

ÉCOLE NATIONALE DE LA STATISTIQUE
ET DE L'ANALYSE DE L'INFORMATION



STAGE DE FIN D'ÉTUDE
pour l'entreprise DataStorm

Estimation adaptive en analyse des données fonctionnelles

rédigé par
Hugo Brunet
Tuteur
Hassan Maissoro

Avril—Septembre 2023

Résumé

Les séries temporelles sont des données omniprésentes dans l'analyse et la prédiction de données. Elles concernent de nombreux secteurs critiques allant du secteur de l'énergie à la finance. Leur étude systématique depuis 1927 (Yule) est ainsi motivée par leur importance et utilité pour la mise en production.

Les données fonctionnelles quant à elles sont particulièrement présentes dans les données de capteurs ou à composante temporelle. Elles permettent grâce au point de vue qu'elles offrent, d'obtenir notamment de meilleures estimation sur le long terme que le point de vue réel multivarié classique. Cependant, la littérature jusqu'alors ne prenait pas en compte les différences de régularité des données traitées, ce qui pose problème pour des données peu régulières pourtant fréquemment observées.

Ce stage porte sur l'estimation de la régularité locale des trajectoires des séries temporelles de données fonctionnelles afin d'obtenir une meilleure estimation de leur fonction moyenne et de l'opérateur d'autocovariance. Plus spécifiquement, le stage consiste à étudier le comportement d'un hyper-paramètre utilisé lors de l'estimation de la régularité locale, et à proposer une méthode de sélection de ce dernier. Enfin cette méthode sera appliquée sur des données réelles du secteur énergétique.

avant-propos

Le lecteur saura excuser, si toutefois il se trouve déjà familier avec certaines notions (telle que la dépendance faible), de les voir réintroduites et ré-expliquées parfois de façon très détaillée (en annexe) car leur (ou plutôt ma) compréhension était importante pour le stage.

correctif



[ENSAI-stage_fin_etude-datastorm_fda_regularite-rapport/issues](#)

contact



mail étudiant : hugo.brunet@eleve.ensai.fr

Notations

Notation	Signification
Abréviations	
CDC	Courbe de Charge
FDC	Facteur de Charge
MPV	Maissoro - Patilea - Vimond
FDA	Analyse de Données Fonctionnelles (Functional Data Analysis)
FPCA	Analyse par Composantes Principales Fonctionnelle(Functional Principal Component Analysis)
KL	Karhunen-Loève
Analyse	
x_0	une valeur spécifique de x
$\mathcal{V}(x_0)$	un voisinage de x_0
$\mathcal{C}^0(E, F)$	fonction continue de E dans F
$\mathcal{H}_{\mathcal{V}(x_0)}(\alpha_{x_0}, L_{\alpha_{x_0}})$	Classe de Hölder de paramètre $\alpha_{x_0}, L_{\alpha_{x_0}}$ sur un voisinage de x_0
$\text{mfBm}(H, L)$	Ensemble des mouvements browniens multi-fractionnaires de fonction de Hurst $H : t \mapsto H_t$ et constante de Hölder locale $L : t \mapsto L_t$: ce sont les « paramètres de Hölder » des mouvements browniens considérés.
Algèbre	
$\text{sp}_{\mathbb{K}}(\varphi)$	valeurs propres d'un opérateur ou endomorphisme linéaire φ sur le corps \mathbb{K}
$\text{sp}(\varphi)$	valeurs propres d'un opérateur ou endomorphisme linéaire φ sous-entendu sur le corps \mathbb{R}
$\overrightarrow{\text{sp}}_{\ \cdot\ }(\varphi)$	vecteurs propres d'un opérateur ou endomorphisme linéaire φ formant une famille orthogonale
$\overrightarrow{\text{sp}}_{\ \cdot\ }^{[1,p]}(\varphi)$	p premiers vecteurs propres d'un opérateur ou endomorphisme linéaire φ formant une famille orthogonale
Statistique	
X	La « vraie » distribution
\tilde{X}	Quantité intangible/inobservable
\hat{X}	un estimation empirique X
$X_{(k)}$	Statistique d'ordre de X , k^{eme} terme : $X_{(k)} \leq X_{(k+1)}$
$Y_n^{(k)}$	Valeur observée au k^{eme} temps (au sens de la relation d'ordre \leq) sur le support du processus X_n : $Y_n^{(k)} \underset{\text{déf}}{\equiv} X_n(T_{(k)}) + \eta_n(T_{(k)})$
Probabilités	
$C_X(s, t)$	Covariance du processus X entre le temps s et le temps t
$c[f]$	opérateur de covariance évalué en f
$\text{VAR}[E]$	Variable aléatoire à valeur dans E : $\text{VAR}[E] \underset{\text{déf}}{\equiv} \mathbf{m}((\Omega, \mathcal{F}, \mathbb{P}), (E, \mathcal{A}, \mu))$

Notation	Signification
Spécifique au stage	
ξ	Bruit blanc gaussien provenant d'un mouvement Brownien multi-fractionnaire
η	bruit blanc gaussien provenant de l'erreur de mesure
M_n	Nombre de points observés sur la trajectoire de la donnée fonctionnelle X_n
B	nombre de burn-in pour atteindre la stationnarité du FAR
N	nombre de courbes observées
G	Nombre de points de la grille du calcul numérique d'une intégrale
λ	$\mathbb{E} [M_n] (M_n \sim \mathcal{P}(\lambda))$
$\hat{\lambda}$	Nombre moyen de points observés par courbe : $\hat{\lambda} = \frac{1}{N} \sum_{i=1}^N M_i$
Temps particuliers ($t \in \mathcal{T}$)	
\mathcal{T}	support du processus X , ici $\mathcal{T} = [0, 1]$
$T_n[m]$	m^{eme} temps (du m^{eme} sampling du phénomène aléatoire) observé sur la trajectoire de la donnée fonctionnelle X_n
$T_n^{(m)}$	m^{eme} temps (au sens de la relation d'ordre \leq) observé sur la trajectoire de la donnée fonctionnelle X_n
t_0 ou t_2	point où l'on souhaite estimer la régularité, t_0 est plus courant comme notation pour fixer un point mais t_2 est utilisé dans l'implémentation pour signaler sa centralité sur J_Δ
$J_\Delta(t_0)$	« voisinage » du point t_0 que l'on utilise pour estimer la régularité, implémenté comme l'intervalle $[t_1, t_3]$
$t_1(\Delta)$	$t_2 - \Delta/2$ lorsque Δ est quelconque, abbrégé en t_1
$t_3(\Delta)$	$t_2 + \Delta/2$ lorsque Δ est quelconque, abbrégé en t_3
g_k	point de la grille du calcul numérique d'une intégrale, $k \in \llbracket 1, G \rrbracket$
ϕ	Relation auto-régressive intégrale
β	Noyau de l'opérateur intégral
$\theta(u, v)$	$= \mathbb{E} [X(v) - X(u) ^2]$
Utilisés dans les algorithmes	
\mathbb{T}	Ensemble des points générés dans la simulation du mouvement brownien multi-fractionnaire : Points observés (aléatoire), Point de la grille d'approximation de l'intégrale de la relation FAR, points utilisés pour l'estimation de la régularité locale

Table des matières

1 Motivations	1
1.1 Difficulté de modélisation de la dynamique des données par un modèle de série temporelle classique	2
1.2 Les données fonctionnelles comme solution à cette difficulté	2
1.3 Importance de l'estimation de la régularité	4
1.4 Importance du choix du voisinage utilisé pour l'estimation de la régularité	5
1.5 Objectif du stage	6
2 Méthodologie	7
2.1 Données Fonctionnelles : l'essentiel	7
2.1.1 Définitions et propriétés informelles	7
2.1.2 Résumé de l'intérêt de la modélisation fonctionnelle	9
2.1.3 Cas non indépendant : séries temporelles de données fonctionnelles	10
2.2 Estimation de la régularité locale des trajectoires	12
2.2.1 Ce qu'on entend par régularité locale	12
2.2.2 Modèle considéré	13
2.2.3 Deux méthodes d'obtention de la régularité locale des trajectoires	13
2.2.4 Préissage	16
2.2.5 Résumé de la méthodologie d'estimation de la régularité locale	17
2.3 Simulation	18
2.3.1 Objectifs de la simulation	18
2.3.2 Mouvement Brownien Multi-Fractionnaire (mfBm)	18
2.3.3 génération d'un FAR(1)	19
2.4 Critère de sélection du Δ	20
3 Détermination du diamètre optimal des intervalles à considérer pour l'estimation de la régularité locale	21
3.1 Choix des paramètres de la simulation des FAR(1) localement Hölderiennes	21
3.1.1 Nombre de simulations	22
3.1.2 Fonction de Hurst	22
3.1.3 Constante de Hölder	22
3.1.4 Moyenne	23
3.1.5 Noyau de la relation FAR(1)	23
3.1.6 Nombre de courbes	23
3.1.7 Nombre moyen de points observés par courbe	24
3.1.8 Ensemble des Δ testés	24
3.1.9 Bruit blanc	24
3.1.10 Résumé des Paramètres	25
3.2 Préissage des données simulées	25
3.2.1 Les courbes obtenues	25
3.2.2 Pré-Lissage	25

3.3	Détermination du Δ optimal à choisir pour l'estimation de la régularité	26
3.4	Discussion	27
4	Application	29
4.1	Généralités	29
4.1.1	estimation adaptative informelle	29
4.1.2	Les estimateurs	30
4.2	Contrôle de la procédure sur des données simulées	31
4.2.1	Détermination du Δ en utilisant la procédure	31
4.2.2	Estimation de la fonction moyenne	32
4.2.3	Conclusion sur la qualité de la détermination du Δ via l'utilisation de la procé- dure	32
4.3	Application sur les données réelles de courbes de charge éolienne et photovol- taïque	33
4.3.1	Présentation des jeux de données	33
4.3.2	Pré-traitement des données	34
4.3.3	Pré-lissage et estimation de la régularité locale	34
4.3.4	Estimation de la fonction moyenne	34
4.4	Conclusion	34
A	Détails techniques et théoriques	35
A.1	Données fonctionnelles : formellement	35
A.2	Régularité Locale	38
A.2.1	Définition formelle des fonctions Höldériennes sur un intervalle & régularité lo- cale formellement	38
A.2.2	Des processus Höldériens ?	38
A.3	Dépendance Faible et LGN version faible	39
A.4	Continuité de Kolmogorov	43
A.5	Estimation adaptative	44
A.5.1	Estimation adaptative de la fonction moyenne	44
A.5.2	Estimation adaptative de l'opérateur de covariance	45
A.5.3	Estimation adaptative de l'auto-corrélation des séries temporelles fonctionnelles	46
A.6	Mouvement Brownien	46
A.6.1	Mouvement Brownien	46
A.6.2	Mouvement Brownien Fractionnaire	46
A.6.3	Mouvement Brownien multi-fractionnaire	46
A.7	Théorie de la base d'ondelettes	46
A.7.1	Transformée en ondelettes	46
A.7.2	Propriétés principales des ondelettes	47
A.7.3	Discussion : potentiel du lissage ondelette pour le prélassage	48
B	Plus de détails sur l'étude du Risque	49
B.1	Qualité de l'estimation des couples d'incrément utilisés dans l'estimation de la ré- gularité & choix du couple à estimer	49
B.1.1	Choix du Risque	51
B.2	Récapitulatif de la qualité des estimations pour la stratégie de « lissage global »	52
B.2.1	Récapitulatif concernant le risque quadratique sur l'estimation de la régularité	52
B.2.2	Récapitulatif concernant le risque quadratique sur les θ individuellement	52
B.2.3	Récapitulatif concernant le risque euclidien pour les couple Θ	53
B.2.4	Conclusion des tables	53

B.3	Etude de l'impact de la méthode de sélection de la fenêtre de pré-lissage sur le risque d'estimation des couples Θ	53
B.3.1	Rappel de la méthodologie utilisée	53
B.3.2	Le cas sparse	53
B.3.3	Densité de points sur les courbes observées	55
B.3.4	Peut-on considérer que tous les Δ conviennent ?	55
B.3.5	couple $\Theta : 1 \rightarrow 3 / 1 \rightarrow 2$	57
B.3.6	couple $\Theta : 1 \rightarrow 3 / 2 \rightarrow 3$	58
B.4	Etude de l'impact de la méthode de prélassage sur l'estimation de la régularité et le Δ optimal	66
B.4.1	Ondelettes	67
C	Un peu d'Histoire	69
C.1	Histoire des séries temporelles	69
C.2	Histoire des données fonctionnelles	70
C.3	Histoire du mouvement brownien et de ses applications	71
D	Algorithmes & Implémentations	72
D.1	Algorithmes de simulation	72

Table des figures

1.1	Courbes de charges éoliennes sur 3 premiers parcs éoliens	1
1.2	Différence entre donnée fonctionnelle et donnée réelle	3
1.3	Comparaison entre une courbe \mathcal{C}^2 et une courbe non dérivable	5
2.1	Donnée fonctionnelle : relation fonctionnelle avec plusieurs paramètres	8
2.2	Résumé des motivations du de l'estimation de la régularité locale des trajectoires	9
2.3	Schéma grossièrement récapitulatif : Estimation de la régularité pour une série temporelle fonctionnelle	11
2.4	Exemple de courbe dont on souhaiterait déterminer la régularité locale, et visualisation de J_Δ : estimation intérieure / au bord	14
2.5	Schéma résumé de la méthode d'estimation de la régularité	17
2.6	Schéma représentant les différentes approximations du couple d'incrément	20
3.1	Hurst Function : Logistic	22
3.2	Graphique du noyau intégral pour la relation FAR(1)	23
3.3	Visualisation des données générées : $\lambda = 255, N = 200$, 30 valeurs de Δ	25
A.1	Schéma du découpage du contrôle des erreurs	44
A.2	Schéma du découpage du contrôle des erreurs	45
B.1	Schéma représentant les différentes approximations du couple d'incrément	51
B.2	Distribution des risques et aperçu d'une courbe pour un échantillon de monte carlo extrême sur le risque euclidien.	60
B.3	Densité de points observés sur $[0.65, 0.95]$ pour $\lambda = 60$ sur un échantillon de monte carlo extrême, en un Δ problématique.	61
B.4	Densité des points observés correspondant à la courbe présentée sur la figure B.2.	61
B.5	Risque Euclidien pour $N = 200, \lambda = 210$ en un point régulier selon la méthode utilisée pour la fenêtre de lissage	62
B.6	Graphe des risques euclidiens dans les cas « sparse » et « raisonnablement dense », ayant enlevé les observations extrêmes ($\leq 20\%$ des échantillons de monte carlo)	64
B.7	Graphe des risques euclidiens relatifs à $\tilde{\Theta}(\Delta)$ dans les cas « sparse » et « raisonnablement dense », ayant enlevé les observations extrêmes ($\leq 20\%$ des échantillons de monte carlo)	65
B.8	Transformée de Fourier à court terme d'une fonction	68

Liste des tableaux

2.1	Tableau récapitulatif du modèle considéré	13
3.1	Hyper-paramètres de la simulation Monte-Carlo	25
4.1	Paramètres de simulation des données de test	31
B.1	Tableau récapitulatif des Δ optimaux : Risque sur H_t	52
	Tables récapitulatives de l'analyse des risques euclidiens non relatifs	52
B.2	Tableau récapitulatif des Θ optimaux : Risque individuel sur $\tilde{\theta}(u, v)$	52
B.3	Tableau récapitulatif des Δ optimaux : Risque euclidien sur $\tilde{\Theta}$ fenêtre de préissage globale pour $\lambda \geq 120$	53
	Différences entre la méthode de lissage « individuelle » et « globale »	53
B.4	Ordre de grandeur des différences entre le risque euclidien minimum et maximum pour $\Delta \in [0.01, 0.2]$ et la norme de la cible	56
	Comparaison des ordres de grandeur de la norme de la quantité ciblée $\tilde{\Theta}$ et de différence entre le risque optimal et maximal.	56
B.5	Comparaison de la médiane et de la variance du Risque entre lissage global et lissage individuel	57
	Comparaison de la distribution des risques entre la méthode de lissage « individuelle » et « globale »	57
B.6	Comparaison de la médiane et de la variance du Risque entre lissage global et lissage individuel	58
B.7	Quelques statistiques sur la distribution du risque euclidien en fonction de la méthode de sélection de la fenêtre de lissage	59
	Données aberrantes dans les échantillons de monte carlo pour le risque euclidien (non relatif)	59

List of Algorithms

1	Get_single_mc_sim : Génération de FAR pour chaque valeur sur la grille	72
2	Simulation de Monte Carlo	73
3	far_sim : Simulation d'un FAR(1)	74

Chapitre 1

Motivations

Contents

1.1	Difficulté de modélisation de la dynamique des données par un modèle de série temporelle classique	2
1.2	Les données fonctionnelles comme solution à cette difficulté	2
1.3	Importance de l'estimation de la régularité	4
1.4	Importance du choix du voisinage utilisé pour l'estimation de la régularité . .	5
1.5	Objectif du stage	6

Dans le cadre de ce stage, les données que l'on traite sont des données du secteur de l'énergie, et plus particulièrement des données de production électrique. On dispose ainsi de plusieurs éoliennes identifiées par le tag "id_(identifiant de l'éolienne)" dont l'énergie produite est mesurée toutes les deux heures, et ce pendant 4 ans (de 2014 à 2017). Cette énergie produite est dénommée la courbe de charge (que l'on abrégera par **FDC** par la suite). Il est cependant plus utile de s'intéresser au facteur de charge (ou **FDC**) qui est défini comme

Courbe de Charge
Facteur de Charge = $\frac{\text{Courbe de Charge}}{\text{Puissance Installée}}$.

On en déduit que **FDC** doit nécessairement être compris entre 0 et 1. C'est entre autre aussi une manière de détecter des anomalies et données atypiques comme la surproduction d'énergie par rapport à ce qui était attendu de la part d'un parc éolien ou encore un défaut de capteur (tension / intensité, ...) qui mesure la courbe de charge.

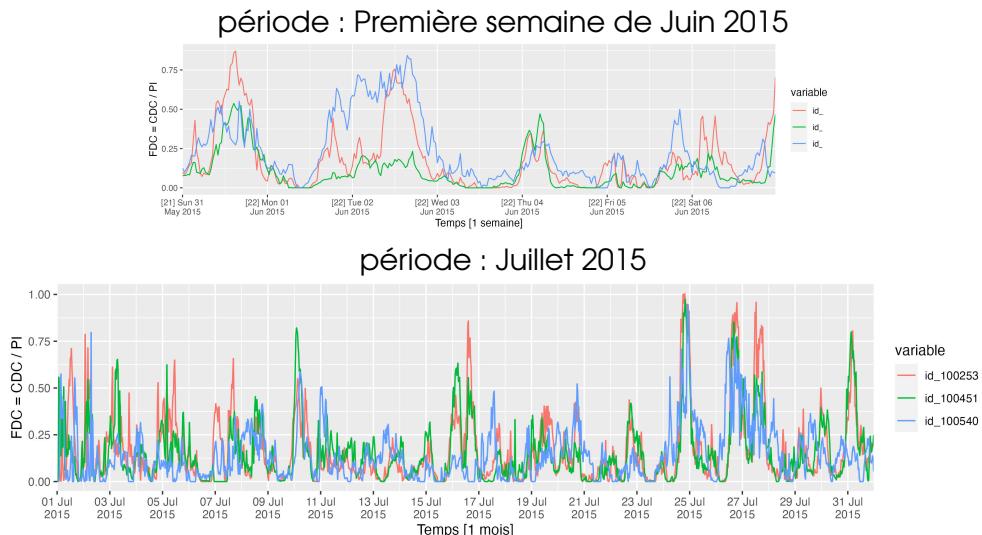


FIGURE 1.1 – Courbes de charges éoliennes sur 3 premiers parcs éoliens

Ainsi, les données qui sont traitées dans le cadre de ce stage sont, entre autres, des courbes de charge éoliennes observées chaque demie-heure. Le schéma d'observation est donc le « common-design ». C'est-à-dire que les temps d'observations sont ici déterministes à intervalle de temps fixe.

1.1 Difficulté de modélisation de la dynamique des données par un modèle de série temporelle classique

Une première idée serait d'utiliser un modèle de série temporelle ARIMA afin de modéliser la dynamique des courbes de charge. Bien que de nombreux outils aient été développés pour les séries temporelles¹, ces modèles présentent des limites en termes de prédiction à long terme, les rendant moins utiles lorsque l'objectif est de prédire à moyen ou long terme. De plus, ils partagent avec la plus part des modèles de machine learning populaires le fait d'estimer les données courbe par courbe ce qui ne tire pas profit du fait que les observations aient une forme similaire entre les courbes.

Même si naturelle, l'utilisation d'un modèle ARIMA ne permet de modéliser la dynamique du phénomène que l'on s'est donné à étudier. En effet, la sélection d'un modèle ARIMA sur le critère du BIC résultait, peu importe le parc éolien, en un modèle auto-régressif d'ordre 0. Ainsi le modèle sélectionné considérait les irrégularités de la courbe de charge, dont on attend que le processus duquel elle est issue soit irrégulier (de par sa complexité), comme étant du bruit. On en conclut que ces modèles peuvent ne pas capturer efficacement la structure complexe des données.

Afin de mieux modéliser nos données, et ainsi prédire sur le long terme, nous allons donc adopter une approche basée sur les données fonctionnelles pour capturer la structure de la courbe de charge. Cette approche permettra de d'exploiter une information clé : la similarité entre les courbes observées.

1.2 Les données fonctionnelles comme solution à cette difficulté



Qu'est-ce qu'une donnée fonctionnelle ?

Une donnée est dite fonctionnelle lorsque la variable aléatoire qui nous intéresse n'est plus une variable aléatoire à valeur dans \mathbb{R}^d , comme le statisticien a l'habitude de manipuler, mais une variable aléatoire à valeur dans un espace de fonction. Concrètement, chaque réalisation n'est plus un nombre mais bien une courbe toute entière indexée (le plus souvent) sur un intervalle \mathcal{T} .

1. cf annexe C sur l'histoire des séries temporelles

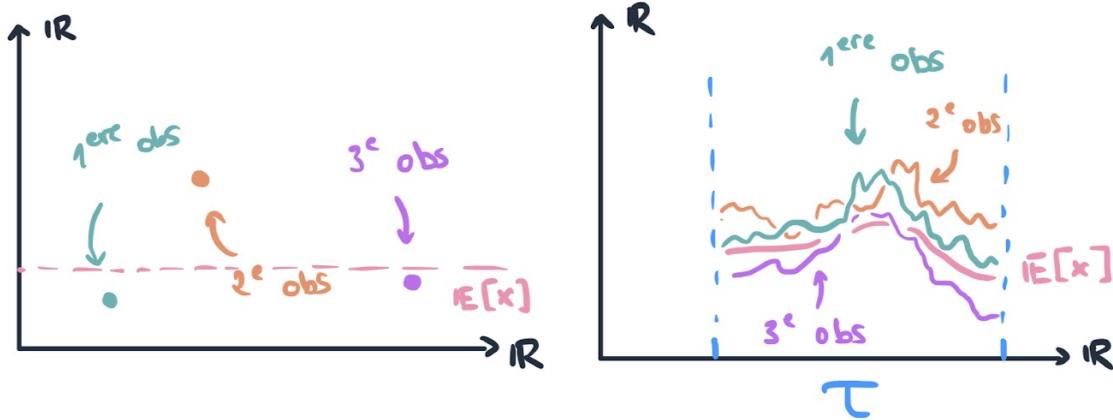


FIGURE 1.2 – Différence entre donnée fonctionnelle et donnée réelle

Si le statisticien est déjà à l'aise avec l'idée qu'une variable aléatoire réelle identiquement distribué puisse modéliser une expérience répétable provenant d'un même phénomène, il pourra se convaincre que les données fonctionnelles permettent elles aussi de modéliser des expériences en lien (fonctionnel) avec un certain paramètre. Et c'est le lien entre les deux valeurs, cette fois-ci, qui provient d'un même phénomène.

Donnons en un exemple : observons la consommation électrique d'un foyer dans une journée. Lorsque l'on travaille sur \mathbb{R} , on s'intéresse à sa consommation électrique disons en l'instant $t = 12h$. Formellement :

$$\mathcal{T} \underset{\text{déf}}{\equiv} [0, 24[= 1 \text{ jour avec } t \text{ en heure}$$

La consommation du foyer i à midi, notée y_i , suit la loi d'un phénomène général Y , comme une loi normale $\mathcal{N}(0.27 \text{ kWh}, 0.1^2)$ ² par exemple. Travailler sur des données fonctionnelles dans ce cadre c'est étudier non plus la consommation y_i à midi, mais regarder l'ensemble de sa consommation en même temps sur toute la journée $y_i(t) = x_i(t)$ avec $t \in \mathcal{T}$.

On remarque ainsi que toutes les consommations électriques le long de la journée d'un foyer à l'autre suivent la même tendance : on consomme plus le matin avant le travail et le soir alors que pendant la journée on consomme moins car on est au travail. Ainsi c'est la fonction $x_i : \mathcal{T} \rightarrow \mathbb{R}$ qui suit la loi d'un phénomène X général. Ce que l'on vient de dire c'est que la **relation** entre le temps $t \in \mathcal{T}$ et la consommation électrique $y_i(t)$ est elle-même sujet à une loi plus générale. Grossièrement, les courbes auront la même allure, mais chaque individu a sa consommation propre.

Plus formellement : comme on a défini une variable aléatoire réelle comme une application :

$$\begin{aligned} \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto x = X(\omega) \end{aligned}$$

On définit de même une donnée fonctionnelle comme une application :

2. ordre de grandeur de la consommation électrique d'un foyer en France calculé à partir des données d'ENGIE disponibles librement (7). **La variance est arbitraire**, tout comme le choix de la loi juste afin de servir d'exemple

$$\begin{aligned}\Omega &\longrightarrow \mathcal{C}^0(\mathcal{T}, \mathbb{R}) \\ \omega &\longmapsto x = X(\omega)\end{aligned}$$

Ce que l'on observe sont donc les valeurs des paramètres $t \in \mathcal{T}$ ainsi que l'image de t par $x : y = x(t)$. Les points que le statisticien observe sont donc les couples de la forme $(t_k^{(\text{individu } i)}, y_k^{(\text{individu } i)})_{i \in [1, m]}$, générés par le processus aléatoire X dont la réalisation est la véritable courbe x_i de l'individu i que l'on souhaite estimer pour travailler avec.



Certaines ressources sur l'analyse de données fonctionnelles définissent les données fonctionnelles de la manière suivante :

$$\begin{aligned}\Omega \times \mathcal{T} &\longrightarrow \mathbb{R} \\ (\omega, t) &\longmapsto X(\omega, t) = y\end{aligned}$$

Cependant, cette représentation ne permet pas une interprétation clé en main du concept mais est certainement plus commode à manipuler pour les mathématiciens.

Maintenant que l'on possède une meilleure intuition de ce que sont les données fonctionnelles, il est naturel de se demander pourquoi le choix de modéliser notre phénomène par des données fonctionnelles serait particulièrement judicieux. Pour cela, rappelons nous les difficultés que l'on avait rencontrées dans le cadre de nos données de production électrique en utilisant un modèle de série temporelle classique :



Rappel :

"[...] le modèle (arima) sélectionné considérait les irrégularités comme étant du bruit [...] Afin de prédire sur le long terme, nous allons donc adopter une approche basée sur les données fonctionnelles pour capturer la structure de la courbe de charge [...]"



Pourquoi est-ce que l'on s'intéresse autant à la régularité des données que l'on étudie ici ? Et surtout, en quoi est ce que les données fonctionnelles vont nous permettre de mieux capturer la régularité ?

