}} \left(\sum_{F \in \mathcal{P}_l} \frac{1}{n^r} \mathbb{E} \left(\left| \sum_{k=1}^n \mathbf{1}_{X_k \in 2^m F \cap B_m} - \mu(2^m F \cap B_m) \right|^r \right) \right)^{1/r}. \tag{4.12}
\end{aligned}$$

Applying Proposition 4.2 to the variables $Y_k = \mathbf{1}_{X_k \in 2^m F \cap B_m} - \mu(2^m F \cap B_m)$ and the filtration $\mathcal{F}_k = \sigma(X_i, i \leq k)$, taking $a = -\mu(2^m F \cap B_m)$, and using that $|Y_0|^r \leq |Y_0|$, we obtain

$$\begin{aligned}
&\frac{1}{n^r} \mathbb{E} \left(\left| \sum_{k=1}^n \mathbf{1}_{X_k \in 2^m F \cap B_m} - \mu(2^m F \cap B_m) \right|^r \right) \\
&\leq \frac{2^{2-r}}{n^{r-1}} \mu(2^m F \cap B_m) + \frac{r 2^{2-r}}{n^{r-1}} \sum_{k=1}^n \mathbb{E}[\mathbf{1}_{X_0 \in 2^m F \cap B_m} |\mathbb{E}(\mathbf{1}_{X_k \in 2^m F \cap B_m} | \mathcal{F}_0) - \mu(2^m F \cap B_m)|].
\end{aligned}$$

Using (4.9), we get

$$\frac{1}{n^r} \mathbb{E} \left(\left| \sum_{k=1}^n \mathbf{1}_{X_k \in 2^m F \cap B_m} - \mu(2^m F \cap B_m) \right|^r \right) \ll \frac{1}{n^{r-1}} \mu(2^m F \cap B_m) + \frac{1}{n^{r-1}} \sum_{k=1}^n \mathbb{E}(\mathbf{1}_{X_0 \in 2^m F \cap B_m} b(k)).$$

Consequently

$$\begin{aligned} \sum_{F \in \mathcal{P}_l} \frac{1}{n^r} \mathbb{E} \left(\left| \sum_{k=1}^n \mathbb{1}_{X_k \in 2^m F \cap B_m} - \mu(2^m F \cap B_m) \right|^r \right) &\ll \frac{1}{n^{r-1}} \mu(B_m) + \frac{1}{n^{r-1}} \sum_{k=1}^n \mathbb{E}(\mathbb{1}_{X_0 \in B_m} b(k)) \\ &\ll \frac{1}{n^{r-1}} \sum_{k=0}^n \mu(B_m) \wedge \tilde{\beta}(k). \end{aligned} \quad (4.13)$$

From (4.12) and (4.13), we obtain

$$\sum_{F \in \mathcal{P}_l} \|(\mu_n(2^m F \cap B_m) - \mu(2^m F \cap B_m))\|_r \leq \frac{2^{\frac{ld(r-1)}{r}}}{n^{\frac{(r-1)}{r}}} \left(\sum_{k=0}^n \mu(B_m) \wedge \tilde{\beta}_k \right)^{1/r}. \quad (4.14)$$

Hence, in the case $r \in (1, 2)$, proposition 4.1 follows from (4.3), (4.4), (4.5) and (4.14).

Case $r > 2$. For these values of r , we apply the Burkholder type inequality given in Dedecker and Doukhan (2003) [21] in Proposition 4 page 13, to the variables $Y_k = \mathbb{1}_{X_k \in 2^m F \cap B_m} - \mu(2^m F \cap B_m)$ and the filtration $\mathcal{F}_k = \sigma(X_i, i \leq k)$. We obtain

$$\frac{1}{n^r} \mathbb{E} \left(\left| \sum_{k=1}^n \mathbb{1}_{X_k \in 2^m F \cap B_m} - \mu(2^m F \cap B_m) \right|^r \right) \ll \frac{1}{n^{r/2}} \mathbb{E} \left(|Y_0|^{r/2} \left| \sum_{k=0}^n \mathbb{E}(Y_k | \mathcal{F}_0) \right|^{r/2} \right).$$

Since $|Y_0|^{r/2} \leq |Y_0|$, we infer that

$$\begin{aligned} \frac{1}{n^r} \mathbb{E} \left(\left| \sum_{k=1}^n \mathbb{1}_{X_k \in 2^m F \cap B_m} - \mu(2^m F \cap B_m) \right|^r \right) &\ll \frac{1}{n^{r/2}} \mu(2^m F \cap B_m) \mathbb{E} \left(\left| \sum_{k=0}^n \mathbb{E}(Y_k | \mathcal{F}_0) \right|^{r/2} \right) \\ &+ \frac{1}{n^{r/2}} \mathbb{E} \left(\mathbb{1}_{X_0 \in 2^m F \cap B_m} \left| \sum_{k=0}^n \mathbb{E}(Y_k | \mathcal{F}_0) \right|^{r/2} \right). \end{aligned}$$

Using (4.9), we derive that

$$\begin{aligned} \frac{1}{n^r} \mathbb{E} \left(\left| \sum_{k=1}^n \mathbb{1}_{X_k \in 2^m F \cap B_m} - \mu(2^m F \cap B_m) \right|^r \right) &\ll \frac{1}{n^{r/2}} \mu(2^m F \cap B_m) \mathbb{E} \left(\left| \sum_{k=0}^n b(k) \right|^{r/2} \right) \\ &+ \frac{1}{n^{r/2}} \mathbb{E} \left(\mathbb{1}_{X_0 \in 2^m F \cap B_m} \left| \sum_{k=0}^n b(k) \right|^{r/2} \right). \end{aligned}$$

Consequently

$$\begin{aligned} \sum_{F \in \mathcal{P}_l} \frac{1}{n^r} \mathbb{E} \left(\left| \sum_{k=1}^n \mathbb{1}_{X_k \in 2^m F \cap B_m} - \mu(2^m F \cap B_m) \right|^r \right) &\ll \frac{1}{n^{r/2}} \mu(B_m) \mathbb{E} \left(\left| \sum_{k=0}^n b(k) \right|^{r/2} \right) \\ &+ \frac{1}{n^{r/2}} \mathbb{E} \left(\mathbb{1}_{X_0 \in B_m} \left| \sum_{k=0}^n b(k) \right|^{r/2} \right). \end{aligned} \quad (4.15)$$

Let us give some upper bounds for the two terms on right hand in (4.15). Let $B_n = \sum_{k=0}^n b(k) \mathbf{1}_{X_0 \in B_m}$ and

$$Z = \frac{B_n^{(r-2)/2}}{\|B_n\|_{r/2}^{(r-2)/2}},$$

in such a way that

$$\mathbb{E}(ZB_n) = \left\| \mathbf{1}_{X_0 \in B_m} \left| \sum_{k=0}^n b(k) \right| \right\|_{r/2}.$$

Now

$$\mathbb{E}(ZB_n) = \sum_{k=0}^n \mathbb{E}(Zb(k) \mathbf{1}_{X_0 \in B_m}).$$

Applying the covariance inequality of Dedecker and Doukhan (2003) [21] (see the inequality (4.1) in Proposition 1 of section 4 and using that $|b(k) \mathbf{1}_{X_0 \in B_m}| \leq 1$), we get

$$\mathbb{E}(ZB_n) \leq \sum_{k=0}^n \int_0^{\|b(k) \mathbf{1}_{X_0 \in B_m}\|_1} Q_Z(u) du \leq \sum_{k=0}^n \int_0^{\mu(B_m) \wedge \tilde{\beta}(k)} Q_Z(u) du,$$

where Q_Z is the càdlàg inverse of the tail function $t \rightarrow \mathbb{P}(Z > t)$. Let $H(u) = \sum_{k=0}^n \mathbf{1}_{u \leq \mu(B_m) \wedge \tilde{\beta}(k)}$. Then

$$\begin{aligned} \mathbb{E}(ZB_n) &\leq \int_0^1 H(u) Q_Z(u) du \leq \left(\int_0^1 H(u)^{r/2} du \right)^{2/r} \left(\int_0^1 Q_Z(u)^{r/(r-2)} du \right)^{(r-2)/r} \\ &\leq \left(\int_0^1 H(u)^{r/2} du \right)^{2/r}, \end{aligned} \quad (4.16)$$

the last inequality being true because $\int_0^1 Q_Z(u)^{r/(r-2)} du = \mathbb{E}(Z^{r/(r-2)}) = 1$. Now, from the proof of Lemma 2 in Dedecker and Doukhan (2003) [21], we get

$$\int_0^1 H(u)^{r/2} du \ll \sum_{k=0}^n (k+1)^{(r-2)/2} (\mu(B_m) \wedge \tilde{\beta}(k)),$$

so that

$$\mathbb{E} \left(\left| \mathbf{1}_{X_0 \in B_m} \sum_{k=0}^n b(k) \right|^{r/2} \right) \ll \int_0^1 H(u)^{r/2} du \ll \sum_{k=0}^n (k+1)^{(r-2)/2} (\mu(B_m) \wedge \tilde{\beta}(k)). \quad (4.17)$$

In the same way,

$$\mathbb{E} \left(\left| \sum_{k=0}^n b(k) \right|^{r/2} \right) \leq \sum_{k=0}^n (k+1)^{(r-2)/2} \tilde{\beta}(k). \quad (4.18)$$

From (4.15), (4.17) and (4.18), we obtain that

$$\sum_{F \in \mathcal{P}_l} \frac{1}{n^r} \mathbb{E} \left(\left| \sum_{k=1}^n \mathbf{1}_{X_k \in 2^m F \cap B_m} - \mu(2^m F \cap B_m) \right|^r \right) \ll \frac{1}{n^{r/2}} \sum_{k=0}^n (k+1)^{(r-2)/2} (\mu(B_m) \wedge \tilde{\beta}(k)). \quad (4.19)$$

From (4.12) and (4.19), we get

$$\sum_{F \in \mathcal{P}_l} \|(\mu_n(2^m F \cap B_m) - \mu(2^m F \cap B_m))\|_r \leq \frac{2^{\frac{ld(r-1)}{r}}}{\sqrt{n}} \left(\sum_{k=0}^n (k+1)^{(r-2)/2} (\mu(B_m) \wedge \tilde{\beta}(k)) \right)^{1/r}. \quad (4.20)$$

Hence, in the case $r > 2$, Proposition 4.1 follows from (4.3), (4.4), (4.5) and (4.20). \square

4.3 Upper bounds under polynomial mixing conditions

In this section, we give our main results.

From the last proposition 4.1, it is clear that the upper bound for the quantity $W_p^p(\mu_n, \mu)$ depends on p , d , behaviour of the sequences $\tilde{\beta}(k)$ and $\mu(B_m)$. In the following section, we give some upper bounds under weak moment conditions on X_0 and under polynomial rates for $\tilde{\beta}(k)$.

In the sequel, we denote by X a random variable distributed as X_0 , so that X has distribution μ .

For $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, let $|x| = \max\{|x_1|, \dots, |x_d|\}$. For $b \geq 1$ the weak moment of order b of X is defined by

$$\|X\|_{b,w}^b := \sup_{t>0} t^b \mathbb{P}(|X| > t).$$

We first consider the moment of order $r = 2$.

4.3.1 Moment of order $r = 2$

4.3.1.1 Case $p > d/2$

Proposition 4.3. *Assume that $\|X\|_{b,w} < +\infty$ for $b > 2p$, and assume that $\tilde{\beta}(k) = O(k^{-a})$ for some $a > 0$.*

• *If $p > d/2$ and $a > 1$, then*

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \begin{cases} \frac{\|X\|_{b,w}^p}{\sqrt{n}} & \text{if } b > \frac{2ap}{a-1}, \\ \frac{\|X\|_{b,w}^p \log n}{\sqrt{n}} & \text{if } b = \frac{2ap}{a-1}, \\ \frac{\|X\|_{b,w}^p}{n^{a(\frac{1}{2}-\frac{p}{b})}} & \text{if } b < \frac{2ap}{a-1}. \end{cases}$$

• *If $p > d/2$ and $a < 1$, then*

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \frac{\|X\|_{b,w}^p}{n^{a(\frac{1}{2}-\frac{p}{b})}}.$$

• *If $p > d/2$ and $a = 1$, then*

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \frac{\|X\|_{b,w}^p \sqrt{\log n}}{n^{a(\frac{1}{2}-\frac{p}{b})}}.$$

Remark 4.1. *The case $a > 1$ corresponds to the short memory case, in the sense that partial sums of bounded functions of the sequence $(X_i)_{i \in \mathbb{Z}}$ satisfy the central limit theorem. In this case, we can hope to obtain rates comparable to those of the iid case, provided that the variables have sufficiently high order moments. This is indeed the case here, since when $b > 2ap/(a-1)$, we obtain a decrease of order $n^{-1/2}$ for $\|W_p^p(\mu_n, \mu)\|_2$, as in theorem 5.4 of Dedecker and Merlevède (2019) [25] (case $p > d/2$). We are not able, for the moment, to show that the moment condition that we impose to get the rate $n^{-1/2}$ is optimal. However, in the case $d = 1$, we obtain the same moment condition as in the paper [24] (starting from inequality (3.9) of that paper). In fact, we can easily see that the rates that we obtain from inequality (3.9) in [24], by assuming that $\|X\|_{b,w} < +\infty$ and $\tilde{\beta}(k) = O(k^{-a})$, are exactly the same as ours, whatever $a > 0$ and $b > 2p$ (note however that the dependence coefficients used in [24] are slightly less restrictive than the coefficients $\tilde{\beta}(k)$).*

Proof. Without loss of generality, one may assume that $\|X\|_{b,w} = 1$. The general case follows by using the homogeneity of the distance W_p .

Case $a > 1$. First, we start by bounding the quantity $\sqrt{\sum_{k=0}^n \mu(B_m) \wedge \tilde{\beta}(k)}$ mentioned in (4.1) for $r \in (1, 2)$ (cf. proposition 4.1). We split the sum according as $\mu(B_m) < \frac{1}{k^a}$ or $\mu(B_m) \geq \frac{1}{k^a}$. This gives the upper bound

$$\sum_{k=0}^n \mu(B_m) \wedge \tilde{\beta}_k = \sum_{k=0}^{[\mu(B_m)^{-1/a}]} \mu(B_m) + \sum_{[\mu(B_m)^{-1/a}]+1}^n \tilde{\beta}(k).$$

Clearly, since $a > 1$, the two sums are of the same order, that is $O(\mu(B_m)^{1-\frac{1}{a}})$. Hence

$$\sqrt{\sum_{k=0}^n \mu(B_m) \wedge \tilde{\beta}(k)} \ll (\mu(B_m))^{\frac{1}{2}(1-\frac{1}{a})} \ll (\mu(B_m))^{\frac{1}{2}-\frac{1}{2a}}.$$

So the inequality (4.1) can be rewritten as follows

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \sum_{m \geq 0} 2^{pm} \mu(B_m)^{\frac{1}{2}-\frac{1}{2a}} \sum_{l \geq 0} 2^{-pl} \left(\mu(B_m)^{\frac{1}{2a}} \wedge \frac{2^{\frac{ld}{2}}}{\sqrt{n}} \right). \quad (4.21)$$

By definition of $\mu(B_m)$ and since $\|X\|_{b,w} = 1$, we get that

$$\mu(B_m) \leq P(|X| > 2^{m-1}) \leq \frac{2^b}{2^{mb}} \quad (4.22)$$

Hence, we infer from (4.21) that

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \sum_{m \geq 0} 2^{pm} \frac{1}{2^{mb(\frac{1}{2}-\frac{1}{2a})}} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{2a}}} \wedge \frac{2^{\frac{ld}{2}}}{\sqrt{n}} \right). \quad (4.23)$$

We shall now split the sum over m according as $2^{\frac{mb}{2a}} > \sqrt{n}$ or $2^{\frac{mb}{2a}} \leq \sqrt{n}$. This gives the

following cutoff: $m \leq \lceil \log_2(n^{a/b}) \rceil$ or $m > \lceil \log_2(n^{a/b}) \rceil$. Let then

$$I_1 := \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 2^{pm} \frac{1}{2^{mb(\frac{1}{2}-\frac{1}{2a})}} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{2a}}} \wedge \frac{2^{\frac{ld}{2}}}{\sqrt{n}} \right). \quad (4.24)$$

$$I_2 := \sum_{m > \lceil \log_2(n^{a/b}) \rceil} 2^{pm} \frac{1}{2^{mb(\frac{1}{2}-\frac{1}{2a})}} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{2a}}} \wedge \frac{2^{\frac{ld}{2}}}{\sqrt{n}} \right). \quad (4.25)$$

We first deal with the term I_1 . Clearly, since $p > d/2$,

$$I_1 \leq \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 2^{m(p-b(\frac{1}{2}-\frac{1}{2a}))} \sum_{l \geq 0} \frac{2^{\frac{ld}{2}}}{2^{pl} \sqrt{n}} \ll \frac{1}{\sqrt{n}} \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 2^{m(p-b(\frac{1}{2}-\frac{1}{2a}))}.$$

Here, we must distinguish three cases:

★ If $b > \frac{2ap}{a-1}$, then

$$I_1 \ll \frac{1}{\sqrt{n}} \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 2^{m(p-b(\frac{1}{2}-\frac{1}{2a}))} \ll \frac{1}{\sqrt{n}}. \quad (4.26)$$

★ If $b = \frac{2ap}{a-1}$, then

$$I_1 \ll \frac{1}{\sqrt{n}} \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 1 \ll \frac{\log n}{\sqrt{n}}. \quad (4.27)$$

★ If $b < \frac{2ap}{a-1}$, then

$$I_1 \ll \frac{1}{\sqrt{n}} \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 2^{m(p-b(\frac{1}{2}-\frac{1}{2a}))} \ll \frac{1}{\sqrt{n}} 2^{\log_2(n^{\frac{a}{b}})(p-b(\frac{1}{2}-\frac{1}{2a}))} \ll \frac{1}{\sqrt{n}} n^{\frac{a}{b}(p-b(\frac{1}{2}-\frac{1}{2a}))} \ll \frac{1}{n^{a(\frac{1}{2}-\frac{p}{b})}}. \quad (4.28)$$

We now deal with the term I_2 . Clearly

$$I_2 \leq \sum_{m > \lceil \log_2(n^{a/b}) \rceil} 2^{m(p-\frac{b}{2})} \sum_{l \geq 0} \frac{1}{2^{pl}} \ll 2^{\log_2(n^{a/b})(p-\frac{b}{2})} \ll \frac{1}{n^{a(\frac{1}{2}-\frac{p}{b})}}. \quad (4.29)$$

Note that the rate in (4.29) is $O(n^{-1/2})$ if $b \geq 2ap/(a-1)$. Hence, in the case $a > 1$, Proposition 4.3 follows from (4.23), (4.26), (4.27), (4.28) and (4.29).

Case $a < 1$. Since $\tilde{\beta}(k) = O(k^{-a})$, we infer that

$$\sqrt{\sum_{k=0}^n \mu(B_m) \wedge \tilde{\beta}_k} \leq \sqrt{\sum_{k=0}^n \tilde{\beta}_k} \ll n^{\frac{1}{2}-\frac{a}{2}}.$$

Hence, we infer from (4.1) and (4.22) that

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \sum_{m \geq 0} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{2}}} \wedge \frac{2^{\frac{ld}{2}}}{n^{\frac{a}{2}}} \right). \quad (4.30)$$

We split the sum over m into two terms according as $m \leq \lceil \log_2(n^{a/b}) \rceil$ or $m > \lceil \log_2(n^{a/b}) \rceil$, where \log_2 is the binary logarithm (see the proof of the case $a > 1$ for more details on the choice of the cutoff). Let then

$$J_1 = \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{2}}} \wedge \frac{2^{\frac{ld}{2}}}{n^{\frac{a}{2}}} \right). \quad (4.31)$$

$$J_2 = \sum_{m > \lceil \log_2(n^{a/b}) \rceil} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{2}}} \wedge \frac{2^{\frac{ld}{2}}}{n^{\frac{a}{2}}} \right). \quad (4.32)$$

We first deal with the term J_1 . Clearly, since $p > d/2$,

$$J_1 \leq \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 2^{pm} \sum_{l \geq 0} \frac{2^{\frac{ld}{2}}}{2^{pl} n^{\frac{a}{2}}} \ll \frac{1}{n^{\frac{a}{2}}} \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 2^{pm} \ll \frac{2^{\log_2(n^{a/b})p}}{n^{\frac{a}{2}}} \ll \frac{1}{n^{a(\frac{1}{2} - \frac{p}{b})}}. \quad (4.33)$$

We now deal with the term J_2 . Clearly

$$J_2 \leq \sum_{m > \lceil \log_2(n^{a/b}) \rceil} 2^{m(p - \frac{b}{2})} \sum_{l \geq 0} \frac{1}{2^{pl}} \ll 2^{\log_2(n^{a/b})(p - \frac{b}{2})} \ll \frac{1}{n^{a(\frac{1}{2} - \frac{p}{b})}}. \quad (4.34)$$

Hence, in the case $a < 1$, proposition 4.3 follows from (4.30), (4.33) and (4.34).