1.3 Importance de l'estimation de la régularité

Comme mentionné auparavant, la production électrique est un phénomène très irrégulier (figure 1.1) étant influencé par la consommation, la météo, etc. Par conséquent, la prévision de ces courbes de charge doit prendre en compte la nature fondamentalement irrégulière du phénomène afin de proprement le modéliser et, en définitive, mieux le prédire. Ce qui est notamment contraire à des nombreux modèles populaires parmi les statisticiens qui utilisent des fonctions de classe \mathcal{C}^2 pour lisser les points observés en données fonctionnelles, ce qui limite la prédiction à des courbes de nature \mathcal{C}^2 . Cela est d'autant plus critique lorsque l'on cherche à estimer le processus moyen ou l'opérateur de covariance du processus, car ces derniers sont estimés à partir des courbes lissées, qui détruisent toute l'information irrégulière si elle n'est pas prise en compte, impactant significativement l'estimation des objets qui nous intéressent en tant que statisticien.

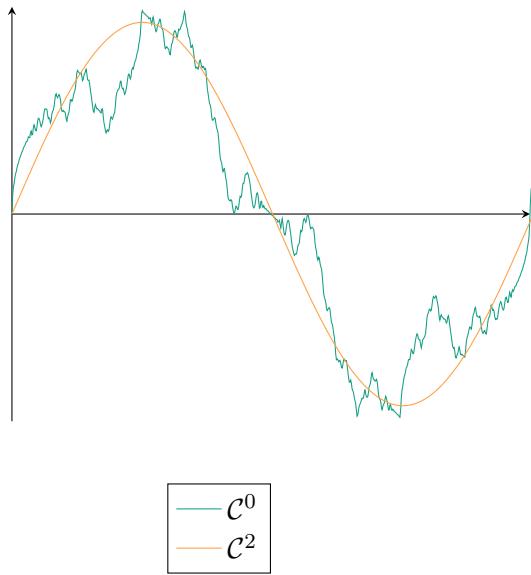


FIGURE 1.3 – Comparaison entre une courbe C^2 et une courbe non dérivable

Il est ainsi important pour des phénomènes de nature irrégulière de ne pas négliger des précautions lors du lissage afin de ne pas perdre l'information irrégulière. L'idée est donc d'estimer dans un premier temps la régularité de notre processus afin de lisser nos données de manière adaptée pour débruiter et prédire des valeurs non observées tout en préservant les informations irrégulières critiques pour la bonne estimation du processus moyen et de l'opérateur de covariance. L'approche fonctionnelle est clé dans l'estimation de cette régularité, car c'est la **réPLICATION DE COURBES** de même nature qui permet in-fine d'**ESTIMER LA RÉGULARITÉ** du phénomène, et il est donc important de bien savoir l'estimer.

1.4 Importance du choix du voisinage utilisé pour l'estimation de la régularité



L'estimation de la régularité des trajectoires est certes importante mais comment l'estime-t-on en pratique ?

L'étude de la convergence des estimateurs des paramètres de régularité locale a été établie par Golovkine et Maisoro-Patilea-Vimond³ (9, 17). En supposant la dépendance dans les données suffisamment faible, il est possible d'estimer la régularité locale du processus ponctuellement en utilisant les informations d'un voisinage arbitrairement donné. Cependant bien que l'estimateur soit convergent en utilisant un voisinage quelconque, il n'est pas spécifié de quelle taille devrait être ce voisinage pour avoir une bonne estimation des paramètres de régularité locale⁴. On appelle le diamètre du voisinage que l'on considère pour effectuer les calculs Δ .



Si la convergence des estimateurs est déjà déterminée pour un Δ donné arbitraire, pourquoi ne pas simplement en prendre un de façon arbitraire ?

3. qui seront désormais mentionnés par « MPV »

4. On entend par bonne estimation une estimation qui comporte les caractéristiques suivantes : une bonne vitesse de convergence, un compromis biais-variance adapté à l'application souhaitée de notre estimateur

Choisir un diamètre de voisinage non approprié mènerait à utiliser des informations non pertinentes pour estimer la régularité car celle-ci peut être variable sur l'ensemble de la trajectoire. De plus, il est naturel de penser que différents niveaux de régularités requièrent de regarder des informations d'une proximité différente.⁵ On introduirait alors un biais significatif dans l'estimation des paramètres de régularité locale, dont on a vu qu'il était important de bien estimer.

1.5 Objectif du stage

Choisir le bon diamètre du voisinage (Δ) que l'on considère pour estimer la régularité locale est donc un problème important, et c'est ce que l'on va étudier lors de ce stage. L'objectif est d'obtenir une procédure de détermination du Δ que le praticien devra choisir pour l'estimation de la régularité locale en fonction de quantités facilement estimables, comme nombre moyen de points observés par courbe par exemple.

Pour cela, on simulera 200 réplications indépendantes de monte-carlo de séries temporelles fonctionnelles FAR(1) dont les courbes sont des mouvements browniens multi-fractionnaires de régularité variable connue. Les estimateurs de régularité fournis par MPV (17) seront ensuite utilisés pour estimer la régularité (connue) de ces courbes. La procédure de sélection du Δ sera alors déterminée en s'appuyant sur l'analyse du comportement d'un risque d'estimation de la régularité en fonction du Δ choisi. Enfin la procédure déterminée sera testée sur les données simulées avant d'être appliquée sur des données réelles pour estimer de façon adaptative la fonction moyenne, l'auto-covariance et effectuer une analyse factorielle des données.

5. Considérer $|x - x_0| = \Delta$ dans la définition de la régularité que l'on considère en A.2.1

Chapitre 2

Méthodologie

Contents

2.1	Données Fonctionnelles : l'essentiel	7
2.1.1	Définitions et propriétés informelles	7
2.1.2	Résumé de l'intérêt de la modélisation fonctionnelle	9
2.1.3	Cas non indépendant : séries temporelles de données fonctionnelles	10
2.2	Estimation de la régularité locale des trajectoires	12
2.2.1	Ce qu'on entend par régularité locale	12
2.2.2	Modèle considéré	13
2.2.3	Deux méthodes d'obtention de la régularité locale des trajectoires	13
2.2.4	Prélissage	16
2.2.5	Résumé de la méthodologie d'estimation de la régularité locale	17
2.3	Simulation	18
2.3.1	Objectifs de la simulation	18
2.3.2	Mouvement Brownien Multi-Fractionnaire (mfBm)	18
2.3.3	génération d'un FAR(1)	19
2.4	Critère de sélection du Δ	20

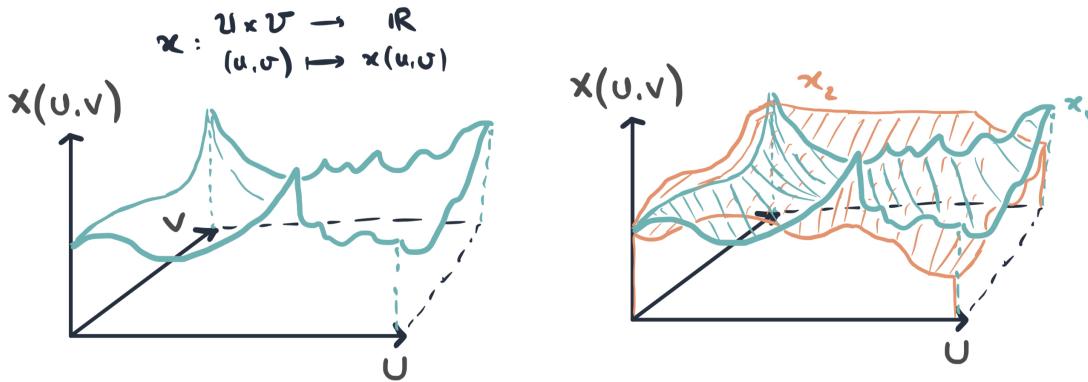
2.1 Données Fonctionnelles : l'essentiel

2.1.1 Définitions et propriétés informelles

Nous allons dans cette section introduire la notion de donnée fonctionnelle ainsi que les propriétés les plus utiles lorsqu'on les manipule. On y regroupe l'ensemble des messages essentiels à retenir des données fonctionnelles pour la pratique, sans alourdir les notions avec des notations mathématiques. Le cadre formel est traité en annexe A.1.

Définition (données fonctionnelles — informel) Les données fonctionnelles sont des données dont les observations sont des fonctions, c'est-à-dire des courbes, des surfaces, des images, ...

i.e : toute donnée ayant une dépendance de type "relation fonctionnelle" avec un ou plusieurs paramètres.



Gauche : exemple de surface

Droite : échantillon de deux observations de la surface suivant une loi fonctionnelle

FIGURE 2.1 – Donnée fonctionnelle : relation fonctionnelle avec plusieurs paramètres

Maintenant introduites, les théorèmes suivant permettent de manipuler ces données à la fois pour la théorie et la pratique :

Théorème (Karhunen-Loeve — informel)

Il est possible pour une large classe de données fonctionnelles de les décomposer dans une base de fonctions adaptée aux données (au sens de la covariance) que l'on appelle base ACP fonctionnelle (FPCA).

Remarque : La classe de fonctions pouvant être décomposées est large, puisqu'elle regroupe l'ensemble des processus qui nous intéressent la plus part du temps en tant que statisticien : celles qui sont à support sur un intervalle, admettant une covariance continue et finie sur le support.

On en déduit que pour travailler avec des données fonctionnelles, il suffit de les décomposer dans la base ACP fonctionnelle puis de travailler sur les composantes de chaque élément de la base. On travaille désormais avec des réels et non plus des fonctions, ce qu'on aime manipuler. On peut alors faire de la statistique traditionnelle avec les outils que l'on connaît.

Propriété (intérêt de la base FPCA — informel)

la base ACP fonctionnelle est la plus économique, c'est à dire qu'elle explique au mieux la covariance des données pour un nombre de composantes fixées, ce qui est utile car on ne sait manipuler numériquement que des objets de dimension finie.

Pour avoir une bonne représentation de ces données, on doit donc s'assurer de bien estimer la covariance. Pour cela, on a mentionné qu'il serait judicieux de lisser les observations en tenant compte de la régularité du processus dont est issu nos données. La question est désormais la suivante :



Est-il possible de récupérer la régularité locale des trajectoires à partir des données ? Si oui, comment ?

C'est ce qu'affirme le théorème suivant provenant des travaux de Golovkine et MPV :

Théorème (Regularité locale — informel)

Les données fonctionnelles permettent de récupérer la régularité locale des trajectoires. Les estimateurs définis **ponctuellement** convergent.

Remarque (Continuité de Kolmogorov) : Un théorème (Continuité de Kolmogorov) permet à partir de l'espérance d'incrément d'un processus aléatoire de déduire sa régularité. C'est pourquoi les estimateurs sont définis à partir des incrément quadratiques. C'est entre autres *la raison pour laquelle les données fonctionnelles permettent de récupérer la régularité locale des trajectoires*.

2.1.2 Résumé de l'intérêt de la modélisation fonctionnelle

Les données fonctionnelles permettent de travailler sur un modèle où la *relation* entre plusieurs quantités est sujet à une loi¹. Ce point de vue de réplication de courbes est notamment utile car il permet d'extraire des observations leur régularité². L'estimation de cette régularité permet, entre autres, de lisser les courbes de façon appropriée en fonction de la quantité que l'on souhaite estimer, telle que la moyenne et la covariance avec une plus grande précision³.

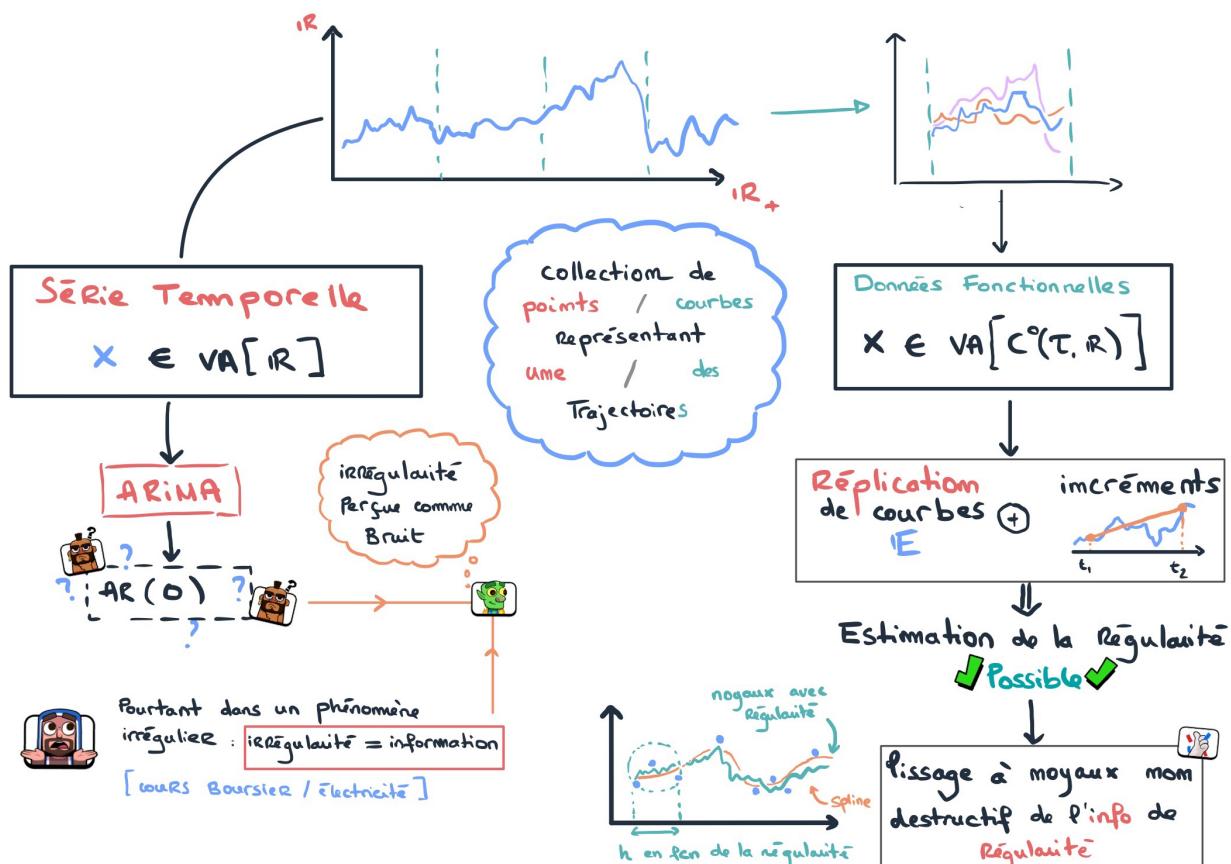


FIGURE 2.2 – Résumé des motivations du de l'estimation de la régularité locale des trajectoires

1. cf données fonctionnelles — informel : 2.1.1
2. cf Continuité de Kolmogorov, Regularité locale — informel : 2.1.1
3. cf Estimateurs de la moyenne et de la covariance — informel (8) : 4.1.1

2.1.3 Cas non indépendant : séries temporelles de données fonctionnelles

Il est commode théorie des données fonctionnelles de supposer que l'on observe des courbes $X_i : \Omega \rightarrow \mathcal{C}^0(I, \mathbb{R})$ **indépendantes** et identiquement distribuées. Cependant une partie non négligeable des données que l'on observe ont des dépendances avec les valeurs passées. Par exemple, il est raisonnable de penser que la consommation électrique d'un foyer au cours d'une année croît avec l'ajout successif de nouveau appareils électroniques. L'hypothèse d'indépendance entre les données n'est donc plus pertinente pour les données que l'on traite et il devient important de considérer des processus autorégressifs adaptés aux données fonctionnelles. Si dans le cadre des données de \mathbb{R} cette relation de dépendance *linéaire* avec le passé pouvait s'écrire sous la forme suivante $X_n = \sum_{k=1}^{n-1} \varphi_k X_k + \xi_n$ où $\varphi_k \in \mathbb{R}$ et $\xi_n \begin{cases} \in \text{VA}(\mathbb{R}) \\ \perp\!\!\!\perp \sigma(X_i)_{1:n-1} \end{cases}$, dans le cadre fonctionnel on capture la même idée en considérant $X_n = \sum_{k=1}^{n-1} \phi_k(X_k) + \xi_n$ où ϕ_k est un opérateur linéaire de $\mathbb{L}^2(I, \mathbb{R})$, le plus souvent intégral.



Il s'agit d'une généralisation naturelle de la relation dans le cadre réel, puisqu'on peut démontrer que sur l'espace des nombres réels l'ensemble des fonctions linéaires $\phi : \mathbb{R} \rightarrow \mathbb{R}$ sont de la forme $x \mapsto ax$ avec $a \in \mathbb{R}$. La relation sur \mathbb{R} que l'on a vue juste avant peut alors se ré-écrire de façon similaire à la version fonctionnelle.

On considère lors de ce stage des données fonctionnelles sous forme de données indépendantes mais aussi sous forme de série temporelle : sur les données éoliennes chaque indice représente un parc éolien différent éloigné géographiquement (donc indépendants), là où les données photovoltaïques présentent clairement une dépendance ne serait-ce que géographique car les panneaux sont proches les uns des autres.



On fera donc très attention à l'appellation historique « *série temporelle* », qui représente ici juste l'idée de dépendance d'un indice à l'autre. Il se peut que l'indice ait ou non une signification temporelle.



Pourquoi se soucier en particulier des séries temporelles fonctionnelles lorsque l'on souhaite incorporer la régularité du processus dont est issu nos données dans l'estimation des quantités qui nous intéressent ?

Rappelons-nous que les données fonctionnelles sont la clé pour déterminer la régularité, et que cela est en réalité permis par le théorème de Continuité de Kolmogorov (que nous n'avons pas énoncé en détails, mais mentionné dans la section 2.1.1). Malheureusement, dans le monde réel où vit le praticien, nous n'avons pas accès à l'espérance de la loi dont sont issues nos données. Il nous faut donc estimer cette espérance, et c'est là que les séries temporelles fonctionnelles entrent en jeu. Puisque l'estimateur usuel de l'espérance est la moyenne empirique, qui nous est fourni par la loi des grands nombres, cela devient très problématiques lorsque l'on dispose de données corrélées. L'hypothèse d'indépendance faible nous permet de tout de même utiliser l'estimateur usuel de l'espérance. Alors, les estimateurs des paramètres de régularité convergents ponctuellement vers ceux du processus dont sont issues nos données.

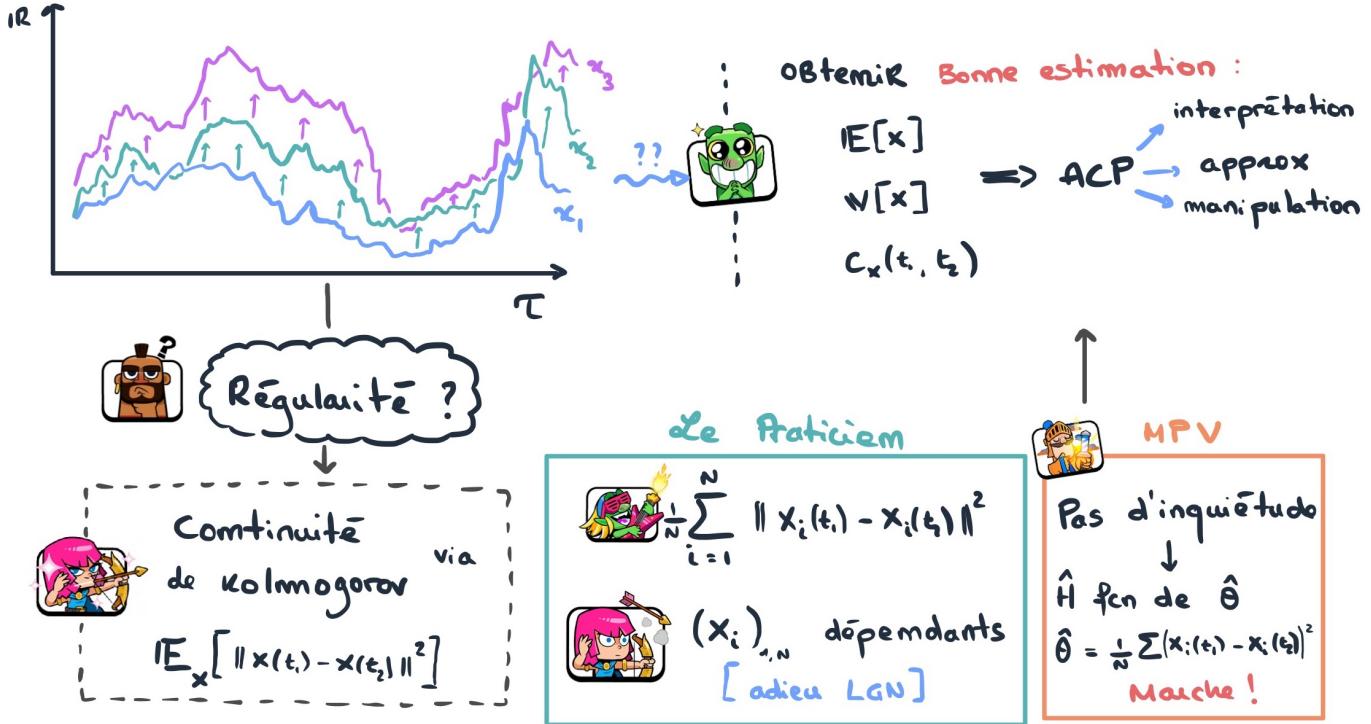


FIGURE 2.3 – Schéma grossièrement récapitulatif : Estimation de la régularité pour une série temporelle fonctionnelle



Il faut faire attention lorsque l'on manipule ou interprète des séries temporelles fonctionnelles. (comme par exemple tout résultat utilisant la loi de $\sum_n X_n, \dots$)

Une série temporelle discrète est le fait que l'observation suivante dépend linéairement de l'observation précédente, dans le cadre fonctionnel *l'observation est une fonction*. La dépendance se fait sur l'indice de la fonction, et non pas sur l'argument de la fonction interprété dans notre cas comme étant le temps.

Dans certains jeux de données c'est d'autant plus trompeur de parler de temps car on observe des courbes sur une année : à la fois l'indice de la fonction et l'argument de la fonction ont des interprétations temporelles.

dans l'expression « $X_n(t)$ », la série temporelle (discrète) concerne bien l'indice n et non pas l'argument t .

Etant donné que l'on souhaite estimer la régularité locale du processus il est naturel de se demander :



Lorsque l'on a une dépendance dans les observations fonctionnelles $\{X_1 \dots X_n\}$, possède-t-on une dépendance dans les observations ponctuelles à t fixé $\{X_1(t) \dots X_n(t)\}$? Est-ce que l'on sait l'identifier?

Et la réponse, c'est qu'**on ne sait pas**. En tout cas, dans le cadre général. Il y a en effet plusieurs façons de définir ce qu'on appelle par « dépendance ». Toutes les définitions de dépendance ne mènent pas à cette conclusion, mais celle adoptée par (MPV) permet de passer de la dépendance fonctionnelle à une dépendance locale. De manière générale, lorsque l'on traite des

données avec de la dépendance, il convient d'être extrêmement précautionneux avec les théorèmes et « faits » que l'on invoque.⁴

La dépendance faible comme définie dans l'article de MPV(17) nous permet de travailler localement : On peut travailler localement sur les trajectoires tout en utilisant des hypothèses fonctionnelles (que ce soit pour la dépendance ou autres) pour obtenir la régularité.

2.2 Estimation de la régularité locale des trajectoires

2.2.1 Ce qu'on entend par régularité locale

Longtemps, il était cru que les fonctions continues étaient dérивables presque partout. C'est notamment Weierstrass qui a démontré qu'il existe des fonctions continues partout mais dérivable nulle part. Poincaré notamment disait de tels objets qu'ils n'existaient que pour contredire le travail des pères. Cependant, des objets manipulés tous les jours comme le monde de la finance notamment traitent des processus qui sont fondamentalement irréguliers⁵ (au point de vue de l'analyse, où l'on traite souvent des fonctions au moins dérivables). Il est donc important de pouvoir quantifier la régularité d'une fonction de façon plus fine que le nombre de dérivées qu'elle possède.

Nous allons repasser rapidement en revue les différents concepts de régularité pour mettre l'emphase sur ce que l'on considère comme régularité locale. Afin de savoir à quel niveau de régularité nous souhaitons estimer, il est important de garder en tête un ordre de différents niveaux de régularité résumé par les relations suivantes :

Lipschitz \Rightarrow Hölder \Rightarrow Localement Hölder \Rightarrow Uniformément continue \Rightarrow Continue
ce qui nous intéresse

Si le lecteur souhaite discerner ce que chaque propriété signifie, et quelles sont les différences entre chaque niveau de régularité, il est possible de se rappeler rapidement les définitions de ces propriétés disponibles en annexe A.2.1.



Pourquoi se concentrer sur des processus localement Hölder ?

La nature des phénomènes rencontrés dans la vie réelle est souvent complexe. Influencés par de nombreux phénomènes, certains d'entre eux sont, comme mentionnés précédemment, irréguliers. C'est notamment le cas des courbes de charge électriques, qui dépendent de multitudes de phénomènes physiques ou comportementaux, dont on peut attendre une certaine régularité, mais qui ne sont pas nécessairement uniformes tant sur leur niveau régularité que l'intervalle de temps sur lequel ils ont une influence. On pourrait par exemple attendre une différence de régularité de la production électrique en plein été (soleil et température stables ...) comparé au mois de mars (plus grande instabilité des conditions climatiques).

4. Le détail théorique de la validité de l'utilisation de la moyenne empirique comme estimateur de l'espérance sous hypothèse d'indépendance faible (proprement définie et motivée) est disponible en annexe A.3.

5. les fonctions dérивables nulle part sont même denses dans les fonctions continues pour la topologie de la convergence uniforme (10). A epsilon près on rencontre toujours une fonction dérivable nulle part lorsque l'on considère la distance maximale réalisée entre deux fonctions continues sur leur support I ...

De plus, les fonctions Hölderiennes représentent une classe suffisamment large de fonctions. L'espace de fonctions sur lequel on travail est donc devrait être en pratique suffisamment grand pour inclure l'ensemble des processus qui nous intéressent. Enfin les fonctions que le praticien sera amené à manipuler seront des fonctions d'un intervalle dans \mathbb{R} , qui lorsque continues sont automatiquement uniformément continues en vertu du théorème de Heine. Il est donc naturel de se concentrer sur des fonctions localement Hölderiennes.⁶

2.2.2 Modèle considéré

On dispose désormais de tous les ingrédients pour expliciter le modèle considéré pendant l'ensemble du stage :

Nom	Objet	Définition
Régularité : constante locale	L	: $[0, 1] \rightarrow \mathbb{R}$ $t \mapsto L_t$
Régularité : puissance locale de l'incrément	H	: $[0, 1] \rightarrow [0, 1]$ $t \mapsto H_t$
donnée fonctionnelle	X	$\in \text{VA}[\mathbb{L}^2 \cap \mathcal{H}(H, L)]$
N -échantillon de la loi de X	$(X_n)_{n \in \llbracket 1, N \rrbracket}$	$X_n \sim X$
Nombre de points sur la trajectoire de X_n	M_n	$\sim \mathcal{P}(\lambda)$
Temps observés	$(T_n[m])_{m \in \llbracket 1 : M_n \rrbracket}$	$\sim \mathcal{U}([0, 1])^{\otimes M_n}$
écart type de l'erreur erreur	σ η	$\in \mathbb{R}_+^*$ $\sim \mathcal{N}(0, \sigma^2)$
noyau de l'opérateur intégral	β	$\in \mathbb{L}^2([0, 1])$
relation auto-régressive intégrale	ϕ	: $\mathbb{L}^2([0, 1], \mathbb{R}) \rightarrow \mathbb{L}^2([0, 1], \mathbb{R})$ $f \mapsto \int_0^1 \beta(u, \cdot) f(u) du$
FAR(1)	X_{n+1}	$= \phi(X_n) + \xi_{n+1}$
observation observation	$Y_n[m]$ $(T_n[m], Y_n[m])_{n, m}$	$= X_n(T_n[m]) + \xi_n(T_n[m])$ $\in [0, 1] \times \mathbb{R}$

TABLE 2.1 – Tableau récapitulatif du modèle considéré

2.2.3 Deux méthodes d'obtention de la régularité locale des trajectoires

Il existe deux méthodes différentes pour estimer la régularité des trajectoires. Si la clé des deux méthodes pour extraire la régularité locale est le théorème de continuité de Kolmogorov⁷, les deux méthodes diffèrent par les points $t \in \mathcal{T}$ considérés dans l'estimation des accroissements quadratiques $\mathbb{E} [|X(u) - X(v)|^2]$ utilisés pour l'estimation de la régularité locale.

La méthode de Golovkine et al. (9, pages : 7—9) n'utilise que les points observés, et construit un estimateur des incrémentations quadratiques à base de statistique d'ordre.

6. Afin de ne pas alourdir l'essence du propos, une simplification par rapport à l'article de MPV (17) a été faite, si le lecteur souhaite aller dans le détail, il est possible de se référer à l'Annexe A.2.

7. cf Annexe A.4

$$\theta(T_{(l)}, T_{(k)}) = \mathbb{E} \left[|X(T_{(l)}) - X(T_{(k)})|^2 \right] \underset{\substack{\text{Hölder} \\ + C^0 \text{ Kol.}}}{\approx} \boxed{\frac{1}{N} \sum_{n=1}^N \left| Y_n^{(2k-1)} Y_n^{(k)} \right|^2 \underset{\text{déf}}{\equiv} \hat{\theta}_k}$$

et on obtient ainsi l'estimateur suivant :

$$\hat{H}_t(k) = \begin{cases} \frac{\log(\hat{\theta}_{4k-3} - \hat{\theta}_{2k-1}) - \log(\hat{\theta}_{2k-1} - \hat{\theta}_k)}{2 \log 2} & \hat{\theta}_{4k-3} > \hat{\theta}_{2k-1} > \hat{\theta}_k \\ 1 & \text{sinon} \end{cases}$$



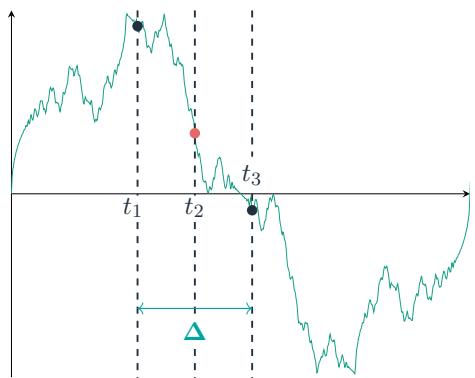
Cette méthode peut s'avérer spécifiquement utile lorsque l'on traite un flux de données, car l'arrivée de nouvelles données ne nécessite pas spécifiquement de recalculer les incrémentations quadratiques sur l'ensemble des points observés.

L'autre méthode proposée dans les articles (8, 17), elle se base sur l'utilisation de points non observés, inférés par lissage des courbes, à une distance $\Delta/2$ les uns des autres pour estimer les incrémentations quadratiques. Cette dernière méthode implique le choix d'un hyper-paramètre lors de l'estimation Δ et pourrait être sensible à la qualité du lissage de la courbe. Etant donné que l'objectif de la détermination de la régularité locale est de pouvoir faire un lissage à noyaux adaptatif en fonction de l'objet que l'on souhaite estimer, on appelle le lissage effectué pour estimer la régularité « pré-lissage ».

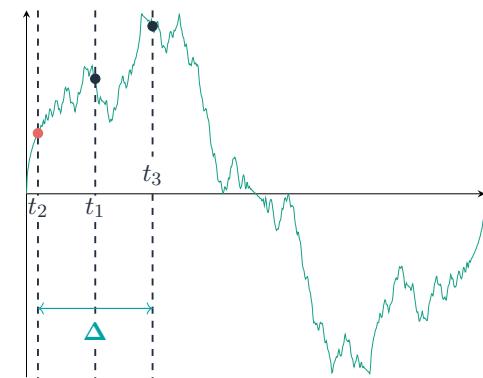
On se donne un $\Delta \in]0, 1[$, arbitraire pour le moment, comme diamètre de l'intervalle J_Δ que l'on considère pour évaluer la régularité en t_0 .

Il est naturel de définir les points d'estimation de la régularité de la façon suivante :

$$\begin{aligned} t_1 &\underset{\text{déf}}{\equiv} t_0 - \frac{\Delta}{2} & J_\Delta &= [t_1, t_3] \\ t_2 &\underset{\text{déf}}{\equiv} t_0 & \text{avec } t_0 &\text{ le point en lequel on souhaite estimer la} \\ t_3 &\underset{\text{déf}}{\equiv} t_0 + \frac{\Delta}{2} & &\text{régularité.} \end{aligned}$$



X
● Point d'estimation de la régularité locale



X
● Point d'estimation de la régularité locale

FIGURE 2.4 – Exemple de courbe dont on souhaiterait déterminer la régularité locale, et visualisation de J_Δ : estimation intérieure / au bord

i

Remarque : Rien n'empêche dans la théorie d'avoir les points t_1, t_2, t_3 non ordonnés dans le temps, mais dans la pratique, on considère naturellement que $t_1 < t_2 < t_3$. Mais cet ordre n'est pas obligatoire. Ainsi aux bords, si l'on souhaite estimer la régularité au point t_0 tel que la définition précédente nous donne un point t_1 en dehors de $[0, 1]$, on peut tout à fait à la place considérer :

$$t_2 \underset{\text{déf}}{\equiv} t_0$$

$$t_1 \underset{\text{déf}}{\equiv} t_0 + \frac{\Delta}{2}$$

$$t_3 \underset{\text{déf}}{\equiv} t_0 + \Delta$$

on pourra se référer à la 2^e image de la figure 2.2.3

i

Le point t_0 , où l'on souhaite estimer la régularité, étant dans la majorité des cas le point central de l'intervalle J_Δ considéré ; il sera à présent mentionné comme le point t_2 . Il s'agit à la fois d'un moyen de se rappeler dans les formules suivantes que l'on considère le point central de l'intervalle J_Δ et d'être au plus proche des noms de variables considérés dans l'implémentation.

alors on approche $\theta(t_1, t_3) = \mathbb{E} [|X(t_3) - X(t_1)|^2]$ par :

$$\tilde{\theta}(t_1, t_3) = \frac{1}{N} \sum_{n=1}^N |X_n(t_3) - X_n(t_1)|^2$$

qui n'est pas observable, étant donné qu'il n'est pas garanti d'observer $X(t_1)$ et $X(t_3)$, et qu'il faut donc lisser dans un premier temps les courbes pour pouvoir évaluer X en t_1 et t_3 . L'estimateur que l'on considère est donc une approximation de $\tilde{\theta}_{13}$, et est défini par :

$$\hat{\theta}(t_1, t_3) = \frac{1}{N} \sum_{n=1}^N |\hat{X}_n(t_3) - \hat{X}_n(t_1)|^2$$

où \hat{X} est la courbe lissée à partir des observations :

$$(T_n[m], Y_n[m])_{n \in 1:N, m \in 1:M_n}$$

avec :

— N : Nombre de courbes observées

— M_n : Nombre de points observés (aléatoire) sur la trajectoire de X_n

Les estimateurs des paramètres de régularité sont alors les suivants (17) :

Définition 1 (estimateurs des paramètres de régularité) .

$$\hat{H}_{t_2} = \frac{\log \hat{\theta}(t_1, t_3) - \log \hat{\theta}(t_1, t_2)}{2 \log 2}$$

ou

$$\hat{H}_{t_2} = \frac{\log \hat{\theta}(t_1, t_3) - \log \hat{\theta}(t_2, t_3)}{2 \log 2}$$

$$\begin{aligned} \hat{L}_{t_2} &= \frac{\hat{\theta}(t_1, t_3)}{\Delta^{2\hat{H}_{t_2}}} \\ &= \frac{\hat{\theta}(t_1, t_2)}{\Delta^{2\hat{H}_{t_2}}} \\ &\text{ou} \\ &= \frac{\hat{\theta}(t_2, t_3)}{\Delta^{2\hat{H}_{t_2}}} \end{aligned}$$

2.2.4 Prélassage

Comme mentionné précédemment, l'estimation de la régularité locale nécessite l'évaluation de notre processus observé X en 3 points. Il est possible de ne pas observer ces points, qui sont de plus bruités dû à l'erreur de mesure de X . C'est pourquoi nous décidons de lisser les courbes comme « pré-lissage » pour pouvoir estimer la régularité locale.



Pourquoi parle-t-on de **pré-lissage**? Le but de considérer la régularité n'était-il pas justement de l'utiliser dans le lissage des trajectoires? Lisser avant même d'estimer la régularité n'est-il pas contre-productif?

L'objectif de l'obtention des paramètres de régularité des trajectoires est de pouvoir effectuer un lissage de ces trajectoires qui préserve les irrégularités fondamentales du processus dont elles sont issues, tout en éliminant le bruit. Les paramètres de régularité sont estimés avec des quantités temporellement équidistantes, qui sont donc potentiellement non observées. L'estimation de la régularité fait donc usage de trajectoires lissées. Les paramètres estimés sont ensuite utilisés pour effectuer un **nouveau lissage** à noyaux en utilisant, cette fois, une fenêtre de lissage appropriée qui dépend de ces paramètres de régularité.

En d'autres termes, le pré lissage utilise un lissage à noyaux tel que la fenêtre de lissage cross-validation nous donne :

$$h_{\text{pre}}^{*[cv]} \text{ estimateur de } h_{\mathcal{R}_{\text{quadr}}}^*(t) = \mathcal{O}_{\mathbb{P}} \left(\lambda^{-\frac{1}{2H_t+1}} \right) \quad (2.1)$$

à partir duquel on peut lisser les courbes observées $(T_i^{[n]}, Y_i^{[n]})_{n \in 1:N, i \in 1:M_n}$ pour estimer la régularité locale donnée par H_t et L_t . On peut désormais obtenir la fenêtre de lissage adaptée à la quantité que l'on souhaite estimer :

$$h_{\mu}^*(t) = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \mathcal{R}_{\mu} \left(\underbrace{t}_{\substack{\text{Régularité, sparsity, ...} \\ \rightarrow H_t, L_t, \mathcal{W}_t}}, h \right) \quad (2.2)$$

Le coeur de ce stage est la détermination du comportement de l'hyper-paramètre Δ , diamètre de l'intervalle que l'on considère dans lequel on vient prendre la valeur de notre processus en 3 points régulièrement espacés. MPV affirme déjà que pour un Δ donné, on a bien la convergence ponctuelle des estimateurs. (17) Toutefois, le praticien est en droit de se demander quel Δ explicitement choisir? Existe-t-il une procédure simple pour déterminer la valeur optimale de Δ qu'il faut choisir pour obtenir un biais le plus petit possible pour l'estimation des paramètres de régularité?

Remarque (Qu'en-est-il de la méthode de lissage?) : Si le stage se concentre sur l'étude du comportement du Δ essentiellement sur un pré-lissage non paramétrique à noyaux, on peut se poser la question suivante :



la méthode de pré-lissage a-t-elle une importance? Si oui, laquelle faut-il choisir?

Cette question a été étudié brièvement dans le cadre de ce stage, les résultats obtenus sont disponibles en annexe B.4. Trois méthodes de lissage y sont comparées.

2.2.5 Résumé de la méthodologie d'estimation de la régularité locale

Résumons rapidement la méthode d'estimation de la régularité en un point $t_2 \in \mathcal{T}$. La procédure d'obtention de la régularité est ainsi la suivante :

- 1 : Pré-lissage de la courbe
- 2 : Calcul des incrément quadratiques sur la courbe lissée
- 3 : Moyennage des incrément (estimateur de l'espérance)
- 4 : Utilisation de l'estimateur

Ce que nous cherchons désormais à déterminer est désormais la réponse à la question suivante :



Quel Δ choisir pour obtenir la meilleure estimation de H en t_2 ?

Pour étudier cela, nous allons simuler des données FAR(1) Höldériennes de régularité connue et allons étudier quel Δ fournit la meilleure estimation des paramètres de régularité en fonction de λ, N, H_t, \dots ⁸

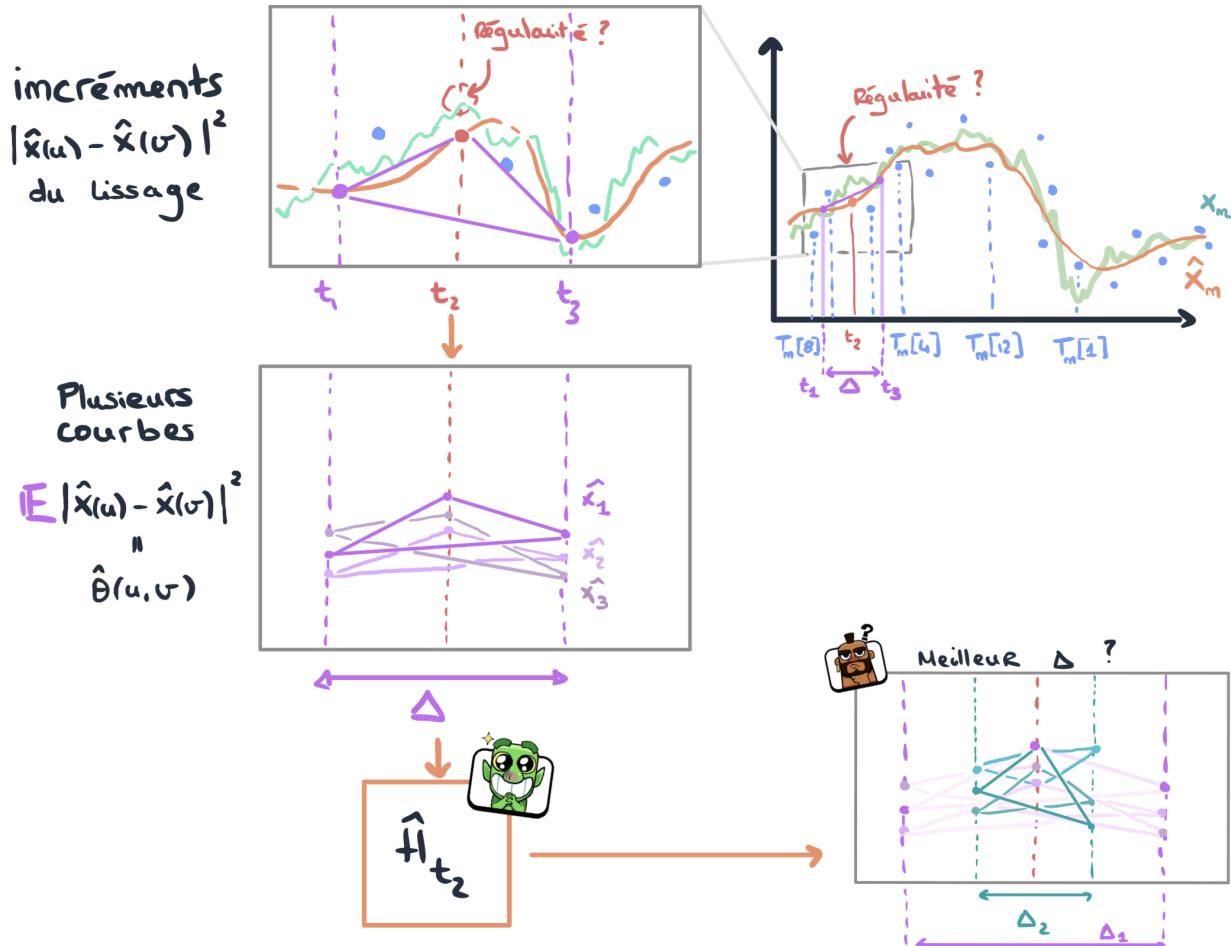


FIGURE 2.5 – Schéma résumé de la méthodologie d'estimation de la régularité

8. on pourra se référer à 2.1 pour la signification des notations

2.3 Simulation

2.3.1 Objectifs de la simulation

L'objectif de la simulation est de pouvoir analyser le comportement des estimateurs des paramètres de régularité lorsque l'on fait varier Δ , le diamètre du voisinage J_Δ dans lequel on vient utiliser de l'information pour capter la régularité. Cela permettra ensuite d'analyser cette fois le comportement de Δ^* , le Δ optimal pour l'estimation des paramètres de régularité.

Les questions jusqu'alors on s'est surtout préoccupé d'expliquer l'intérêt et la méthodologie associée à l'obtention de la régularité. Toutefois différentes questions se posent vis-à-vis de cette estimation, notamment en lien avec le choix du Δ :



- ① : Le risque associé à l'estimation des paramètres de régularité, est-il une fonction convexe (voire strictement convexe) de Δ ? Si non, l'est-elle au moins au voisinage du Δ^* optimal pour l'estimation de la régularité?
- ② : Quel lien, si il en existe un, y-a-t-il entre Δ^* et d'autres quantités susceptibles d'affecter la vitesse de convergence de l'estimateur : $N, \lambda = \mathbb{E}[M_n], H_{t_2}, \dots$?
- ③ : Peut-on fournir une procédure simple de détermination d'un Δ proche du Δ^* optimal, pouvant être obtenu à partir des données pour être utilisé en pratique par les statisticiens?

2.3.2 Mouvement Brownien Multi-Fractionnaire (mfBm)

Afin de simuler un processus Höldérien non dérivable, on choisit de simuler un mouvement Brownien Multi-Fractionnaire. En effet il s'agit d'un processus qui, presque-sûrement, est continu mais différentiable nulle part. Le mouvement brownien multi-fractionnaire est d'autant plus intéressant dans le cadre de la simulation car on sait le générer en contrôlant localement sa régularité sur un voisinage V de t_0 .

$$\forall u, v \in V \quad \mathbb{E} [|\xi_n(u) - \xi_n(v)|^2] \simeq L_{H_\xi(t_0)} |u - v|^{2H_\xi(t_0)}$$

Le lecteur pourra, si il le souhaite trouver plus de ressources, définitions et propriétés formelles sur le mouvement brownien fractionnaire et multi-fractionnaire en annexe A.6. Le point essentiel exploité par l'algorithme utilisé pour la simulation est le suivant : le mouvement brownien multi-fractionnaire est un processus gaussien dont on sait expliciter la covariance en fonction de la régularité des instants considérés :

$$C(t, s) = D(H_s, H_t) [s^{H_s+H_t} + t^{H_s+H_t} - |t - s|^{H_s+H_t}]$$

avec :

$$D(x, y) = \frac{\sqrt{\Gamma(2x+1)\Gamma(2y+1)} \sin(\pi x) \sin(\pi y)}{2\Gamma(x+y+1) \sin(\pi(x+y)/2)}$$

il nous suffit donc de déterminer la matrice suivante :

$$\Sigma = \begin{bmatrix} \ddots & & \\ & D(H(t_i), H(t_j)) \cdot (t_i^{H(t_i)+H(t_j)} + t_j^{H(t_i)+H(t_j)} - |t_i - t_j|^{H(t_i)+H(t_j)}) & \\ & \ddots & \end{bmatrix}$$

avec :

$$t_i, t_j \in \mathbb{T} = \left\{ \underbrace{\begin{array}{c} t_1(\Delta, t), t_2(t), t_3(\Delta, t) \\ t \in \vec{t} \quad \Delta \in \vec{\Delta} \end{array}}_{\text{estimateur de régularité}}, \underbrace{g_1 \dots g_G}_{\text{grille pour l'intégration de la relation FAR}}, \underbrace{T_n[1] \dots T_n[M_n]}_{\text{points observés (aléatoire)}} \right\}$$

Pour simuler un mouvement brownien multi-fractionnaire en les points désirés, il convient donc de simuler une loi normale multivariée d'espérance nulle et de covariance Σ .

Remarque (complexité de la simulation) : La méthode de génération du mouvement brownien multi-fractionnaire via une loi normale multi-variee implique l'inversion d'une matrice de covariance. La méthode utilisée pour l'inversion de cette matrice par la fonction MASS::mvrnorm utilisée pour la simulation est via la décomposition spectrale de celle-ci.



The matrix decomposition is done via `eigen`; although a Choleski decomposition might be faster, the eigendecomposition is stabler.

- Documentation du package R MASS (22)

L'inversion de la matrice de covariance via sa décomposition spectrale est un algorithme de complexité $\mathcal{O}((\text{card } T)^3)$. Ce qui implique qu'évaluer de plus en plus de point sur le même processus devient rapidement cher en calcul.

2.3.3 génération d'un FAR(1)

Parmi les avantages de l'utilisation d'un mouvement brownien multi-fractionnaire à simuler pour analyser le comportement du Δ se trouve le fait que l'on peut dériver facilement la régularité des courbes d'une relation FAR(1) basée sur ce processus. En effet, si l'opérateur linéaire de cette relation ϕ est un opérateur intégral $X \mapsto \int \beta(u, t)X(u)du$, il suffit alors que $H_\beta > H_\xi$ pour que le FAR(1) hérite de la régularité du mouvement brownien multi-fractionnaire. Ainsi on dispose directement de la régularité de notre FAR(1) en chaque point du support, ce qui s'avère très utile pour analyser les résultats.

Afin de générer la $N^{\text{ème}}$ observation d'un FAR défini par la relation auto-regressive suivante :

$$X_N(t) = \int \beta(u, t)X_{N-1}(u)du + \varepsilon_N(t) \quad (2.3)$$

Ainsi, il nous suffit de générer N mouvements browniens multi-fractionnaires indépendants aux points dont on a besoin l'évaluation du mouvement brownien :

$$(\xi_1, \xi_2, \dots, \xi_N) \sim \text{mfBm}(H, L)$$

que l'on utilise comme innovations dans la relation FAR(1) [eq. 2.3]. La méthode de calcul numérique de l'intégrale sélectionnée est la *méthode des rectangles au point médian* car il s'agit de la méthode de calcul numérique d'intégrale de plus grand ordre avec 1 unique point d'évaluation requis pour le calcul, ce qui est primordial vis à vis de la remarque formulée dans la section 2.3.2.

Afin d'atteindre la stationnarité, on effectue une période de « Burn-in » où l'on génère un nombre d'éléments (100) non retenus dans l'échantillon sauvegardé. En appelant B le nombre d'étape de burn-in, on génère donc $N + B$ mouvements browniens $(\xi_1, \dots, \xi_B, \dots, \xi_{N+B})$ multi-fractionnaires indépendants aux points désirés, on applique à chaque itération la relation FAR(1) et on ne retient que les N dernières itérations.

L'algorithme utilisé pour la génération du FAR(1) basée sur un opérateur intégral comme défini par la relation 2.3 est disponible en annexe : Algorithme 1.

2.4 Critère de sélection du Δ

L'estimation des paramètres de régularité locale en t_2 (H_{t_2} et L_{t_2}) utilise les incrémentations quadratiques θ entre les différents points t_1, t_2, t_3 situés dans un voisinage de diamètre Δ autour de t_2 . En observant que L est estimée par une expression impliquant θ et Δ^{2H} , une estimation précise de H paraît plus cruciale pour la bonne estimation conjointe des deux quantités.



Doit-t-on se concentrer sur l'estimation de H ou de θ ?

Les incrémentations quadratiques θ sont utilisées à la fois pour l'estimation de H et de L . Nous allons donc chercher à déterminer un Δ adapté à l'estimation des incrémentations quadratiques. En effet, un mauvaise estimation des incrémentations induirait une mauvaise estimation de H . Puisque l'on a déterminé qu'une bonne estimation de H était cruciale, et que cette quantité utilise deux valeurs d'incrémentations distinctes, nous allons nous focaliser sur l'estimation d'un couple de θ que nous noterons Θ .

La métrique que nous allons sélectionner est *la distance euclidienne entre une estimation d'un tel couple Θ relative à la cible de l'estimation $\tilde{\Theta}$, que l'on nomme $\hat{\Theta}$ et l'estimateur de l'espérance via la moyenne empirique des courbes non bruitées, observées pleinement, puisqu'il s'agit du meilleur estimateur de l'espérance que l'on pourrait espérer atteindre (du biais est introduit lors du lissage non paramétrique des courbes bruitées)*. On considère donc le risque suivant :

$$\widehat{\mathcal{R}}(\Theta, \Delta) = \frac{1}{mc} \sum_{p=1}^{mc} \frac{\|\hat{\Theta}[p] - \tilde{\Theta}[p]\|_2^2}{\|\tilde{\Theta}[p]\|_2^2} \quad (2.4)$$

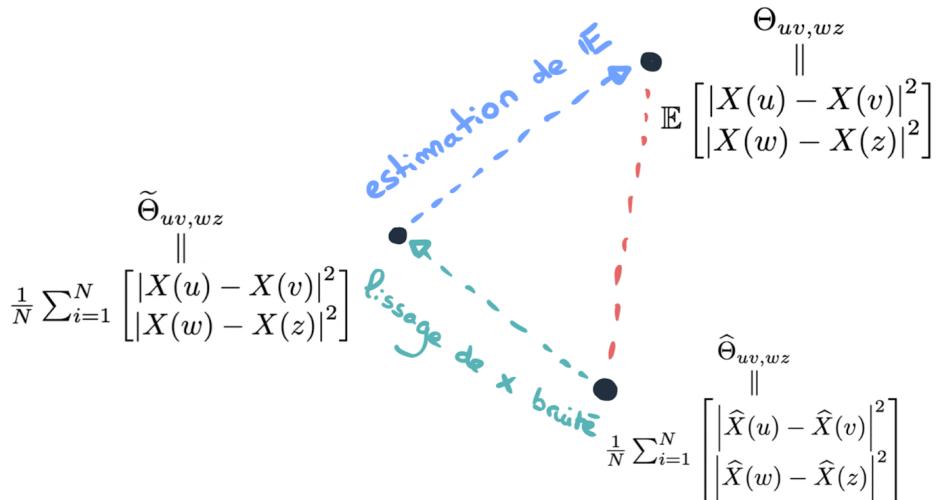


FIGURE 2.6 – Schéma représentant les différentes approximations du couple d'incrémentations

Chapitre 3

Détermination du diamètre optimal des intervalles à considérer pour l'estimation de la régularité locale

Contents

3.1	Choix des paramètres de la simulation des FAR(1) localement Hölderiennes	21
3.1.1	Nombre de simulations	22
3.1.2	Fonction de Hurst	22
3.1.3	Constante de Hölder	22
3.1.4	Moyenne	23
3.1.5	Noyau de la relation FAR(1)	23
3.1.6	Nombre de courbes	23
3.1.7	Nombre moyen de points observés par courbe	24
3.1.8	Ensemble des Δ testés	24
3.1.9	Bruit blanc	24
3.1.10	Résumé des Paramètres	25
3.2	Prélissage des données simulées	25
3.2.1	Les courbes obtenues	25
3.2.2	Pré-Lissage	25
3.3	Détermination du Δ optimal à choisir pour l'estimation de la régularité	26
3.4	Discussion	27

Nous avons désormais établi que la génération d'un FAR(1) basé sur un mouvement brownien multi-fractionnaire permettait de contrôler de bout-en-bout la régularité du processus. Ceci va nous permettre de pouvoir analyser correctement le comportement du risque d'estimation de la régularité en fonction de Δ , ainsi que le comportement du Δ^* optimal.

3.1 Choix des paramètres de la simulation des FAR(1) localement Hölderiennes



Il est conseillé de se référer aux tableaux (Notations-Spécifique au stage) et 2.1 pour la signification des notations déjà introduites et utilisées dans ce chapitre.

3.1.1 Nombre de simulations

Afin d'étudier la relation entre le Δ optimum et différentes quantités caractéristiques aux données, on va effectuer une simulation de monte carlo.

On décide de générer $mc = 200$ simulations de Monte-Carlo, afin d'obtenir les résultats les plus robustes possibles pour l'estimation du risque $\mathbb{E} [\|\hat{\Theta} - \tilde{\Theta}\|_2]$ ¹, tout en gardant un temps de calcul raisonnable. On fait varier λ de 30 à 480 en incrémentant de 15 à chaque fois. L'idée et de pouvoir regarder si il existe une relation entre Δ^* et la position du nombre moyen de points observés aprè courbe (λ) par rapport au nombre de courbes (N).