Case $a = 1$. Since $\tilde{\beta}(k) = O(k^{-1})$, we infer that

$$\sqrt{\sum_{k=0}^n \mu(B_m) \wedge \tilde{\beta}_k} \leq \sqrt{\sum_{k=0}^n \tilde{\beta}_k} \ll \sqrt{\log n}.$$

Hence, we infer from (4.1) and (4.22) that

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \sum_{m \geq 0} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{2}}} \wedge \frac{2^{\frac{ld}{2}} \sqrt{\log n}}{\sqrt{n}} \right). \quad (4.35)$$

We split the sum over m into two terms according as $m \leq \lceil \log_2(n^{1/b}) \rceil$ or $m > \lceil \log_2(n^{1/b}) \rceil$, where \log_2 is the binary logarithm (see the proof of the case $a > 1$ for more details on the choice of the cutoff). Let then

$$Q_1 := \sum_{m=0}^{\lceil \log_2(n^{1/b}) \rceil} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{2}}} \wedge \frac{2^{\frac{ld}{2}} \sqrt{\log n}}{\sqrt{n}} \right). \quad (4.36)$$

$$Q_2 := \sum_{m > \lceil \log_2(n^{1/b}) \rceil} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{2}}} \wedge \frac{2^{\frac{ld}{2}} \sqrt{\log n}}{\sqrt{n}} \right). \quad (4.37)$$

The rest of the proof is exactly the same as in the case $a < 1$, replacing the term J_1 by Q_1 . We first deal with the term Q_1 . Clearly, since $p > d/2$,

$$Q_1 \leq \sum_{m=0}^{\lfloor \log_2(n^{1/b}) \rfloor} 2^{pm} \sum_{l \geq 0} \frac{2^{\frac{ld}{2}} \sqrt{\log n}}{2^{pl} \sqrt{n}} \ll \frac{\sqrt{\log n}}{\sqrt{n}} \sum_{m=0}^{\lfloor \log_2(n^{1/b}) \rfloor} 2^{pm} \ll \frac{\sqrt{\log n}}{\sqrt{n}} \times 2^{\log_2(n^{1/b})p} \ll \frac{\sqrt{\log n}}{n^{(\frac{1}{2} - \frac{p}{b})}}. \quad (4.38)$$

We now deal with the term Q_2 . It is the same computation as for the term J_2 . Clearly

$$Q_2 \leq \sum_{m > \lfloor \log_2(n^{1/b}) \rfloor} 2^{m(p - \frac{b}{2})} \sum_{l \geq 0} \frac{1}{2^{pl}} \ll 2^{\log_2(n^{1/b})(p - \frac{b}{2})} \ll \frac{1}{n^{(\frac{1}{2} - \frac{p}{b})}}. \quad (4.39)$$

Hence, in the case $a = 1$, proposition 4.3 follows from (4.35), (4.38) and (4.39). \square

4.3.1.2 Case $p < d/2$

Proposition 4.4. *Assume that $\|X\|_{b,w} < +\infty$ for $b > 2p$, and assume that $\tilde{\beta}(k) = O(k^{-a})$ for some $a > 0$.*

• If $p < d/2$ and $a > 1$, then

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \begin{cases} \frac{\|X\|_{b,w}^p}{n^{p/d}} & \text{if } b > \frac{2adp}{ad - 2p}, \\ \frac{\|X\|_{b,w}^p \log n}{n^{p/d}} & \text{if } b = \frac{2adp}{ad - 2p}, \\ \frac{\|X\|_{b,w}^p}{n^{a(\frac{1}{2} - \frac{p}{b})}} & \text{if } b < \frac{2adp}{ad - 2p}. \end{cases}$$

• If $p < d/2$ and $a < 1$, then

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \begin{cases} \frac{\|X\|_{b,w}^p}{n^{ap/d}} & \text{if } b > \frac{2dp}{d - 2p}, \\ \frac{\|X\|_{b,w}^p \log n}{n^{ap/d}} & \text{if } b = \frac{2dp}{d - 2p}, \\ \frac{\|X\|_{b,w}^p}{n^{a(\frac{1}{2} - \frac{p}{b})}} & \text{if } b < \frac{2dp}{d - 2p}. \end{cases}$$

• If $p < d/2$ and $a = 1$, then

$$\|W_p^p(\mu_n, \mu)\|_2 \ll \begin{cases} \frac{\|X\|_{b,w}^p \sqrt{\log n}}{n^{p/d}} & \text{if } b > \frac{2dp}{d - 2p}, \\ \frac{\|X\|_{b,w}^p (\log n)^{3/2}}{n^{p/d}} & \text{if } b = \frac{2dp}{d - 2p}, \\ \frac{\|X\|_{b,w}^p \sqrt{\log n}}{n^{(\frac{1}{2} - \frac{p}{b})}} & \text{if } b < \frac{2dp}{d - 2p}. \end{cases}$$

Remark 4.2. When $a > 1$ and $b > 2adp/(ad - 2p)$, we obtain a decrease of order $n^{-p/d}$ for $\|W_p^p(\mu_n, \mu)\|_2$, as in theorem 5.4 of Dedecker and Merlevède (2019) [25] (case $p < d/2$). We are not able, for the moment, to show that the moment condition that we impose to get the rate $n^{-p/d}$ is optimal.

Proof. Without loss of generality, one may assume that $\|X\|_{b,w} = 1$. The general case follows by using the homogeneity of the distance W_p .

Case $a > 1$. We start from the inequality (4.23) and we split the sum in two terms, as for the case $p > d/2$. We obtain

$$\|W_p^p(\mu_n, \mu)\|_2 \ll I_1 + I_2, \quad (4.40)$$

where I_1 and I_2 are defined in (4.24) and (4.25) respectively.

We first deal with the term I_1 . We split the sum over l in I_1 according as $l \leq \left\lceil \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) \right\rceil$ or $l > \left\lceil \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) \right\rceil$. This gives the equality

$$I_1 = \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 2^{pm} \frac{1}{2^{mb(\frac{1}{2} - \frac{1}{2a})}} \left(\frac{1}{\sqrt{n}} \sum_{l=0}^{\left\lceil \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) \right\rceil} 2^{l(\frac{d}{2} - p)} + \frac{1}{2^{\frac{mb}{2a}}} \sum_{l > \left\lceil \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) \right\rceil} 2^{-pl} \right).$$

First, we deal with the second sum on the right hand side.

$$\frac{1}{2^{\frac{mb}{2a}}} \sum_{l > \left\lceil \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) \right\rceil} 2^{-pl} \ll \frac{1}{2^{\frac{mb}{2a}}} 2^{-p \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right)} \ll \frac{1}{2^{\frac{mb}{2a}}} \frac{n^{-p/d}}{2^{-\frac{mbp}{ad}}} \ll \frac{1}{n^{\frac{p}{d}}} \frac{1}{2^{\frac{mb}{a}(\frac{1}{2} - \frac{p}{d})}}.$$

Now, we deal with the first sum on the right hand side.

$$\frac{1}{\sqrt{n}} \sum_{l=0}^{\left\lceil \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) \right\rceil} 2^{l(\frac{d}{2} - p)} \ll \frac{1}{\sqrt{n}} 2^{\log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) (\frac{d}{2} - p)} \ll \frac{1}{\sqrt{n}} \frac{n^{(\frac{1}{2} - \frac{p}{d})}}{2^{(\frac{mb}{2a} - \frac{mbp}{ad})}} \ll \frac{1}{n^{\frac{p}{d}}} \frac{1}{2^{\frac{mb}{a}(\frac{1}{2} - \frac{p}{d})}}.$$

From these computations, we infer that

$$I_1 \ll \frac{1}{n^{\frac{p}{d}}} \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 2^{pm} \frac{1}{2^{mb(\frac{1}{2} - \frac{1}{2a})}} \frac{1}{2^{\frac{mb}{a}(\frac{1}{2} - \frac{p}{d})}} \ll \frac{1}{n^{\frac{p}{d}}} \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} \frac{2^{pm}}{2^{mb(\frac{1}{2} - \frac{p}{ad})}}.$$

Here, we must distinguish three cases:

★ If $b > \frac{2adp}{ad-2p}$, then

$$I_1 \ll \frac{1}{n^{\frac{p}{d}}}. \quad (4.41)$$

★ If $b = \frac{2adp}{ad-2p}$, then

$$I_1 \ll \frac{\log n}{n^{\frac{p}{d}}}. \quad (4.42)$$

★ If $b < \frac{2adp}{ad-2p}$, then

$$I_1 \ll \frac{1}{n^{\frac{p}{d}}} \sum_{m=0}^{\lfloor \log_2(n^{a/b}) \rfloor} 2^{m(p-b(\frac{1}{2}-\frac{p}{ad}))} \ll \frac{1}{n^{\frac{p}{d}}} n^{\frac{a}{b}(p-b(\frac{1}{2}-\frac{p}{ad}))} \ll \frac{1}{n^{a(\frac{1}{2}-\frac{p}{b})}}. \quad (4.43)$$

Now, the term I_2 can be bounded exactly as in the case $p > d/2$: we obtain the upper bound (4.29). Hence, in the case $a > 1$, Proposition 4.4 follows from (4.40), (4.41), (4.42), (4.43) and (4.29).

Case $a < 1$. We start from the inequality (4.30) and we split the sum in two terms, as for the case $p > d/2$. We obtain

$$\|W_p^p(\mu_n, \mu)\|_2 \ll J_1 + J_2, \quad (4.44)$$

where J_1 and J_2 are defined in (4.31) and (4.32) respectively.

We first deal with the term J_1 . This term is again splitted into two terms:

$$J_1 = \sum_{m=0}^{\lfloor \log_2(n^{a/b}) \rfloor} 2^{pm} \left(\frac{1}{n^{\frac{a}{2}}} \sum_{l=0}^{\lfloor \log_2(n^{a/d} 2^{-mb/d}) \rfloor} 2^{l(\frac{d}{2}-p)} + \frac{1}{2^{\frac{mb}{2}}} \sum_{l > \lfloor \log_2(n^{a/d} 2^{-mb/d}) \rfloor} 2^{-pl} \right).$$

Now, we calculate the second sum on the right hand side.

$$\frac{1}{2^{\frac{mb}{2}}} \sum_{l > \lfloor \log_2(n^{a/d} 2^{-mb/d}) \rfloor} 2^{-pl} \ll \frac{1}{2^{\frac{mb}{2}}} 2^{-p \log_2(n^{a/d} 2^{-mb/d})} \ll \frac{1}{2^{\frac{mb}{2}}} \frac{n^{-ap/d}}{2^{-\frac{mbp}{d}}} \ll \frac{1}{n^{\frac{ap}{d}}} \frac{1}{2^{mb(\frac{1}{2}-\frac{p}{d})}},$$

Then, we calculate the first sum on the right hand side.

$$\frac{1}{n^{\frac{a}{2}}} \sum_{l=0}^{\lfloor \log_2(n^{a/d} 2^{-mb/d}) \rfloor} 2^{l(\frac{d}{2}-p)} \ll \frac{1}{n^{\frac{a}{2}}} 2^{\log_2(n^{a/d} 2^{-mb/d})(\frac{d}{2}-p)} \ll \frac{1}{n^{\frac{a}{2}}} \frac{n^{a(\frac{1}{2}-\frac{p}{d})}}{2^{(\frac{mb}{2}-\frac{mbp}{d})}} \ll \frac{1}{n^{\frac{ap}{d}}} \frac{1}{2^{mb(\frac{1}{2}-\frac{p}{d})}}.$$

From these computations, we infer that

$$J_1 \ll \frac{1}{n^{\frac{ap}{d}}} \sum_{m=0}^{\lfloor \log_2(n^{a/b}) \rfloor} \frac{2^{pm}}{2^{mb(\frac{1}{2}-\frac{p}{d})}}.$$

Here, we must distinguish three cases:

★ If $b > \frac{2dp}{d-2p}$, then

$$J_1 \ll \frac{1}{n^{\frac{ap}{d}}}. \quad (4.45)$$

★ If $b = \frac{2dp}{d-2p}$, then

$$J_1 \ll \frac{\log n}{n^{\frac{ap}{d}}}. \quad (4.46)$$

★ If $b < \frac{2dp}{d-2p}$, then

$$J_1 \ll \frac{1}{n^{\frac{ap}{d}}} \sum_{m=0}^{\lfloor \log_2(n^{a/b}) \rfloor} 2^{m(p-b(\frac{1}{2}-\frac{p}{d}))} \ll \frac{1}{n^{\frac{ap}{d}}} n^{\frac{a}{b}(p-b(\frac{1}{2}-\frac{p}{d}))} \ll \frac{1}{n^{a(\frac{1}{2}-\frac{p}{b})}}. \quad (4.47)$$

Now, the term J_2 can be bounded exactly as in the case $p > d/2$: we obtain the upper bound (4.34). Hence, in the case $a < 1$, Proposition 4.4 follows from (4.44), (4.45), (4.46), (4.47) and (4.34).

Case $a = 1$. The proof is exactly the same as in the case $a < 1$, replacing the term J_1 by the term Q_1 defined in (4.36), changing a by 1 and adding $\sqrt{\log n}$. To be clear we will give the full details for the control of Q_1 in the case $a = 1$.

We start from the inequality (4.35) and we split the sum in two terms, as for the case $p > d/2$. We obtain

$$\|W_p^p(\mu_n, \mu)\|_2 \ll Q_1 + Q_2, \quad (4.48)$$

where Q_1 and Q_2 are defined in (4.36) and (4.37) respectively.

The term Q_1 is again splitted into two terms according as $l \leq \left\lceil \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) \right\rceil$ or $l > \left\lceil \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) \right\rceil$. This gives the equality

$$Q_1 = \sum_{m=0}^{\lfloor \log_2(n^{1/b}) \rfloor} 2^{pm} \left(\frac{\sqrt{\log n}}{\sqrt{n}} \sum_{l=0}^{\left\lceil \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) \right\rceil} 2^{l(\frac{d}{2}-p)} + \frac{1}{2^{\frac{mb}{2}}} \sum_{l > \left\lceil \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) \right\rceil} 2^{-pl} \right).$$

Now, we calculate the second sum on the right hand side.

$$\frac{1}{2^{\frac{mb}{2}}} \sum_{l > \left\lceil \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) \right\rceil} 2^{-pl} \ll \frac{1}{2^{\frac{mb}{2}}} 2^{-p \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right)} \ll \frac{1}{2^{\frac{mb}{2}}} \frac{n^{-p/d}}{2^{-\frac{mbp}{d}}} \ll \frac{1}{n^{\frac{p}{d}}} \frac{1}{2^{mb(\frac{1}{2}-\frac{p}{d})}},$$

Then, we control the first sum on the right hand side.

$$\begin{aligned} \frac{\sqrt{\log n}}{\sqrt{n}} \sum_{l=0}^{\left\lceil \log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) \right\rceil} 2^{l(\frac{d}{2}-p)} &\ll \frac{\sqrt{\log n}}{\sqrt{n}} 2^{\log_2 \left(\frac{n^{1/d}}{2^{-\frac{mb}{ad}}} \right) (\frac{d}{2}-p)} \ll \frac{\sqrt{\log n}}{\sqrt{n}} \frac{n^{(\frac{1}{2}-\frac{p}{d})}}{2^{(\frac{mb}{2}-\frac{mbp}{d})}} \\ &\ll \frac{\sqrt{\log n}}{n^{\frac{p}{d}}} \frac{1}{2^{mb(\frac{1}{2}-\frac{p}{d})}}. \end{aligned} \quad (4.49)$$

From these computations, we infer that

$$Q_1 \ll \frac{\sqrt{\log n}}{n^{\frac{p}{d}}} \sum_{m=0}^{\lfloor \log_2(n^{1/b}) \rfloor} \frac{2^{pm}}{2^{mb(\frac{1}{2}-\frac{p}{d})}}.$$

Here, we must distinguish three cases:

★ If $b > \frac{2dp}{d-2p}$, then

$$Q_1 \ll \frac{\sqrt{\log n}}{n^{\frac{p}{d}}}. \quad (4.50)$$

★ If $b = \frac{2dp}{d-2p}$, then

$$Q_1 \ll \frac{(\log n)^{3/2}}{n^{\frac{p}{d}}}. \quad (4.51)$$

★ If $b < \frac{2dp}{d-2p}$, then

$$Q_1 \ll \frac{\sqrt{\log n}}{n^{\frac{p}{d}}} \sum_{m=0}^{\lfloor \log_2(n^{1/b}) \rfloor} 2^{m(p-b(\frac{1}{2}-\frac{p}{d}))} \ll \frac{\sqrt{\log n}}{n^{\frac{p}{d}}} n^{\frac{1}{b}(p-b(\frac{1}{2}-\frac{p}{d}))} \ll \frac{\sqrt{\log n}}{n^{(\frac{1}{2}-\frac{p}{b})}}. \quad (4.52)$$

Now, the term Q_2 can be bounded exactly as in the case $p > d/2$: we obtain the upper bound (4.39). Hence, in the case $a = 1$, proposition 4.4 follows from (4.48), (4.50), (4.51), (4.52) and (4.39). □

4.3.2 Moment of order $r \in (1, 2)$

In this subsection, we will study the moment of $W_p^p(\mu_n, \mu)$ of order $r \in (1, 2)$. As in subsection 4.3.1, we study the case $p > d(r-1)/r$ and the case $p < d(r-1)/r$. The proofs of the next 4 Proposition will be done in section 4.4.

4.3.2.1 Case $p > d(r-1)/r$

Proposition 4.5. *Assume that $\|X\|_{b,w} < +\infty$ for $b > rp$, and assume that $\tilde{\beta}(k) = O(k^{-a})$ for some $a > 0$.*

• If $p > d(r-1)/r$ and $a > 1$, then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \begin{cases} \frac{\|X\|_{b,w}^p}{n^{\frac{(r-1)}{r}}} & \text{if } b > \frac{arp}{a-1}, \\ \frac{\|X\|_{b,w}^p \log n}{n^{\frac{(r-1)}{r}}} & \text{if } b = \frac{arp}{a-1}, \\ \frac{\|X\|_{b,w}^p}{n^{a(r-1)(\frac{1}{r}-\frac{p}{b})}} & \text{if } b < \frac{arp}{a-1}. \end{cases}$$

• If $p > d(r-1)/r$, $a < 1$ and $2-r < a < 1$, then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \frac{\|X\|_{b,w}^p}{n^{(a+r-2)(\frac{1}{r}-\frac{p}{b})}}.$$

• If $p > d(r-1)/r$ and $a = 1$, then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \frac{\|X\|_{b,w}^p (\log n)^{1/r}}{n^{(r-1)(\frac{1}{r}-\frac{p}{b})}}.$$

4.3.2.2 Case $p < d(r-1)/r$

Proposition 4.6. Assume that $\|X\|_{b,w} < +\infty$ for $b > rp$, and assume that $\tilde{\beta}(k) = O(k^{-a})$ for some $a > 0$.

• If $p < d(r-1)/r$ and $a > 1$, then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \begin{cases} \frac{\|X\|_{b,w}^p}{n^{p/d}} & \text{if } b > \frac{adr(r-1)p}{ad(r-1) - pr}, \\ \frac{\|X\|_{b,w}^p \log n}{n^{p/d}} & \text{if } b = \frac{adr(r-1)p}{ad(r-1) - pr}, \\ \frac{\|X\|_{b,w}^p}{n^{a(r-1)(\frac{1}{r} - \frac{p}{b})}} & \text{if } b < \frac{adr(r-1)p}{ad(r-1) - pr}. \end{cases}$$

• If $p < d(r-1)/r$, $a < 1$ and $2 - r < a < 1$, then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \begin{cases} \frac{\|X\|_{b,w}^p}{n^{p(a+r-2)/d(r-1)}} & \text{if } b > \frac{dr(r-1)p}{d(r-1) - pr} \\ \frac{\|X\|_{b,w}^p \log n}{n^{p(a+r-2)/d(r-1)}} & \text{if } b = \frac{dr(r-1)p}{d(r-1) - pr}, \\ \frac{\|X\|_{b,w}^p}{n^{(a+r-2)(\frac{1}{r} - \frac{p}{b})}} & \text{if } b < \frac{dr(r-1)p}{d(r-1) - pr}. \end{cases}$$

• If $p < d(r-1)/r$ and $a = 1$, then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \begin{cases} \frac{\|X\|_{b,w}^p (\log n)^{1/r}}{n^{p/d}} & \text{if } b > \frac{dr(r-1)p}{d(r-1) - pr}, \\ \frac{\|X\|_{b,w}^p (\log n)^{1+r/r}}{n^{p/d}} & \text{if } b = \frac{dr(r-1)p}{d(r-1) - pr}, \\ \frac{\|X\|_{b,w}^p (\log n)^{1/r}}{n^{(r-1)(\frac{1}{r} - \frac{p}{b})}} & \text{if } b < \frac{dr(r-1)p}{d(r-1) - pr}. \end{cases}$$

4.3.3 Moment of order $r > 2$

4.3.3.1 Case $p > d(r-1)/r$

Proposition 4.7. Assume that $\|X\|_{b,w} < +\infty$ for $b > rp$, and assume that $\tilde{\beta}(k) = O(k^{-a})$ for some $a > 0$.