Il est possible de voir comment les paramètres que l'on va définir sont utilisés dans l'implémentation en annexe D.

3.1.2 Fonction de Hurst

On appelle $H : t \mapsto H_t$ la fonction de Hurst.

Celle qui a été choisie est la suivante :

$$H_{\text{logistic}}^{[h_l, h_r, s, pos]} : \begin{matrix} [0, 1] \\ t \end{matrix} \longrightarrow \begin{matrix} [h_l, h_r] \\ h_l + \frac{(h_r - h_l)}{1 + e^{-s(t - pos)}} \end{matrix}$$

La fonction de Hurst retenue est la suivante :

$$H_{\text{logistic}}^{[0.4, 0.8, 5, 0.5]}$$

On dispose donc d'une régularité locale qui varie sur \mathcal{T} , tout en ayant une évolution pas trop brusque. Nous allons étudier le comportement du Δ lors de l'estimation de la régularité locale en les points suivants :

$$\vec{t} = \begin{bmatrix} 0.3 \\ 0.4 \\ 0.5 \\ 0.6 \\ 0.7 \\ 0.8 \end{bmatrix} \quad H(\vec{t}) = \begin{bmatrix} 0.51 \\ 0.55 \\ 0.6 \\ 0.65 \\ 0.69 \\ 0.73 \end{bmatrix}$$

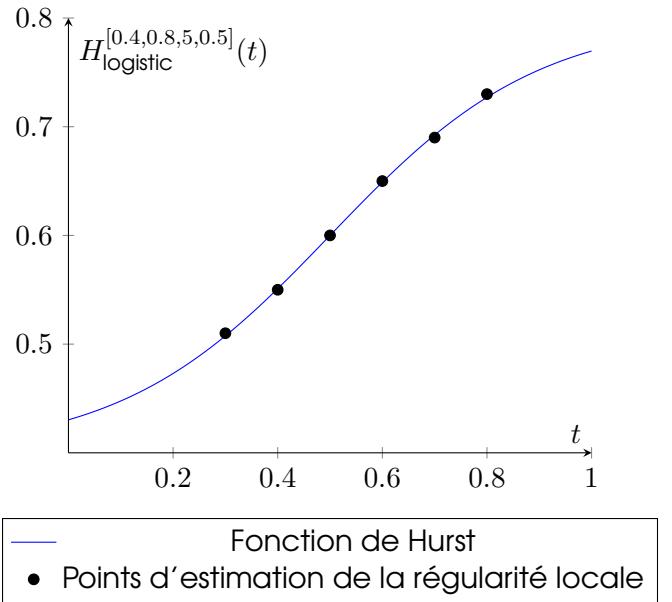


FIGURE 3.1 – Hurst Function : Logistic

3.1.3 Constante de Hölder

On décide de simuler des mouvements browniens multi-fractionnaires de constante de Hölder L_t identique sur tout le support.

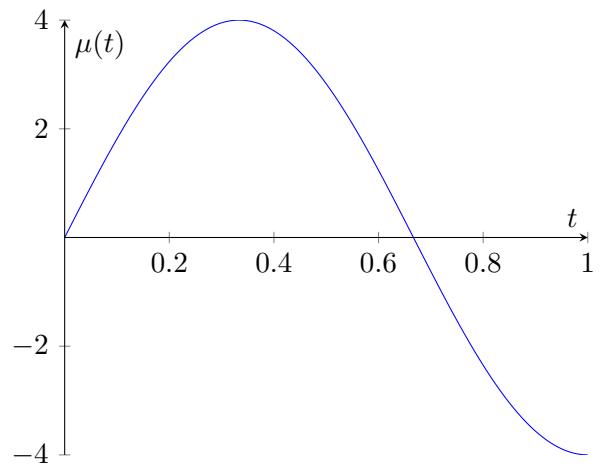
$$\forall t \quad L_t = 1$$

1. le raisonnement pour le choix du risque utilisé sera explicité en section 2.4

3.1.4 Moyenne

La fonction moyenne du processus utilisée pour la simulation est la suivante :

$$\mu : \begin{array}{ccc} [0, 1] & \longrightarrow & \mathbb{R} \\ t & \mapsto & 4 \cdot \sin\left(\frac{3}{2}\pi \cdot t\right) \end{array}$$



3.1.5 Noyau de la relation FAR(1)

On décide de simuler un FAR(1) basé sur un opérateur linéaire intégral :

$$X_{n+1} = \phi(X_n) + \varepsilon_{n+1}$$

avec : $\phi : f \mapsto \int_{\mathcal{T}} f(u) \beta(u, \cdot) du$

C'est une modélisation fréquente des FAR(1). Le noyaux que l'on considère dans l'opérateur intégral pour les simulations est le suivant :

$$\beta : \begin{array}{ccc} [0, 1]^2 & \longrightarrow & \mathbb{R} \\ (t, s) & \mapsto & \frac{9}{4}t\sqrt{s(1-s)} \end{array}$$

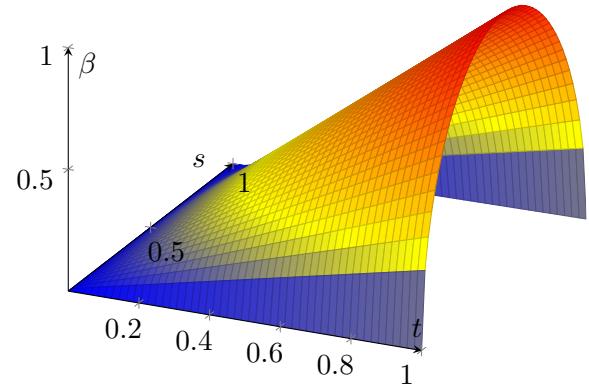


FIGURE 3.2 – Graphique du noyau intégral pour la relation FAR(1)

On notera que le noyaux utilisé pour la relation de FAR(1) est une fonction lisse. Ainsi il remplit aisément la condition pour que le FAR(1) hérite de la régularité du mouvement brownien multi-fractionnaire généré.

3.1.6 Nombre de courbes

Afin d'étudier le lien potentiel qu'il pourrait y avoir entre le nombre de courbes observées et le Δ optimal pour l'estimation de la régularité locale, on choisit plusieurs valeurs de nombres de courbes observées de telle sorte à avoir un « petit » et un « grand » nombre de courbes observées.

On choisit les valeurs suivantes concernant le nombre de courbes observées :

$$\vec{N} = [100, 200, 300, 400]$$

Ainsi on traite les cas de ce qu'on pourrait considérer comme la limite avant d'entrer dans un cas « sparse » (en terme du nombre d'observations de courbe), jusqu'à un nombre de courbe que l'on peut considérer important.

3.1.7 Nombre moyen de points observés par courbe

Le nombre de points observés sur la courbe X_n est défini comme étant la variable aléatoire M_n . Dans le cadre de notre simulation, M_n suit une loi de poisson de paramètre λ . Ainsi, $\mathbb{E}[M_n] = \lambda$ dans le cadre de notre simulation.

On effectue ainsi une simulation d'un échantillon de série temporelle FAR(1) par nombre moyen de points que l'on souhaite observer sur les courbes (λ). Afin de traiter différents cas, d'observation « dense » à observation « sparse » (dans le sens du nombre de points par courbe), on fait varier λ de 30 points par courbe en moyenne à 480 points. L'idée est de voir ensuite si il y a une relation entre le Δ^* et le fait que l'on ait λ petit, similaire ou grand par rapport à N .

3.1.8 Ensemble des Δ testés

On souhaite obtenir plusieurs graphiques avec Δ sur l'axe des abscisses afin de pouvoir étudier le comportement de diverses quantités, dont le risque euclidien, lorsque l'on fait varier Δ avec certains paramètres fixés (nombre de courbes observées, nombre moyen de points observés par courbe, ...). Toutefois plus on va considérer de Δ , et plus la simulation sera coûteuse. En effet, on a vu en section 2.3.2 que l'ordre de complexité de la simulation du mfBm est de $\mathcal{O}(\text{card } T^3)$, avec T les points où l'on doit évaluer nos $(X_i)_{1,n}$. Dans notre cas, le nombre de points considérés pour la simulation est :

$$\underbrace{\dim \vec{\Delta}}_{30} \times \underbrace{3}_{t_1/t_2/t_3} \times \underbrace{\dim \vec{t}}_6 + \underbrace{n_{Grid_f}}_{100} + \underbrace{\lambda}_{\leq 480} \leq \underbrace{640}_{fixe} + \underbrace{480}_{pts\ aleat} = 1120$$

Pour assurer un équilibre entre le nombre de points et les temps de simulation, nous choisissons 30 valeurs uniformément réparties entre 0.01 et 0.2 pour Δ . Au-delà de cette plage, la largeur des intervalles pour l'évaluation de la régularité devient disproportionnée par rapport à la taille du support, rendant inappropriée la notion de « régularité locale ».

$$\vec{\Delta} = [0.01 \cdots 0.2]_{30}$$

3.1.9 Bruit blanc

Une fois que l'on a simulé :

$$(X_i)_{1,n} \quad \text{vérifiant} \quad X_{n+1} = \phi(X_n) + \xi_n$$

on doit désormais reproduire l'erreur de mesure, pour cela chaque courbe est ensuite bruitée en rajoutant un bruit blanc :

$$\eta \sim \mathcal{N}(0, 0.04)$$

Il est important d'avoir un bruit blanc d'écart type d'un ordre de grandeur en dessous de celui des valeurs prises par le processus, sinon l'estimation serait mauvaise quoi qu'il arrive. En effet le bruit écraserait à lui tout seul toute l'information fine de régularité.

3.1.10 Résumé des Paramètres

	nombre de valeurs testées	de	jusqu'à	valeur
Δ	30			
λ	30			
N	4			
Erreur de mesure : (σ_η)		0.01	0.2	
nb simulations MC		30	480	
		100	400	
				0.2
				200

TABLE 3.1 – Hyper-paramètres de la simulation Monte-Carlo

3.2 Pré-lissage des données simulées

3.2.1 Les courbes obtenues

Voici l'exemple d'une courbe obtenue (courbe 1, 99^{eme} simulation de monte-carlo) :

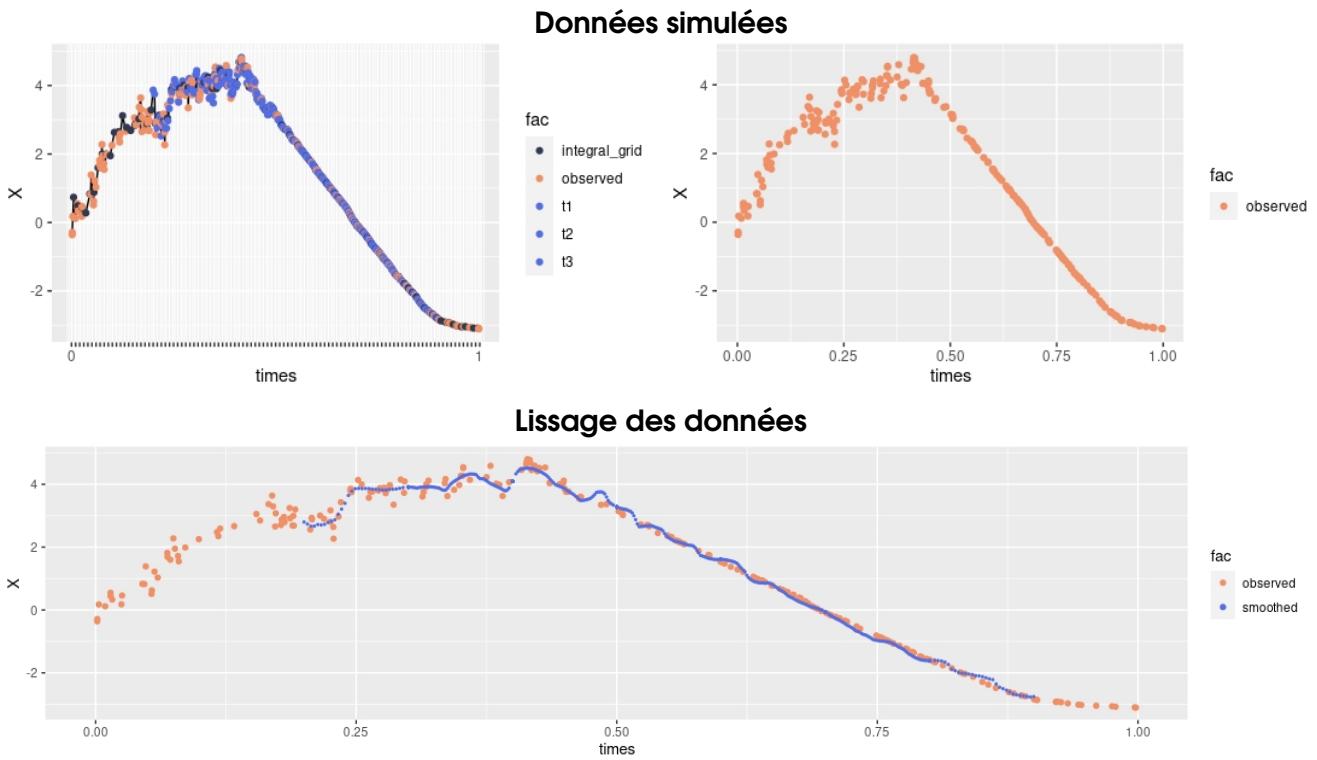


FIGURE 3.3 – Visualisation des données générées : $\lambda = 255$, $N = 200$, 30 valeurs de Δ

3.2.2 Pré-Lissage

Le pré-lissage des courbes a été fait en utilisant un lissage non paramétrique à noyaux². Comme mentionné dans la section 2.1, chaque courbe est lissée en utilisant une fenêtre par validation croisée avec la grille $\mathcal{H} = \{0.01 \dots 0.2\}_{50}$ avec pour métrique une estimation du risque quadratique $\mathbb{E} \left[|\hat{Y}_{(-i)} - Y|^2 \right]$. On ne regarde pas de fenêtre au delà de $\Delta = 0.2$ car il serait difficile de justifier

2. ainsi qu'un lissage utilisant des splines pénalisées, et des ondelettes mais on ne se concentrera sur les autres lissages qu'en Annexe B.4

pour l'estimation de la régularité que l'on lisse en regardant plus de 20% du support alors que la régularité évolue sur l'ensemble de l'intervalle.

L'obtention de la fenêtre de lissage a été réalisée en réalisant une validation croisée sur une grille de fenêtre étalées sur une échelle de puissance entre $h_{min} = 1/\hat{\lambda}^3$ et $h_{max} = 2/\hat{\lambda}$.

$$\begin{aligned}\mathcal{H} &= \{h_k, k \in \llbracket 1, K \rrbracket\} \\ \forall k \in \llbracket 1, K \rrbracket, \quad h_k &= h_{min} e^{-a \cdot k} \\ \text{avec} \quad a &= \frac{\log(h_{max}) - \log(h_{min})}{K} \\ \text{de valeur max} \quad h_K &= h_{max} = h_{min} e^{-a \cdot K} \\ \text{et} \quad K &= 30\end{aligned}$$

Cela est dû au fait que $h_{R_{\text{quadr}}}^* = \mathcal{O}_{\mathbb{P}}\left(\lambda^{-\frac{1}{1+2H_t}}\right)$ avec $0 < H_t < 1$, comme vu dans la section 2.2.4. De plus, on souhaite que dans notre fenêtre de lissage, en moyenne, se trouvent 2 points au minimum.

3.3 Détermination du Δ optimal à choisir pour l'estimation de la régularité

L'étude des courbes de risques obtenues :

$$\widehat{\mathcal{R}}(\Theta, \Delta) = \frac{1}{mc} \sum_{p=1}^{mc} \frac{\left\| \widehat{\Theta}[p] - \widetilde{\Theta}[p] \right\|_2^2}{\left\| \widetilde{\Theta}[p] \right\|_2^2} \quad (3.1)$$

où $[p]$ signifie que la quantité a été calculée à partir de la p^{eme} réPLICATION de monte carlo de la simulation.

indiquent la sélection du Δ par la procédure suivante³ :

— **Détermination de la fenêtre de pré-lissage :**

- calculer $\hat{\lambda}$, le nombre moyen de points par courbe et effectuer une validation croisée sur une grille d'échelle comprise entre $\frac{1}{\hat{\lambda}}$ et $\frac{1}{\hat{\lambda}^3}$
- Il est important de prendre en compte les « trous » lors du lissage à noyau des courbes. Il convient donc de ne pas sélectionner les fenêtres de lissage où le lissage à noyau a échoué sur une partie du support.⁴
- Pour les données « **sparses** » en nombre moyen d'observations par courbe ($\hat{\lambda} \leq 100 - 150$) : Bien lisser les courbes individuellement en déterminant la fenêtre de lissage par validation croisée. Pour les données « **denses** » en nombre moyen d'observations par courbes, économiser du temps de calcul est possible en effectuant un lissage global médian sur les fenêtres cross-validées des premières courbes comme indiquent les tables B.5 et B.7.

— **Choix du couple :**

3. Des détails sur la détermination de la procédure de sélection du Δ en annexe B.1

4. On pourra se référer à l'annexe B.3.2 pour plus de détails.

- **on dispose d'une information a priori sur une idée des zones plus ou moins régulières de données :** Il est conseillé d'estimer le couple dont la deuxième composante (c'est à dire soit θ_{12} soit θ_{23}) est celle qui pointe vers l'information la plus régulière. Même si les risques entre les couples $\Theta_{1 \rightarrow 2}$ et $\Theta_{2 \rightarrow 3}$ étaient proches, celui qui utilisait θ_{23} était souvent meilleur. Et-ce, que ce soit pour le risque relatif ou le risque absolu d'ailleurs. Dans le cadre de notre simulation θ_{23} correspondait à l'information plus régulière (car $t \mapsto H_t$ était croissante).
- **On ne dispose d'aucune information à priori sur la régularité :** Les deux couples ne disposent pas d'un risque qui diffère grandement (*cf* graphiques B.3.6)
- **Choix du Δ :** En se référant aux graphiques en annexe B.3.6, on recommande l'utilisation d'un Δ relativement grand vis-à-vis du support. Les graphes des risques indiquent que les Δ de la taille de 10 à 20% du support étaient à risques relatifs équivalents, et ce pour les différents niveaux de régularité (0.5 à 0.7). C'est pourquoi dans l'optique d'une estimation de la régularité « locale » on recommande plutôt de se situer aux environs de 10-15% du support pour bénéficier du plateau autour, tout en restant « local ».⁵
- **Si l'on tient au risque euclidien (non relatif) :** Il est recommandé pour le risque euclidien non relatif après avoir analysé les graphes B.5 en ayant retiré les points extrêmes, de prendre un Δ proche de 0.01 pour les zones moins régulières ($H_t < 0.6$) et Δ proche de 0.2 pour les zones plus régulières ($H_t \geq 0.6$). Si l'on ne possède aucune information, il est possible d'effectuer un premier lissage qui estime la régularité avec $\Delta = 0.01$, puis avec cette estimation ré-effectuer un lissage en évaluant cette fois la régularité avec le Δ qui aura été recommandé dans cette section pour le risque euclidien (non relatif).⁶

3.4 Discussion

Il est important de garder en tête le modèle dans lequel on s'est placé pour étudier le comportement du Δ . La recommandation qui est faite pour la sélection du Δ est valable pour :



- FAR(1) construit à partir d'un $\text{mfBm}(H, L)$
- la régularité donnée par $H(t)$, qui dans notre cas : $H \in \mathcal{C}^\infty([0, 1])$
- le noyau de la relation auto-régressive β est une fonction de classe \mathcal{C}^∞ sur $]0, 1]$ et continue en 0
- La dérivée de H est $H' : t \mapsto \frac{2e^{-5(t-0.5)}}{(1+e^{-5(t-0.5)})^2}$, la variation maximale de la régularité est atteinte en $\underset{t \in [0, 1]}{\operatorname{argmax}} H'(t) = \frac{1}{2}$ avec $H'(\frac{1}{2}) = \frac{1}{2}$
- La régularité est monotone et strictement croissante sur $[0, 1]$
- la constante « locale » de Hölder $L : t \mapsto L_t$ est en réalité constante sur l'ensemble du support dans le cadre de nos simulations

Trop s'éloigner de ces hypothèses pourrait demander d'analyser de nouveau le comportement du Δ dû aux propriétés de certaines de ces quantités qui auraient pu influencer le résultat.

5. sera détaillé en discussion

6. Lorsque $\Delta = 0.2$ est optimal, pour le risque euclidien non relatif à la norme de la cible, $\Delta = 0.01$ n'engendre pas —dans le cadre de la simulation— la pire valeur de risque possible, là où 0.2 donne le pire risque lorsque 0.01 est optimal. Il est donc plus prudent d'effectuer une première estimation avec $\Delta = 0.01$.

C'est pourquoi on recommande avec des affirmations du type « vers le point le plus régulier » plutôt qu'avec la position.

Un exemple de l'importance de ces hypothèses et du cadre de simulation est que le Δ optimal obtenu pour le risque euclidien relatif était dans la grande majorité des cas $\Delta^* = 0.2$. Ce qui peut sembler surprenant pour l'estimation d'une régularité *locale*. Toutefois, la fonction de Hurst est très régulière et varie même lentement sur la majorité du support. Ainsi il est plus raisonnable de penser que l'on peut aller chercher de l'information plus loin, sans trop perdre l'information locale.

De plus, il est à noter que le choix du risque apportait une conclusion différente sur la détermination de Δ à choisir en pratique. Le choix du risque euclidien relatif à la norme de la cible semblait être un choix qui fait plus sens à la fois d'un point de vue méthodologique (étant donné que la norme de la cible varie avec Δ) mais aussi par chance qui se trouve être un choix qui mène à une règle de détermination assez simple du Δ avec une courbe de risque qui semble mieux se comporter.

On pourra enfin noter les différents pics présents à certaines valeurs de Δ dans les graphes de risque. Après investigation des courbes des réplications de monte-carlo présentant des cas extrêmes (comme par exemple un risque euclidien non relatif s'élevant jusqu'à 80 — comme sur la figure B.2), il semblerait que cela soit dû à l'estimation de $X(t_1), X(t_3)$ non loin, ou bien même dans un « trou d'observation ». Le praticien doit donc faire attention lors de l'estimation de la régularité si il obtient des résultats peu concluants, si les courbes initiales possèdent peu de points (ou pas de points) proches des temps $t_1(\Delta), t_2(\Delta)$ ou $t_3(\Delta)$. Auquel cas, pour estimer la régularité deux pistes se présentent :

- estimer la régularité par les statistiques d'ordre, pour le voisinage des points problématiques via la méthode de Golovkine et al. Il s'agit donc de la méthode recommandée pour les cas pathologiques.
- (**A**conjuncture) Investiguer sur le potentiel de lisser via une base de fonction, qui ne détruit pas trop les informations de régularité, comme la base d'ondelettes. Ce qui rend la base d'ondelettes particulièrement intéressante est d'ailleurs explicité en annexe A.7.

Chapitre 4

Application

Contents

4.1	Généralités	29
4.1.1	estimation adaptative informelle	29
4.1.2	Les estimateurs	30
4.2	Contrôle de la procédure sur des données simulées	31
4.2.1	Détermination du Δ en utilisant la procédure	31
4.2.2	Estimation de la fonction moyenne	32
4.2.3	Conclusion sur la qualité de la détermination du Δ via l'utilisation de la procédure	32
4.3	Application sur les données réelles de courbes de charge éolienne et photovoltaïque	33
4.3.1	Présentation des jeux de données	33
4.3.2	Pré-traitement des données	34
4.3.3	Pré-lissage et estimation de la régularité locale	34
4.3.4	Estimation de la fonction moyenne	34
4.4	Conclusion	34

4.1 Généralités

4.1.1 estimation adaptative informelle

Les motivations de l'obtention de la régularité étaient en partie de pouvoir mieux estimer les quantités qui nous intéressent dont la fonction moyenne du processus, ainsi que son opérateur de covariance. Ce qui est à la fois important pour l'analyse (via l'interprétation de la base ACP déterminée par la covariance) et pour la prédiction. On peut alors se demander si il existe des estimateurs de la moyenne et de la covariance prenant en compte la régularité locale. C'est ce qu'affirme les théorèmes suivants :



demander à Hassan la dernière version de son papier car la partie d'estimation adaptative a beaucoup changé

Théorème (Estimateurs de la moyenne et de la covariance — informel (8))

Il est possible en lissant les observations par méthode à noyaux avec une largeur de bande spécifique à l'objet que l'on souhaite estimer, de dériver des estimateurs de la moyenne et de la covariance qui convergent. La largeur de bande optimale pour l'objet que l'on souhaite estimer est celle qui minimise un risque qui effectue un compromis biais-variance, qui dépend de la régularité locale du processus, en pénalisant les largeurs de bande menant à des "trous" dans les fonctions lissées. On parle d'« estimation adaptative ».

Cependant, bien qu'une largeur de bande optimale existe, elle est inconnue. Il est donc important de savoir si le praticien peut l'estimer, et avec quelle précision (c'est à dire à quel point l'estimateur sera biaisé ou non). C'est ce que nous affirme le théorème suivant :

Théorème (expression de la largeur de bande optimale — informel (8))

Sous certaines hypothèses de régularité du processus, et d'indépendance des temps observés, la largeur de bande optimale peut être approchée (avec forte probabilité de bonne approximation) par une expression ne dépendant que du nombre de courbes observées, du nombre moyen de temps observés par courbe, et de la régularité locale du processus. Ce biais de l'estimateur de la fonction moyenne est alors contrôlé en fonction de ces mêmes quantités.

Sous des hypothèses un peu plus fortes sur le nombre d'observations par courbe, et le nombre de courbe on dispose de résultats similaires pour l'estimateur de la covariance.

Enfin, on peut se demander ce qu'il en est des estimateurs dans le cadre où l'on dispose de la dépendance dans les données (ce qui est le cas pour les données éoliennes notamment). Ce cas est traité par le théorème suivant dérivé par MPV :

Théorème (Estimation adaptative de séries temporelles fonctionnelles — informel (17))

On peut estimer la régularité d'une série temporelle de données fonctionnelles à condition que la mémoire temporelle de la série soit courte. (La décroissance de la dépendance temporelle doit être au moins aussi rapide qu'une décroissance géométrique)

4.1.2 Les estimateurs



Plus de précisions théoriques sur les estimateurs utilisés sont disponibles dans l'annexe A.5.

On utilisera les estimateurs considérés par MPV (17) pour la fonction moyenne, la fonction d'autocorrélation ainsi que l'auto-covariance qui sont définis de la façon suivante :

On notera l'indicatrice de l'évènement « il y a suffisamment de points autour de t pour le lissage à noyaux de la courbe i » : $\pi_i(t | h)$ ¹.

Le nombre d'observations exploitables dans la bande J_Δ est donc : $P_N^* = P_N(t | h_\mu^*(\Delta^*)) = \sum_{i=1}^N \pi_i(t | h_\mu^*(\Delta^*))$, et de manière équivalente on obtient le nombre d'observations exploitables pour l'estimation conjointe sur les courbes i et $i + \ell$ en t : $P_{i,\ell}^*$.

1. On ne demande la présence que d'un unique point, un autre seuil plus strict peut être fixé. Il a été aperçu empiriquement que l'amélioration de l'estimation n'est pas significative, un point suffit pour éviter la dégénérescence.

$$\widehat{\mu_{\text{adapt}}^*}(t) = \frac{1}{P_N^*} \sum_{i=1}^N \left[\pi_i \cdot \widehat{X}_i^{[NW]} \right] (t | h_\mu^*(\Delta^*)) \quad (4.1)$$

$$\widehat{\gamma_{\text{adapt}}^*}(s, t, \ell) = \frac{1}{P_{i,\ell}^*} \sum_{i=1}^{N-\ell} \left[\left[\pi_i \cdot \widehat{X}_i^{[NW]} \right] (s | h_\mu^*(\Delta^*)) \left[\pi_{i+\ell} \cdot \widehat{X}_{i+\ell}^{[NW]} \right] (t | h_\mu^*(\Delta^*)) \right] \quad (4.2)$$

On établit désormais si la procédure de détermination du Δ établie en section (3.3) permet bel et bien d'obtenir une bonne estimation des paramètres de régularité sur de nouvelles données générées.