• If $p > d(r-1)/r$ and $a > r/2$, then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \begin{cases} \frac{\|X\|_{b,w}^p}{\sqrt{n}} & \text{if } b > \frac{2apr}{2a-r}, \\ \frac{\|X\|_{b,w}^p \log n}{\sqrt{n}} & \text{if } b = \frac{2apr}{2a-r}, \\ \frac{\|X\|_{b,w}^p}{n^{a(\frac{1}{r} - \frac{p}{b})}} & \text{if } b < \frac{2apr}{2a-r}. \end{cases}$$

• If $p > d(r-1)/r$ and $a < r/2$, then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \frac{\|X\|_{b,w}^p}{n^{a(\frac{1}{r}-\frac{p}{b})}}.$$

• If $p > d(r-1)/r$ and $a = r/2$, then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \frac{\|X\|_{b,w}^p (\log n)^{1/r}}{n^{\frac{1}{2}(1-\frac{pr}{b})}}.$$

4.3.3.2 Case $p < d(r-1)/r$

Proposition 4.8. Assume that $\|X\|_{b,w} < +\infty$ for $b > rp$, and assume that $\tilde{\beta}(k) = O(k^{-a})$ for some $a > 0$.

• If $p < d(r-1)/r$ and $a > r/2$, then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \begin{cases} \frac{\|X\|_{b,w}^p}{n^{\frac{rp}{2d(r-1)}}} & \text{if } b > \frac{2adpr(r-1)}{2ad(r-1) - r^2p}, \\ \frac{\|X\|_{b,w}^p \log n}{n^{\frac{rp}{2d(r-1)}}} & \text{if } b = \frac{2adpr(r-1)}{2ad(r-1) - r^2p}, \\ \frac{\|X\|_{b,w}^p}{n^{a(\frac{1}{r}-\frac{p}{b})}} & \text{if } b < \frac{2adpr(r-1)}{2ad(r-1) - r^2p}. \end{cases}$$

• If $p < d(r-1)/r$ and $a < r/2$, then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \begin{cases} \frac{\|X\|_{b,w}^p}{n^{ap/d(r-1)}} & \text{if } b > \frac{dr(r-1)p}{d(r-1) - rp}, \\ \frac{\|X\|_{b,w}^p \log n}{n^{ap/d(r-1)}} & \text{if } b = \frac{dr(r-1)p}{d(r-1) - rp}, \\ \frac{\|X\|_{b,w}^p}{n^{a(\frac{1}{r}-\frac{p}{b})}} & \text{if } b < \frac{dr(r-1)p}{d(r-1) - rp}. \end{cases}$$

• If $p < d(r-1)/r$ and $a = r/2$, then

$$\|W_p^p(\mu_n, \mu)\|_r \ll \begin{cases} \frac{\|X\|_{b,w}^p (\log n)^{1/r}}{n^{rp/2d(r-1)}} & \text{if } b > \frac{dr(r-1)p}{d(r-1) - rp}, \\ \frac{\|X\|_{b,w}^p (\log n)^{\frac{1+r}{r}}}{n^{rp/2d(r-1)}} & \text{if } b = \frac{dr(r-1)p}{d(r-1) - rp}, \\ \frac{\|X\|_{b,w}^p (\log n)^{1/r}}{n^{\frac{1}{2}(1-\frac{rp}{b})}} & \text{if } b < \frac{dr(r-1)p}{d(r-1) - rp}. \end{cases}$$

4.4 Proofs

In this section, we prove the propositions 4.5, 4.6, 4.7 and 4.8. The proofs of these propositions are very similar to the proofs of the propositions 4.3 and 4.4.

4.4.1 Proof of proposition 4.5

Without loss of generality, one may assume that $\|X\|_{b,w} = 1$. The general case follows by using the homogeneity of the distance W_p . We follow the lines of proof of proposition 4.3.

Case $a > 1$. We proceed as in the proof of proposition 4.3. Since $\tilde{\beta}(k) = O(k^{-a})$, we infer that

$$\left(\sum_{k=0}^n \mu(B_m) \wedge \tilde{\beta}(k) \right)^{1/r} \ll (\mu(B_m))^{\frac{1}{r} - \frac{1}{ar}}.$$

So the inequality (4.1) can be rewritten as follow

$$\|W_p^p(\mu_n, \mu)\|_r \ll \sum_{m \geq 0} 2^{pm} (\mu(B_m))^{\frac{1}{r} - \frac{1}{ar}} \sum_{l \geq 0} 2^{-pl} \left((\mu(B_m))^{\frac{1}{ar}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{n^{\frac{(r-1)}{r}}} \right). \quad (4.53)$$

As in proposition 4.3, we split the sum over m into two terms according as $m \leq [\log_2(n^{a(r-1)/b})]$ or $m > [\log_2(n^{a(r-1)/b})]$. Hence, we infer from (4.22) that

$$K_1 := \sum_{m=0}^{[\log_2(n^{a(r-1)/b})]} 2^{pm} \frac{1}{2^{mb(\frac{1}{r} - \frac{1}{ar})}} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{ml}{ar}}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{n^{\frac{(r-1)}{r}}} \right). \quad (4.54)$$

$$K_2 := \sum_{m > [\log_2(n^{a(r-1)/b})]} 2^{pm} \frac{1}{2^{mb(\frac{1}{r} - \frac{1}{ar})}} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{ml}{ar}}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{n^{\frac{(r-1)}{r}}} \right). \quad (4.55)$$

We first deal with the term K_1 . Clearly, since $p > d(r-1)/r$,

$$K_1 \leq \sum_{m=0}^{[\log_2(n^{a(r-1)/b})]} 2^{pm} \frac{1}{2^{mb(\frac{1}{r} - \frac{1}{ar})}} \sum_{l \geq 0} \frac{2^{\frac{ld(r-1)}{r}}}{2^{pl} n^{\frac{(r-1)}{r}}} \ll \frac{1}{n^{\frac{(r-1)}{r}}} \sum_{m=0}^{[\log_2(n^{a(r-1)/b})]} 2^{m(p-b(\frac{1}{r} - \frac{1}{ar}))}.$$

Here, we must distinguish three cases:

★ If $b > \frac{arp}{a-1}$, then

$$K_1 \ll \frac{1}{n^{\frac{r-1}{r}}} \sum_{m=0}^{[\log_2(n^{a(r-1)/b})]} 2^{m(p-b(\frac{1}{r} - \frac{1}{ar}))} \ll \frac{1}{n^{\frac{r-1}{r}}}. \quad (4.56)$$

★ If $b = \frac{arp}{a-1}$, then

$$K_1 \ll \frac{1}{n^{\frac{r-1}{r}}} \sum_{m=0}^{[\log_2(n^{a(r-1)/b})]} 1 \ll \frac{\log n}{n^{\frac{r-1}{r}}}. \quad (4.57)$$

★ If $b < \frac{arp}{a-1}$, then

$$\begin{aligned} K_1 &\ll \frac{1}{n^{\frac{(r-1)}{r}}} \sum_{m=0}^{[\log_2(n^{a(r-1)/b})]} 2^{m(p-b(\frac{1}{r} - \frac{1}{ar}))} \ll \frac{1}{n^{\frac{(r-1)}{r}}} 2^{\log_2(n^{\frac{a(r-1)}{b}})(p-b(\frac{1}{r} - \frac{1}{ar}))} \\ &\ll \frac{1}{n^{\frac{(r-1)}{r}}} n^{\frac{a(r-1)}{b}(p-b(\frac{1}{2} - \frac{1}{2a}))} \ll \frac{1}{n^{a(r-1)(\frac{1}{r} - \frac{p}{b})}}. \end{aligned} \quad (4.58)$$

We now deal with the term K_2 . Clearly

$$K_2 \leq \sum_{m > \lceil \log_2(n^{a(r-1)/b}) \rceil} 2^{m(p-\frac{b}{r})} \sum_{l \geq 0} \frac{1}{2^{pl}} \ll 2^{\log_2(n^{a(r-1)/b})(p-\frac{b}{r})} \ll \frac{1}{n^{a(r-1)(\frac{1}{r}-\frac{p}{b})}}. \quad (4.59)$$

Note that the rate in (4.59) is $O(n^{-\frac{a(r-1)}{r}})$ if $b \geq arp/(a-1)$. Hence, in the case $a > 1$, proposition 4.5 follows from (4.53), (4.56), (4.57), (4.58) and (4.59).

Case $a < 1$. Since $\tilde{\beta}(k) = O(k^{-a})$, we infer that

$$\left(\sum_{k=0}^n \mu(B_m) \wedge \tilde{\beta}_k \right)^{1/r} \leq \left(\sum_{k=0}^n \tilde{\beta}_k \right)^{1/r} \ll n^{\frac{1-a}{r}}. \quad (4.60)$$

Hence, we infer from (4.1) and (4.60) that

$$\|W_p^p(\mu_n, \mu)\|_r \ll \sum_{m \geq 0} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{r}}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{n^{\frac{r+a-2}{r}}} \right). \quad (4.61)$$

We split the sum over m into two terms according as $m \leq \lceil \log_2(n^{\frac{r+a-2}{b}}) \rceil$ or $m > \lceil \log_2(n^{\frac{r+a-2}{b}}) \rceil$. We infer from (4.22) that

$$L_1 := \sum_{m=0}^{\lceil \log_2(n^{(r+a-2)/b}) \rceil} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{r}}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{n^{\frac{r+a-2}{r}}} \right). \quad (4.62)$$

$$L_2 := \sum_{m > \lceil \log_2(n^{(r+a-2)/b}) \rceil} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{r}}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{n^{\frac{r+a-2}{r}}} \right). \quad (4.63)$$

We first deal with the term L_1 . Clearly, since $p > d(r-1)/r$,

$$\begin{aligned} L_1 &\leq \sum_{m=0}^{\lceil \log_2(n^{(r+a-2)/b}) \rceil} 2^{pm} \sum_{l \geq 0} \frac{2^{\frac{ld(r-1)}{r}}}{2^{pl} n^{\frac{r+a-2}{r}}} \ll \frac{1}{n^{\frac{r+a-2}{r}}} \sum_{m=0}^{\lceil \log_2(n^{a(r-1)/b}) \rceil} 2^{pm} \\ &\ll \frac{2^{\log_2(n^{(r+a-2)/b})p}}{n^{\frac{r+a-2}{r}}} \ll \frac{1}{n^{(r+a-2)(\frac{1}{r}-\frac{p}{b})}}. \end{aligned} \quad (4.64)$$

We now deal with the term L_2 . Clearly

$$L_2 \leq \sum_{m > \lceil \log_2(n^{(r+a-2)/b}) \rceil} 2^{m(p-\frac{b}{r})} \sum_{l \geq 0} \frac{1}{2^{pl}} \ll 2^{\log_2(n^{(r+a-2)/b})(p-\frac{b}{r})} \ll \frac{1}{n^{(r+a-2)(\frac{1}{r}-\frac{p}{b})}}. \quad (4.65)$$

Hence, in the case $a < 1$, proposition 4.5 follows from (4.61), (4.64) and (4.65).

Case $a = 1$. Since $\tilde{\beta}(k) = O(k^{-1})$, we infer that

$$\left(\sum_{k=0}^n \mu(B_m) \wedge \tilde{\beta}_k \right)^{1/r} \leq \left(\sum_{k=0}^n \tilde{\beta}_k \right)^{1/r} \ll (\log n)^{1/r}. \quad (4.66)$$

Hence, we infer from (4.1) and (4.66) that

$$\|W_p^p(\mu_n, \mu)\|_r \ll \sum_{m \geq 0} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{r}}} \wedge \frac{2^{\frac{ld(r-1)}{r}(\log n)^{1/r}}}{n^{\frac{r-1}{r}}} \right). \quad (4.67)$$

We split the sum over m into two terms according as $m \leq \lceil \log_2(n^{(r-1)/b}) \rceil$ or $m > \lceil \log_2(n^{(r-1)/b}) \rceil$. Let then

$$M_1 := \sum_{m=0}^{\lceil \log_2(n^{(r-1)/b}) \rceil} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{r}}} \wedge \frac{2^{\frac{ld(r-1)}{r}(\log n)^{1/r}}}{n^{\frac{r-1}{r}}} \right). \quad (4.68)$$

$$M_2 := \sum_{m > \lceil \log_2(n^{(r-1)/b}) \rceil} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{r}}} \wedge \frac{2^{\frac{ld(r-1)}{r}(\log n)^{1/r}}}{n^{\frac{r-1}{r}}} \right). \quad (4.69)$$

We first deal with the term M_1 . Clearly, since $p > d(r-1)/r$,

$$\begin{aligned} M_1 &\leq \sum_{m=0}^{\lceil \log_2(n^{\frac{r-1}{b}}) \rceil} 2^{pm} \sum_{l \geq 0} \frac{2^{\frac{ld(r-1)}{r}(\log n)^{1/r}}}{2^{pl} n^{\frac{r-1}{r}}} \ll \frac{(\log n)^{1/r}}{n^{\frac{r-1}{r}}} \sum_{m=0}^{\lceil \log_2(n^{\frac{r-1}{b}}) \rceil} 2^{pm} \\ &\ll \frac{(\log n)^{1/r} 2^{\log_2(n^{\frac{r-1}{b}})p}}{n^{\frac{r-1}{r}}} \ll \frac{(\log n)^{1/r}}{n^{(r-1)(\frac{1}{r} - \frac{p}{b})}}. \end{aligned} \quad (4.70)$$

We now deal with the term M_2 . Clearly

$$M_2 \leq \sum_{m > \lceil \log_2(n^{(r-1)/b}) \rceil} 2^{m(p - \frac{b}{2})} \sum_{l \geq 0} \frac{1}{2^{pl}} \ll 2^{\log_2(n^{\frac{r-1}{b}})(p - \frac{b}{2})} \ll \frac{1}{n^{(r-1)(\frac{1}{r} - \frac{p}{b})}}. \quad (4.71)$$

Hence, the case $a = 1$, in proposition 4.5 follows from (4.67), (4.70) and (4.71).

4.4.2 Proof of proposition 4.6

Without loss of generality, one may assume that $\|X\|_{b,w} = 1$. The general case follows by using the homogeneity of the distance W_p . We follow the line of proof of proposition 4.4.

Case $a > 1$. We start from the inequality (4.53) and we split the sum in two terms, as for the case $p > d(r-1)/r$. We obtain

$$\|W_p^p(\mu_n, \mu)\|_r \ll K_1 + K_2, \quad (4.72)$$

where K_1 and K_2 are defined in (4.54) and (4.55) respectively.

We first deal with the term K_1 . This term is again splitted into two terms as follows:

$$K_1 = \sum_{m=0}^{\lceil \log_2(n^{\frac{a(r-1)}{b}}) \rceil} 2^{pm} \frac{1}{2^{mb(\frac{1}{r} - \frac{1}{ar})}} \left(\frac{1}{n^{\frac{r-1}{r}}} \sum_{l=0}^{\lceil \log_2(n^{\frac{1}{d} 2^{\frac{-mb}{ad(r-1)}})} \rceil} 2^{l(\frac{d}{2} - p) + \frac{1}{2^{\frac{mb}{ar}}}} \sum_{l > \lceil \log_2(n^{\frac{1}{d} 2^{\frac{-mb}{ad(r-1)}})} \rceil} 2^{-pl} \right).$$

Now

$$\frac{1}{2^{\frac{mb}{ar}}} \sum_{l > \lceil \log_2 \left(n^{\frac{1}{d}} 2^{\frac{-mb}{ad(r-1)}} \right) \rceil} 2^{-pl} \ll \frac{1}{2^{\frac{mb}{ar}}} 2^{-p \log_2 \left(n^{\frac{1}{d}} 2^{\frac{-mb}{ad(r-1)}} \right)} \ll \frac{1}{2^{\frac{mb}{ar}}} \frac{n^{-p/d}}{2^{\frac{-mbp}{ad(r-1)}}} \ll \frac{1}{n^{\frac{p}{d}}} \frac{1}{2^{\frac{mb}{a} \left(\frac{1}{r} - \frac{p}{d(r-1)} \right)}},$$

and

$$\frac{1}{n^{\frac{r-1}{r}}} \sum_{l=0}^{\lceil \log_2 \left(n^{\frac{1}{d}} 2^{\frac{-mb}{ad(r-1)}} \right) \rceil} 2^{l \left(\frac{d(r-1)}{r} - p \right)} \ll \frac{1}{n^{\frac{r-1}{r}}} 2^{\log_2 \left(n^{\frac{1}{d}} 2^{\frac{-mb}{ad(r-1)}} \right) \left(\frac{d(r-1)}{r} - p \right)} \ll \frac{1}{n^{\frac{r-1}{r}}} \frac{n^{\left(\frac{r-1}{r} - \frac{p}{d} \right)}}{2^{\left(\frac{mb}{ar} - \frac{mbp}{ad(r-1)} \right)}} \ll \frac{1}{n^{\frac{p}{d}}} \frac{1}{2^{\frac{mb}{a} \left(\frac{1}{r} - \frac{p}{d(r-1)} \right)}}. \quad (4.73)$$

From these computations, we infer that

$$K_1 \ll \frac{1}{n^{\frac{p}{d}}} \sum_{m=0}^{\lceil \log_2 \left(n^{\frac{a(r-1)}{b}} \right) \rceil} 2^{pm} \frac{1}{2^{mb \left(\frac{1}{r} - \frac{1}{ar} \right)}} \frac{1}{2^{\frac{mb}{a} \left(\frac{1}{r} - \frac{p}{d(r-1)} \right)}} \ll \frac{1}{n^{\frac{p}{d}}} \sum_{m=0}^{\lceil \log_2 \left(n^{\frac{a(r-1)}{b}} \right) \rceil} \frac{2^{pm}}{2^{mb \left(\frac{1}{r} - \frac{p}{ad(r-1)} \right)}}.$$

Here, we must distinguish three cases:

★ If $b > \frac{adr(r-1)p}{ad(r-1)-pr}$, then

$$K_1 \ll \frac{1}{n^{\frac{p}{d}}}. \quad (4.74)$$

★ If $b = \frac{adr(r-1)p}{ad(r-1)-pr}$, then

$$K_1 \ll \frac{\log n}{n^{\frac{p}{d}}}. \quad (4.75)$$

★ If $b < \frac{adr(r-1)p}{ad(r-1)-pr}$, then

$$K_1 \ll \frac{1}{n^{\frac{p}{d}}} \sum_{m=0}^{\lceil \log_2 \left(n^{a(r-1)/b} \right) \rceil} 2^{m \left(p - b \left(\frac{1}{r} - \frac{p}{ad(r-1)} \right) \right)} \ll \frac{1}{n^{\frac{p}{d}}} n^{\frac{a(r-1)}{b} \left(p - b \left(\frac{1}{r} - \frac{p}{ad(r-1)} \right) \right)} \ll \frac{1}{n^{a(r-1) \left(\frac{1}{r} - \frac{p}{b} \right)}}. \quad (4.76)$$

Now, the term K_2 can be bounded exactly as in the case $p > d(r-1)/r$: we obtain the upper bound (4.59). Hence, in the case $a > 1$, proposition 4.6 follows from (4.72), (4.74), (4.75), (4.76) and (4.59).

Case $a < 1$. We start from the inequality (4.61) and we split the sum in two terms, as for the case $p > d(r-1)/r$. We obtain

$$\|W_p^p(\mu_n, \mu)\|_r \ll L_1 + L_2, \quad (4.77)$$

where L_1 and L_2 are defined in (4.62) and (4.63) respectively.