4.2 Contrôle de la procédure sur des données simulées

Il est important que les données de test n'aient pas été utilisées pour déterminer la procédure. Pour vérifier que la procédure fonctionne comme prévu, on génère de nouvelles données aux caractéristiques suivantes :

Nombre moyen d'observations par courbe	symbole	variation	valeur
Nombre moyen d'observations	λ	sparse	80
		moyen	180
		dense	300
Nombre de courbes	N	sparse	100
		moyen	200
		dense	300
Nombre de simulations de monte carlo	mc		200
Points d'estimation de la régularité	\vec{t}		[0.3, 0.6, 0.8]
Fonction de Hurst	H_1	même que sim	$t \mapsto H_{\text{logistic}}^{[0.4, 0.8, 5, 0.5]}(t)$
	H_2	pente plus abrupte	$t \mapsto H_{\text{logistic}}^{[0.4, 1, 16, 0.6]}(t)$
Valeurs de régularité testées	$H_1(\vec{t})$	même que sim	[0.51, 0.65, 0.73]
	$H_2(\vec{t})$	pente plus abrupte	[0.40, 0.70, 0.98]
constante de régularité locale	L_t	constante	3

TABLE 4.1 – Paramètres de simulation des données de test

Puis on estimera la régularité de chacune de ces courbes, et on comparera les résultats obtenus avec les valeurs théoriques.

4.2.1 Détermination du Δ en utilisant la procédure

Nous choisissons, en accord avec la section précédente, le Δ suivant :

$$\Delta^{proc} = 15\% \text{ du support}$$

4.2.2 Estimation de la fonction moyenne

On compare désormais les fonctions moyennes estimées de façon adaptative avec la fonction moyenne théorique en ayant effectué une estimation de la régularité avec $\Delta = \Delta^{proc}$ et $h = h^{[cv(\hat{\lambda})]}$.

 (graphe de la fonction moyenne estimée contre la véritable fonction moyenne)

4.2.3 Conclusion sur la qualité de la détermination du Δ via l'utilisation de la procédure

En fonction de l'application souhaitée, il peut être plus adapté de s'intéresser à un critère global qui prend en compte l'ensemble de l'intervalle T , ou un critère local qui a d'autant plus de sens du fait que l'on prend en compte la régularité locale des trajectoires.

De la même manière que dans l'annexe ??, on retire sur les 200 échantillons de monte carlo simulés, les échantillons « extrêmes » où une observation peut potentiellement faire exploser le risque (auquel cas on rappelle que l'on conseille la méthode de Golovkine par la statistique d'ordre si les résultats sont insatisfaisants sur le voisinage problématique). On note mc_{Ret} le nombre de simulations de monte carlo retenues dans le calcul après filtrage des « extrêmes ».

Les risques affichés par la suite sont pour $N = 200$, $\lambda = 180$.² On considère l'estimation du couple

$$\widetilde{\Theta}_{\overset{1}{2} \rightarrow 3} = \begin{bmatrix} \widetilde{\theta}(t_1, t_3) \\ \widetilde{\theta}(t_2, t_3) \end{bmatrix}$$

4.2.3 □ A) Qualité d'estimation du couple d'incrément $\Theta_{\overset{1}{2} \rightarrow 3}$: risque relatif

t	H_t	$\widehat{\mathcal{R}}(\Theta, \Delta)$	mc_{Ret}	$\text{med } \widehat{\mathcal{R}}_{mc}(\Theta, \Delta)$	$V[\widehat{\mathcal{R}}_{mc}(\Theta, \Delta)]$
Moins Régulier	$H_1 : 0.51$	$2.9 \cdot 10^{-3}$:	$2.3 \cdot 10^{-3}$	$5.0 \cdot 10^{-6}$
	$H_2 : 0.40$	$7.8 \cdot 10^{-3}$:	$5.1 \cdot 10^{-3}$	$1.1 \cdot 10^{-4}$
Inflexion	$H_1 : 0.65$	$9.0 \cdot 10^{-5}$:	$4.9 \cdot 10^{-5}$	$1.2 \cdot 10^{-8}$
	$H_2 : 0.70$	$1.5 \cdot 10^{-3}$	198 (/200)	$5.0 \cdot 10^{-5}$	$2.3 \cdot 10^{-4}$
Plus régulier	$H_1 : 0.73$	$1.1 \cdot 10^{-4}$:	$6.7 \cdot 10^{-5}$	$1.5 \cdot 10^{-8}$
	$H_2 : 0.98$	$2.3 \cdot 10^{-3}$:	$7.8 \cdot 10^{-5}$	$3.1 \cdot 10^{-4}$



On peut constater que les risques relatifs observés sur l'estimation du couple $\Theta_{\overset{1}{2} \rightarrow 3}$ pour de tous nouveaux paramètres de simulation sur lesquels la procédure de sélection du Δ n'a pas été déterminée (H_2) sont aussi très bons ce qui donne confiance en la procédure.

2. Les tableaux de risque pour les autres valeurs mentionnées précédemment sont disponibles en annexe

4.2.3 □ B) Qualité d'estimation de la fonction moyenne : Critère Global

t	H_t	$\widehat{\mathcal{R}}$ avec $\mathcal{R} =$ MISE	$\mathbb{V}[\widehat{\mathcal{R}}_{mc}]$	$med(\widehat{\mathcal{R}}_{mc})$	$\min \widehat{\mathcal{R}}(t)$	$\max \widehat{\mathcal{R}}(t)$
Moins régulier	$H_1 : 0.51$ $H_2 : 0.40$					
Inflexion	$H_1 : 0.65$ $H_2 : 0.70$					
Plus régulier	$H_1 : 0.73$ $H_2 : 0.98$					

4.2.3 □ C) Qualité d'estimation de la fonction moyenne : Critère Local

t	H_t	$\widehat{\mathcal{R}}(t)$ avec $\mathcal{R} =$ MSE	$med(\widehat{\mathcal{R}}_{mc}(t))$	$\mathbb{V}[\widehat{\mathcal{R}}_{mc}(t)]$
Moins régulier	$H_1 : 0.51$ $H_2 : 0.40$			
Inflexion	$H_1 : 0.65$ $H_2 : 0.70$			
Plus régulier	$H_1 : 0.73$ $H_2 : 0.98$			

Les résultats permettent de conclure sur la méthode de sélection du Δ .

4.3 Application sur les données réelles de courbes de charge éolienne et photovoltaïque

4.3.1 Présentation des jeux de données

Les données que l'on traite sont des courbes de charge provenants de différents moyens de production : éolien ou photovoltaïque. Nous modélisons le jeu de données de la façon suivante :

Données éoliennes

support	identifié comme $\mathcal{T} = [0, 1]$
modèle fonctionnel	$\forall i \in I \quad E_i : \begin{array}{ccc} \Omega \times [0, 1] & \longrightarrow & \mathbb{R}_+ \\ (\omega, t) & \mapsto & e_i(t) \end{array}$
indices	éolienne individuelle : $I = \{ \text{id de l'éolienne} \}$
observations	$\hat{E} = \{ (T_i[m], Y_i[m]) : i \in \llbracket 1, N \rrbracket, m \in \llbracket 1, M_i \rrbracket \} : Y_i[m] = E_i(T_i[m]) + \eta_i(T_i[m])$
erreur de mesure	η_i : indépendants 2 à 2
dépendance	données indépendantes car piochées aléatoirement sur des éoliennes de géo-localisation éloignées.

Données Photovoltaïques

support	identifié comme $\mathcal{T} = [0, 1]$
modèle fonctionnel	$V_i : \begin{array}{ccc} \Omega \times [0, 1] & \longrightarrow & \mathbb{R}_+ \\ (\omega, t) & \mapsto & v_i(t) \end{array}$
indices	panneau individuel : $I = \{ \text{id du panneau} \}$
observations	$\hat{V} = \{ (T_i[m], Z_i[m]) : i \in \llbracket 1, N \rrbracket, m \in \llbracket 1, M_i \rrbracket \} : Z_i[m] = V_i(T_i[m]) + \eta_i(T_i[m])$
erreur de mesure	η_i : indépendants 2 à 2
dépendance	données dépendantes car proches géographiquement.

4.3.2 Pré-traitement des données

4.3.3 Pré-lissage et estimation de la régularité locale

On effectue un lissage à noyaux en utilisant une fenêtre de lissage déterminée par validation croisée, dont la grille suit la même échelle que celle établie dans la section 2.1.

 (Montrer graphe de courbes non lissées et lissées)

On utilise la procédure de détermination du Δ déterminée auparavant pour estimer la régularité locale des données.

4.3.4 Estimation de la fonction moyenne

On détermine la fonction moyenne du processus en utilisant la méthode d'estimation adaptative proposée par (8).

4.3.4 □ A) Interprétation

4.4 Conclusion

Annexe A

Détails techniques et théoriques

Contents

A.1	Données fonctionnelles : formellement	35
A.2	Régularité Locale	38
A.2.1	Définition formelle des fonctions Höldériennes sur un intervalle & régularité locale formellement	38
A.2.2	Des processus Höldériens ?	38
A.3	Dépendance Faible et LGN version faible	39
A.4	Continuité de Kolmogorov	43
A.5	Estimation adaptative	44
A.5.1	Estimation adaptative de la fonction moyenne	44
A.5.2	Estimation adaptative de l'opérateur de covariance	45
A.5.3	Estimation adaptative de l'auto-corrélation des séries temporelles fonctionnelles	46
A.6	Mouvement Brownien	46
A.6.1	Mouvement Brownien	46
A.6.2	Mouvement Brownien Fractionnaire	46
A.6.3	Mouvement Brownien multi-fractionnaire	46
A.7	Théorie de la base d'ondelettes	46
A.7.1	Transformée en ondelettes	46
A.7.2	Propriétés principales des ondelettes	47
A.7.3	Discussion : potentiel du lissage ondelette pour le préissage	48

A.1 Données fonctionnelles : formellement

A.1.0 A) Définition formelle

Pour éviter d'alourdir les notations, on se place dans le cas où les fonctions sont à valeurs dans \mathbb{R} et à support sur un intervalle fermé I de \mathbb{R} . Toutefois, on peut très bien considérer des fonctions à valeurs dans \mathbb{R}^d et à support sur un compact K de \mathbb{R}^p sans perte de généralités.

Définition 2 (données fonctionnelles) On appelle données fonctionnelles, un échantillon $(x_i)_{1,n}$ de fonctions continues $x_i : I \rightarrow \mathbb{R}^d$ issues d'un processus X défini comme ci-dessous :

$$\begin{aligned} X : \Omega &\longrightarrow \mathcal{C}(I, \mathbb{R}) \\ \omega &\longmapsto X(\omega) = x \end{aligned}$$

A.1.0 □ B) Résultats importants pour les données fonctionnelles



Cette section motive le souhait de bien estimer la régularité locale et la fonction moyenne en prenant compte de la régularité locale car utile pour l'estimation de la covariance. Cette covariance est centrale en données fonctionnelles. **Le lecteur n'a pas besoin de lire cette partie pour comprendre le sujet** mais elle reste un complément intéressant pour les motivations et la compréhension de l'utilisation pratique des données fonctionnelles.

Si le stage se concentre sur l'estimation du couple de risques quadratiques utilisés pour estimer la régularité locale et son application pour l'estimation de la moyenne, la bonne estimation de la covariance des données fonctionnelles est essentielle. En effet, les résultats énoncés ci-dessous permettent de travailler sur des données fonctionnelles à partir de nombres réels que l'on sait bien mieux manipuler. Ils permettent enfin d'approximer les objets de notre modèle (les fonctions continues) vivant dans un espace vectoriel de dimension infinie en établissant une base qui représente au mieux, avec un nombre fixé de composantes nos données.

On énonce désormais le théorème central de l'analyse de données fonctionnelles qui n'est autre que la décomposition dans la base FPCA de notre processus.

Remarque : on notera que dans le cadre des données fonctionnelles, on ne travaille pas de façon générale avec la covariance :

$$C_X : (s, t) \mapsto \mathbb{E} [[X - \mu] (s) \cdot [X - \mu] (t)]$$

On travaille plutôt avec l'**opérateur** de covariance :

$$c : \begin{array}{ccc} \mathbb{L}^2 & \longrightarrow & \mathbb{L}^2 \\ f & \longmapsto & \int_I f(u) C_X(u, \cdot) du \end{array}$$

C'est parce que cet opérateur est linéaire continu (car Hilbert-Schmidt donc borné pour la norme d'opérateur) symétrique semi-défini positif (pour le produit scalaire de \mathbb{L}^2) et que l'on peut donc en faire une décomposition spectrale sur une base orthonormale de vecteurs propres associés à des valeurs propres positives. Cette décomposition est à la base des approximations que le praticien effectuera ainsi qu'à la base de la dérivation de nombreux théorèmes et propriétés.

Etant donné que l'on traite des données fonctionnelles, on considère la géométrie usuelle de $\mathbb{L}^2(\mathbb{R}, \lambda)$ et on note ainsi

$$\langle \cdot | \cdot \rangle_{\mathbb{L}^2} : \begin{array}{ccc} \mathbb{L}^2 \times \mathbb{L}^2 & \longrightarrow & \mathbb{R} \\ (f, g) & \longmapsto & \int f(u)g(u) d\lambda(u) \end{array}$$

le produit scalaire que l'on considère pour manipuler les données fonctionnelles.

Théorème 1 (Karhunen-Loeve)

référence : (15, pages : 238-239-241)

Hypothèses :

- $X \in \mathbb{L}^2(\Omega, \mathcal{C}(I, \mathbb{R}))$
- covariance : $C : \begin{array}{ccc} \mathbb{L}^2(\Omega, \mathcal{C}(I, \mathbb{R})) & \longrightarrow & \mathcal{C}(I^2, \mathbb{R}) \\ X & \mapsto & C_X \end{array}$
- ie : $C_X : (s, t) \mapsto C_X(s, t)$ est continue
- * opérateur covariance $c_X[\cdot] : \begin{array}{ccc} \mathcal{C}(I, \mathbb{R}) & \longrightarrow & \mathcal{C}(I, \mathbb{R}) \\ f & \mapsto & \int_I f(s) C_X(s, \cdot) ds \end{array}$
- valeurs propres ordonnées : $\forall p \geq 1, \lambda_{p+1} \leq \lambda_p \quad \lambda_p, \lambda_{p+1} \in \text{sp}(c_X)$
- * on pose $\vec{sp}_{\|\cdot\|}^{[1,p]}(c_X) \stackrel{\text{déf}}{=} \left\{ \phi_k \in \vec{sp}_{\|\cdot\|}^{[1,p]}(c_X) \text{ associé à } \lambda_k, k \in \llbracket 1, p \rrbracket \right\}$

alors :

- $\forall p \geq 1 \quad \underset{u_k \in \mathcal{C}(I, \mathbb{R})}{\operatorname{argmin}} \mathbb{E} \left\| X - \sum_{k=1}^p \langle X - \mu | u_k \rangle_{\mathbb{L}^2} u_k \right\|^2 = \vec{sp}_{\|\cdot\|}^{[1,p]}(c_X)$
- $$X = \mu + \sum_{k=1}^{+\infty} \langle X - \mu | \phi_k \rangle \phi_k$$

avec $\phi_k \in \vec{sp}_{\|\cdot\|}^{[1,p]}(c_X)$

✿ preuve informelle. La covariance est un opérateur bilinéaire symétrique défini positif, on peut donc appliquer le théorème de Mercer (équivalent du théorème spectral) qui nous donne une base orthonormale de \mathbb{L}^2 sur laquelle on va décomposer notre processus centré. □

Remarque : pour pouvoir ordonner les valeurs propres dans l'ordre décroissant, et sélectionner les composantes principales les plus informatives, il faut pouvoir réarranger l'ordre de la somme. Pour cela il faut que les valeurs propres forment une famille sommable, une condition suffisante et souvent utilisée est que $\mathbb{E}\|X\|^2 < \infty$

Remarque : la propriété de la section précédente sur l'aspect économe de la base FPCA découle directement de l'assertion

$$\forall p \geq 1 \quad \underset{u_k \in \mathcal{C}(I, \mathbb{R})}{\operatorname{argmin}} \mathbb{E} \left\| X - \sum_{k=1}^p \langle X - \mu | u_k \rangle u_k \right\|^2 = \vec{sp}_{\|\cdot\|}^{[1,p]}(c_X)$$

dans le théorème de Karhunen-Loeve.

A.2 Régularité Locale

A.2.1 Définition formelle des fonctions Höldériennes sur un intervalle & régularité locale formellement

Il est important de se référer aux définitions formelles pour garder à l'esprit les propriétés des objets que l'on manipule. Voici en version formelle la différence entre les différents « modes de régularité » traités en section 2.2.1 (Ce qu'on entend par régularité locale)

— Continuité :

$$(\forall \varepsilon > 0) (\forall x) (\exists \delta_{\text{blue}} > 0) (\forall y) |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon$$

— Uniforme Continuité :

$$(\forall \varepsilon > 0) (\exists \delta > 0) (\forall x, y) |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon$$

— Lipschitz :

$$\exists L_I \quad (\forall x, y \in I) \quad |f(x) - f(y)| < L_I |x - y|$$

— Hölder :

$$\exists \alpha \in (0, 1] \quad \exists L_{\alpha(I)} \quad (\forall x, y \in I) \quad |f(x) - f(y)| < L_{\alpha(I)} |x - y|^{\alpha}$$



une fonction lipschitz est une fonction Holderienne avec $\alpha = 1$

— Localement Hölder :

$$\forall x_0 \in I \quad \exists V \in \mathcal{V}(x_0) \quad \text{tq} \quad \exists \alpha(x_0), L_{\alpha(x_0)}(x_0) \quad \begin{cases} (\forall x \in V) \quad |f(x) - f(x_0)| < L_{\alpha(x_0)} |x - x_0|^{\alpha(x_0)} \\ 0 < \alpha(x_0) \leq 1 \end{cases}$$

A.2.2 Des processus Höldériens ?

Nous avons mentionné que les processus auxquels on allait s'intéresser étaient les processus localement Höldériens de paramètres $(\alpha(t), L_{\alpha}(t))$. Ce n'est pas tout à fait vrai. Si le coeur de ce que l'on considère sont bel et bien les processus Höldériens, on élargit encore plus la classe des processus que l'on considère en considérant les processus qui sont **presque** Höldériens.



Qu'est ce qu'on entend exactement par presque Höldérien ?

pour u et v dans un voisinage de t de diamètre Δ :

Ce que l'on demandait pour un processus X est que pour tout u, v dans un voisinage de t de diamètre Δ , il existe L_t et H_t telle qu'on ait :

$$\theta(u, v) \underset{\text{déf}}{\equiv} \mathbb{E} \left[|X(u) - X(v)|^2 \right] \leq L_t^2 |u - v|^{2H_t}$$

on peut alors retrouver la régularité du processus comme un processus Höldérien de paramètres (H_t, L_t) d'après le théorème de Continuité de Kolmogorov.

en réalité il suffit que $\theta(u, v)$ soit suffisamment proche d'un processus localement Höldérien de paramètres (H_t, L_t) et que l'on puisse contrôler l'écart entre les deux. Cet écart dépend de Δ et de la régularité. C'est ce qu'affirme les deux hypothèses suivantes qui sont en fait les hypothèses de régularité qui sont considérées par MPV(17).

$$|\theta(u, v) - L_t^2|u - v|^{2H_0}| \leq S_t^2|u - v|^{2H_0}\Delta^{2\beta_0}$$

(17, H6)

$$\left| \nu_2 \left(\nabla^\delta X(u) - \nabla^\delta X(v) \right)^2 - L_{\delta,t}^2 |u - v|^{2H_\delta} \right| \leq S_{\delta,t}^2 |u - v|^{2H_\delta} \Delta^{2\beta_\delta}$$

(17, D1-7)

On remarquera que si le processus est localement Höldérien, alors on a un contrôle optimal de l'écart entre $\theta(u, v)$ et $L_t^2|u - v|^{2H_t}$.

L'auteur saura donc reconnaître, que bien que ce qui ait été exposé ne soit pas la forme exacte, cela ne change rien à l'idée générale. De plus, cela alourdirait considérablement la rédaction et rendrait la compréhension bien plus difficile de l'objectif du stage.

A.3 Dépendance Faible et LGN version faible

Il y a dans un premier temps ce qu'on appelle la dépendance « forte », comme la dépendance dite de « α -mixing » comme définie dans (24) :

Définition (α -mixing) une suite $X = (X_i)_{i \geq 0}$ de variables aléatoires est dite α -mixing si pour tout $n \in \mathbb{N}$

$$\alpha(n) \xrightarrow{n \rightarrow \infty} 0$$

$$\text{avec : } \alpha(n) = \sup_k \{ |\mathbb{P}[A \cap B] - \mathbb{P}[A]\mathbb{P}[B]| \quad | \quad A \in \sigma(X_{1:k}), B \in \sigma(X_{k+n:\infty}) \}$$

en d'autres termes, la « dépendance » ($|\mathbb{P}[A \cap B] - \mathbb{P}[A]\mathbb{P}[B]|$) entre les variables aléatoires X_k et X_{k+n} tend vers 0 lorsque n tend vers l'infini.

Ce point de vue est « fort » dans le sens où l'on manipule directement les tribus, et que l'on regarde leur degré d'indépendance via la mesure de probabilité. Il ne s'agit pas de l'approche considérée par MPV qui est plus faible, en se reposant non pas sur l'indépendance des tribus engendrées par la série temporelle mais en exploitant la qualité d'approximation de la série temporelle que l'on étudie par un autre processus, indépendant de la série temporelle étudiée à partir d'un certain rang. La définition de dépendance temporelle est alors dite « faible ». Le point de vue faible offre un comportement plus sympathique pour l'aspect *local* dans l'estimation de la régularité : qui est le cœur de l'approche de MPV.

Les processus qui nous intéressent et ceux auxquels on va se limiter dans un premier temps sont les processus causaux. Comme dans le cas réel, on peut étudier les séries temporelles en posant l'opérateur :

$$B : x_n \mapsto x_{n-1}$$

et la relation de dépendance encodée par :

$$X_{n-1} = \phi(X_n) + \xi_n \quad \phi \text{ linéaire}$$

Si le processus est inversible, on peut écrire X_n comme le développement en série entière suivant :

$$\begin{aligned}
X_n &= \phi \circ B(X_n) + \xi_n \\
[I - (\phi \circ B)](X_n) &= \xi_n \\
X_n &\underset{\|\phi \circ B\| < 1}{=} [\phi \circ B]^{-1}(\xi_n) \\
X_n &\underset{\sum \epsilon}{=} \sum_{k=0}^{\infty} \underbrace{[\phi \circ B]^k}_{\phi^k \circ B^k}(\xi_n)
\end{aligned}$$

En effet, les opérateurs ϕ et B commutent car :

$$x = (x_n)_{n \in \mathbb{Z}} = (\dots, x_0, x_1, x_2, \dots)$$

$$\phi(x) = (\dots, \phi(x_0), \phi(x_1), \phi(x_2), \dots)$$

on a bien $\phi \circ B = B \circ \phi$

$$\begin{aligned}
\phi \circ B(x) &= (\dots, \phi \circ B(x_0), \phi \circ B(x_1), \phi \circ B(x_2), \dots) \\
&= (\dots, \phi(x_{-1}), \phi(x_0), \phi(x_1), \dots) \\
&= (\dots, B(\phi(x_0)), B(\phi(x_1)), \dots) \\
&= B(\phi(x))
\end{aligned}$$

et ainsi

$$X_n = \sum_{k=0}^{\infty} \phi^k(\xi_{n-k}) = f(\dots \xi_{n-k} \dots \mid k \geq 0)$$

Définition 3 (copie indépendante) on appelle V une copie indépendante de U si $V \sim U \sim \mathcal{L}$ ET $V \perp\!\!\!\perp U$.

i.e : U et V sont de même loi et indépendantes. Exemple : même étude réalisée à deux laboratoires différents avec des patients différents.

soit maintenant

$$\Xi_n \underset{\text{déf}}{\equiv} \{\xi_n\}_{-\infty:n} \text{ la suite de bruits blancs dans l'inversion précédente}$$

on va regarder le niveau de dépendance de X_n à l'ordre a . pour cela nous allons commencer par effectuer une copie indépendante du bruit pour chaque ordre a que nous allons regarder. L'idée est que l'on ne va garder que les a derniers termes de notre processus dont on souhaite savoir jusqu'à combien de termes la dépendance avec le passé est significative. Les termes qui les précèdent seront remplacés par une copie indépendante qui n'a donc pas pu avoir d'influence sur les a derniers termes (par copie *indépendante*) : les termes que l'on a conservé ne peuvent pas dépendre de la copie.

$$\begin{aligned}
\Xi^{[1]} &= \underset{\perp\!\!\!\perp}{\text{copy}} \Xi \\
&\vdots \quad \vdots \\
\Xi^{[a]} &= \underset{\perp\!\!\!\perp}{\text{copy}} \Xi \\
&\vdots \quad \vdots \\
\Xi^{[\infty]} &= \underset{\perp\!\!\!\perp}{\text{copy}} \Xi
\end{aligned}
\qquad
X_n^{(a)} = f \left(\underbrace{\xi_n, \xi_{n-1}, \dots}_{a \text{ termes}}, \overbrace{\xi_{n-a}^{[a]}, \dots, \xi_1^{[a]}}^{\substack{a^{\text{ème}} \text{ copy de } (\Xi_n) \\ \text{tronqué } a \text{ derniers termes}}} \right)$$

Ensuite il nous suffit de regarder si on a perdu beaucoup d'information sur le processus en le comparant au processus initial, dont on souhaite déterminer l'ordre de dépendance. On regarde le pire cas pour $t \in \mathcal{T}$:

$$L_p(X_n|a) = \mathbb{E}\|X_n - X_n^{[a]}\|_{\infty(\mathcal{T})}^p$$

On parle alors de $\mathbb{L}^p - a$ approximation en étudiant la convergence de la série :

$$\sum_{a=1}^{\infty} L_p(X_n|a)^{\frac{1}{p}} = \sum_{a=1}^{\infty} \left(\mathbb{E}\|X_n - X_n^{[a]}\|_{\infty(\mathcal{T})}^p \right)^{\frac{1}{p}}$$

Définition 4 ($\mathbb{L}^p - a$ approximation) une suite de variables aléatoires $(X_i)_{i \geq 0}$ est dite $\mathbb{L}^p - a$ approximable si la série $\sum_{a=1}^{\infty} L_p(X_n|a)^{\frac{1}{p}}$ converge.

Il s'agit de la définition de dépendance faible proposée pour les données fonctionnelles par Hörmann et Kokoszka(11). Une autre définition est aussi populaire : au lieu de remplacer tout le passé par la copie, on ne remplace que ξ_0 par la $a^{\text{ème}}$ copie.