We first deal with the term L_1 . Once again, we split the sum over l in two terms as follows:

$$L_1 = \sum_{m=0}^{\lceil \log_2(n^{a(r-1)/b}) \rceil} 2^{pm} \left(\frac{1}{n^{\frac{a+r-2}{r}}} \sum_{l=0}^{\lceil \log_2 \left(n^{\frac{a+r-2}{d(r-1)}} 2^{\frac{-mb}{d(r-1)}} \right) \rceil} 2^{l(\frac{d(r-1)}{r}-p)} + \frac{1}{2^{\frac{mb}{r}}} \sum_{l > \lceil \log_2 \left(n^{\frac{a+r-2}{d(r-1)}} 2^{\frac{-mb}{d(r-1)}} \right) \rceil} 2^{-pl} \right).$$

Now

$$\frac{1}{2^{\frac{mb}{r}}} \sum_{l > \lceil \log_2 \left(n^{\frac{a+r-2}{d(r-1)}} 2^{\frac{-mb}{d(r-1)}} \right) \rceil} 2^{-pl} \ll \frac{1}{2^{\frac{mb}{r}}} 2^{-p \log_2 \left(n^{\frac{a+r-2}{d(r-1)}} 2^{\frac{-mb}{d(r-1)}} \right)} \ll \frac{1}{n^{\frac{a+r-2}{d(r-1)}p}} \frac{1}{2^{mb(\frac{1}{r}-\frac{p}{d(r-1)})}},$$

and

$$\begin{aligned} \frac{1}{n^{\frac{a+r-2}{r}}} \sum_{l=0}^{\lceil \log_2 \left(n^{\frac{a+r-2}{d(r-1)}} 2^{\frac{-mb}{d(r-1)}} \right) \rceil} 2^{l(\frac{d(r-1)}{r}-p)} &\ll \frac{1}{n^{\frac{a+r-2}{r}}} 2^{\log_2 \left(n^{\frac{a+r-2}{d(r-1)}} 2^{\frac{-mb}{d(r-1)}} \right) (\frac{d(r-1)}{r}-p)} \\ &\ll \frac{1}{n^{\frac{a+r-2}{d(r-1)}r}} \frac{1}{2^{mb(\frac{1}{r}-\frac{p}{d(r-1)})}}. \end{aligned}$$

From these computations, we infer that

$$L_1 \ll \frac{1}{n^{\frac{a+r-2}{d(r-1)}}} \sum_{m=0}^{\lceil \log_2 \left(n^{\frac{a(r-1)}{b}} \right) \rceil} \frac{2^{pm}}{2^{mb(\frac{1}{r}-\frac{p}{d(r-1)})}}.$$

Here, we must distinguish three cases:

★ If $b > \frac{dr(r-1)p}{d(r-1)-rp}$, then

$$L_1 \ll \frac{1}{n^{\frac{p(a+r-2)}{d(r-1)}}}. \quad (4.78)$$

★ If $b > \frac{dr(r-1)p}{d(r-1)-rp}$ then

$$L_1 \ll \frac{\log n}{n^{\frac{p(r+a-2)}{d(r-1)}}}. \quad (4.79)$$

★ If $b > \frac{dr(r-1)p}{d(r-1)-rp}$, then

$$L_1 \ll \frac{1}{n^{\frac{a+r-2}{d(r-1)}}} \sum_{m=0}^{\lceil \log_2 \left(n^{\frac{a+r-2}{b}} \right) \rceil} 2^{m(p-b(\frac{1}{r}-\frac{p}{d(r-1)}))} \ll \frac{1}{n^{\frac{a+r-2}{d(r-1)}}} n^{\frac{a+r-2}{b}(p-b(\frac{1}{r}-\frac{p}{d(r-1)}))} \ll \frac{1}{n^{(a+r-2)(\frac{1}{r}-\frac{p}{b})}}. \quad (4.80)$$

Now, the term L_2 can be bounded exactly as in the case $p > d(r-1)/r$: we obtain the upper bound (4.65). Hence, in the case $a < 1$, proposition 4.6 follows from (4.77), (4.78), (4.79), (4.80) and (4.65).

Case $a = 1$.

We start from the inequality (4.67) and we split the sum in two terms, as for the case $p > d(r-1)/r$. We obtain

$$\|W_p^p(\mu_n, \mu)\|_r \ll M_1 + M_2, \quad (4.81)$$

where M_1 and M_2 are defined in (4.68) and (4.69) respectively.

We first deal with the term M_1 . This term is again splitted into two terms as follows:

$$M_1 = \sum_{m=0}^{\lfloor \log_2(n^{\frac{r-1}{b}}) \rfloor} 2^{pm} \left(\frac{(\log n)^{1/r}}{n^{\frac{r-1}{r}}} \sum_{l=0}^{\lfloor \log_2(n^{\frac{1}{d}} 2^{\frac{-mb}{d(r-1)}}) \rfloor} 2^{l(\frac{d(r-1)}{r}-p)} + \frac{1}{2^{\frac{mb}{r}}} \sum_{l > \lfloor \log_2(n^{\frac{1}{d}} 2^{\frac{-mb}{d(r-1)}}) \rfloor} 2^{-pl} \right).$$

Now

$$\frac{1}{2^{\frac{mb}{r}}} \sum_{l > \lfloor \log_2(n^{\frac{1}{d}} 2^{\frac{-mb}{d(r-1)}}) \rfloor} 2^{-pl} \ll \frac{1}{2^{\frac{mb}{r}}} 2^{-p \log_2(n^{1/d} 2^{\frac{-mb}{d(r-1)}})} \ll \frac{1}{2^{\frac{mb}{r}}} \frac{n^{-p/d}}{2^{\frac{-mbp}{d(r-1)}}} \ll \frac{1}{n^{\frac{p}{d}}} \frac{1}{2^{mb(\frac{1}{r} - \frac{p}{d(r-1)})}},$$

and

$$\begin{aligned} \frac{(\log n)^{1/r}}{n^{\frac{r-1}{r}}} \sum_{l=0}^{\lfloor \log_2(n^{\frac{1}{d}} 2^{\frac{-mb}{d(r-1)}}) \rfloor} 2^{l(\frac{d(r-1)}{r}-p)} &\ll \frac{(\log n)^{1/r}}{n^{\frac{r-1}{r}}} 2^{\log_2(n^{\frac{1}{d}} 2^{\frac{-mb}{d(r-1)}})(\frac{d(r-1)}{r}-p)} \\ &\ll \frac{(\log n)^{1/r}}{n^{\frac{r-1}{r}}} \frac{n^{\frac{r-1}{r} - \frac{p}{d}}}{2^{(\frac{mb}{r} - \frac{mbp}{d(r-1)})}} \ll \frac{(\log n)^{1/r}}{n^{\frac{p}{d}}} \frac{n^{-\frac{p}{d}}}{2^{mb(\frac{1}{r} - \frac{p}{d(r-1)})}}. \end{aligned} \quad (4.82)$$

From these computations, we infer that

$$M_1 \ll \frac{(\log n)^{1/r}}{n^{\frac{p}{d}}} \sum_{m=0}^{\lfloor \log_2(n^{(r-1)/b}) \rfloor} \frac{2^{pm}}{2^{mb(\frac{1}{r} - \frac{p}{d(r-1)})}}.$$

Here, we must distinguish three cases:

★ If $b < \frac{dr(r-1)p}{d(r-1)-pr}$, then

$$M_1 \ll \frac{(\log n)^{1/r}}{n^{\frac{p}{d}}}. \quad (4.83)$$

★ If $b = \frac{dr(r-1)p}{d(r-1)-pr}$, then

$$M_1 \ll \frac{(\log n)^{(1+r)/r}}{n^{\frac{p}{d}}}. \quad (4.84)$$

★ If $b > \frac{dr(r-1)p}{d(r-1)-pr}$, then

$$M_1 \ll \frac{(\log n)^{1/r}}{n^{\frac{p}{d}}} \sum_{m=0}^{\lfloor \log_2(n^{(r-1)/b}) \rfloor} 2^{m(p-b(\frac{1}{r} - \frac{p}{d(r-1)}))} \ll \frac{(\log n)^{1/r}}{n^{\frac{p}{d}}} n^{\frac{r-1}{b}(p-b(\frac{1}{r} - \frac{p}{d(r-1)}))} \ll \frac{(\log n)^{1/r}}{n^{(r-1)(\frac{1}{r} - \frac{p}{b})}}. \quad (4.85)$$

Now, the term M_2 can be bounded exactly as in the case $p > d(r-1)/r$: we obtain the upper bound (4.71). Hence, in the case $a = 1$, proposition 4.6 follows from (4.81), (4.83), (4.84), (4.85) and (4.71).

4.4.3 Proof of proposition 4.7

Without loss of generality, one may assume that $\|X\|_{b,w} = 1$. The general case follows by using the homogeneity of the distance W_p . We follow the same lines of proof as proposition 4.3.

Case $a > \frac{r}{2}$. As in the proof of proposition 4.3, we write

$$\sum_{k=0}^n (k+1)^{(r-2)/2} \left(\mu(B_m) \wedge \tilde{\beta}_k \right) = \sum_{k=0}^{\lfloor \mu(B_m)^{-1/a} \rfloor} (k+1)^{(r-2)/2} \mu(B_m) + \sum_{k=\lfloor \mu(B_m)^{-1/a} \rfloor + 1}^n (k+1)^{(r-2)/2} \tilde{\beta}(k).$$

Clearly, the two sums on right hand are of the same order, that is $O(\mu(B_m)^{1-\frac{r}{2a}})$.

It follows that

$$\left(\sum_{k=0}^n (k+1)^{(r-2)/2} \left(\mu(B_m) \wedge \tilde{\beta}(k) \right) \right)^{1/r} \ll (\mu(B_m))^{\frac{1}{r} - \frac{1}{2a}}.$$

So the inequality (4.2) can be rewritten as follow

$$\begin{aligned} \|W_p^p(\mu_n, \mu)\|_r &\ll \sum_{m \geq 0} 2^{pm} \sum_{l \geq 0} 2^{-pl} \min \left((\mu(B_m))^{1/r}, \frac{2^{\frac{ld(r-1)}{r}}}{\sqrt{n}} (\mu(B_m))^{\frac{1}{r} - \frac{1}{2a}} \right) \\ &\ll \sum_{m \geq 0} 2^{pm} (\mu(B_m))^{\frac{1}{r} - \frac{1}{2a}} \sum_{l \geq 0} 2^{-pl} \left((\mu(B_m))^{\frac{1}{2a}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{\sqrt{n}} \right). \end{aligned} \quad (4.86)$$

Hence, we infer from (4.86) and (4.22) that

$$\|W_p^p(\mu_n, \mu)\|_r \ll \sum_{m \geq 0} 2^{pm} \frac{1}{2^{mb(\frac{1}{r} - \frac{1}{2a})}} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{2a}}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{\sqrt{n}} \right). \quad (4.87)$$

As in the proof of proposition 4.3, we split the sum over m into two terms according as $m \leq \lfloor \log_2(n^{\frac{a}{b}}) \rfloor$ or $m > \lfloor \log_2(n^{\frac{a}{b}}) \rfloor$, where \log_2 is the binary logarithm. We obtain from (4.22) and (4.87) that

$$N_1 := \sum_{m=0}^{\lfloor \log_2(n^{\frac{a}{b}}) \rfloor} 2^{pm} \frac{1}{2^{mb(\frac{1}{r} - \frac{1}{2a})}} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{2a}}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{\sqrt{n}} \right). \quad (4.88)$$

$$N_2 := \sum_{m > \lfloor \log_2(n^{\frac{a}{b}}) \rfloor} 2^{pm} \frac{1}{2^{mb(\frac{1}{r} - \frac{1}{2a})}} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{2a}}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{\sqrt{n}} \right). \quad (4.89)$$

We first deal with the term N_1 . Clearly, since $p > d(r-1)/r$,

$$N_1 \leq \sum_{m=0}^{\lfloor \log_2(n^{\frac{a}{b}}) \rfloor} 2^{m(p-b(\frac{1}{r} - \frac{1}{2a}))} \sum_{l \geq 0} \frac{2^{\frac{ld(r-1)}{r}}}{2^{pl} \sqrt{n}} \ll \frac{1}{\sqrt{n}} \sum_{m=0}^{\lfloor \log_2(n^{\frac{a}{b}}) \rfloor} 2^{m(p-b(\frac{2a-r}{2ar}))}.$$

Here, we must distinguish three cases:

★ If $b > \frac{2ap}{2a-r}$, then

$$N_1 \ll \frac{1}{\sqrt{n}} \sum_{m=0}^{\lfloor \log_2(n^{\frac{a}{b}}) \rfloor} 2^{m(p-b(\frac{2a-r}{2ar}))} \ll \frac{1}{\sqrt{n}}. \quad (4.90)$$

★ If $b = \frac{2ap}{2a-r}$, then

$$N_1 \ll \frac{1}{\sqrt{n}} \sum_{m=0}^{\lfloor \log_2(n^{\frac{a}{b}}) \rfloor} 1 \ll \frac{\log n}{\sqrt{n}}. \quad (4.91)$$

★ If $b < \frac{2ap}{2a-r}$, then

$$N_1 \ll \frac{1}{\sqrt{n}} \sum_{m=0}^{\lfloor \log_2(n^{\frac{a}{b}}) \rfloor} 2^{m(p-b(\frac{2a-r}{2ar}))} \ll \frac{1}{\sqrt{n}} 2^{\log_2(n^{\frac{a}{b}})(p-b(\frac{2a-r}{2ar}))} \ll \frac{1}{\sqrt{n}} n^{\frac{a}{b}(p-b(\frac{2a-r}{2ar}))} \ll \frac{1}{n^{a(\frac{1}{r}-\frac{p}{b})}}. \quad (4.92)$$

We now deal with the term N_2 . Clearly

$$N_2 \leq \sum_{m > \lfloor \log_2(n^{a/b}) \rfloor} 2^{m(p-\frac{b}{r})} \sum_{l \geq 0} \frac{1}{2^{pl}} \ll 2^{\log_2(n^{\frac{a}{b}})(p-\frac{b}{r})} \ll \frac{1}{n^{a(\frac{1}{r}-\frac{p}{b})}}. \quad (4.93)$$

Note that the rate in (4.93) is $O(n^{-1/2})$ if $b \geq \frac{2ap}{2a-r}$. Hence, in the case $a > r/2$, proposition 4.7 follows from (4.87), (4.90), (4.91), (4.92) and (4.93).

Case $a < \frac{r}{2}$. Since $\tilde{\beta}(k) = O(k^{-a})$, we obtain

$$\left(\sum_{k=0}^n (k+1)^{\frac{r}{2}-1} \left(\mu(B_m) \wedge \tilde{\beta}_k \right) \right)^{1/r} \leq \left(\sum_{k=0}^n (k+1)^{\frac{r}{2}-1} \tilde{\beta}_k \right)^{1/r} \ll \left(\sum_{k=1}^n k^{\frac{r}{2}-1-a} \right)^{1/r} \ll n^{\frac{1}{2}-\frac{a}{r}}.$$

Hence, we infer from (4.2) that

$$\|W_p^p(\mu_n, \mu)\|_r \ll \sum_{m \geq 0} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{r}}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{n^{\frac{a}{r}}} \right). \quad (4.94)$$

We split the sum over m into two terms according as $m \leq \lfloor \log_2(n^{a/b}) \rfloor$ or $m > \lfloor \log_2(n^{a/b}) \rfloor$, where \log_2 is the binary logarithm. Let then

$$O_1 := \sum_{m=0}^{\lfloor \log_2(n^{a/b}) \rfloor} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{r}}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{n^{\frac{a}{r}}} \right). \quad (4.95)$$

$$O_2 := \sum_{m > \lfloor \log_2(n^{a/b}) \rfloor} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{r}}} \wedge \frac{2^{\frac{ld(r-1)}{r}}}{n^{\frac{a}{r}}} \right). \quad (4.96)$$

We first deal with the term O_1 . Clearly, since $p > d(r-1)/r$,

$$O_1 \leq \sum_{m=0}^{\lfloor \log_2(n^{a/b}) \rfloor} 2^{pm} \sum_{l \geq 0} \frac{2^{\frac{ld(r-1)}{r}}}{2^{pl} n^{\frac{a}{r}}} \ll \frac{1}{n^{\frac{a}{r}}} \sum_{m=0}^{\lfloor \log_2(n^{a/b}) \rfloor} 2^{pm} \ll \frac{2^{\log_2(n^{a/b})p}}{n^{\frac{a}{r}}} \ll \frac{1}{n^{a(\frac{1}{r} - \frac{p}{b})}}. \quad (4.97)$$

We now deal with the term O_2 . Clearly

$$O_2 \leq \sum_{m > \lfloor \log_2(n^{a/b}) \rfloor} 2^{m(p - \frac{b}{r})} \sum_{l \geq 0} \frac{1}{2^{pl}} \ll 2^{\log_2(n^{a/b})(p - \frac{b}{r})} \ll \frac{1}{n^{a(\frac{1}{r} - \frac{p}{b})}}. \quad (4.98)$$

Hence, in the case $a < r/2$, proposition 4.7 follows from (4.94), (4.97) and (4.98).

Case $a = \frac{r}{2}$. Since $\tilde{\beta}(k) = O(k^{-r/2})$, we infer that

$$\left(\sum_{k=0}^n (k+1)^{\frac{r}{2}-1} \left(\mu(B_m) \wedge \tilde{\beta}_k \right) \right)^{1/r} \leq \left(\sum_{k=0}^n (k+1)^{\frac{r}{2}-1} \tilde{\beta}_k \right)^{1/r} \ll \left(\sum_{k=1}^n k^{-1} \right)^{1/r} \ll (\log n)^{1/r}.$$

Hence, we infer from (4.2) that

$$\|W_p^p(\mu_n, \mu)\|_r \ll \sum_{m \geq 0} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{r}}} \wedge \frac{2^{\frac{ld(r-1)}{r}(\log n)^{1/r}}}{\sqrt{n}} \right). \quad (4.99)$$

We split the sum over m into two terms according as $m \leq \lfloor \log_2(n^{r/2b}) \rfloor$ or $m > \lfloor \log_2(n^{r/2b}) \rfloor$, where \log_2 is the binary logarithm. Let then

$$P_1 := \sum_{m=0}^{\lfloor \log_2(n^{r/2b}) \rfloor} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{r}}} \wedge \frac{2^{\frac{ld(r-1)}{r}(\log n)^{1/r}}}{\sqrt{n}} \right). \quad (4.100)$$

$$P_2 := \sum_{m > \lfloor \log_2(n^{r/2b}) \rfloor} 2^{pm} \sum_{l \geq 0} 2^{-pl} \left(\frac{1}{2^{\frac{mb}{r}}} \wedge \frac{2^{\frac{ld(r-1)}{r}(\log n)^{1/r}}}{\sqrt{n}} \right). \quad (4.101)$$

We first deal with the term P_1 . Clearly, since $p > d(r-1)/r$,

$$P_1 \leq \sum_{m=0}^{\lfloor \log_2(n^{r/2b}) \rfloor} 2^{pm} \sum_{l \geq 0} \frac{2^{\frac{ld(r-1)}{r}(\log n)^{1/r}}}{2^{pl} \sqrt{n}} \ll \frac{(\log n)^{1/r}}{\sqrt{n}} \sum_{m=0}^{\lfloor \log_2(n^{r/2b}) \rfloor} 2^{pm} \ll \frac{(\log n)^{1/r}}{\sqrt{n}} 2^{\log_2(n^{r/2b})p} \ll \frac{(\log n)^{1/r}}{n^{\frac{1}{2}(1 - \frac{rp}{b})}}. \quad (4.102)$$

We now deal with the term P_2 . Clearly

$$P_2 \leq \sum_{m > \lfloor \log_2(n^{r/2b}) \rfloor} 2^{m(p - \frac{b}{r})} \sum_{l \geq 0} \frac{1}{2^{pl}} \ll 2^{\log_2(n^{r/2b})(p - \frac{b}{r})} \ll \frac{1}{n^{\frac{1}{2}(1 - \frac{rp}{b})}}. \quad (4.103)$$

Hence, in the case $a = r/2$, proposition 4.7 follows from (4.99), (4.102) and (4.103).

4.4.4 Proof of proposition 4.8

Without loss of generality, one may assume that $\|X\|_{b,w} = 1$. The general case follows by using the homogeneity of the distance W_p . We follow the same line of proof as for the proof of Proposition 4.4.

Case $a > r/2$. We start from the inequality (4.87) and we split the sum in two terms, as for the case $p > d(r-1)/r$. We obtain

$$\|W_p^p(\mu_n, \mu)\|_r \ll N_1 + N_2, \quad (4.104)$$

where N_1 and N_2 are defined in (4.88) and (4.89) respectively.

We first deal with the term N_1 . As usual, the sum over l in N_1 is splitted into two terms as follows:

$$N_1 = \sum_{m=0}^{\lfloor \log_2(n^{a/b}) \rfloor} 2^{pm} \frac{1}{2^{mb(\frac{1}{r}-\frac{1}{2a})}} \left(\frac{1}{\sqrt{n}} \sum_{l=0}^{\left\lfloor \log_2 \left(n^{\frac{r}{2d(r-1)}} 2^{\frac{-mbr}{2ad(r-1)}} \right) \right\rfloor} 2^{l(\frac{d(r-1)}{r}-p)} + \frac{1}{2^{\frac{mb}{2a}}} \sum_{l > \left\lfloor \log_2 \left(n^{\frac{r}{2d(r-1)}} 2^{\frac{-mbr}{2ad(r-1)}} \right) \right\rfloor} 2^{-pl} \right). \quad (4.105)$$

Now

$$\frac{1}{2^{\frac{mb}{2a}}} \sum_{l > \left\lfloor \log_2 \left(n^{\frac{r}{2d(r-1)}} 2^{\frac{-mbr}{2ad(r-1)}} \right) \right\rfloor} 2^{-pl} \ll \frac{1}{2^{\frac{mb}{2a}}} 2^{-p \log_2 \left(n^{\frac{r}{2d(r-1)}} 2^{\frac{-mbr}{2ad(r-1)}} \right)} \ll \frac{1}{2^{\frac{mb}{2a}}} \frac{n^{\frac{-rp}{2d(r-1)}}}{2^{\frac{-mbr}{2ad(r-1)}}} \ll \frac{1}{n^{\frac{rp}{2d(r-1)}}} \frac{1}{2^{\frac{mb}{2a} \left(1 - \frac{rp}{d(r-1)}\right)}}, \quad (4.106)$$

and

$$\frac{1}{\sqrt{n}} \sum_{l=0}^{\left\lfloor \log_2 \left(n^{\frac{r}{2d(r-1)}} 2^{\frac{-mbr}{2ad(r-1)}} \right) \right\rfloor} 2^{l(\frac{d(r-1)}{r}-p)} \ll \frac{1}{\sqrt{n}} 2^{\log_2 \left(n^{\frac{r}{2d(r-1)}} 2^{\frac{-mbr}{2ad(r-1)}} \right) \left(\frac{d(r-1)}{r} - p \right)} \ll \frac{1}{\sqrt{n}} \frac{n^{\left(\frac{1}{2} - \frac{rp}{2d(r-1)}\right)}}{2^{\left(\frac{mb}{2a} - \frac{mbrp}{2ad(r-1)}\right)}} \ll \frac{1}{n^{\frac{rp}{2d(r-1)}}} \frac{1}{2^{\frac{mb}{2a} \left(1 - \frac{rp}{d(r-1)}\right)}}. \quad (4.107)$$

From these computations, we infer that

$$N_1 \ll \frac{1}{n^{\frac{rp}{2d(r-1)}}} \sum_{m=0}^{\lfloor \log_2(n^{a/b}) \rfloor} \frac{2^{pm}}{2^{mb(\frac{1}{r}-\frac{1}{2a})}} \frac{1}{2^{\frac{mb}{2a} \left(1 - \frac{rp}{d(r-1)}\right)}} \ll \frac{1}{n^{\frac{rp}{2d(r-1)}}} \sum_{m=0}^{\lfloor \log_2(n^{a/b}) \rfloor} \frac{2^{pm}}{2^{mb(\frac{1}{r}-\frac{rp}{2ad(r-1)})}} \ll \frac{1}{n^{\frac{rp}{2d(r-1)}}} \sum_{m=0}^{\lfloor \log_2(n^{a/b}) \rfloor} 2^m \left(p - b \left(\frac{1}{r} - \frac{rp}{2ad(r-1)} \right) \right). \quad (4.108)$$

Here, we must distinguish three cases:

★ If $b > \frac{2adpr(r-1)}{2ad(r-1)-r^2p}$, then

$$N_1 \ll \frac{1}{n^{\frac{rp}{2d(r-1)}}}. \quad (4.109)$$

★ If $b = \frac{2adpr(r-1)}{2ad(r-1)-r^2p}$, then

$$N_1 \ll \frac{\log n}{n^{\frac{rp}{2d(r-1)}}}. \quad (4.110)$$

★ If $b < \frac{2adpr(r-1)}{2ad(r-1)-r^2p}$, then

$$N_1 \ll \frac{1}{n^{\frac{rp}{2d(r-1)}}} \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 2^m \left(p-b \left(\frac{1}{r} - \frac{rp}{2ad(r-1)} \right) \right) \ll \frac{1}{n^{\frac{rp}{2d(r-1)}}} n^{\frac{a}{b}} \left(p-b \left(\frac{1}{r} - \frac{rp}{2ad(r-1)} \right) \right) \ll \frac{1}{n^{a(\frac{1}{r} - \frac{p}{b})}}. \quad (4.111)$$

Now, the term N_2 can be bounded exactly as in the case $p > d(r-1)/r$: we obtain the upper bound (4.93). Hence, in the case $a > r/2$, Proposition 4.8 follows from (4.104), (4.109), (4.110), (4.111) and (4.93).

Case $a < r/2$. We start from the inequality (4.61) and we split the sum in two terms, as for the case $p > d(r-1)/r$. We obtain

$$\|W_p^p(\mu_n, \mu)\|_r \ll O_1 + O_2, \quad (4.112)$$

where O_1 and O_2 are defined in (4.95) and (4.96) respectively.

We first deal with the term O_1 . As in the case $a > r/2$, the sum over l in O_1 is again splitted into two terms as follows:

$$O_1 = \sum_{m=0}^{\lceil \log_2(n^{a/b}) \rceil} 2^{pm} \left(\frac{1}{n^{\frac{a}{r}}} \sum_{l=0}^{\lceil \log_2 \left(n^{\frac{a}{d(r-1)}} 2^{\frac{-mb}{d(r-1)}} \right) \rceil} 2^{l \left(\frac{d(r-1)}{r} - p \right)} + \frac{1}{2^{\frac{mb}{r}}} \sum_{l > \lceil \log_2 \left(n^{\frac{a}{d(r-1)}} 2^{\frac{-mb}{d(r-1)}} \right) \rceil} 2^{-pl} \right).$$

Now

$$\frac{1}{2^{\frac{mb}{r}}} \sum_{l > \lceil \log_2 \left(n^{\frac{a}{d(r-1)}} 2^{\frac{-mb}{d(r-1)}} \right) \rceil} 2^{-pl} \ll \frac{1}{2^{\frac{mb}{r}}} 2^{-p \log_2 \left(n^{\frac{a}{d(r-1)}} 2^{\frac{-mb}{d(r-1)}} \right)} \ll \frac{1}{n^{\frac{ap}{d(r-1)}}} \frac{1}{2^{mb \left(\frac{1}{r} - \frac{p}{d(r-1)} \right)}},$$

and

$$\begin{aligned} \frac{1}{n^{\frac{a}{r}}} \sum_{l=0}^{\lceil \log_2 \left(n^{\frac{a}{d(r-1)}} 2^{\frac{-mb}{d(r-1)}} \right) \rceil} 2^{l \left(\frac{d(r-1)}{r} - p \right)} &\ll \frac{1}{n^{\frac{a}{r}}} 2^{\log_2 \left(n^{\frac{a}{d(r-1)}} 2^{\frac{-mb}{d(r-1)}} \right) \left(\frac{d(r-1)}{r} - p \right)} \\ &\ll \frac{1}{n^{\frac{a}{r}}} \frac{n^{\frac{a}{r} - \frac{ap}{d(r-1)}}}{2^{\frac{mb}{r} - \frac{mbp}{d(r-1)}}} \ll \frac{1}{n^{\frac{ap}{d(r-1)}}} \frac{1}{2^{mb \left(\frac{1}{r} - \frac{p}{d(r-1)} \right)}}. \end{aligned}$$

From these computations, we infer that

$$O_1 \ll \frac{1}{n^{\frac{ap}{d(r-1)}}} \sum_{m=0}^{\lceil \log_2 \left(n^{\frac{a}{b}} \right) \rceil} \frac{2^{pm}}{2^{mb \left(\frac{1}{r} - \frac{p}{d(r-1)} \right)}}.$$

Here, we must distinguish three cases:

★ If $b > \frac{dr(r-1)p}{d(r-1)-rp}$, then

$$O_1 \ll \frac{1}{n^{\frac{ap}{d(r-1)}}}. \quad (4.113)$$

★ If $b > \frac{dr(r-1)p}{d(r-1)-rp}$ then

$$O_1 \ll \frac{\log n}{n^{\frac{ap}{d(r-1)}}}. \quad (4.114)$$

★ If $b > \frac{dr(r-1)p}{d(r-1)-rp}$, then

$$O_1 \ll \frac{1}{n^{\frac{ap}{d(r-1)}}} \sum_{m=0}^{\lceil \log_2 \left(n^{\frac{a}{b}} \right) \rceil} 2^{m(p-b \left(\frac{1}{r} - \frac{p}{d(r-1)} \right))} \ll \frac{1}{n^{\frac{ap}{d(r-1)}}} n^{\frac{a}{b} \left(p - b \left(\frac{1}{r} - \frac{p}{d(r-1)} \right) \right)} \ll \frac{1}{n^{a \left(\frac{1}{r} - \frac{p}{b} \right)}}. \quad (4.115)$$

Now, the term O_2 can be bounded exactly as in the case $p > d(r-1)/r$: we obtain the upper bound (4.98). Hence, in the case $a < r/2$, Proposition 4.8 follows from (4.112), (4.113), (4.114), (4.115) and (4.98).

Case $a = r/2$. We start from the inequality (4.99) and we split the sum in two terms, as for the case $p > d(r-1)/r$. We obtain

$$\|W_p^p(\mu_n, \mu)\|_r \ll P_1 + P_2, \quad (4.116)$$

where P_1 and P_2 are defined in (4.100) and (4.101) respectively.

We first deal with the term P_1 . As usual, we split the sum over l in P_1 into to terms. Hence,

$$P_1 = \sum_{m=0}^{\lceil \log_2(n^{r-1/b}) \rceil} 2^{pm} \left(\frac{(\log n)^{1/r}}{\sqrt{n}} \sum_{l=0}^{\lceil \log_2(n^{\frac{r}{2d(r-1)}} 2^{\frac{-mb}{d(r-1)}}) \rceil} 2^{l \left(\frac{d(r-1)}{r} - p \right) + \frac{1}{2 \frac{mb}{r}}} \sum_{l > \lceil \log_2(n^{\frac{r}{2d(r-1)}} 2^{\frac{-mb}{d(r-1)}}) \rceil} 2^{-pl} \right).$$

Now

$$\begin{aligned} \frac{1}{2^{\frac{mb}{r}}} \sum_{l > \lceil \log_2(n^{\frac{r}{2d(r-1)}} 2^{\frac{-mb}{d(r-1)}}) \rceil} 2^{-pl} &\ll \frac{1}{2^{\frac{mb}{r}}} 2^{-p \log_2(n^{\frac{r}{2d(r-1)}} 2^{\frac{-mb}{d(r-1)}})} \ll \frac{1}{2^{\frac{mb}{r}}} \frac{n^{\frac{-rp}{2d(r-1)}}}{2^{\frac{-mbp}{d(r-1)}}} \\ &\ll \frac{1}{n^{\frac{rp}{2d(r-1)}}} \frac{1}{2^{mb \left(\frac{1}{r} - \frac{p}{d(r-1)} \right)}}, \end{aligned}$$

and

$$\begin{aligned} \frac{(\log n)^{1/r}}{\sqrt{n}} \sum_{l=0}^{\lfloor \log_2(n^{\frac{r}{2d(r-1)}} 2^{\frac{-mb}{d(r-1)}}) \rfloor} 2^{l(\frac{d(r-1)}{r}-p)} &\ll \frac{(\log n)^{1/r}}{\sqrt{n}} 2^{\log_2(n^{\frac{r}{2d(r-1)}} 2^{\frac{-mb}{d(r-1)}})(\frac{d(r-1)}{r}-p)} \\ &\ll \frac{(\log n)^{1/r}}{\sqrt{n}} \frac{n^{\frac{1}{2}-\frac{rp}{2d(r-1)}}}{2^{(\frac{mb}{r}-\frac{mbp}{d(r-1)})}} \ll \frac{(\log n)^{1/r} n^{\frac{-rp}{2d(r-1)}}}{2^{mb(\frac{1}{r}-\frac{p}{d(r-1)})}}. \end{aligned} \quad (4.117)$$

From these computations, we infer that

$$P_1 \ll \frac{(\log n)^{1/r}}{n^{\frac{rp}{2d(r-1)}}} \sum_{m=0}^{\lfloor \log_2(n^{r/2b}) \rfloor} \frac{2^{pm}}{2^{mb(\frac{1}{r}-\frac{p}{d(r-1)})}}.$$

Here, we must distinguish three cases:

★ If $b < \frac{dr(r-1)p}{d(r-1)-pr}$, then

$$P_1 \ll \frac{(\log n)^{1/r}}{n^{\frac{rp}{2d(r-1)}}}. \quad (4.118)$$

★ If $b = \frac{dr(r-1)p}{d(r-1)-pr}$, then

$$P_1 \ll \frac{(\log n)^{(1+r)/r}}{n^{\frac{rp}{2d(r-1)}}}. \quad (4.119)$$

★ If $b > \frac{dr(r-1)p}{d(r-1)-pr}$, then

$$P_1 \ll \frac{(\log n)^{1/r}}{n^{\frac{rp}{2d(r-1)}}} \sum_{m=0}^{\lfloor \log_2(n^{r/2b}) \rfloor} 2^{m(p-b(\frac{1}{r}-\frac{p}{d(r-1)}))} \ll \frac{(\log n)^{1/r}}{n^{\frac{rp}{2d(r-1)}}} n^{\frac{r}{2b}(p-b(\frac{1}{r}-\frac{p}{d(r-1)}))} \ll \frac{(\log n)^{1/r}}{n^{\frac{1}{2}(1-\frac{rp}{b})}}. \quad (4.120)$$

Now, the term P_2 can be bounded exactly as in the case $p > d(r-1)/r$: we obtain the upper bound (4.103). Hence, in the case $a = r/2$, Proposition 4.8 follows from (4.116), (4.118), (4.119), (4.120) and (4.103).

4.5 Appendix

In this section, we prove Proposition 4.2. This proposition follows from the more general version given below:

Proposition 4.9. *Let $r \in (1, 2]$. Let $(Y_i)_{i \geq 0}$ be a sequence of real-valued random variables adapted to a non decreasing filtration $(\mathcal{G}_i)_{i \geq 0}$. Assume that all the variables Y_i have a moment order r and that $\mathbb{E}(Y_i) = 0$. Let $T_n = Y_1 + \dots + Y_n$, and let $(a_i)_{i \geq 0}$ be a sequence of real numbers. Then*

$$\mathbb{E}(\|T_n\|_r^r) \leq 2^{2-r} \left(\sum_{i=1}^n \mathbb{E}(|Y_i|^r) + r \sum_{i=1}^n \sum_{k=1}^{i-1} \mathbb{E}(|Y_k - a_k|^{r-1} |\mathbb{E}(Y_i | \mathcal{G}_k)|) \right).$$

Proof. We proceed as in the proof of Proposition 10 of Cuny, Dedecker and Jan (2018) [18]. For $r \in (1, 2]$, let ψ_r be the function from \mathbb{R} to \mathbb{R}^+ defined by $\psi_r(x) = |x|^r$. We start from the elementary decomposition (using the convention $T_0 = 0$):

$$\begin{aligned} |T_n|^r &= \psi_r(T_n) = \sum_{i=1}^n \psi_r(T_i) - \psi_r(T_{i-1}) = \sum_{i=1}^n Y_i \int_0^1 \psi'_r(T_{i-1} + tY_i) dt \\ &= \sum_{i=1}^n Y_i \int_0^1 (\psi'_r(T_{i-1} + tY_i) - \psi'_r(T_{i-1})) dt + \sum_{i=1}^n Y_i \psi'_r(T_{i-1}). \end{aligned}$$

Let $A_0 = 0$ and $A_i = \sum_{j=1}^i a_j$ for $i \geq 1$. Since $\mathbb{E}(Y_i) = 0$, we infer that

$$\mathbb{E}(|T_n|^r) = \sum_{i=1}^n \mathbb{E} \left(Y_i \int_0^1 (\psi'_r(T_{i-1} + tY_i) - \psi'_r(T_{i-1})) dt \right) + \sum_{i=1}^n \mathbb{E}(Y_i(\psi'_r(T_{i-1}) - \psi'_r(A_{i-1}))).$$

Let $Z_{k,i-1} = A_{i-1} + \sum_{j=1}^k (Y_j - a_j)$ and $Z_{0,i-1} = A_{i-1}$. Then

$$\mathbb{E}(Y_i(\psi'_r(T_{i-1}) - \psi'_r(A_{i-1}))) = \mathbb{E} \left(\sum_{k=1}^{i-1} Y_i(\psi'_r(Z_{k,i-1}) - \psi'_r(Z_{k-1,i-1})) \right)$$

Using the fact that

$$|\psi'_r(x) - \psi'_r(y)| \leq r2^{2-r}|x - y|^{r-1},$$

and taking the conditional expectation, we obtain

$$\left| \mathbb{E} \left(\sum_{i=1}^n \sum_{k=1}^{i-1} Y_i(\psi'_r(Z_{k,i-1}) - \psi'_r(Z_{k-1,i-1})) \right) \right| \leq r2^{2-r} \sum_{i=1}^n \sum_{k=1}^{i-1} \mathbb{E} \left(|Y_k - a_k|^{r-1} |\mathbb{E}(Y_i | \mathcal{G}_k)| \right).$$

We also have

$$\left| \sum_{i=1}^n \mathbb{E} \left(Y_i \int_0^1 \psi'_r(T_{i-1} + tY_i) - \psi'_r(T_{i-1}) dt \right) \right| \leq 2^{2-r} \sum_{i=1}^n \mathbb{E}(|Y_i|^r) \int_0^1 r t^{r-1} dt \leq 2^{2-r} \sum_{i=1}^n \mathbb{E}(|Y_i|^r).$$

Consequently,

$$\mathbb{E}(|T_n|^r) \leq 2^{2-r} \left(\sum_{i=1}^n \mathbb{E}(|Y_i|^r) + r \sum_{i=1}^n \sum_{k=1}^{i-1} \mathbb{E} \left(|Y_k - a_k|^{r-1} |\mathbb{E}(Y_i | \mathcal{G}_k)| \right) \right).$$

□

Bibliography

- [1] M. Ajtai, J. Komlós, and G. Tusnády. On optimal matchings. *Combinatorica*, 4(4):259–264, 1984.
- [2] S. Ammous. Testing Kendall’s τ for a large class of dependent sequences. *Statistics*, 54(4):686–713, 2020.
- [3] S. Ammous, O. Bouaziz, A. Dedecker, J. Dedecker, J. El Methni, M. Mellouk, and F. Muri. The robustest package: two-sample tests revisited. *arXiv preprint arXiv:2211.08784*, 2022.
- [4] D. W. Andrews. Non-strong mixing autoregressive processes. *Journal of Applied Probability*, 21(4):930–934, 1984.
- [5] D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991.
- [6] D. W. K. Andrews and J. C. Monahan. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60(4):953–966, 1992.
- [7] N. Asin and J. Johannes. Adaptive nonparametric estimation in the presence of dependence. *Journal of Nonparametric Statistics*, 29(4):694–730, 2017.
- [8] A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probability theory and related fields*, 113(3):301–413, 1999.
- [9] F. Barthe and C. Bordenave. Combinatorial optimization over two random point sets. In *Séminaire de Probabilités XLV*, volume 2078 of *Lecture Notes in Math.*, pages 483–535. Springer, Cham, 2013.
- [10] H. C. P. Berbee. *Random walks with stationary increments and renewal theory*, volume 112 of *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam, 1979.
- [11] K. Bertin and N. Klutchnikoff. Pointwise adaptive estimation of the marginal density of a weakly dependent process. *Journal of Statistical Planning and Inference*, 187:115–129, 2017.
- [12] L. Birgé and P. Massart. From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, New York, 1997.
- [13] S. Borovkova, R. Burton, and H. Dehling. Limit theorems for functionals of mixing processes with applications to U -statistics and dimension estimation. *Trans. Amer. Math. Soc.*, 353(11):4261–4318, 2001.
- [14] R. C. Bradley. Basic properties of strong mixing conditions. In *Dependence in probability and statistics (Oberwolfach, 1985)*, volume 11 of *Progr. Probab. Statist.*, pages 165–192. Birkhäuser Boston, Boston, MA, 1986.