L'idée est qu'après inversion du processus causal on obtient :

$$\begin{aligned}
X_n &= \sum_{k=0}^{a-1} \phi^k(\xi_{n-k}) + \sum_{k=a}^{\infty} \phi^k(\xi_{n-k}) \\
X_n^{[a]} &= \sum_{k=0}^{a-1} \phi^k(\xi_{n-k}) + \sum_{k=a}^{\infty} \phi^k(\xi_{n-k}^{[a]}) \\
X_n^{[a]} &= \sum_{k=n}^{\infty} \phi^k(\xi_{n-k}) + \phi^n(\xi_0^{[a]})
\end{aligned}$$

Le reste dans l'approximation $\mathbb{L}^p - a$ ($X_n - X_n^{[a]}$) devient alors le suivant :

$$\begin{aligned}
X_n &= \sum_{k=0}^{a-1} \phi^k(\xi_{n-k}) + \sum_{k=a}^{\infty} \phi^k(\xi_{n-k}) \\
R_n^{[a]} &\underset{\phi \parallel n}{=} \sum_{k=a}^{\infty} \phi^k(\xi_{n-k}^{[a]} - \xi_{n-k}) \\
R_n^{[a]} &\underset{\phi \parallel n}{=} \phi^n(\xi_0^{[a]} - \xi_0)
\end{aligned}$$

et on peut alors montrer que pour une certaines métrique ν_2 basée sur la norme \mathbb{L}^2 ,

$$\nu_2 \left(R_n^{[A]} \right) \leq C \sum_{a \in A} \nu_2 \left(R_n^{[a]} \right)$$

ce qui fait de la dernière version introduite est une version plus forte. Avec la dernière définition introduite, il avait été démontré différentes inégalités qui se trouvent très utiles pour déterminer les bornes de concentration de différents estimateurs. La question est désormais la suivante :

« est ce que ces inégalités restent vraies pour la définition $X_n^{[a]}_{[k \leq a]}$? »

La réponse, déterminée par MPV (17) est oui. C'est important de l'avoir aussi pour cette définition car MPV a réussi à étendre la notion de $\mathbb{L}^p - a$ approximation au cas \mathbb{L}^∞ (17) pour avoir un héritage local de la notion de dépendance définie sur les trajectoires.



N'est-il pas bizarre qu'une norme infinie permette de définir une notion de dépendance locale ?

Il semble en effet plus que contre-intuitif qu'une norme infinie, c'est à dire une norme invoquant le supremum sur un intervalle, permette d'obtenir une notion de dépendance locale.

en notant $\nu_p : x \mapsto \mathbb{E}[|x|^p]^{\frac{1}{p}}$

$$\sum_n \mathbb{E} [|X_n(t) - X_n^{[a]}(t)|^p] \leq \sum_n \mathbb{E} \left[\left\| X_n - X_n^{[a]} \right\|_{\infty(\mathcal{T})}^p \right]$$

La somme des $\nu_P(|\cdot(t)|)$ étant bornée par la somme des $\nu_P(\|\cdot\|_\infty)$, la dépendance locale (ie à t fixé) est directement héritée. Si la démarche consistait juste à obtenir une notion de dépendance locale, on remarque que ce qui la fait marcher est le fait que l'on a la convergence en considérant les pires cas sur chaque trajectoire.



Démontrer que $\sum_n \mathbb{E} [|X_n(t) - X_n^{[a]}(t)|^p] < \infty$ t par t ne suffit pas pour que les résultats sur l'obtention de la régularité découlent :
il est important de définir les hypothèses de données fonctionnelles sur les fonctions et non pas sur les valeurs prises par les fonctions. Puisque c'est la réplication des courbes qui est la clé.

L'idéal serait d'avoir une notion de dépendance faible qui permettrait d'obtenir une inégalité du genre :

$$\sum_n \nu_p(\left\| X_n - X_n^{[a]} \right\|_{\text{hypothétique}_{inf}}) \leq \sum_n \nu_p(|X_n(t) - X_n^{[a]}(t)|) \leq \sum_n \nu_p(\left\| X_n - X_n^{[a]} \right\|_{\text{hypothétique}_{sup}})$$

Qui donnerait une sorte d'équivalence entre le point de vu fonctionnel et le point de vue local en terme de dépendance, mais à ce jour, et à notre connaissance, il n'existe pas de telle notion de dépendance.



Si l'on souhaite juste regarder l'ordre de dépendance, en remplaçant l'information après le $a^{\text{ème}}$ dernier terme par quelquechose dont le processus qui nous intéresse ne dépend pas, pourquoi s'embêter avec des copies indépendantes au lieu de simplement tronquer (c'est-à-dire remplacer par des 0) ?

Il s'avère que les deux définitions sont en quelques sorte « équivalentes » mais que celles avec les copies est plus générale et donc est évidemment privilégiée pour plus de flexibilité et de puissance dans les résultats dérivés.

$$\begin{aligned} X_n &= \sum_{k=0}^{a-1} \phi^k(\xi_{n-k}) + \sum_{k=a}^{\infty} \phi^k(\xi_{n-k}) \\ X_n^{[a]} &= \sum_{k=0}^{a-1} \phi^k(\xi_{n-k}) + \sum_{k=a}^{\infty} \phi^k(\xi_{n-k}^{[a]}) \\ X_n^{[a]} &= \sum_{k=0}^{a-1} \phi^k(\xi_{n-k}) + 0 \end{aligned}$$

et ainsi lorsque l'on va regarder

$$\|X_n - X_n^{[a]}\|_{\mathbb{L}^\infty}^p = \left\| \sum_{k=a}^p \phi^k(\xi_{n-k} - \xi_{n-k}^{[a]}) \right\|_{\mathbb{L}^\infty}^p$$

que ce soit avec une méthode ou l'autre, on remarque que lorsque l'on va développer les sommes, les termes en $\|\xi \cdot \xi^{[a]}\|_{\mathbb{L}^\infty}$ seront nuls.

A.4 Continuité de Kolmogorov

Le théorème de continuité de Kolmogorov nous permet de dérivé la régularité d'un processus, au sens de la classe de Hölder, à partir de l'espérance de ses incrément.

Théorème 2 (Continuité de Kolmogorov)

référence : (2, thm : 2.197 | page : 145)

- $X : \begin{matrix} \mathbb{R}_+ \times \Omega & \longrightarrow & \mathbb{R} \\ (t, \omega) & \longmapsto & X(t, \omega) = x(t) \end{matrix}$ séparable
- $\exists r, c, \varepsilon, \delta \in \mathbb{R}_+$ $(\forall h < \delta)(\forall t \in \mathbb{R}_+) \quad \mathbb{E}[|X(t+h) - X(t)|^r] \leq c \cdot h^{1+\varepsilon}$

\Downarrow

* X est continu en $t \in \mathbb{R}_+$ pour presque tout $\omega \in \Omega$

ie : il existe une version \tilde{X} de X continue en t telle que $\mathbb{P}[\tilde{X}(t) = X(t)] = 1$

* \tilde{X} est γ -Hölderienne en t pour tout $0 < \gamma < \frac{\varepsilon}{r}$

Etant donné que notre estimateur utilise les incrément quadratiques, on se place dans le cas où $r = 2$. Dans notre cas, $\varepsilon = 1$.

C'est ce théorème qui est exploité pour récupérer la régularité locale de nos données en utilisant un estimateur de $\mathbb{E}[|X(u) - X(v)|^2]$, qui est entre autres, la moyenne empirique qui converge bien vers la quantité souhaitée sous hypothèse de dépendance faible comme vu en A.3.

A.5 Estimation adaptative

Dans la section précédente, nous avons déterminé comment obtenir des estimateurs de la régularité locale des trajectoires. Cette régularité locale nous permet désormais de lisser les courbes observées de manière à ne pas détruire l'information irrégulière. L'obtention d'un tel lissage était motivé notamment par l'obtention de quantités capitales pour l'analyse de nos données, l'interprétation et la prise de décision : la moyenne, la covariance, et l'auto-corrélation des séries temporelles fonctionnelles observées.

Un meilleur lissage nous donne ainsi une meilleure estimation de ces quantités. Toutefois, il est possible d'aller plus loin dans l'adaptation de notre lissage. En effet, il faut dans un premier temps constater que les différentes quantités que l'on souhaite estimer représentent des concepts différents, préférant chacun un lissage différent.

A.5.1 Estimation adaptative de la fonction moyenne

L'idée du lissage adaptatif est que chaque quantité évaluée en un point $t \in \mathcal{T}$ tire parti différemment des informations du voisinage de t . Il semble intuitif que le processus moyen ($\mu = \mathbb{E}[X]$) considère des informations d'un voisinage assez large du processus et que celui-ci soit « assez lisse ». Pour déterminer une fenêtre adaptée à l'estimation de la moyenne, on définit une grille de fenêtres à évaluer $\mathfrak{H} = (h_i)_{1:r}$ que l'on choisit en minimisant un risque spécifiquement adapté :

$$\widehat{h}_\mu^* = \underset{h \in \mathfrak{H}}{\operatorname{argmin}} R_\mu(t, h)$$

Déterminons maintenant ce risque.

A.5.1 A) Méthode Golovkine et al. : indépendance

Dans le cadre de données indépendantes, on peut invoquer la Loi des Grands Nombres pour approximer l'espérance par la moyenne empirique.

On effectue une suite d'approximations de la façon suivante :

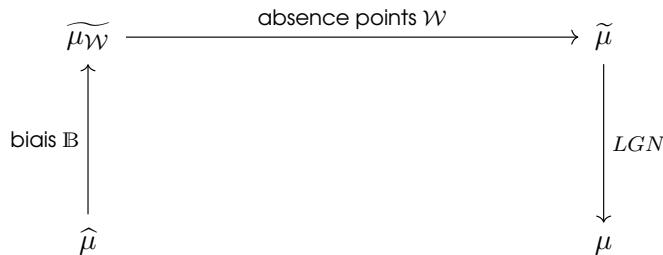


FIGURE A.1 – Schéma du découpage du contrôle des erreurs

On détermine alors fenêtre de lissage en minimisant le risque suivant (8) :

$$R_\mu^{[\text{Golovk.}]}(t, h) = \underbrace{q_1^2 h^{2H_t}}_{\text{contrôle du biais}} + \underbrace{\frac{q_2^2}{\mathcal{N}_\mu(t, h)}}_{\text{contrôle de la variance}} + \underbrace{q_3^2 \left[\frac{1}{\sum_k w_k} - \frac{1}{n} \right]}_{\text{pénalise absence de points}}$$



Il est tout à fait possible de regarder directement l'erreur d'approximation entre $\widetilde{\mu_W}$ et μ . Toutefois, le choix de Golovkine est avant tout un choix pédagogique, pour signaler et renforcer l'idée qu'il faut faire attention à l'erreur d'approximation entre l'inobservable et le véritable processus (\mathbb{E} vs $\frac{1}{N} \sum X_i \neq \frac{1}{N} \sum \widehat{X}_i$)

Afin de prendre en compte la dépendance, que l'on doit contrôler aussi, on raisonne plutôt de la façon suivante.

A.5.1 □ B) Méthode MPV : dépendance

Lorsque l'on traite le cas de la dépendance, il est tout de suite plus délicat d'obtenir la convergence d'estimateurs de moments d'une loi. MPV utilise ce découpage du risque pour déterminer une fenêtre de lissage adaptée à l'estimation de la fonction moyenne :

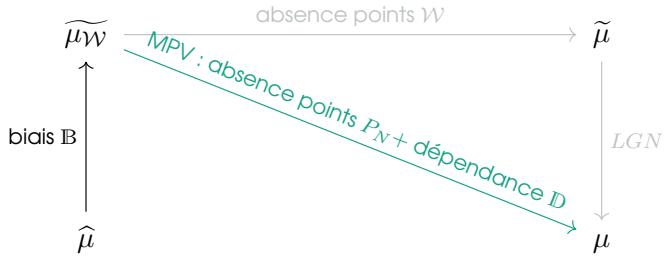


FIGURE A.2 – Schéma du découpage du contrôle des erreurs

On détermine cette fois-ci la fenêtre de lissage en minimisant le risque suivant (17) :

$$R_\mu(t, h) = \underbrace{L_t^2 h^{2H_t} B(t, h, 2H_t)}_{\text{contrôle du biais}} + \underbrace{\sigma^2 V_\mu(t, h)}_{\text{contrôle de la variance}} + \underbrace{\frac{D_\mu(t)}{P_N(t, h)}}_{\text{contrôle de la dépendance}}$$

A.5.2 Quelques remarques sur les applications

Bien que la théorie suppose des données dépendantes, il est tout à fait possible d'appliquer les estimateurs dérivés pour des données dépendantes sur des données indépendantes. Il ne devrait pas y avoir en pratique de différences significatives de performances lors de l'utilisation des estimateurs dérivés par MPV sur des données indépendantes. Lors de l'estimation de la fonction moyenne lors du stage, l'estimateur pour les données dépendantes a été utilisé sur l'ensemble des données traitées.

A.6 Mouvement Brownien

A.6.1 Mouvement Brownien

Le concept du Mouvement Brownien, initialement découvert par Robert Brown en 1827, revêt une signification profonde dans le contexte des phénomènes aléatoires et a trouvé des applications éminentes dans les domaines de la physique, des mathématiques et au-delà. Il constitue un modèle fondamental pour décrire le comportement erratique et imprévisible des particules immergées dans un fluide, où chaque particule suit une trajectoire chaotique.

A.6.2 Mouvement Brownien Fractionnaire

Le Mouvement Brownien Fractionnaire, une extension du modèle classique et offre une perspective plus riche pour modéliser des phénomènes complexes. Son étude systématique s'est trouvée être fructueuse notamment dans des milieux tels que la finance. Il permet notamment d'étudier un phénomène non différentiable, Höldérien de paramètres $H_I(H, L)$

A.6.2 □ A) Construction du Mouvement Brownien Fractionnaire

Le lecteur pourra, si il le souhaite, trouver une définition du mouvement brownien fractionnaire ainsi que différentes méthodes de simulations de ce derniers dans la thèse doctorale de Ton Dieker (2004) (6).



Un mouvement Brownien fractionnaire normalisé $B_H = \{B_H(t) : t \in \mathbb{R}_+, H \in]0, 1[\}$ est caractérisé de façon unique par :

les incrément de $B_H(t)$ sont stationnaires

$$B_H(0) = 0$$

$$\forall t \in \mathbb{R}_+ \quad \mathbb{E}[B_H(t)] = 0$$

$$\forall t \in \mathbb{R}_+ \quad \mathbb{E}|B_H(t)|^2 = t^{2H} = \sigma_H^2(t)$$

$$\forall t > 0 \quad B_H(t) \sim \mathcal{N}(0, \sigma_H^2(t))$$

$$C_{B_H}(u, v) = \mathbb{E}[B_H(u)B_H(v)] = \frac{1}{2}[u^{2H} + v^{2H} + |u - v|^{2H}]$$

source : Diecker, 2004 (6)

A.6.3 Mouvement Brownien multi-fractionnaire

L'expression explicite de leur covariance a été dérivée notamment par Stoev et Taqqu en 2006 (25). Les processus browniens multi-fractionnaires sont aussi intéressants pour leur « richesse » : On peut pour chaque fonction $H : t \mapsto H_t$ observer « une diversité infinie de processus browniens multi-fractionnaires de manière générale ». (25)

A.7 Théorie de la base d'ondelettes

A.7.1 Transformée en ondelettes

Introduisons maintenant de façon plus formelle les ondelettes et regardons leurs propriétés intéressantes dans le cadre du lissage de trajectoires.

on définit la transformée en ondelettes vis à vis de l'ondelette mère ψ d'une fonction f par :

$$F : \begin{array}{ccc} \mathbb{R} \times \mathbb{R}_+ & \longrightarrow & \mathbb{R} \\ (t, s) & \longmapsto & \frac{1}{\sqrt{|s|}} \int_{\mathbb{R}} f(\textcolor{blue}{u}) \psi \left(\frac{\textcolor{blue}{u} - t}{s} \right) d\textcolor{blue}{u} \end{array}$$



on peut remarquer que la formule de la transformée en ondelettes ressemble à une projection : $\frac{\langle f, \psi_{t,s} \rangle_{L^2}}{\|\psi_{t,s}\|}$. Cela vient en quelque sorte motiver la section suivante

A.7.1 □ A) Base d'ondelettes

Proposition 2.1 (base d'ondelette dichotomique)

$$\left\{ \psi_{k,n} : t \mapsto \frac{1}{\sqrt{2^k}} \psi \left(\frac{t - 2^k n}{2^k} \right) \right\}_{(k,n) \in \mathbb{Z}^2} \text{ est une base } \perp_{\|\cdot\|} \text{ de } \mathbb{L}^2$$

i

notons que les résolutions sont des puissances de 2, ceci est un détail qui demandera une implémentation particulière dans le cadre des données réelles : il faudra faire attention à ce que le nombre de points que l'on donne dans l'algorithme de transformée rapide en ondelettes soit aussi une puissance de 2.

A.7.2 Propriétés principales des ondelettes

A.7.2 □ A) Approximation dans l'espace fréquentiel-temporel

La transofrmée en ondelettes

$$\mathcal{W} : f \mapsto \langle f | \psi_{t,s} \rangle$$

est une isométrie de \mathbb{L}^2 . Etant donné qu'elle est de plus une application linéaire, nous pouvons donc d'affirmer que

$$\|f - \hat{f}\|_{\mathbb{L}^2} = \|\mathcal{W}f - \mathcal{W}\hat{f}\|_{\mathbb{L}^2}$$

Ainsi on peut travailler dans l'espace des ondelettes pour approximer (dans notre cas lisser les trajectoires) des fonctions et contrôler l'approximation directement dans le domaine fréquence-temporel tout en le conservant dans le domaine temporel.

A.7.2 □ B) Propriété de Fast Decay : (ref : (18))

Une caractérisation des fonctions Hölderaines, fournie par Antoniadis et Gijbels en 2002  ( citation requise ) est :

$$f \in \mathcal{H}_{\mathcal{V}(t_0)}(\alpha, L_\alpha) \cap \mathbb{L}^2 \iff \exists P \in \mathbb{R}[X], \deg P \leq \alpha \leq \deg P + 1 \quad \exists f_{loc} \underset{t \rightarrow 0}{=} \mathcal{O}(t^\alpha) \quad f(t_0 + h) \underset{t \rightarrow 0}{=} P(h) + f_{loc}(h)$$

Définition 5 (vanishing moment) on dit qu'une ondelette ψ possède n vanishing-moments si :

$$\forall k < n \left\langle t \mapsto t^k | \psi \right\rangle_{\mathbb{L}^2} = 0 = \int_{\mathbb{R}} t^k \psi(t) dt$$

Proposition 2.2 (vanishing-moment et polynômes)

il suffit donc de choisir une ondelette avec $n > \alpha$ vanishing-moments pour obtenir :

$$\mathcal{W}f|_{\mathcal{V}(t_0)} = \mathcal{W}(P + f_{loc}) = \mathcal{W}P + \mathcal{W}f_{loc} = \mathcal{W}f_{loc}$$

enfin

Théorème 3 (Fast Decay | ref : (18) - thm 6.3)

$$f \in \mathcal{H}_{\mathcal{V}(t_0)}(\alpha, L_\alpha) \cap \mathbb{L}^2 \implies \exists A > 0, |\mathcal{W}f|(t, s) \leq A \cdot s^{\alpha + \frac{1}{2}}$$

et inversement en supposant f bornée (ce qui est le cas pour une fonction continue sur un segment : notre cas) et f Hölder juste après les bords. (C'est à dire que ça ne marche pas pour les points extrémaux $t \in \{0, 1\}$)

Ainsi lorsque $s \in \{2^{-k}\}_{k \in \mathbb{N}}$:

$$|[\mathcal{W}f](t, s)| \leq A \cdot 2^{-k(\alpha + \frac{1}{2})}$$

La magnitude de la transformée en ondelette décroît exponentiellement vers 0, et beaucoup plus rapidement là où f est plus régulière. Ainsi, la transformée en ondelette agit comme un encodeur efficace d'information d'irrégularité

A.7.3 Discussion : potentiel du lissage ondelette pour le prélassage

Annexe B

Plus de détails sur l'étude du Risque

Contents

B.1	Qualité de l'estimation des couples d'incrément utilisés dans l'estimation de la régularité & choix du couple à estimer	49
B.1.1	Choix du Risque	51
B.2	Récapitulatif de la qualité des estimations pour la stratégie de « lissage global »	52
B.2.1	Récapitulatif concernant le risque quadratique sur l'estimation de la régularité	52
B.2.2	Récapitulatif concernant le risque quadratique sur les θ individuellement	52
B.2.3	Récapitulatif concernant le risque euclidien pour les couple Θ	53
B.2.4	Conclusion des tables	53
B.3	Etude de l'impact de la méthode de sélection de la fenêtre de pré-lissage sur le risque d'estimation des couples Θ	53
B.3.1	Rappel de la méthodologie utilisée	53
B.3.2	Le cas sparse	53
B.3.3	Densité de points sur les courbes observées	55
B.3.4	Peut-on considérer que tous les Δ conviennent ?	55
B.3.5	couple $\Theta : 1 \rightarrow 3 / 1 \rightarrow 2$	57
B.3.6	couple $\Theta : 1 \rightarrow 3 / 2 \rightarrow 3$	58
B.4	Etude de l'impact de la méthode de prélassage sur l'estimation de la régularité et le Δ optimal	66
B.4.1	Ondelettes	67

B.1 Qualité de l'estimation des couples d'incrément utilisés dans l'estimation de la régularité & choix du couple à estimer



on rappelle les notations suivantes :

- vrai : $\theta = \mathbb{E}[f(X)]$
- intangible/inobservable : $\tilde{\theta} = \frac{1}{N} \sum_i f(X_i)$
- observable : $\hat{\theta} = \frac{1}{N} \sum_i f(\hat{X}_i)$

L'estimation des paramètres de régularité locale en t_2 , H_{t_2} et L_{t_2} , utilise les incrément quadratiques θ entre les différents points t_1, t_2, t_3 situés dans un voisinage de diamètre Δ autour de t_2 .



Quelle quantité est-il judicieux d'évaluer pour estimer au mieux la régularité locale en t_2 ? Doit-t-on regarder la qualité de l'approximation de θ (qui est une espérance) car il est utilisé pour tous les estimateurs? Ou doit-on regarder la qualité de l'approximation de H_{t_2} et L_{t_2} car ce sont les quantités qui nous intéressent? Ou bien les deux?

Le choix du bon critère d'évaluation est d'une grande importance. Il faut se rappeler l'objectif que l'on cherche à atteindre : déterminer une procédure (simple si possible) de détermination de l'hyper-paramètre Δ utilisé pour l'estimation de la régularité locale en fonction de quantités facilement estimables, ou directement observables par le praticien. En observant que L est estimée par une expression impliquant θ et Δ^{2H} , une estimation précise de H paraît plus cruciale pour la bonne estimation des deux quantités. Bien qu'il existe certainement un compromis entre la bonne estimation de H et de L qui fournit une meilleure estimation adaptative des quantités qui nous intéressent, il est certainement plus probable de dériver une procédure de sélection de Δ simple à implémenter pour le praticien en se basant sur la qualité de l'estimation d'une unique quantité.



Doit-t-on se concentrer sur l'estimation de H ou de θ ?

Les incrémentations sont des quantités importantes dans l'estimation de la régularité, utilisées à la fois pour l'estimation de H et de L , l'approche que l'on considère se base sur cette remarque. Nous allons donc chercher à déterminer un Δ adapté à l'estimation des incrémentations quadratiques. Toutefois, il y a plusieurs possibilités de $u, v \in J_\Delta(t_2)$ que l'on peut considérer pour l'estimation de H et L . C'est pourquoi nous décidons de considérer les couples d'incrémentations utilisés pour l'estimation de H . Ainsi en posant :

$$\Theta_{\substack{1 \rightarrow 3 \\ 1 \rightarrow 2}} = \begin{bmatrix} \theta(t_1, t_3) \\ \theta(t_1, t_2) \end{bmatrix} \quad \Theta_{\substack{1 \rightarrow 3 \\ 2 \rightarrow 3}} = \begin{bmatrix} \theta(t_1, t_3) \\ \theta(t_2, t_3) \end{bmatrix}$$

L'estimateur du paramètre de régularité H_t peut se ré-écrire comme :

$$\hat{H}_t : \Theta = \begin{bmatrix} \Theta_1 \\ \Theta_2 \end{bmatrix} \mapsto \frac{\log \hat{\Theta}_1 - \log \hat{\Theta}_2}{2 \log 2}$$

Le problème c'est qu'on ne dispose pas de la véritable valeur de $\theta(u, v)$, on pourrait exploiter le fait que l'on dispose d'un mouvement brownien multi-fractionnaire qui a été étudié de façon extensive dans la littérature, mais on décide de ne pas l'utiliser pour adopter une approche plus proche d'un cadre général. Le meilleur estimateur que l'on puisse espérer atteindre est l'estimateur de l'espérance par la moyenne empirique des courbes non bruitées :

$$\tilde{H}_t(\Theta_{\substack{1 \rightarrow 3 \\ 1 \rightarrow 2}}) \quad \text{ou} \quad \tilde{H}_t(\Theta_{\substack{1 \rightarrow 3 \\ 2 \rightarrow 3}})$$

$$\text{avec } \tilde{H}_t : \Theta \mapsto \frac{\log \tilde{\Theta}_1 - \log \tilde{\Theta}_2}{2 \log 2}$$

On va donc s'intéresser désormais à l'erreur d'estimation conjointe des deux $\theta(u, v)$ utilisés dans l'estimation de H_t par rapport à cet estimateur en quelque sorte « idéal » de θ comme critère de sélection du diamètre Δ .

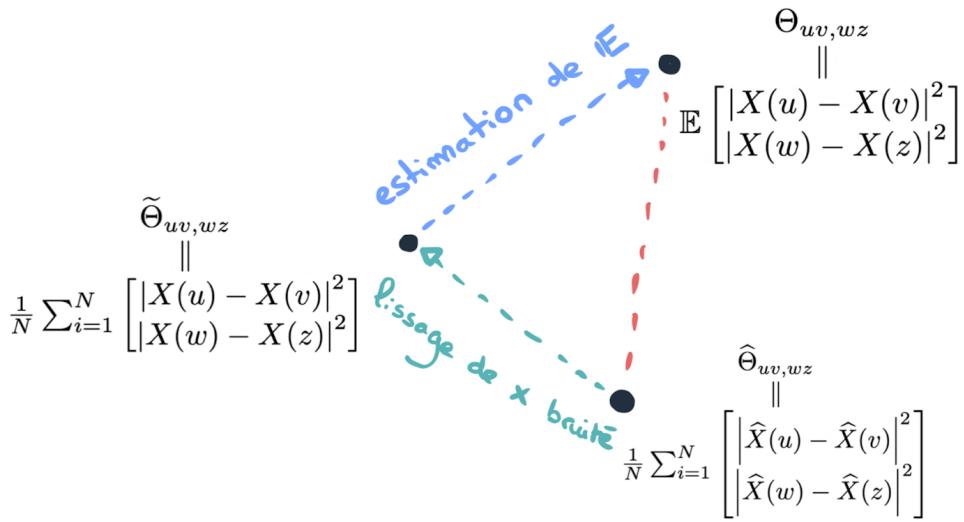


FIGURE B.1 – Schéma représentant les différentes approximations du couple d’incréments

B.1.1 Choix du Risque

B.1.1 □ A) Distance Euclidienne

Afin de quantifier la qualité de l'estimation conjointe du couple de θ , il est raisonnable de considérer la distance euclidienne usuelle pour des vecteurs de \mathbb{R}^2

$$R(\Theta, \Delta) = \left\| \hat{\Theta}(\Delta) - \tilde{\Theta}(\Delta) \right\|_2^2$$

B.1.1 □ B) Distance Euclidienne Relative

On va cependant considérer le risque relatif à la norme de la quantité que l'on cible :

$$R(\Theta, \Delta) = \frac{\left\| \hat{\Theta}(\Delta) - \tilde{\Theta}(\Delta) \right\|_2^2}{\left\| \tilde{\Theta}(\Delta) \right\|_2^2}$$



Pourquoi considérer la distance euclidienne relative à la norme de la cible $\tilde{\Theta}$ plutôt que la distance euclidienne classique qui est plus simple ?

Le risque sert à déterminer la qualité de l'estimation du couple $\tilde{\Theta}$ par $\hat{\Theta}$ à un Δ donné. Il faut cependant garder à l'esprit que Θ est en réalité une fonction de Δ car la valeur de t_1, t_2, t_3 dépendent de Δ . Ainsi la norme de $\tilde{\Theta}$ va varier lorsque l'on fait varier Δ ¹. Les risques obtenus via la norme euclidienne sont des risques qui mesurent une différence absolue, mais alors avoir un risque plus petit qu'un autre n'a pas le même sens pour différents Δ en termes de qualité d'approximation. C'est pourquoi nous considérons le risque relatif dans la détermination du critère du choix du Δ .