- [15] E. Caron, J. Dedecker, and B. Michel. Linear regression with stationary errors : the R package `slm`. *The R Journal*, 13(1):83–100, June 2021. The associated R package 'slm' is available on the CRAN website (<https://cran.r-project.org>) or on the GitHub website (<https://github.com/E-Caron/slm>).
- [16] F. Comte, J. Dedecker, and M. L. Taupin. Adaptive density deconvolution with dependent inputs. *Math. Methods Statist.*, 17(2):87–112, 2008.
- [17] F. Comte and F. Merlevède. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM: Probability and Statistics*, 6:211–238, 2002.
- [18] C. Cuny, J. Dedecker, and C. Jan. Limit theorems for the left random walk on $GL_d(\mathbb{R})$. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(4):1839–1865, 2017.
- [19] K. B. Davis. Mean integrated square error properties of density estimates. *Ann. Statist.*, 5(3):530–535, 1977.
- [20] J. Dedecker. An empirical central limit theorem for intermittent maps. *Probab. Theory Related Fields*, 148(1-2):177–195, 2010.
- [21] J. Dedecker and P. Doukhan. A new covariance inequality and applications. *Stochastic processes and their applications*, 106(1):63–80, 2003.
- [22] J. Dedecker, P. Doukhan, G. Lang, J. R. León R., S. Louhichi, and C. Prieur. *Weak dependence: with examples and applications*, volume 190 of *Lecture Notes in Statistics*. Springer, New York, 2007.
- [23] J. Dedecker, S. Gouëzel, and F. Merlevède. Large and moderate deviations for bounded functions of slowly mixing Markov chains. *Stoch. Dyn.*, 18(2):38 pages, 2018.
- [24] J. Dedecker and F. Merlevède. Behavior of the Wasserstein distance between the empirical and the marginal distributions of stationary α -dependent sequences. *Bernoulli*, 23(3):2083–2127, 2017.
- [25] J. Dedecker and F. Merlevède. Behavior of the empirical Wasserstein distance in \mathbb{R}^d under moment conditions. *Electron. J. Probab.*, 24:Paper No. 6, 32, 2019.
- [26] J. Dedecker, F. Merlevède, and E. Rio. Strong approximation results for the empirical process of stationary sequences. *Ann. Probab.*, 41(5):3658–3696, 2013.
- [27] J. Dedecker and C. Prieur. New dependence coefficients. examples and applications to statistics. *Probability Theory and Related Fields*, 132(2):203–236, 2005.
- [28] J. Dedecker and C. Prieur. An empirical central limit theorem for dependent sequences. *Stochastic Process. Appl.*, 117(1):121–142, 2007.
- [29] J. Dedecker and E. Rio. On the functional central limit theorem for stationary processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 36(1):1–34, 2000.
- [30] H. Dehling, D. Vogel, M. Wendler, and D. Wied. Testing for changes in Kendall's tau. *Econometric Theory*, 33(6):1352–1386, 2017.
- [31] H. Dehling and M. Wendler. Central limit theorem and the bootstrap for U -statistics of strongly mixing data. *J. Multivariate Anal.*, 101(1):126–137, 2010.
- [32] M. Denker and G. Keller. On U -statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete*, 64(4):505–522, 1983.
- [33] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(4):1183–1203, 2013.

- [34] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.
- [35] D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard. Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(2):301–337, 1995.
- [36] D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard. Density estimation by wavelet thresholding. *The Annals of Statistics*, pages 508–539, 1996.
- [37] P. Doukhan, P. Massart, and E. Rio. The functional central limit theorem for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 30(1):63–82, 1994.
- [38] C. Duval and J. Kappus. Adaptive procedure for fourier estimators: application to deconvolution and decompounding. *Electronic Journal of Statistics*, 13(2):3424–3452, 2019.
- [39] S. Efromovich. Data-driven efficient estimation of the spectral density. *J. Amer. Statist. Assoc.*, 93(442):762–769, 1998.
- [40] F. Esscher. On a method of determining correlation from the ranks of the variates. *Scandinavian Actuarial Journal*, 1924(1):201–219, 1924.
- [41] N. Etemadi. On some classical results in probability theory. *Sankhyā Ser. A*, 47(2):215–221, 1985.
- [42] W. Feller. *An introduction to probability theory and its applications. Vol. II.* John Wiley & Sons, Inc., New York-London-Sydney, 1966.
- [43] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, 162(3-4):707–738, 2015.
- [44] I. Gannaz and O. Wintenberger. Adaptive density estimation under weak dependence. *ESAIM: Probability and Statistics*, 14:151–172, 2010.
- [45] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, pages 1608–1632, 2011.
- [46] A. Goldenshluger and O. Lepski. On adaptive minimax density estimation on R^d . *Probab. Theory Related Fields*, 159(3-4):479–543, 2014.
- [47] M. Gordin. Abstract of communication T.1: A-K. *International Conference on Probability Theory, Vilnius*, 1973.
- [48] S. Gouëzel. Central limit theorem and stable laws for intermittent maps. *Probab. Theory Related Fields*, 128(1):82–122, 2004.
- [49] W. Härdle, G. Kerkycharian, D. Picard, and A. Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129. Springer Science & Business Media, 2012.
- [50] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19:293–325, 1948.
- [51] T. Hsing and W. B. Wu. On weighted U -statistics for stationary processes. *Ann. Probab.*, 32(2):1600–1631, 2004.
- [52] I. A. Ibragimov. Some limit theorems for stationary processes. *Teor. Veroyatnost. i Primenen.*, 7:361–392, 1962.

- [53] L. Kantorovitch. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.
- [54] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30:277–283, 1938.
- [55] C. Lacour, P. Massart, and V. Rivoirard. Estimator selection: a new method with applications to kernel density estimation. *Sankhya A*, 79(2):298–335, 2017.
- [56] M. Lerasle. Adaptive density estimation of stationary β -mixing and τ -mixing processes. *Mathematical Methods of statistics*, 18(1):59–83, 2009.
- [57] M. Lerasle. Optimal model selection for density estimation of stationary data under various mixing conditions. *The Annals of Statistics*, 39(4):1852–1877, 2011.
- [58] J. W. Lindeberg. Über die correlation. *VI Skand. Mathematikerkongre i Kobenhavn*, pages 437–446, 1925.
- [59] J. W. Lindeberg. Some remarks of the mean error of the percentage of correlation. *Nordic Statistical Journal*, 1:137–141, 1929.
- [60] P. Massart. *Concentration inequalities and model selection*, volume 6. Springer, 2007.
- [61] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.
- [62] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I: Theory*. Probability and its Applications (New York). Springer-Verlag, New York, 1998.
- [63] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. II: Applications*. Probability and its Applications (New York). Springer-Verlag, New York, 1998.
- [64] E. Rio. *Asymptotic theory of weakly dependent random processes*, volume 80. Springer, 2017.
- [65] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.*, 42:43–47, 1956.
- [66] P. K. Sen. On the properties of U -statistics when the observations are not independent. I. Estimation of non-serial parameters in some stationary stochastic process. *Calcutta Statist. Assoc. Bull.*, 12:69–92, 1963.
- [67] M. Talagrand. Matching random samples in many dimensions. *Ann. Appl. Probab.*, 2(4):846–856, 1992.
- [68] M. Talagrand. The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3 . *Ann. Probab.*, 22(2):919–959, 1994.
- [69] M. Talagrand. New concentration inequalities in product spaces. *Inventiones mathematicae*, 126(3):505–563, 1996.
- [70] K. Tribouley and G. Viennet. Lp adaptive density estimation in a β mixing framework. *Annales de l’Institut Henri Poincaré (B) Probability and Statistics*, 34(2):179–208, 1998.
- [71] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.
- [72] C. Villani. *Optimal transport: Old and new*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

- [73] V. A. Volkonskiĭ and Y. A. Rozanov. Some limit theorems for random functions. I. *Theor. Probability Appl.*, 4:178–197, 1959.
- [74] K.-I. Yoshihara. Limiting behavior of U -statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 35(3):237–252, 1976.
- [75] O. Zhao, M. Woodroffe, and D. Volný. A central limit theorem for reversible processes with nonlinear growth of variance. *J. Appl. Probab.*, 47(4):1195–1202, 2010.

Appendix A

The `robustTest` package: two-sample tests revisited

This chapter was submitted in 2022 to R journal. This is a collaborative work of 7 authors S. Ammous, O. Bouaziz, A. Dedecker, J. Dedecker, J. El Methni, M. Mellouk and F. Muri.

Contents

A.1 Introduction	108
A.2 Robust tests for testing correlation and independence	109
A.2.1 Pearson’s correlation test	109
A.2.2 Kendall’s correlation test	110
A.2.3 Spearman’s correlation test	111
A.2.4 Kolmogorov–Smirnov statistic for testing independence	112
A.2.5 Simulation study	112
A.3 Robust test for the equality of variances	115
A.3.1 Simulation study	117
A.4 Robust test for testing stochastic dominance	118
A.4.1 Simulations	119
A.5 Other tests of the <code>robustTest</code> package for paired two-sample	121
A.6 Testing correlation on a real data set	123

Abstract The R package `robustTest` offers corrected versions of several common tests in bivariate statistics. We point out the limitations of these tests in their classical versions, some of which are well known such as robustness or calibration problems, and provide simple alternatives that can be easily used instead. The classical tests and their robust alternatives are compared through a small simulation study. The latter emphasizes the superiority of robust versions of the test of interest. Finally, an illustration of correlation’s tests on a real data set is also provided.

A.1 Introduction

In this article, we consider several usual tests in bivariate statistics, which are taught in many scientific courses at various levels.

Our first goal is to illustrate, with the help of mathematical considerations and simulations, that many of these tests are either not very robust (i.e. they do not work outside the very strict framework in which they have been defined), or badly calibrated (i.e. we can find simple examples for which the null hypothesis is true, but the type I error rate is not the one announced).

Of course, we are not the only ones or the first ones to have noticed this, and in some cases valid solutions have been proposed (see for example the famous article by Welch [18] about the test of equality of two expectations in the case where the variances are unequal, the Welch test being (asymptotically) robust to non-normality).

Nevertheless, we felt it was important to return to these issues for at least two reasons:

- First, these tests are not only mathematical objects on which students can discover the basic principles of statistical analysis, they are also very often used in practice (for example in biomedical research articles). It is therefore important to come back to the limits of these tests, which are not always well indicated (especially in the case of so-called "non-parametric" tests).

- Secondly, because it is often very easy to modify these tests in order to make them more robust or (asymptotically) well calibrated. The modification presented is each time based on the calculation of the limiting variance of the test statistic under the null hypothesis. Renormalizing by an estimator of the standard deviation, we obtain a robust or asymptotically well-calibrated version thanks to the central limit theorem. Of course, the justification is asymptotic, but we will illustrate that, on simulated examples, it is always interesting to correct these tests, even for relatively small sample sizes. It seems to us that these modified tests, which are easy to describe and to implement, should be systematically pointed out to the students, and also to researchers from other disciplines unfamiliar with these issues.

Our second objective is to present the *robustTest* package, which implements robust modifications of the usual bivariate statistics tests. As we will see, the functions of the *robustTest* package, as well as their syntax, are very close to the functions of the *stats* package, so that regular users of the R software can use and compare them easily.

The article is organised as follows: in the first section, we present the robust versions of the Pearson (see [9]), Kendall (see [14]) and Spearman (see [12]) correlation tests. For completeness, we also provide in the same section the Kolmogorov-Smirnov type test of independence for continuous variables (see [4]). In the second section, we describe the test of equality of conditional variances of X knowing Y , when Y is a categorical variable with several levels. This procedure is based on the James-Welch ANOVA which we briefly recall (see [13] and [19]). In the third section, we present the robust version of the two samples Mann-Whitney test for stochastic dominance (see [16]). In the fourth section, we describes several ways to test the stochastic dominance in the case of paired two samples (see [17]). At the end of each subsection, we provide the R-functions of the *robustTest* package for each corrected test. Finally, in the last section, we illustrate on a real data set the functions of the *robustTest* package to test the correlation, and we compare the outputs to those of the usual tests.

A.2 Robust tests for testing correlation and independence

In the following four subsections we respectively focus on the correlation tests of Pearson, Kendall and Spearman, and on the independence test of Kolmogorov-Smirnov type.

A.2.1 Pearson's correlation test

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent and identically distributed (i.i.d) copies of the random pair (X, Y) , where X, Y are real-valued random variables. We suppose that X_i and Y_i have finite moments of order 2 and we denote $\mathbb{E}(X_i) = \mu_X$ and $\mathbb{E}(Y_i) = \mu_Y$. Our aim is to test

the null hypothesis $H_0 : \rho = 0$ against the alternative hypothesis $H_1 : \rho \neq 0$

where ρ is the correlation coefficient between X_i and Y_i . To this end we often use the Pearson test statistic (see for instance Fisher [9]):

$$T_n = \frac{\hat{\rho}_n}{\sqrt{\frac{1-\hat{\rho}_n^2}{n-2}}} \quad \text{where} \quad \hat{\rho}_n = \frac{\sum_{k=1}^n (X_k - \bar{X}_n)(Y_k - \bar{Y}_n)}{\sqrt{\sum_{k=1}^n (Y_k - \bar{Y}_n)^2} \sqrt{\sum_{k=1}^n (X_k - \bar{X}_n)^2}}$$

is the empirical correlation coefficient and where \bar{X}_n, \bar{Y}_n are the empirical means.

The interest of this statistic being that, in the case where the couple (X_i, Y_i) is Gaussian, its exact distribution under H_0 is known: it is the Student $St(n-2)$ -distribution. If the pair (X_i, Y_i) is not Gaussian, we can also easily show that, if X_i is independent of Y_i (and if X_i and Y_i are not constants), then T_n converges in distribution to the standard normal distribution $\mathcal{N}(0, 1)$. Therefore, if X_i and Y_i are independent, the rejection region $R_{n,\alpha} = \{|T_n| > c_\alpha\}$ where c_α is the quantile of order $1 - (\alpha/2)$ of the $\mathcal{N}(0, 1)$ distribution, is such that its probability tends to α as $n \rightarrow \infty$. But this is not the case in general under H_0 , which of course does not imply independence of the variables.

We can therefore make the following conclusion: in a general context, the Pearson test is not well calibrated to test the null hypothesis $\rho = 0$. See also [8] for a similar observation. Outside the strict framework of the linear model (Gaussian or not), it is therefore preferable to use the intuitive statistic

$$T'_n = \frac{\sum_{k=1}^n (X_k - \bar{X}_n)(Y_k - \bar{Y}_n)}{\sqrt{\sum_{k=1}^n (Z_k - \bar{Z}_n)^2}} \quad \text{where} \quad Z_i = (X_i - \bar{X}_n)(Y_i - \bar{Y}_n).$$

Under the assumption $0 < \mathbb{E}((X_i - \mu_X)^2(Y_i - \mu_Y)^2) < \infty$, a direct application of the central limit theorem and Slutsky's lemma provides the convergence in distribution of T'_n under H_0 to the standard normal distribution. The rejection region of H_0 is $R'_{n,\alpha} = \{|T'_n| > c_\alpha\}$ where c_α is the quantile of order $1 - (\alpha/2)$ of the $\mathcal{N}(0, 1)$ distribution, which provides a test asymptotically well calibrated.

Remark. It is easy to see that, in the Gaussian case and under H_0 , the distribution of the statistic T'_n does not depend on the expectation and variance parameters, and can therefore be tabulated. In practice, in the *robustTest* package, we use this table of quantiles for $n < 130$ and the quantiles of the Student $St(n-2)$ distribution for $n \geq 130$ (simulations show that the quantiles of the Student $St(n-2)$ distribution are close to those of the distribution of the statistic T'_n).

under H_0 into the Gaussian case).

The corresponding R function of the *robustTest* package is: `cortest(,method="pearson")`.

Remark. Starting from the quantity

$$\frac{\sum_{k=1}^n (X_k - \bar{X}_n)(Y_k - \bar{Y}_n) - n\text{Cov}(X, Y)}{\sqrt{\sum_{k=1}^n (Z_k - \bar{Z}_n)^2}},$$

which converges to the $\mathcal{N}(0, 1)$ distribution as $n \rightarrow \infty$, we can easily obtain a confidence interval for $\text{Cov}(X, Y)$, the covariance between X and Y . As an output of the command `cortest(,method="pearson")`, we propose instead a confidence interval for Pearson's ρ coefficient. This confidence interval is based on the central limit theorem for the empirical estimator of $(\text{Cov}(X, Y), \text{Var}(X), \text{Var}(Y))$ and the delta method applied to the function $h(x, y, z) = x/\sqrt{yz}$ from \mathbb{R}^3 to \mathbb{R} . There may be a difference between the result based on the robust Pearson test and whether or not 0 is in the confidence interval of ρ . In this case, the test should be preferred since the confidence interval based on the central limit theorem and the delta method is a priori less precise.

A.2.2 Kendall's correlation test

The context is the same as in the previous paragraph ($X_1, Y_1, \dots, (X_n, Y_n)$ are i.i.d. copies of the pair (X, Y) , where X, Y are real-valued random variables. We assume moreover that the variables are continuous. We want to know if X_i and Y_i tend to vary in the same direction or in the opposite direction. Let then consider or define

$$\tau = 2(\mathbb{P}((X_2 - X_1)(Y_2 - Y_1) > 0) - 0.5)$$

be Kendall's correlation coefficient [14]. As for Pearson's correlation coefficient, τ is between -1 and 1 . X_i and Y_i are positively correlated in the sense of Kendall when $\tau > 0$ and negatively if $\tau < 0$. To test

$$H_0 : \tau = 0 \quad \text{against} \quad H_1 : \tau \neq 0$$

Kendall [14] therefore proposed to count the number of concordant pairs (i.e. for which the product $(X_i - X_j)(Y_i - Y_j)$ is strictly positive), which leads to the statistic

$$T_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left(\mathbf{1}_{(X_i - X_j)(Y_i - Y_j) > 0} - 0.5 \right).$$

It is quite easy to see that, if X_i and Y_i are independent, then the distribution of T_n is distribution-free (i.e. does not depend on the distribution of (X_i, Y_i)), by reducing to two independent sequences U_1, \dots, U_n and V_1, \dots, V_n of i.i.d. random variables with standard uniform distribution $\mathcal{U}([0, 1])$. Consequently, if X_i and Y_i are independent, T_n is distributed according to a known and tabulated distribution. Kendall's test is constructed from the quantiles of this distribution. However if X_i and Y_i are not independent, the distribution of the statistic T_n has no reason to be distribution free under H_0 (it depends on the joint distribution of (X_i, Y_i)). We can therefore make the following conclusion: in a general context, Kendall's test is not well calibrated to test $\tau = 0$.

We can nevertheless solve (asymptotically) this problem, by considering the limiting distribution of $\sqrt{n}T_n$ under H_0 . Starting from the Hoeffding decomposition of the U -statistic T_n (see [11] or [17], example 12.5), we see that, under H_0 , $\sqrt{n}T_n$ converges in distribution to the $\mathcal{N}(0, V)$ distribution, with

$$V = 4\text{Var}(F(X_i, Y_i) + H(X_i, Y_i)),$$

where $F(x, y) = \mathbb{P}(X_i < x, Y_i < y)$ and $H(x, y) = \mathbb{P}(X_i > x, Y_i > y)$. The empirical estimator of V is then

$$V_n = \frac{4}{n-1} \sum_{k=1}^n (F_n(X_k, Y_k) + H_n(X_k, Y_k) - \bar{F}_n - \bar{H}_n)^2$$

where

$$\begin{aligned} F_n(x, y) &= \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k < x, Y_k < y}, & H_n(x, y) &= \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k > x, Y_k > y} \\ \bar{F}_n &= \frac{1}{n} \sum_{k=1}^n F_n(X_k, Y_k), & \bar{H}_n &= \frac{1}{n} \sum_{k=1}^n H_n(X_k, Y_k). \end{aligned}$$

Finally, under H_0 ,

$$K_n = \frac{\sqrt{n}T_n}{\sqrt{V_n}} \text{ converges in distribution to the } \mathcal{N}(0, 1) \text{ distribution.}$$

The rejection region of the corrected Kendall test is then $R_{n,\alpha} = \{|K_n| > c_\alpha\}$ where c_α is the quantile $1 - (\alpha/2)$ of the $\mathcal{N}(0, 1)$ distribution, which provides a test asymptotically well calibrated (see [1] for more details).

The corresponding R function of the *robustTest* package is: `cortest(,method="kendall")`.

A.2.3 Spearman's correlation test

In the same context as for the Kendall test statistic of the previous paragraph, we can also correct Spearman's correlation test (as described by Hotelling and Pabst [12]), which tests

$$H_0 : \rho_S = 0 \quad \text{against} \quad H_1 : \rho_S \neq 0$$

where ρ_S is the correlation coefficient between the variables $F_X(X_i)$ and $F_Y(Y_i)$ uniformly distributed over $[0, 1]$ (here $F_X(x) = \mathbb{P}(X \leq x)$ and $F_Y(y) = \mathbb{P}(Y \leq y)$). Like the Kendall and Pearson tests, this test is not well calibrated if the variables X_i and Y_i are not independent.

As with Kendall's test, Hoeffding [11] showed that Spearman's test statistic can be expressed using a U -statistic, from which he deduced a central limit theorem for the normalized statistic, with an exact expression of the limiting variance. As in the previous paragraph we can estimate this limit variance (by taking the empirical estimator), and then obtain a corrected test which is asymptotically well calibrated. The expression of the limit variance being more complicated than that of Kendall's statistic, we do not give all the details here.

The corresponding R function of the *robustTest* package is: `cortest(,method="spearman")`.

A.2.4 Kolmogorov–Smirnov statistic for testing independence

The context is the same as in the previous subsection: $(X_1, Y_1), \dots, (X_n, Y_n)$ are independent copies of a couple (X, Y) of real-valued random variables. We assume moreover that the variables X and Y are continuous. We want to test

$$H_0 : X \text{ and } Y \text{ are independent} \quad \text{against} \quad H_1 : X \text{ and } Y \text{ are not independent}$$

To answer this question, one can use the Kolmogorov-Smirnov statistic (as described for instance in Blum, Kiefer and Rosenblatt [4])

$$KS_n = \sup_{s, t \in \mathbb{R}} \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq t, Y_i \leq s} - F_{n,X}(t)F_{n,Y}(s) \right|,$$

where

$$F_{n,X}(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq t}, \quad F_{n,Y}(s) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{Y_i \leq s}.$$

It is clear that, under the hypothesis H_0 : X_i is independent of Y_i , the statistic KS_n is distribution-free (since the variables are continuous, one can go back to the case where X_i and Y_i are uniformly distributed over $[0, 1]$). On another hand, one can easily check that KS_n is asymptotically equivalent to

$$KS'_n = \sup_{s, t \in \mathbb{R}} \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n (\mathbf{1}_{X_i \leq t} - F_X(t))(\mathbf{1}_{Y_i \leq s} - F_Y(s)) \right|.$$

Now, as proved in [4], KS'_n (and hence KS_n) converges in distribution under H_0 to the supremum of a Gaussian process. The rejection region of the Kolmogorov-Smirnov test of independence is then $R_{n,\alpha} = \{KS_n > c_\alpha\}$ where c_α is the exact quantile of order $1 - \alpha$ of the distribution of KS_n under H_0 (note that, for small n , the quantity $\mathbb{P}_{H_0}(R_{n,\alpha})$ is in general not exactly equal to α , because the distribution of KS_n under H_0 is a discrete distribution). These quantiles (or the p -value of the test) can be easily estimated via a basic Monte-Carlo procedure.

The corresponding R function of the *robustest* package is: `indeptest()`.

Remark: When the continuous variables X_i and Y_i are observed with too rough a rounding, the statistics of Kendall, Spearman or Kolmogorov-Smirnov can behave badly (because they involve quantities of the type $\mathbf{1}_{X_i > Y_j}$). In the *robustest* package, we add the possibility to correct this problem by a simple randomization procedure (if there is a tie, we toss heads to see if the indicator is 0 or 1); to do so, it suffices to use the argument `ties.break="random"`. This remark is also valid for the Mann-Whitney test which will be presented in one of the following sections.

In order to highlight the differences between the classical tests and the corrected tests, we will consider in the next section two simulation scenarios.

A.2.5 Simulation study

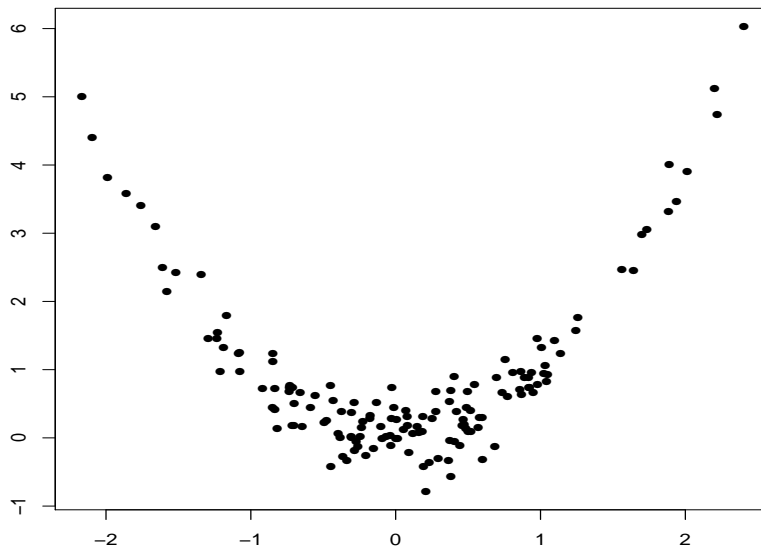
First scenario: we simulate, for different values of n , i.i.d. pairs $(X_i, Y_i)_{1 \leq i \leq n}$ according to the model

$$Y_i = X_i^2 + 0.3\varepsilon_i \tag{A.1}$$

n	30	40	50	60	70	80	90	100	200	300
usual P	0.361	0.376	0.365	0.369	0.38	0.353	0.366	0.362	0.374	0.384
robust P	0.055	0.047	0.048	0.052	0.055	0.05	0.049	0.049	0.051	0.054
usual K	0.174	0.184	0.183	0.201	0.201	0.202	0.187	0.194	0.203	0.194
robust K	0.077	0.067	0.069	0.066	0.064	0.057	0.057	0.052	0.052	0.051
usual S	0.123	0.124	0.132	0.135	0.145	0.131	0.131	0.126	0.139	0.137
robust S	0.113	0.096	0.089	0.082	0.078	0.071	0.068	0.061	0.057	0.056
KS indep	0.732	0.934	0.989	0.999	1	1	1	1	1	1

Table A.1: Frequencies of rejection of the seven tests at level 5% for model (A.1).

where the $(X_i)_{1 \leq i \leq n}$ and the $(\varepsilon_i)_{1 \leq i \leq n}$ are two independent sequences of i.i.d variables with $\mathcal{N}(0, 1)$ distribution (see Figure A.1). One can easily see that, for this model, $\rho = \tau = \rho_S = 0$. Hence X_i and Y_i are not correlated in the sense of Pearson, Kendall or Spearman; but of course, they are not independent.

Figure A.1: Scatter plot of 150 couples (x_i, y_i) drawn according to model (A.1)

We simulate $N = 3000$ samples of size n according to model (A.1), and for each test we indicate the frequency of rejection of H_0 at level 5%. We consider seven tests : the usual Pearson (usual P), Kendall (usual K) and Spearman (usual S), and the Robust Pearson (robust P), Kendall (robust K) and Spearman (robust S), to test $H_0 : \rho = 0$, $H_0 : \tau = 0$ and $H_0 : \rho_S = 0$ respectively; the Kolmogorov-Smirnov independence test (KS indep) to test $H_0 : X$ and Y are independent. The results are given in Table A.1 (level $\alpha = 5\%$).

From Table A.1, we can notice that the three usual correlation tests are poorly calibrated, with rejection frequencies of H_0 around 37% for usual P, 19% for usual K, and 13% for usual

S, instead of the expected 5%. We can also remark on this set of simulations, that it is always preferable to use robust tests, whose rejection frequencies are always lower than those of the usual tests, and close to 5% for $n \geq 100$. In fact, for robust P, the frequencies of rejection are always between 4.7% and 5.5% for $n \geq 30$; for robust K, the frequencies of rejection are always between 5% and 7% for $n \geq 40$; for robust S, the frequencies of rejection are always between 5% and 7% for $n \geq 90$. The fact that the rejection frequencies of robust S converge more slowly towards 5% can perhaps be explained by the fact that the asymptotic variance term of the Spearman statistic is quite complicated, and is therefore more difficult to estimate. To conclude, unsurprisingly, the independence test detects very well the non-independence of X and Y , systematically as soon as $n \geq 70$.

Second scenario: we simulate, for different values of n , i.i.d. pairs $(X_i, Y_i)_{1 \leq i \leq n}$ according to the model

$$Y_i = (X_i \cdot 2(\varepsilon_i - 0.5))^3 \quad (\text{A.2})$$

where $(X_i)_{1 \leq i \leq n}$ and $(\varepsilon_i)_{1 \leq i \leq n}$ are two independent sequences, the X_i 's being i.i.d. random variables with $\mathcal{U}([0, 1])$ distribution, the ε_i 's being i.i.d. random variables with a $\mathcal{B}(0.5)$ Bernoulli distribution (see Figure A.2). Here, again, one can easily see that $\rho = \tau = \rho_S = 0$. Hence X_i and Y_i are not correlated in the sense of Pearson, Kendall or Spearman; but of course, they are not independent.

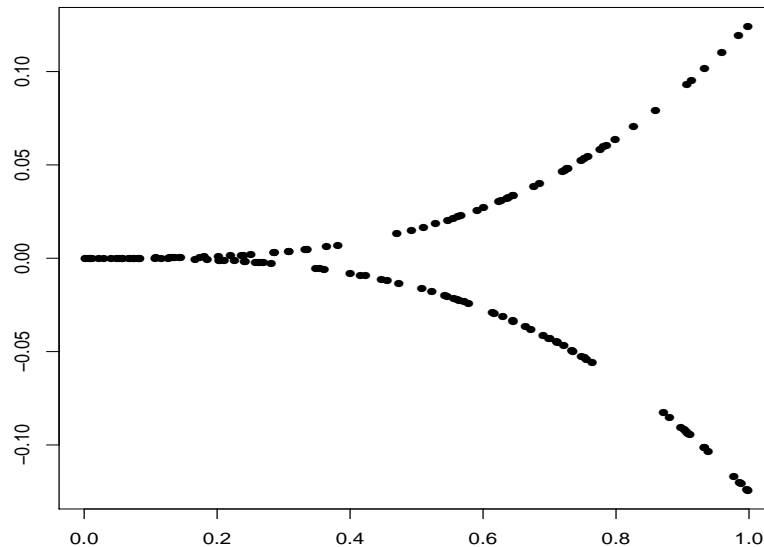


Figure A.2: Scatter plot of 150 couples (x_i, y_i) drawn according to model (A.2)

We simulate $N = 3000$ samples of size n according to model (A.2), and for each test we indicate the frequency of rejection of H_0 at level 5%. We consider the same seven tests as in the first scenario. The results are given in Table A.2 (level $\alpha = 5\%$).

From Table A.2 we can make much the same comments and observations as for scenario 1 (although models (A.1) and (A.2) are quite different). The correlation tests are all poorly cali-

n	30	40	50	60	70	80	90	100	200	300
usual P	0.162	0.159	0.158	0.154	0.148	0.146	0.154	0.153	0.141	0.142
robust P	0.072	0.064	0.06	0.057	0.056	0.058	0.058	0.05	0.049	0.052
usual K	0.231	0.236	0.236	0.242	0.249	0.248	0.255	0.247	0.245	0.246
robust K	0.082	0.074	0.063	0.06	0.06	0.059	0.06	0.059	0.054	0.05
usual S	0.138	0.143	0.133	0.132	0.142	0.139	0.148	0.14	0.142	0.135
robust S	0.113	0.099	0.085	0.079	0.071	0.071	0.07	0.068	0.059	0.056
indep	0.956	1	1	1	1	1	1	1	1	1

Table A.2: Frequencies of rejections of the seven tests at level 5% for model (A.2).

brated. Robust tests always have a lower rejection frequency than usual tests. In fact, for robust P, the frequencies of rejection are always between 5% and 6% for $n \geq 50$; for robust K, the frequencies of rejection are always between 5% and 7% for $n \geq 50$; for robust S, the frequencies of rejection are always between 5% and 7% for $n \geq 90$. Our last observation is that, unsurprisingly, the independence test detects very well the non-independence of X and Y , systematically as soon as $n \geq 40$.

Conclusion: We have seen, through two really different scenarios of simulations, that if we want to test non-correlation, it is preferable to use robust tests. Indeed the usual tests can lead to a false positive rate that is much too high. One could object that the usual correlation tests can be used to test the null hypothesis H_0 : X and Y are independent. But our simulation study reveal that the usual correlation tests are not consistent to test this hypothesis (their rejection frequency does not tend to 1 when n tends to infinity), whereas the Kolmogorov-Smirnov type test is consistent.

Remark: Concerning the particular case of Pearson's test of correlation: if we know that the couple (X, Y) is Gaussian, then the usual test will be well calibrated, and more powerful than the robust version. Note however that, to our knowledge, there is no practical way to test whether the pair (X, Y) is Gaussian (which is a much stronger assumption than assuming that X and Y are Gaussian).

A.3 Robust test for the equality of variances

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent and identically distributed (i.i.d) copies of the random pair (X, Y) , where X is real-valued random variable and Y is a categorical random variable with p different levels L_1, \dots, L_p . Assume that $\mathbb{E}(X^2) < \infty$, and let $\mu_i = \mathbb{E}(X|Y = L_i)$ and $V_i = \text{Var}(X|Y = L_i)$. The problem of testing the hypothesis $H_0 : \mu_1 = \dots = \mu_p$ is called ANOVA (Analysis of Variance) in the literature. If we do not assume that the variances V_i are all equal, this problem has been solved satisfactorily by James [13] and Welch [19]. The procedure proposed by James and Welch works for any sample size n when the conditional distribution of X knowing Y is Gaussian, but also without any additional assumptions when n is large enough (the test is asymptotically well calibrated). More precisely, denoting by JW_n the James-Welch test statistic, we know that : under H_0 , $(p-1)JW_n$ converges in distribution as $n \rightarrow \infty$ to a $\chi^2(p-1)$ distribution. For the ANOVA test of James-Welch (not assuming equal variances), the

corresponding R function of the *stats* package is: `oneway.test()`.

Consider now the problem of testing the equality of conditional variances, that is

$$H_0 : V_1 = \dots = V_p \quad \text{against} \quad H_1 : \exists i, j \text{ such that } V_i \neq V_j$$

We will need some notations in order to proceed. Denote by $L(Y_i)$ the level of the random variable Y_i , by $n(L_k)$ the number of variables Y_i for which $L(Y_i) = L_k$, and by

$$\bar{X}_{L_k} = \frac{1}{n(L_k)} \sum_{i=1}^n X_i \mathbf{1}_{Y_i=L_k}.$$

If $p = 2$, to test the equality of conditional variances, one often uses the Fisher statistic

$$F_{n(L_1),n(L_2)} = \frac{S_{n(L_1)}^2}{S_{n(L_2)}^2}, \quad \text{where} \quad S_{n(L_k)}^2 = \frac{1}{n(L_k) - 1} \sum_{i=1}^n (X_i - \bar{X}_{L_k})^2 \mathbf{1}_{Y_i=L_k}.$$

The interest of this statistic being that, in the case where the conditional distribution of X knowing Y is Gaussian, its exact conditional distribution given $(n(L_1), n(L_2))$ under H_0 is known: it is the Fisher distribution $F(n(L_1) - 1, n(L_2) - 1)$. Fisher's test for equality of variances is constructed from the quantiles of this distribution. But if the conditional distribution of X knowing Y is not Gaussian, the distribution of $F_{n(L_1),n(L_2)}$ under H_0 is not known. For $p \geq 2$, one can use Bartlett's test for equality of variances [3], but again this test (based on the Gaussian likelihood) is not robust to non normality. We can therefore make the following conclusion: in a general context, Fisher's and Bartlett's tests are not well calibrated to test the equality of variances. See also [5] for a similar observation.

An alternative to Bartlett's test often mentioned in the literature is Levene's test (modified by Brown and Forsythe [6]), valid for any $p \geq 2$. This test consists of performing a classic ANOVA on the variables $|X_i - \text{Med}_n(X|L(Y_i))|$, where $\text{Med}_n(X|L_k)$ is the empirical estimator of the conditional median of X given $Y = L_k$. It is quite easy to see that this procedure does not test the equality of conditional variances, but the equality of absolute deviations from the conditional medians (as it is clearly indicated in the R documentation of the package *lawstat*).

Assume now that the random variable X has a finite moment of order 4. In order to test

$$H_0 : V_1 = \dots = V_p \quad \text{against} \quad H_1 : \exists i, j \text{ such that } V_i \neq V_j$$

we propose to perform a James-Welch ANOVA on the variables $Z_i = (X_i - \bar{X}_{L(Y_i)})^2$. Let T_n be the test statistic that we obtain by proceeding in this way. It is then easy to deduce from the James-Welch test statistic that, under H_0 , $(p - 1)T_n$ converges in distribution as $n \rightarrow \infty$ to a $\chi^2(p - 1)$ distribution.

The corresponding R function of the *robustTest* package is: `vartest()`.

Remark: The procedure described above still works within the framework of the "fixed design" ANOVA, i.e. when we observe p independent sequences $(X_{i,1})_{1 \leq i \leq n_1}, \dots, (X_{i,p})_{1 \leq i \leq n_p}$ of i.i.d. variables (in this case $V_j = \text{Var}(X_{1,j})$), provided that all n_i are large enough.

In the next subsection, we will illustrate the differences between the classical test for the equality of variance and the corrected test by a small simulation study.

A.3.1 Simulation study

We simulate, for different values of n , i.i.d. pairs $(X_i, Y_i)_{1 \leq i \leq n}$ according to the following model:

$$Y_i \sim \mathcal{B}(2/3), \quad (\text{A.3})$$

where the conditional distribution of X_i given $Y_i = 0$ is the $\mathcal{N}(0, 1)$ distribution and the conditional distribution of X_i given $Y_i = 1$ is the $\chi^2(2)/2$ distribution. The boxplots of the conditional distributions of (x_i) given (y_i) based on 180 observations drawn according to model (A.3) are shown in Figure A.3.

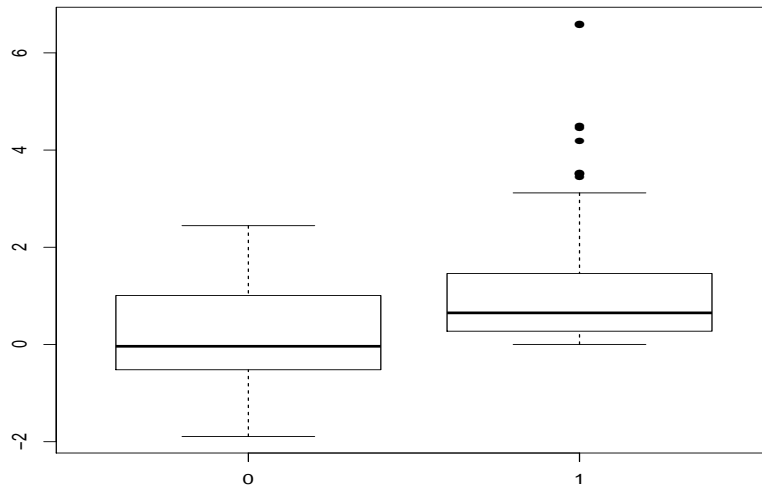


Figure A.3: Boxplots of the conditional distributions of (x_i) given (y_i) based on 180 observations drawn according to model (A.3).

Note that the conditional variances of X given $Y = 0$ or $Y = 1$ are equal; that is, with the notations above, $V_1 = V_2 = 1$ and the hypothesis $H_0 : V_1 = V_2$ is satisfied.

We simulate $N = 3000$ samples of size n according to model (A.3), and for each test we indicate the frequency of rejection of $H_0 : V_1 = V_2$ at level 5%. We consider four tests : Fisher's test, Bartlett's test, Levene's test, and our test based on the James-Welch procedure (VWelch). The results are given in Table A.3 (level $\alpha = 5\%$).

n	60	70	80	90	100	150	200	250	300
Fisher	0.145	0.151	0.144	0.154	0.152	0.153	0.161	0.16	0.161
Bartlett	0.142	0.147	0.143	0.153	0.152	0.153	0.161	0.16	0.161
Levene	0.08	0.083	0.091	0.097	0.106	0.119	0.155	0.172	0.196
VWelch	0.057	0.054	0.056	0.056	0.055	0.053	0.056	0.055	0.053

Table A.3: Frequency of rejection of the 4 tests at level 5% for model (A.3)

From Table A.3, our first observation is that, for this two-sample scenario, Fisher's test and

Bartlett's test behave similarly, with much too high rejection frequencies hovering around 15%. The frequencies of rejection of Levene's test increases with n , from 8% when $n = 60$ to 19.6% when $n = 300$. This is not surprising, because for model (A.3), the absolute deviations from the conditional medians are slightly different, depending on whether $Y = 0$ or $Y = 1$. The test VWelch, based on the James-Welch procedure is well calibrated, with frequencies of rejection between 5% and 6% as soon as $n \geq 60$.

Conclusion: We have seen that Fisher's and Bartlett's tests of equality of variances are not robust to non normality. Levene's test is well suited to test equality of absolute deviations from the median, but not to test the equality of variances. For testing equality of variances, it is then preferable to use the function `vartest()` of the *robustTest* package.

A.4 Robust test for testing stochastic dominance

Let (X_1, \dots, X_{n_1}) and (Y_1, \dots, Y_{n_2}) be two independent sequences of i.i.d. real-valued random variables. Assume moreover that the variables X_i and Y_i are continuous. We want to know if the variables Y_i tend to take larger or smaller values than the variables X_i . Let $\text{Med}(Y - X)$ be the median of $X_i - Y_i$. To answer the question, one can for example test the hypothesis

$$H_0 : \text{Med}(Y - X) = 0 \quad \text{against} \quad H_1 : \text{Med}(Y - X) \neq 0.$$

Mann and Whitney [16] proposed the test statistic

$$T_{n_1, n_2} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\mathbf{1}_{X_i < Y_j} - 0.5).$$

It is quite easy to see, that if X_i and Y_i have the same distribution, then the distribution of T_{n_1, n_2} is distribution-free (i.e. does not depend on the common distribution of X_i and Y_i): it suffices to reduce to two independent sequences U_1, \dots, U_{n_1} and V_1, \dots, V_{n_2} of i.i.d. random variables with $\mathcal{U}([0, 1])$ distributions. Consequently, if X_i and Y_i have the same distribution, T_{n_1, n_2} follows a known and tabulated distribution. The Mann-Whitney test is constructed from the quantiles of this distribution. It is well known that this test is in fact equivalent to Wilcoxon rank-sum test (see [20]). But if X_i and Y_i are not identically distributed, the distribution of statistics T_{n_1, n_2} has no reason to be distribution-free under H_0 . We can therefore make the following conclusion: in a general context, the Mann-Whitney test is not well calibrated to test $\text{Med}(Y - X) = 0$.