On pourra cependant observer la différence entre le risque euclidien et le risque euclidien relatif à la norme de la cible en Δ sur les figures B.3.6 et B.3.6.

1. il est possible d'obtenir plus de détails en annexe B.3.4

B.2 Récapitulatif de la qualité des estimations pour la stratégie de « lissage global »

Les simulations de Monte Carlo permettent d'avoir accès directement à la véritable régularité de la courbe en chaque point. Nous allons dans l'étude du comportement du Δ essayer de tirer profit de cet avantage que ne possède pas le praticien qui utilise des données réelles.



On rappelle que la stratégie de lissage dite globale » lisse par validation croisée courbe par courbe pour $\lambda < 110$ puis calcule une fenêtre de lissage médiane à appliquer sur toutes les courbes pour $\lambda > 110$

B.2.1 Récapitulatif concernant le risque quadratique sur l'estimation de la régularité

		$\lambda < 120$	$\lambda \geq 120$
		$\checkmark \iff \mathcal{R}$	$\simeq \checkmark \iff \mathcal{R}$
$H_t \leq 0.65$	$\checkmark \iff \Delta^*$	$\checkmark \iff \Delta^*$	$\times \iff \Delta^*$
	$\Delta^- \downarrow 0.01$		$\Delta^+ \rightarrow [\leq 0.6]0.1/0.2[\geq 0.6]$
$H_t > 0.65$	$\checkmark \iff \mathcal{R}$		$\Theta_{\substack{1 \rightarrow 3 \\ 1 \rightarrow 2}}$
	$\times \iff \Delta^*$		$\blacktriangle H = 0.7 : \Delta^- = 0.02$
	$\blacktriangle H = 0.7 : \Delta^- = 0.02$		$\blacktriangle H = 0.73 : \Delta^- = 0.2$
	$\blacktriangle H = 0.73 : \Delta^- = 0.2$		

TABLE B.1 – Tableau récapitulatif des Δ optimaux : Risque sur H_t

B.2.2 Récapitulatif concernant le risque quadratique sur les θ individuellement

		$\lambda < 120$	$\lambda \geq 120$
		meilleur : $\Theta_{\substack{1 \rightarrow 2 \\ 2 \rightarrow 3}}$	meilleur : $\Theta_{\substack{1 \rightarrow 3 \\ 1 \rightarrow 2}}$
$H_t < 0.6$	$\Delta^- \rightarrow 0.01$		$\Delta^+ \rightarrow 0.2$
	$\Delta^- \rightarrow 0.2$	meilleur : $\Theta_{\substack{1 \rightarrow 3 \\ 1 \rightarrow 2}}$	meilleur : $\Theta_{\substack{1 \rightarrow 2 \\ 2 \rightarrow 3}}$
$H_t \geq 0.6$			

TABLE B.2 – Tableau récapitulatif des Θ optimaux : Risque individuel sur $\tilde{\theta}(u, v)$

B.2.3 Récapitulatif concernant le risque euclidien pour les couple Θ

	$\lambda < 120$	$\lambda \geq 120$
$H_t < 0.6$	$\checkmark \Leftrightarrow \mathcal{R}, \Delta^*$ $\Delta_-^* = 0.01$ meilleur : $\Theta_{\substack{1 \rightarrow 3 \\ 2 \rightarrow 3}}$ $\Delta_-^* = 0.2$	meilleur : $\Theta_{\substack{1 \rightarrow 3 \\ 2 \rightarrow 3}}$ $\Delta_+^* = 0.2$
	meilleur : $\Theta_{\substack{1 \rightarrow 3 \\ 2 \rightarrow 3}}$ $\Delta_-^* = 0.2$	
$H_t \geq 0.6$	$\blacktriangle H = 0.7 : \Delta^- = 0.01 \oplus \checkmark \Leftrightarrow \mathcal{R}$ $\blacktriangle H = 0.8 : \Theta_{\substack{1 \rightarrow 3 \\ 1 \rightarrow 2}}$	$\checkmark \Leftrightarrow \mathcal{R}, \Delta^*$ $\Delta_+^* = 0.01$

TABLE B.3 – Tableau récapitulatif des Δ optimaux : Risque euclidien sur $\tilde{\Theta}$ | fenêtre de prélassage globale pour $\lambda \geq 120$

B.2.4 Conclusion des tables

Le risque euclidien a une structure plus complexe que le risque relatif sur la détermination du Δ optimal². Concernant le risque euclidien, on peut recommander, en s'appuyant conjointement sur les observations graphiques ayant enlevé les observations extrêmes (figures B.5 et B.3.6)

B.3 Etude de l'impact de la méthode de sélection de la fenêtre de prélassage sur le risque d'estimation des couples Θ

B.3.1 Rappel de la méthodologie utilisée

Afin de simplifier les notations, dans la suite nous nous référons à 2 cas de méthodologie différentes nommées « **individuel** » et « **global** ». Cependant le nom choisi ne reflète pas entièrement la procédure, il faut donc bien garder en tête ce qu'on attend par ces noms de méthodologie :

threshold/méthodologie	individuel	global
$\lambda < 110$	h cross-validé par courbe	h cross-validé par courbe
$\lambda \geq 110$	h cross-validé par courbe	détermination de 50 fenêtres h_i cross-validées sur les 50 premières courbes puis utilise $h = \text{med}_{i \in [1, 50]} h_i^{*-cv}$ sur toutes les courbes

Ce rappel sera important dès la prochaine section portant sur le comportement du cas sparse.

B.3.2 Le cas sparse

les tables B.5 et B.6 indiquent que la différence en terme de performance sur le Risque entre un lissage global (déterminé comme la médiane des fenêtres cross-validées sur les 50 premières

2. cf différence entre les graphes B.3.6 (euclidien) et B.3.6 (relatif)

courbes) et un lissage individuel (la fenêtre de lissage de chaque courbe est déterminée par validation croisée) devient mineure une fois que le nombre moyen de points par courbe devient « dense ». Dans notre cas, on voit qu'à partir de $\lambda = 150$, la différence entre la médiane du risque par lissage à fenêtre globale et par lissage individuel devient faible (de l'ordre de 10^{-6} à 10^{-3}) et une différence de variance presque imperceptible (de l'ordre de 10^{-9} à 10^{-5}).



A la vue de ces résultats il semblerait que la conclusion évidente soit de faire une fenêtre cross-validée individuelle lorsque l'on est dans un cas sparse. Sauf qu'il ne faut pas oublier que dans l'appellation « global » comme mentionné précédemment, les différentes courbes étaient déjà lissées individuellement lorsque $\lambda < 110$.



D'où viennent alors de telles différences ?

Un changement d'algorithme a été opéré entre les lissages effectués avec la méthode « globale » (première méthode utilisée) et la méthode « individuelle » : lorsqu'il existe un endroit où on lisse pour déterminer la régularité tel qu'il n'y a pas assez de points autour, on ne sélectionne pas la fenêtre associée.

Sous R, entre autre, l'implémentation est telle que si le lissage à noyau renvoie `NaN`, on ne choisit pas le h associé :

```

1 cv_error <- sapply(h_grid, function(hi, y, t, K) {
2   yhat <- estimate_nw(y = y, t = t, h = hi, tnew = t, smooth_ker = K)$yhat
3   wmat <- outer(X = t, Y = t, function(u, v) K((u - v) / hi))
4   metric <- (y - yhat) / (1 - K(0) / rowSums(wmat))
5
6   # If there is only one value in the kernel support, it return NaN.
7   error_hi <- mean(metric[!is.nan(metric)]**2)
8 }, y = y, t = t, K = smooth_ker)
9
10 # If cv_error is NaN, do take it into account
11 if (any(is.nan(cv_error))) {
12   h_grid <- h_grid[-which(is.nan(cv_error))]
13   cv_error <- cv_error[!is.nan(cv_error)]
14   hcv <- h_grid[min(which.cv_error)]
15 }
```



① Il semblerait donc que cette différence de traitement du nombre de points autour des points lissés soit ce qui explique la différence entre les deux méthodes pour le cas sparse, ce qui semble raisonnable : on ne voit pas de différence dans le cas dense simplement car ne pas avoir suffisamment de points autour se rarifie extrêmement.



② On notera tout de même que la faible différence de médiane et de variance entre le cas « global » et « individuel » indique qu'il est possible d'économiser en temps de calcul sur des données même corrélées en calculant une fenêtre globale à partir des premières courbes lorsque l'on dispose d'observations denses à moindre coût sur le risque.

B.3.3 Densité de points sur les courbes observées

B.3.3 □ A) Densité moyenne

Les figures B.3 et B.4 affichent la densité de points présents autour de $t_2 = 0.8$ en considérant toutes les courbes pour un individu de monte-carlo associé à un risque extrême. En utilisant un estimateur de Parzen-Rosenblatt de fenêtre Δ utilisé pour calculer $X(t_1(\Delta))$ et $X(t_3(\Delta))$:

$$\hat{f}_T = \frac{1}{N} \sum_{i=1}^N \frac{1}{M_i} \sum_{m=1}^{M_i} \frac{1}{\Delta} K\left(\frac{t - T_i[m]}{\Delta}\right)$$

De ces figures on peut déduire les conclusions suivantes : il semblerait que l'affichage de la densité de points autour du point problématique indique que les valeurs de risque extrêmes interviennent lorsqu'il existe peu ou pas de points sur l'ensemble des courbes en t_1 ou t_3 . Toutefois, on pourra fournir en contre-exemple la figure B.4, qui indique dans un rayon Δ une densité constante de points (ce qui est en accord avec une simulation uniforme de points sur \mathcal{T} lorsque le nombre moyen de points par courbe devient relativement dense : $\lambda = 210$).

B.3.4 Peut-on considérer que tous les Δ conviennent ?

Nous observons sur les différents graphes des risques de l'ordre de grandeur de 10^{-2} ou même de 10^{-3} . La question que l'on se pose désormais est si il est raisonnable de penser que ne pas choisir le Δ^* optimal n'est pas si important dans l'estimation du couple Θ .

Ce que nous allons observer est qu'il est tout de même préférable de bien déterminer le Δ

$$\theta(u, v) = \mathbb{E}_X \left[|X(v) - X(u)|^2 \right] \leq L_{J(\Delta)}^2 |v - u|^{2H_{J(\Delta)}} \quad (\text{B.1})$$

sachant que l'on évalue :

$$\text{soit } \Theta_{1 \rightarrow 3} = \begin{bmatrix} \theta(t_1, t_3) \\ \theta(t_1, t_2) \end{bmatrix} \quad (\text{B.2})$$

$$\text{soit } \Theta_{2 \rightarrow 3} = \begin{bmatrix} \theta(t_1, t_3) \\ \theta(t_2, t_3) \end{bmatrix} \quad (\text{B.3})$$

avec :

$$\begin{aligned} |t_3 - t_1| &= \Delta \\ |t_3 - t_2| &= \frac{\Delta}{2} = |t_2 - t_1| \end{aligned} \quad (\text{B.4})$$

et donc :

$$\begin{aligned} \|\Theta\|_2 &= \sqrt{\theta_{13}^2 + \theta_{12/23}^2} \\ &\leq \sqrt{L_{J(\Delta)}^4 (\Delta^{4H_{J(\Delta)}} [1 + \frac{1}{2}])} \\ &= L_{J(\Delta)}^2 \cdot \Delta^{2H_{J(\Delta)}} \cdot \sqrt{\frac{3}{2}} \end{aligned} \quad (\text{B.5})$$

Et donc :

$$\|\Theta\|_2 \leq L_{J(\Delta)}^2 \cdot \Delta^{2H_{J(\Delta)}} \cdot \sqrt{\frac{3}{2}}$$

On se réfère à ces bornes même si l'on étudie plutôt $\tilde{\Theta}$ car on peut se ramener asymptotiquement à Θ par la loi des grands nombres grâce à la dépendance faible :

$$\left\| \tilde{\Theta} \right\|_2 = \left\| \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} |X_i(t_3) - X_i(t_1)|^2 \\ |X_i(t_3) - X_i(t_2)|^2 \end{bmatrix} \right\|_2 \quad (\text{B.6})$$

$$\xrightarrow[N \rightarrow \infty]{\text{LGN+dep. faible+norme } \mathcal{C}^0(\mathbb{R}^2 \setminus \{0\})} \quad \|\Theta\|_2 \leq L_{J(\Delta)}^2 \cdot \Delta^{2H_{J(\Delta)}} \cdot \sqrt{\frac{3}{2}} \quad (\text{B.7})$$

Remarque : On peut remplacer le couple (t_2, t_3) par (t_1, t_2) dans la deuxième composante, l'argument reste valide, comme explicité dans l'équation B.4

En utilisant les données de la simulation, $L = 1$, on obtient :

$$\begin{aligned} H_{J(\Delta)} = 0.4 &\implies \|\Theta\|_2 \begin{cases} \lesssim 3 \cdot 10^{-2} & \Delta = 0.01 \\ \lesssim 3 \cdot 10^{-1} & \Delta = 0.2 \end{cases} \\ H_{J(\Delta)} = 0.5 &\implies \|\Theta\|_2 \begin{cases} \lesssim 1 \cdot 10^{-2} & \Delta = 0.01 \\ \lesssim 2 \cdot 10^{-1} & \Delta = 0.2 \end{cases} \\ H_{J(\Delta)} = 0.6 &\implies \|\Theta\|_2 \begin{cases} \lesssim 5 \cdot 10^{-3} & \Delta = 0.01 \\ \lesssim 2 \cdot 10^{-1} & \Delta = 0.2 \end{cases} \\ H_{J(\Delta)} = 0.73 &\implies \|\Theta\|_2 \begin{cases} \lesssim 1 \cdot 10^{-3} & \Delta = 0.01 \\ \lesssim 1 \cdot 10^{-1} & \Delta = 0.2 \end{cases} \end{aligned} \quad (\text{B.8})$$

Ainsi, la différence de risque entre l'optimum et le pire cas étant de l'ordre de 10^{-2} dans un cas très sparse comme dans la figure B.3.6 et dans un cas raisonnablement dense on observe même des différences de l'ordre de 10^{-3} pour le plus régulier.

H	λ	Différence $\mathcal{R}_{\max} - \mathcal{R}_{\min}$:	ordre de grandeur de la borne de $\ \Theta\ _2$	$\ \Theta\ _2^2$
0.51	60	$3.3 \cdot 10^{-2}$	$\Delta^* \simeq 0.2 : 10^{-1}$	10^{-2}	
0.51	210	$1.1 \cdot 10^{-2}$	$\Delta^* \simeq 0.2 : 10^{-1}$	10^{-2}	
0.6	60	$4.2 \cdot 10^{-2}$	$\Delta^* \simeq 0.01 : 10^{-3}$	10^{-6}	
0.6	210	$1.2 \cdot 10^{-2}$	$\Delta^* \simeq 0.01 : 10^{-3}$	10^{-6}	
0.73	60	$1.2 \cdot 10^{-2}$	$\Delta^* \simeq 0.01 : 10^{-3}$	10^{-6}	
0.73	210	$5.4 \cdot 10^{-3}$	$\Delta^* \simeq 0.01 : 10^{-3}$	10^{-6}	

TABLE B.4 – Ordre de grandeur des différences entre le risque euclidien minimum et maximum pour $\Delta \in [0.01, 0.2]$ et la norme de la cible

Etant donné que le risque utilisé est homogène à la norme euclidienne au carré, on ne peut dire, du point de vue du risque euclidien, que l'on peut prendre n'importe quel Δ dans $[0.01, 0.2]$ sans trop de conséquences. Il est donc vivement recommandé si l'on souhaite approcher $\tilde{\Theta}$ par $\hat{\Theta}$ en se référant au risque euclidien, de bien suivre la procédure décrite en 3.3. Ce tableau est en parti ce qui nous mène à considérer le risque euclidien relatif à la norme de la cible.

B.3.5 couple $\Theta : 1 \rightarrow 3 / 1 \rightarrow 2$

t	H_t	N	λ	différence med $\mathcal{R}(\Theta, \Delta)$	meilleur	différence $\mathbb{V}\mathcal{R}(\Theta, \Delta)$	meilleur
0.3	0.51	:	:	7.26	indiv	8.36	indiv
0.4	0.55	:	:	6.31	indiv	6.91	indiv
0.5	0.60	:	:	2.32	indiv	0.07	indiv
0.6	0.65	200	45	0.09	indiv	13.68	global
0.7	0.69	:	:	0.01	indiv	15.70	global
0.8	0.73	:	:	0.22	indiv	0.02	indiv
0.3	0.51	:	:	0.72	indiv	0.32	indiv
0.4	0.55	:	:	0.79	indiv	0.36	indiv
0.5	0.60	:	:	0.29	indiv	0.04	indiv
0.6	0.65	200	90	0.01	indiv	9.93	global
0.7	0.69	:	:	$2 \cdot 10^{-3}$	indiv	$5 \cdot 10^{-6}$	indiv
0.8	0.73	:	:	0.03	indiv	$2 \cdot 10^{-3}$	indiv
0.3	0.51	:	:	$1 \cdot 10^{-3}$	global	$1 \cdot 10^{-5}$	indiv
0.4	0.55	:	:	$3 \cdot 10^{-3}$	global	$9 \cdot 10^{-5}$	global
0.5	0.60	:	:	$3 \cdot 10^{-5}$	global	$7 \cdot 10^{-6}$	global
0.6	0.65	200	150	$4 \cdot 10^{-6}$	global	$1 \cdot 10^{-7}$	global
0.7	0.69	:	:	$1 \cdot 10^{-6}$	indiv	$1 \cdot 10^{-8}$	global
0.8	0.73	:	:	$9 \cdot 10^{-6}$	indiv	$4 \cdot 10^{-6}$	global
0.3	0.51	:	:	$4 \cdot 10^{-3}$	indiv	$3 \cdot 10^{-5}$	global
0.4	0.55	:	:	$2 \cdot 10^{-3}$	indiv	$3 \cdot 10^{-6}$	indiv
0.5	0.60	:	:	$5 \cdot 10^{-5}$	global	$5 \cdot 10^{-7}$	global
0.6	0.65	200	270	$2 \cdot 10^{-4}$	global	$2 \cdot 10^{-7}$	global
0.7	0.69	:	:	$4 \cdot 10^{-4}$	global	$4 \cdot 10^{-6}$	global
0.8	0.73	:	:	$2 \cdot 10^{-4}$	global	$2 \cdot 10^{-7}$	indiv
0.3	0.51	:	:	$5 \cdot 10^{-3}$	indiv	$7 \cdot 10^{-5}$	global
0.4	0.55	:	:	$1 \cdot 10^{-3}$	indiv	$4 \cdot 10^{-6}$	indiv
0.5	0.60	:	:	$7 \cdot 10^{-5}$	indiv	$1 \cdot 10^{-7}$	indiv
0.6	0.65	200	405	$6 \cdot 10^{-5}$	global	$5 \cdot 10^{-9}$	global
0.7	0.69	:	:	$6 \cdot 10^{-5}$	global	$5 \cdot 10^{-9}$	global
0.8	0.73	:	:	$6 \cdot 10^{-5}$	global	$3 \cdot 10^{-9}$	indiv

TABLE B.5 – Comparaison de la médiane et de la variance du Risque entre lissage global et lissage individuel

B.3.6 couple $\Theta : 1 \rightarrow 3 / 2 \rightarrow 3$

t	H_t	N	λ	différence med $\mathcal{R}(\Theta, \Delta)$	meilleur	différence $\mathbb{V}\mathcal{R}(\Theta, \Delta)$	meilleur
0.3	0.51	:	:	7.37	indiv	8.20	indiv
0.4	0.55	:	:	6.12	indiv	7.44	indiv
0.5	0.60	:	:	3.80	indiv	0.84	global
0.6	0.65	200	45	0.02	indiv	14.35	global
0.7	0.69	:	:	$1 \cdot 10^{-2}$	indiv	15.33	global
0.8	0.73	:	:	9.90	indiv	0.13	indiv
0.3	0.51	:	:	7.19	indiv	3.40	indiv
0.4	0.55	:	:	7.46	indiv	2.90	indiv
0.5	0.60	:	:	6.37	indiv	0.59	indiv
0.6	0.65	200	90	$5 \cdot 10^{-3}$	indiv	10.53	global
0.7	0.69	:	:	$2 \cdot 10^{-3}$	indiv	$5 \cdot 10^{-6}$	indiv
0.8	0.73	:	:	$1 \cdot 10^{-1}$	indiv	$8 \cdot 10^{-3}$	indiv
0.3	0.51	:	:	$1 \cdot 10^{-3}$	global	$2 \cdot 10^{-5}$	indiv
0.4	0.55	:	:	$3 \cdot 10^{-5}$	global	$2 \cdot 10^{-4}$	global
0.5	0.60	:	:	$2 \cdot 10^{-5}$	indiv	$6 \cdot 10^{-6}$	indiv
0.6	0.65	200	150	$4 \cdot 10^{-6}$	global	$1 \cdot 10^{-7}$	indiv
0.7	0.69	:	:	$1 \cdot 10^{-5}$	global	$2 \cdot 10^{-8}$	indiv
0.8	0.73	:	:	$7 \cdot 10^{-6}$	indiv	$3 \cdot 10^{-6}$	indiv
0.3	0.51	:	:	$2 \cdot 10^{-3}$	indiv	$1 \cdot 10^{-5}$	indiv
0.4	0.55	:	:	$1 \cdot 10^{-3}$	indiv	$5 \cdot 10^{-6}$	global
0.5	0.60	:	:	$7 \cdot 10^{-5}$	global	$4 \cdot 10^{-7}$	global
0.6	0.65	200	270	$3 \cdot 10^{-4}$	global	$1 \cdot 10^{-7}$	global
0.7	0.69	:	:	$4 \cdot 10^{-4}$	global	$4 \cdot 10^{-6}$	global
0.8	0.73	:	:	$2 \cdot 10^{-4}$	global	$7 \cdot 10^{-8}$	indiv
0.3	0.51	:	:	$3 \cdot 10^{-3}$	indiv	$1 \cdot 10^{-5}$	indiv
0.4	0.55	:	:	$4 \cdot 10^{-4}$	indiv	$4 \cdot 10^{-6}$	indiv
0.5	0.60	:	:	$2 \cdot 10^{-5}$	indiv	$3 \cdot 10^{-7}$	indiv
0.6	0.65	200	405	$6 \cdot 10^{-5}$	global	$5 \cdot 10^{-9}$	global
0.7	0.69	:	:	$5 \cdot 10^{-5}$	global	$6 \cdot 10^{-9}$	global
0.8	0.73	:	:	$5 \cdot 10^{-5}$	global	$2 \cdot 10^{-8}$	indiv

TABLE B.6 – Comparaison de la médiane et de la variance du Risque entre lissage global et lissage individuel

statistique sur \mathcal{R}_{eucl}	Individuel	Global	Δ	λ
min	0.000226	0.338	0.1	:
max	36.71	2.448	0.1	60
quantile $q_{97.5\%}$	0.0618	1.862	0.1	:
quantile $q_{95\%}$	0.0457	1.682	0.1	:
min	$3.06 \cdot 10^{-6}$	0.474	0.015	:
max	83.53	2.825	0.015	:
quantile $q_{97.5\%}$	0.00322	1.909	0.015	60
quantile $q_{95\%}$	0.00236	1.794	0.015	:
min	$4 \cdot 10^{-6}$	$6 \cdot 10^{-6}$	0.062	:
max	67.89	0.01	0.062	210
quantile $q_{97.5\%}$	$4.9 \cdot 10^{-3}$	$4.6 \cdot 10^{-3}$	0.062	:
quantile $q_{95\%}$	$3.6 \cdot 10^{-3}$	$3.5 \cdot 10^{-3}$	0.062	:

TABLE B.7 – Quelques statistiques sur la distribution du risque euclidien en fonction de la méthode de sélection de la fenêtre de lissage



Dans l'ensemble des graphes suivants se trouve une erreur de labélisation des graphes : La métrique \mathcal{R} considérée est bel et bien la norme euclidienne **au carré** $\mathcal{R}(\Theta, \Delta) = \left\| \hat{\Theta} - \tilde{\Theta} \right\|_2^2$.

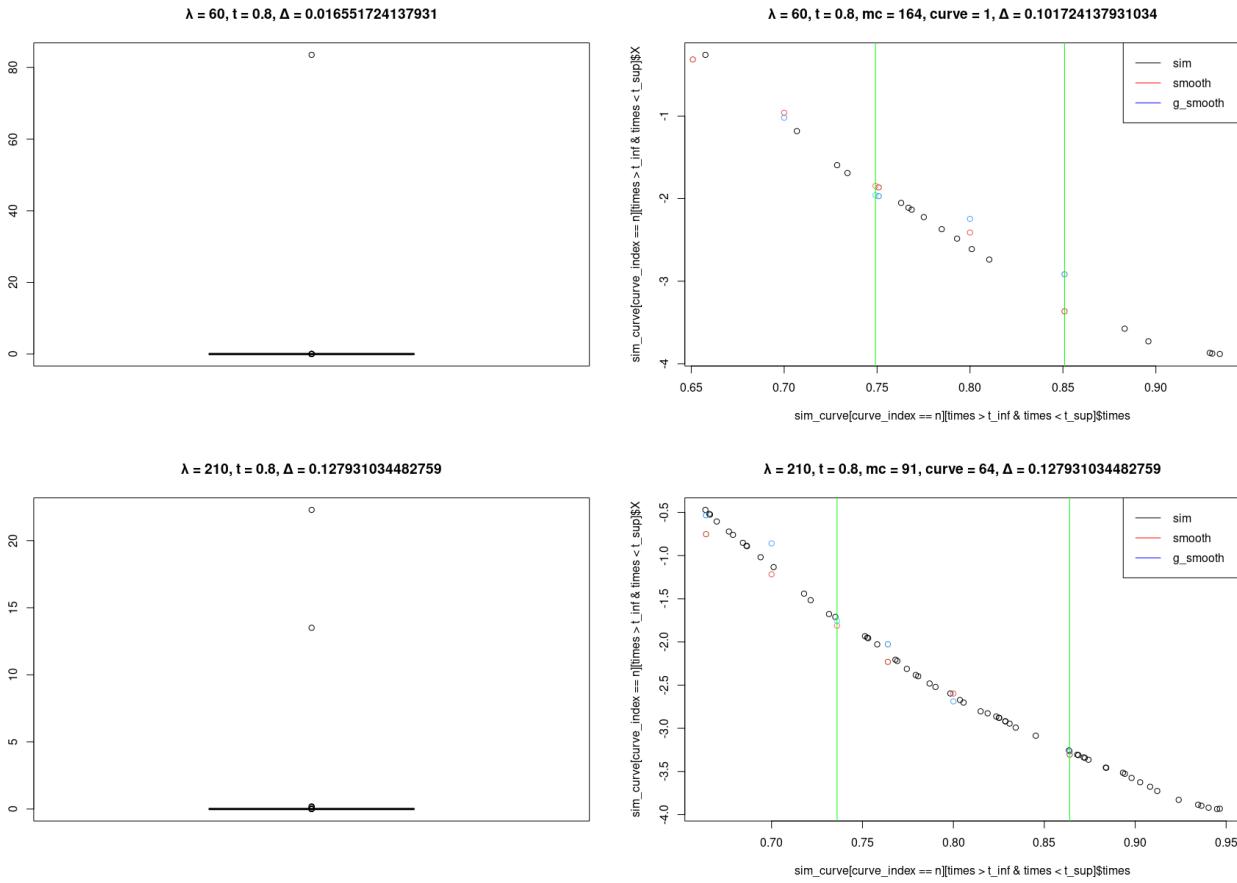


FIGURE B.2 – Distribution des risques et aperçu d'une courbe pour un échantillon de monte carlo extrême sur le risque euclidien.