We can nevertheless solve (asymptotically) this problem, by considering the limiting distribution of $a_{n_1, n_2} T_{n_1, n_2}$ under H_0 , for a suitable normalisation a_{n_1, n_2} . Starting from Hoeffding's decomposition of the U -statistic T_{n_1, n_2} (see for instance [17], example 12.7), we see that, under H_0 ,

$$\frac{T_{n_1, n_2}}{\sqrt{\frac{V_1}{n_1} + \frac{V_2}{n_2}}} \quad \text{converges in distribution as } n_1, n_2 \rightarrow \infty \text{ to the } \mathcal{N}(0, 1) \text{ distribution,}$$

with

$$V_1 = \text{Var}(H_Y(X_1)) \quad \text{and} \quad V_2 = \text{Var}(F_X(Y_1)),$$

where $F_X(x) = \mathbb{P}(X_1 < x)$ and $H_Y(x) = \mathbb{P}(Y_1 > x)$. The empirical estimators of V_1 and V_2 are then

$$V_{1,n_1,n_2} = \frac{1}{n_1 - 1} \sum_{k=1}^{n_1} (H_{n_2}(X_k) - \bar{H}_{n_2})^2 \quad \text{and} \quad V_{2,n_1,n_2} = \frac{1}{n_2 - 1} \sum_{k=1}^{n_2} (F_{n_1}(Y_k) - \bar{F}_{n_1})^2$$

where

$$\begin{aligned} F_{n_1}(x) &= \frac{1}{n_1} \sum_{k=1}^{n_1} \mathbf{1}_{X_k < x} & H_{n_2}(y) &= \frac{1}{n_2} \sum_{k=1}^{n_2} \mathbf{1}_{Y_k > y} \\ \bar{F}_{n_1} &= \frac{1}{n_2} \sum_{k=1}^{n_2} F_{n_1}(Y_k) & \bar{H}_{n_2} &= \frac{1}{n_1} \sum_{k=1}^{n_1} H_{n_2}(X_k). \end{aligned}$$

Finally, under H_0 ,

$$MW_{n_1,n_2} = \frac{T_{n_1,n_2}}{\sqrt{\frac{V_{1,n_1,n_2}}{n_1} + \frac{V_{2,n_1,n_2}}{n_2}}}$$

converges in distribution as $n_1, n_2 \rightarrow \infty$ to the $\mathcal{N}(0, 1)$ distribution.

The rejection region of H_0 of the corrected Mann-Whitney test is therefore of the form $R_{n_1,n_2,\alpha} = \{|MW_{n_1,n_2}| > c_\alpha\}$ where c_α is the quantile of order $1 - (\alpha/2)$ of the $\mathcal{N}(0, 1)$ distribution, which provides a test asymptotically well calibrated.

The corresponding R function of the *robustTest* package is: `wilcoxtest(,paired=FALSE)`.

Remark: The procedure described above still works when both samples are obtained from n i.i.d random variables $(X_i, Y_i)_{1 \leq i \leq n}$, where the X_i 's are continuous random variables and the Y_i 's are categorical variables with two levels L_1, L_2 , considering the two sub-samples of continuous variables obtained by conditioning with respect to $Y_i = L_1$ or $Y_i = L_2$.

In the next subsection, we will illustrate the differences between the classical test for testing stochastic dominance and the corrected test by a small simulation study.

A.4.1 Simulations

We simulate $N = 3000$ samples of size n_1 according to the distribution $\mathcal{U}([-0.5, 0.5])$, and 3000 samples of size n_2 according to the distribution $\mathcal{N}(0, (0.04)^2)$ (with standard deviation equal to 0.04), with $n_2 = 3n_1$ (see Figure A.4). For each test, we indicate the frequency of rejection of H_0 at level 5%. We consider here 4 tests: the robust Mann-Whitney (robust M-W) test based on the MW_{n_1,n_2} statistic, the uncorrected Mann-Whitney (M-W) test (both for testing $H_0 : \text{Med}(Y - X) = 0$, which is satisfied here), Welch's test (which tests the equality of the expectations without assuming the equality of the variances, see the previous section), and the two-sample Kolmogorov-Smirnov (K-S) test for testing the equality of distributions (see for instance [7], pages 309-314). The results are given in Table A.4 (level $\alpha = 5\%$).

On this set of simulations (Table A.4), we see that, for the robust M-W test and all the proposed sample sizes, the rejection frequency of H_0 is below 7%. It is between 5% and 6% as soon as $n_1 \geq 50, n_2 \geq 150$. Welch's test is also well calibrated, with rejection frequencies

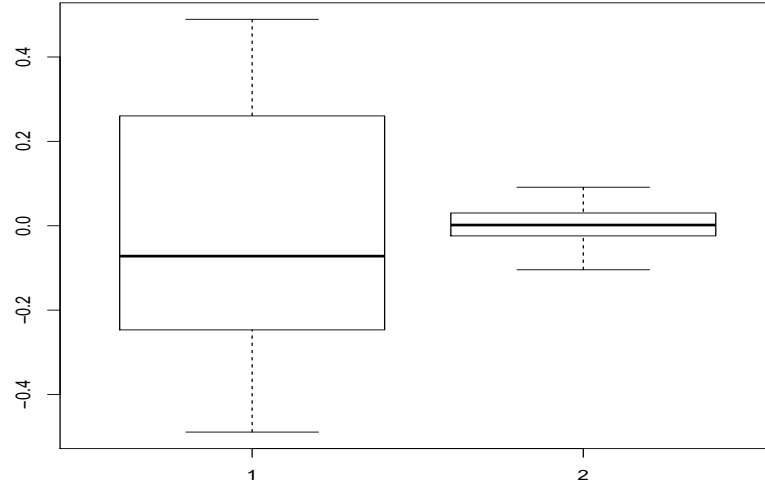


Figure A.4: Boxplots of 1: 60 observations drawn according to the distribution $\mathcal{U}([-0.5, 0.5])$, and 2: 180 observations drawn according to the distribution $\mathcal{N}(0, (0.04)^2)$.

$n_1; n_2$	20;60	30;90	40;120	50;150	60;180	70;210	80;240	90;270	100;300
robust M-W	0.061	0.064	0.057	0.058	0.054	0.056	0.055	0.054	0.051
M-W	0.172	0.171	0.18	0.166	0.17	0.186	0.168	0.179	0.161
Welch	0.057	0.048	0.051	0.057	0.49	0.57	0.053	0.051	0.051
K-S	0.998	1	1	1	1	1	1	1	1

Table A.4: Frequency of rejection of the four tests at level 5% for n_1 observations drawn according to the distribution $\mathcal{U}([-0.5, 0.5])$, and n_2 observations drawn according to the distribution $\mathcal{N}(0, (0.04)^2)$.

all between 4.8% and 5.7%. The uncorrected Mann-Whitney test is poorly calibrated, with rejection frequencies around 17%. As expected, the two-sample Kolmogorov-Smirnov test detects very well the difference between the two distributions, systematically as soon as $n_1 \geq 30, n_2 \geq 90$.

Conclusion If we want to test if $\text{Med}(Y - X) = 0$, it is a priori preferable to use the robust Mann-Whitney test, the usual test can lead to a false positive rate that is much too high. One could object that the usual Mann-Whitney test can be used to test the hypothesis H_0 : X and Y have the same distribution. But our simulations reveal that the usual Mann-Whitney test is not consistent to test this hypothesis (its rejection frequency does not tend to 1 when n tends to infinity), whereas the two-sample Kolmogorov-Smirnov type test is consistent.

We also investigated how to test stochastic dominance in the case of paired samples, which will be the subject of the next section.

A.5 Other tests of the *robustTest* package for paired two-sample

In the *robustTest* package, we also provide functions to test stochastic dominance in the case of paired samples. The context is as follows: let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent copies of the pair (X, Y) , where X, Y are real-valued random variables. We also assume that the variables are continuous. As in the previous section, we want to know if Y tends to take larger or smaller values than X . But here, the variables X and Y are a priori not independent. Classically, we consider the series of differences $D_i = Y_i - X_i$, and we are therefore reduced to a problem of univariate statistics. We can then use the confidence interval for the median $\text{Med}(D)$ described in Example 21.8 of the book by van der Vaart [17], and based on the order statistics $(D_{(i)})_{1 \leq i \leq n}$. Let us describe this confidence interval: for $\alpha \in (0, 1)$, let

$$k_{n,\alpha} = \left\lceil -c_\alpha \frac{\sqrt{n}}{2} + \frac{n}{2} \right\rceil, \quad \ell_{n,\alpha} = \left\lceil c_\alpha \frac{\sqrt{n}}{2} + \frac{n}{2} \right\rceil,$$

the square brackets denoting the integer part, and c_α being the quantile of order $1 - (\alpha/2)$ of the $\mathcal{N}(0, 1)$ distribution. Then, the interval

$$IC_{1-\alpha} = [D_{(k_{n,\alpha})}, D_{(\ell_{n,\alpha})}], \quad (\text{A.4})$$

is a confidence interval of $\text{Med}(D)$ with asymptotic confidence level $1 - \alpha$. As usual, a test of $H_0 : \text{Med}(D) = 0$ against $H_1 : \text{Med}(D) \neq 0$ can be deduced from (A.4). This test is an alternative to the sign test.

For these confidence interval and test, the R function of the package *robustTest* is `mediantest`. This function can be used for a single sample, as well as in the case of paired two-sample (in this case, the confidence interval and the test are for the quantity $\text{Med}(Y - X)$).

Note that we could have implemented the non-asymptotic confidence interval for $\text{Med}(D)$ (which is also described in Example 21.8 of [17]). However, some simulations suggest that, even for small samples ($25 \leq n \leq 50$), the asymptotic confidence interval behaves as well as the non asymptotic one.

As another alternative to the sign test, Wilcoxon's sign and rank test is often proposed, whose statistic is written

$$W_n = \sum_{i=1}^n R_i \mathbf{1}_{D_i > 0},$$

where R_i is the rank of $|D_i|$ in the sample $(|D_k|)_{1 \leq k \leq n}$. If the variable D is symmetric (which means that D has the same distribution as $-D$), Wilcoxon [20] showed that $|D|$ and $\mathbf{1}_{D > 0}$ are independent; we can then easily deduce that the distribution of W_n is distribution free (i.e. does not depend on the distribution of D). Therefore, if D is symmetric, the distribution of W_n is known and tabulated. Wilcoxon's sign and rank test is constructed from the quantiles of this distribution. But if the variable D is not symmetric, the distribution of W_n has no reason to be distribution free under $H_0 : \text{Med}(D) = 0$. We can therefore make the following conclusion: in a general context, Wilcoxon's sign and rank test is not correctly calibrated to test $\text{Med}(D)=0$. In fact, the sign and rank test allows to test

$$H'_0 : \text{Med}(D_1 + D_2) = 0 \quad \text{against} \quad H_1 : \text{Med}(D_1 + D_2) \neq 0.$$

But then again, it is not well calibrated to test this hypothesis. We can nevertheless solve (asymptotically) this problem, by considering the limiting distribution of $\sqrt{n}(2U_n/(n(n-1)) - 0.5)$

under H'_0 , with

$$U_n = \sum_{i=1}^n \sum_{j=1}^{i-1} \mathbf{1}_{D_i + D_j > 0}.$$

Note that

$$W_n = U_n + \sum_{i=1}^n \mathbf{1}_{D_i > 0},$$

the second term on the right in the equality being asymptotically negligible with respect to U_n . Starting from the Hoeffding decomposition of the U -statistic U_n (see for example [17], example 12.4), we see that the variable $\sqrt{n}(2U_n/(n(n-1)) - 0.5)$ converges in distribution under H'_0 to the distribution $\mathcal{N}(0, V)$, where

$$V = 4\text{Var}(F(-D))$$

and F is the distribution function of the variables D_i . The empirical estimator of V is therefore

$$V_n = \frac{1}{n-1} \sum_{i=1}^n (F_n(-D_i) - \bar{F}_n)^2 \quad \text{where } F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{D_i \leq x} \quad \text{and} \quad \bar{F}_n = \frac{1}{n} \sum_{i=1}^n F_n(-D_i).$$

Finally, under H'_0 ,

$$W'_n = \frac{\sqrt{n}}{\sqrt{V_n}} \left(\frac{2U_n}{n(n-1)} - 0.5 \right)$$

converges in distribution when $n \rightarrow \infty$ to the distribution $\mathcal{N}(0, 1)$.

The rejection region of H'_0 of the corrected sign and rank test is therefore of the form $R_{n,\alpha} = \{|W'_n| > c_\alpha\}$ where c_α is the quantile of order $1 - (\alpha/2)$ of the distribution $\mathcal{N}(0, 1)$, which provides a test asymptotically well calibrated.

For the robust Wilcoxon sign and rank test, the correspondind R function of the *robustTest* package is: `wilcoxtest(,paired=TRUE)`.

Remark. As we see, the Wilcoxon sign and rank test is not well suited to test $H_0 : \text{Med}(D) = 0$ (calibration and consistency issues) nor $H_0 : "D \text{ is symmetric}"$ (consistency issues). If one wants to test the symmetry, one can rather use the Kolmogorov-Smirnov type statistic:

$$K_n = \sqrt{n} \sup_{t \in \mathbb{R}} |F_n(t) - F_{n,-}(t)|,$$

where

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{D_i \leq t} \quad \text{and} \quad F_{n,-}(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{-D_i \leq t}.$$

One can easily check that the test statistic K_n is distribution-free under $H_0 : "D \text{ is symmetric}"$, and that it converges in distribution under H_0 to the supremum of a Gaussian process.

In the last section, we illustrate on a real data set the functions of the *robustTest* package to test the correlation, and we compare the outputs to those of the usual tests.

A.6 Testing correlation on a real data set

We will illustrate our results on the Evans data set which comes from the *lbref* package (see also [15]). These are data from a cohort study in which white males in Evans County were followed for 7 years, with coronary heart disease as the outcome of interest. The data set is a data frame with 609 rows and 9 variables. We focus on the three variables :

CDH : outcome variable; 1 = coronary disease

CHL : cholesterol, mg/dl

DBP : diastolic blood pressure, mmHg

In this example, we restrict ourselves to the sub-sample made up of the 71 men having coronary heart disease. We propose to test the correlation between the CHL and DBP variables in men affected by this disease. The scatter plot is drawn in Figure A.5.

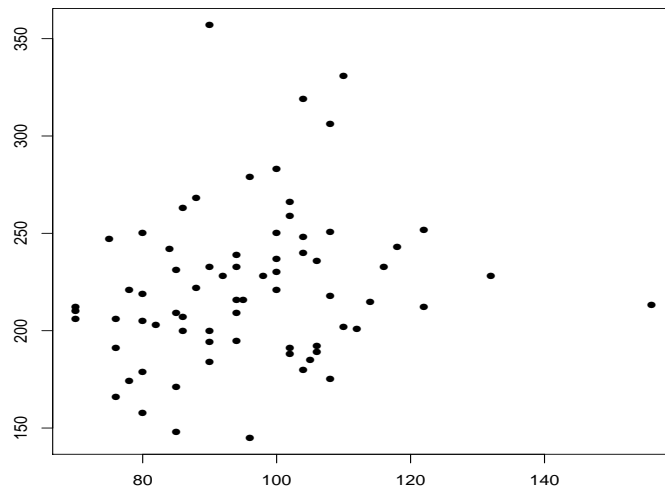


Figure A.5: Scatter plot of cholesterol as a function of diastolic blood pressure in patients with coronary artery disease

To test Pearson's correlation with the `cor.test` function of the package *stats*, type

```
cor.test(CHL[CDH==1],DBP[CDH==1])
```

We get the output

```
Pearson's product-moment correlation
```

```
data: CHL[CDH == 1] and DBP[CDH == 1]
```

```
t = 1.706, df = 69, p-value = 0.09251
```

```
alternative hypothesis: true correlation is not equal to 0
```

```
95 percent confidence interval:
```

```
-0.03370795  0.41500321
```


sample estimates:

cor

0.2011779

To test Pearson's correlation with the `cortest` function of the package *robustest*, type

```
cortest(CHL[CDH==1],DBP[CDH==1])
```

We get the output

Corrected Pearson correlation test

t = 2.4126, p-value = 0.0174

alternative hypothesis: true correlation is not equal to 0

95 percent asymptotic confidence interval for the correlation coefficient:

0.0331 0.3692

sample estimates:

cor

0.2011779

We can see that the results of the two tests are very different, with a p-value of 9.25% for the usual Pearson test, and 1.74% for the robust Pearson test. The robust test not requiring any hypothesis on the distribution of the pair of variables, it is a priori the one that must be retained. Note that Shapiro's test on the variables CHL and DBP each time gives a p-value lower than 1%, which implies that the observations do not come a priori from a pair of Gaussian variables. At the 5% risk level, we conclude that there is indeed a significant correlation in the sense of Pearson.

Let us now try to see the differences between the usual versions of the Kendall and Spearman tests and their corrected versions. Using the command `tiebreak` of the package *robustest*, we see that ties are present in both series of observations. To compare the usual tests and their corrected versions, we get rid of this problem by using

```
X=tiebreak(CHL[CDH==1])
```

```
Y=tiebreak(DBP[CDH==1])
```

To test Kendall's correlation with the `cor.test` function of the package *stats*, type

```
cor.test(X,Y, method="kendall")
```

We get the output

Kendall's rank correlation tau

data: X and Y

z = 2.0003, p-value = 0.04546

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

0.162173

To test Kendall's correlation with the `cortest` function of the package *robustTest*, type

```
cortest(X,Y, method="kendall")
```

We get the output

```
Corrected Kendall correlation test
```

```
t = 2.4122, p-value = 0.0159
alternative hypothesis: true tau is not equal to 0
95 percent confidence interval:
0.0304  0.2939
sample estimates:
tau
0.162173
```

We see a clear difference between the p-values of the two tests. That of the usual test is close to 5%, while that of the robust test is 1.6%. Note that these values remain globally stable when several attempts are made with the `tiebreak` function. The comments are exactly the same for the usual and robust Spearman tests (using `method="spearman"`): the p-value of the usual test is 4.45% while that of the robust test is 1.72%. Note the the p-values of the three robust tests (Pearson, Kendall, Spearman) are all below 2%.

Bibliography

- [1] S. Ammous (2020). Testing Kendall's τ for a large class of dependent sequences. *Statistics*, 54, 686–713.
- [2] F. J. Anscombe (1973). Graphs in Statistical Analysis. *The American Statistician*, 27, 17–21.
- [3] M. S. Bartlett (1937). Properties of sufficiency and statistical tests. *Proceeding of the Royal Statistical Society*, 160, 268–282.
- [4] J. R. Blum, J. Kiefer and M. Rosenblatt (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, 32, 485–498.
- [5] G. E. P. Box (1953). Non-Normality and Tests on Variances. *Biometrika*, 40, 318–335.
- [6] M. B. Brown and A. B. Forsythe (1974). Robust Tests for the equality of Variances. *Journal of the American Statistical Association*, 69, 363–367.
- [7] W. J. Conover (1971). Practical Nonparametric Statistics. *Wiley & Sons, New York*.
- [8] S. E. Edgell and S. M. Noon (1984). Effect of violation of normality on the t test of the correlation coefficient. *Psychological Bulletin*, 95, 576–583.
- [9] R. A. Fisher (1925). Applications of "Student" Distribution. *Metron*, 5, 90–104.
- [10] R. A. Fisher (1935). The design of experiments. *Oliver & Boyd, Oxford, England*.
- [11] W. Hoeffding (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19, 293–325.
- [12] H. Hotelling and M. R. Pabst (1936). Rank correlation and test of significance involving no assumptions of normality. *Ann. Math. Statistics*, 7, 29–43.
- [13] G. S. James (1951). The Comparison of Several Groups of Observations When the Ratios of the Population Variances are Unknown. *Biometrika*, 38, 324–329.
- [14] M. Kendall (1938). A New Measure of Rank Correlation. *Biometrika*, 30, 81–89.
- [15] D. G. Kleinbaum and M. Klein (2002). Logistic regression: A self-learning text. *Springer Edition*.
- [16] H. B. Mann and D. R. Whitney (2002). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, 18, 50–60.

- [17] A. W. van der Vaart (1998). Asymptotic statistics. *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge.
- [18] B. L. Welch (1947). The generalization of "Student" problem when several different population variances are involved. *Biometrika*, 34, 28–35.
- [19] B. L. Welch (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, 38, 330–336.
- [20] F. Wilcoxon (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1, 80–83.