$\lambda = 60, t = 0.8, mc = 164, \Delta = 0.101724137931034$

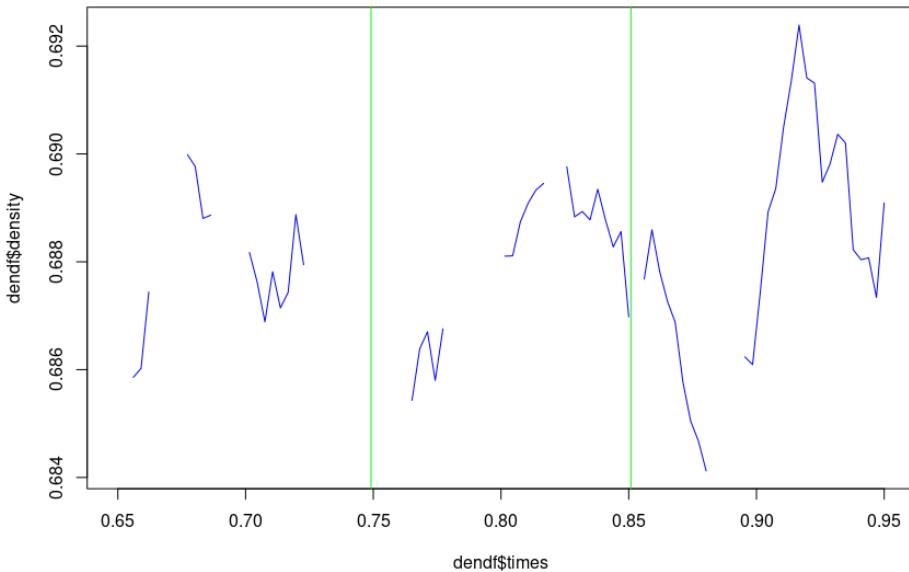


FIGURE B.3 – Densité de points observés sur $[0.65, 0.95]$ pour $\lambda = 60$ sur un échantillon de monte carlo extrême, en un Δ problématique.

$\lambda = 210, t = 0.8, mc = 6, \Delta = 0.0624137931034483$

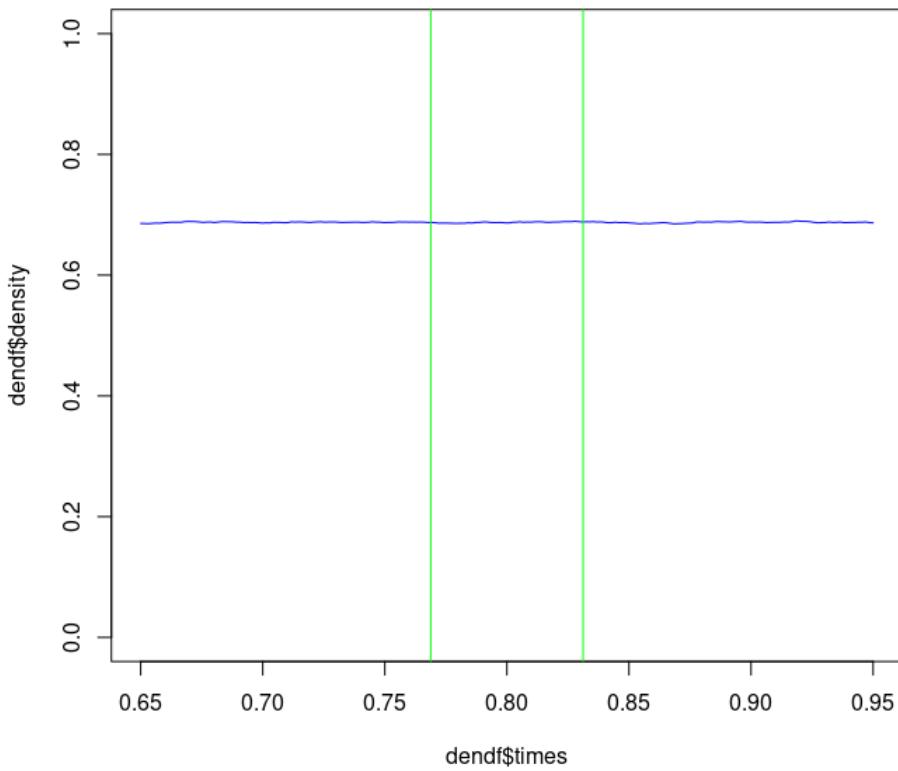
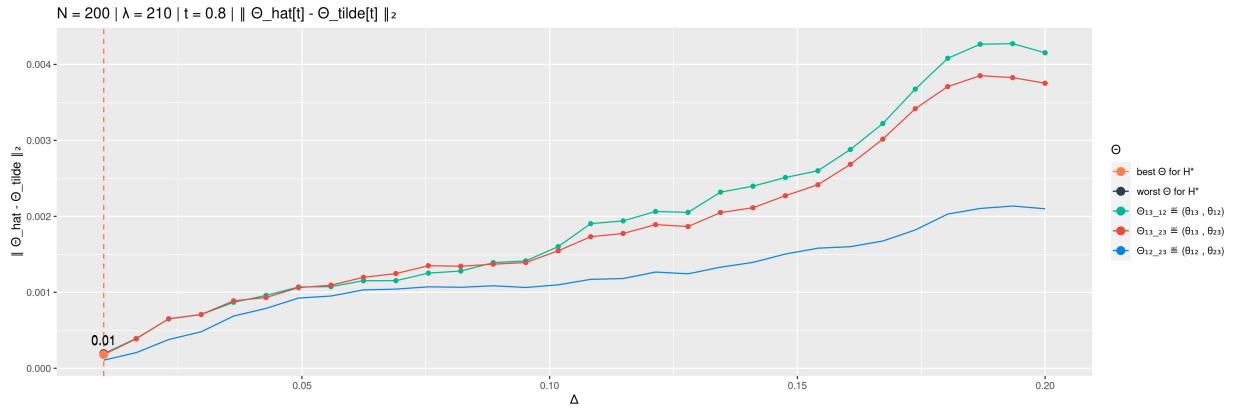
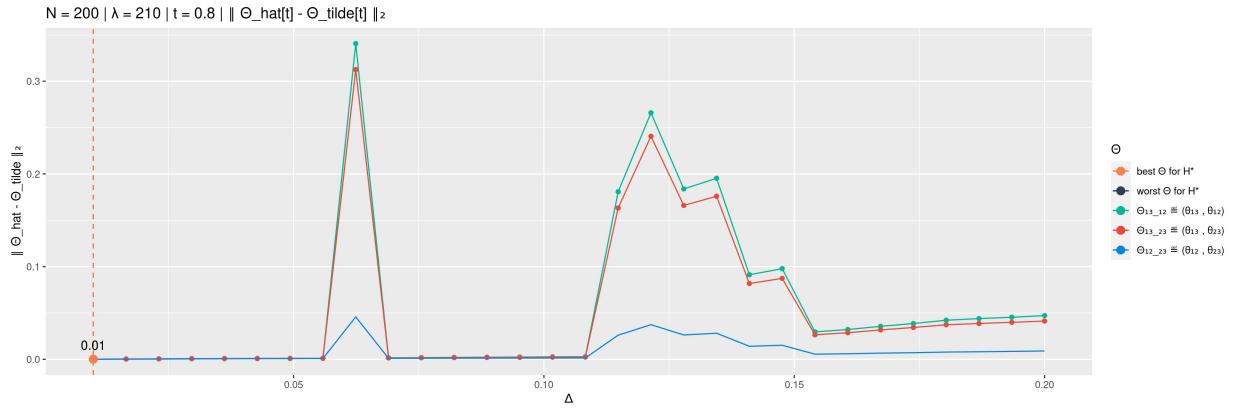


FIGURE B.4 – Densité des points observés correspondant à la courbe présentée sur la figure B.2.

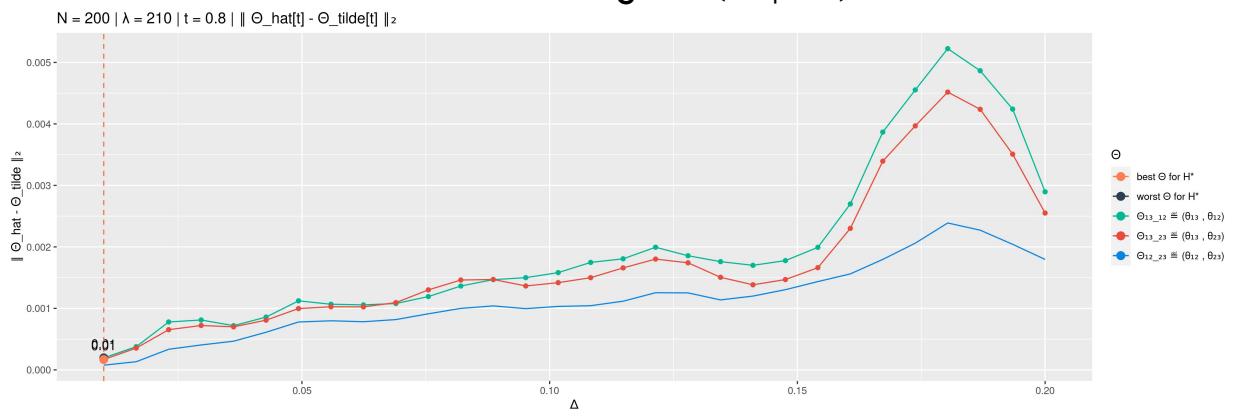
avec extrêmes : global



avec extrêmes : individuel



sans extrêmes : global (- top 2%)



sans extrêmes : individuel (- top 2%)

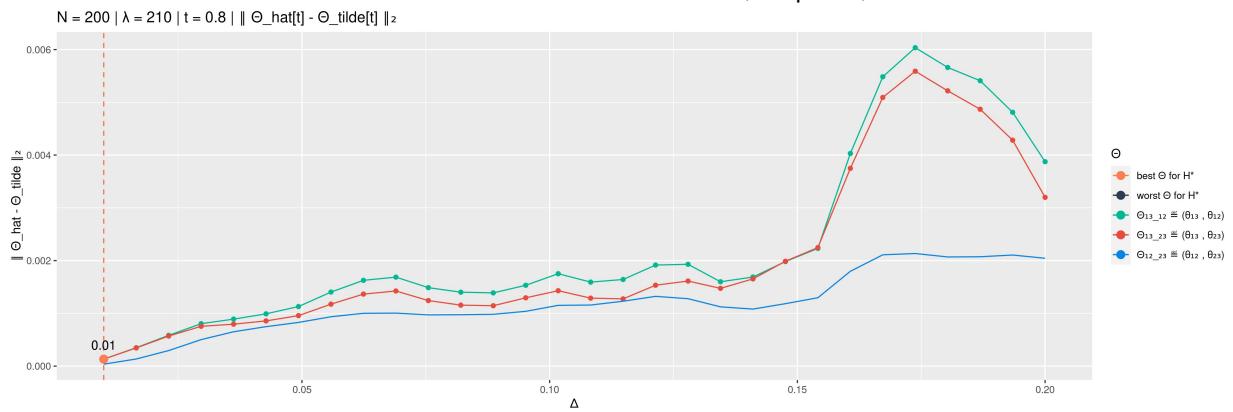
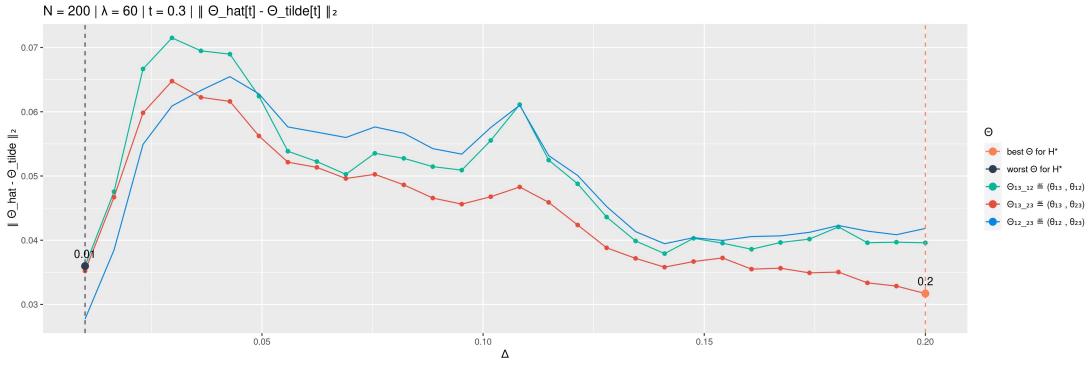


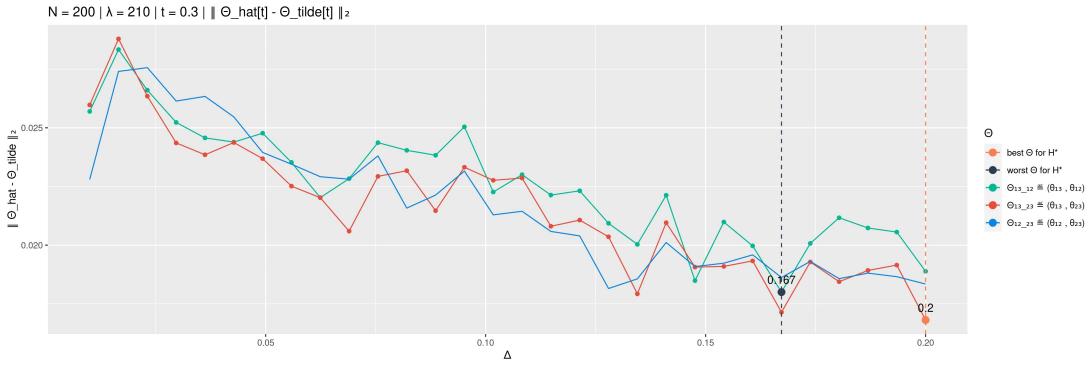
FIGURE B.5 – Risque Euclidien pour $N = 200$, $\lambda = 210$ en un point régulier selon la méthode utilisée pour la fenêtre de lissage

H = 0.51

Sparse :

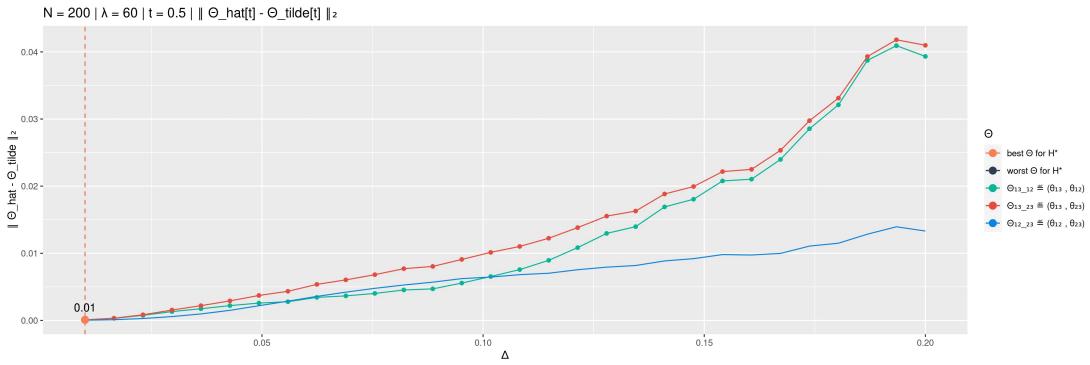


Dense :

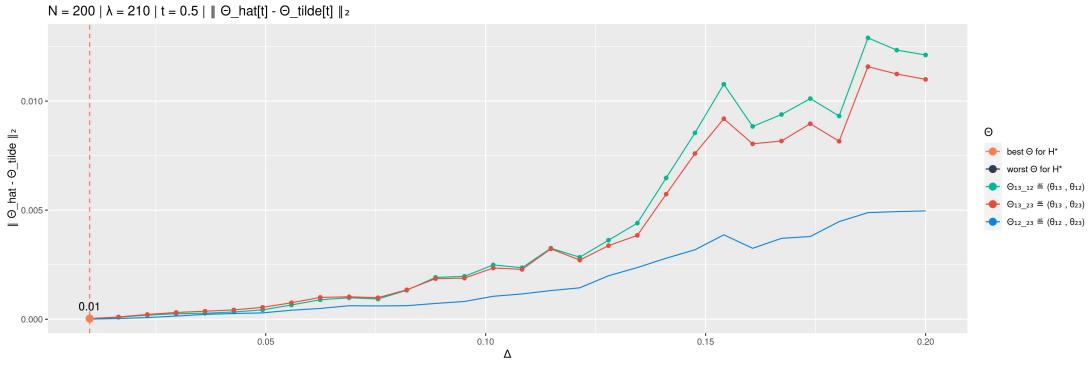


H = 0.6

Sparse :

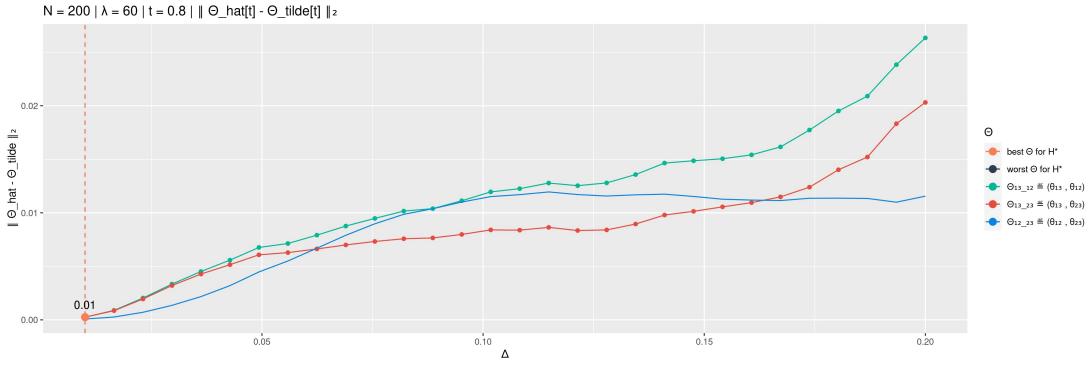


Dense :



H = 0.73

Sparse :



Dense :

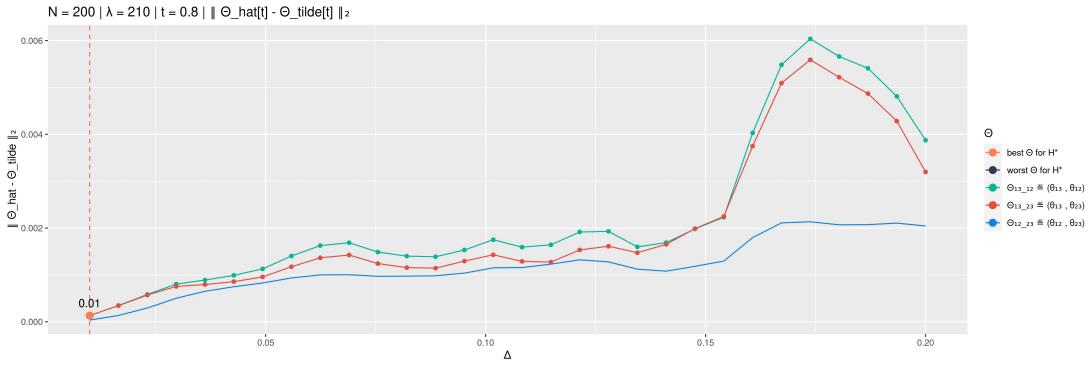
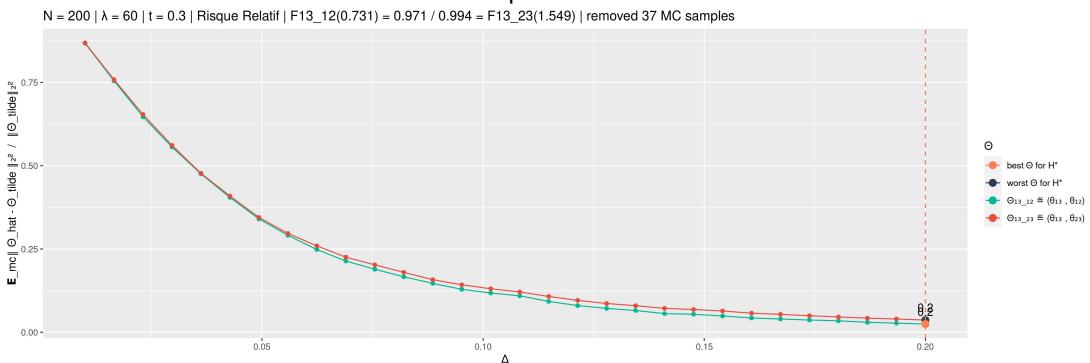


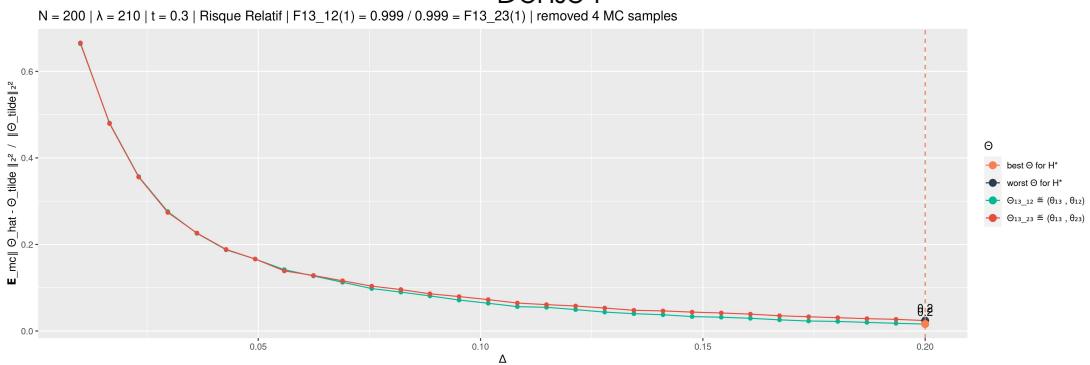
FIGURE B.6 – Graphe des risques **euclidiens** dans les cas « sparse » et « raisonnablement dense », ayant enlevé les observations extrêmes (< 20% des échantillons de monte carlo)

H = 0.51

Sparse :

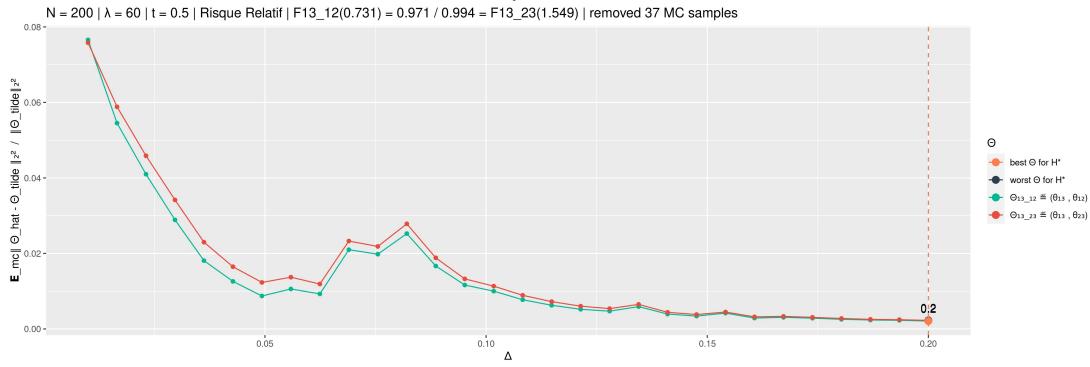


Dense :

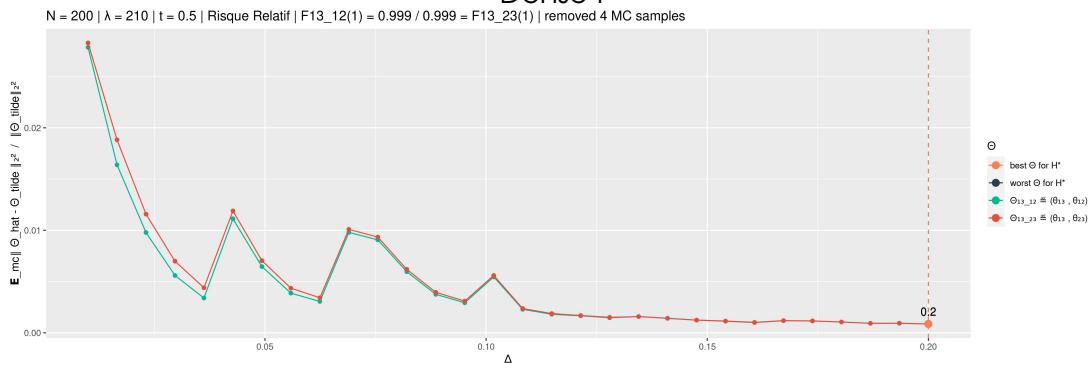


H = 0.6

Sparse :

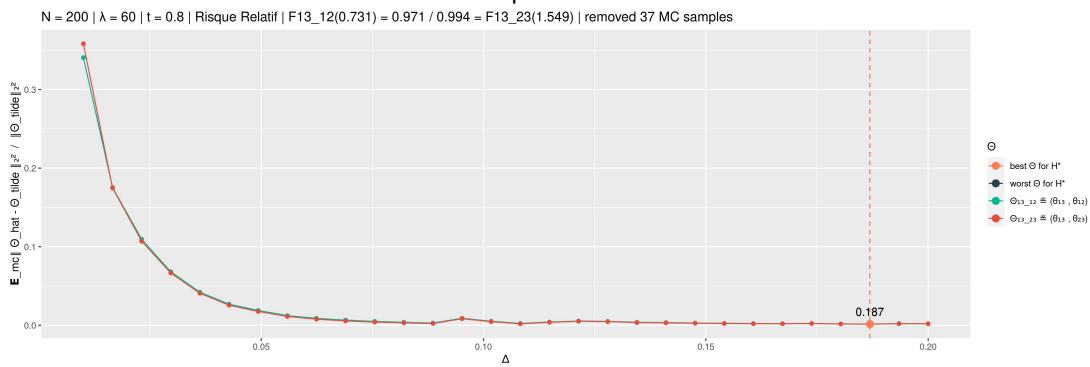


Dense :



H = 0.73

Sparse :



Dense :

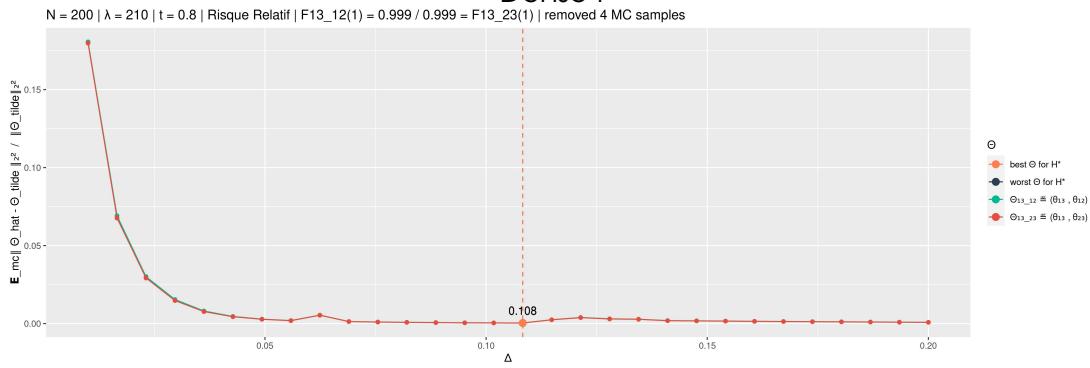


FIGURE B.7 – Graphe des **risques euclidiens relatifs** à $\tilde{\Theta}(\Delta)$ dans les cas « sparse » et « raisonnablement dense », ayant enlevé les observations extrêmes ($\leq 20\%$ des échantillons de monte carlo)

B.4 Etude de l'impact de la méthode de pré-lissage sur l'estimation de la régularité et le Δ optimal

On peut aussi se poser la question suivante, traitée ici en annexe pour ne pas flouter la mission principale de ce stage :



④ : La méthode de pré-lissage utilisée possède-t-elle un impact important sur l'estimation de la régularité et donc sur la stratégie de sélection du Δ en pratique ?



Peut-on quantifier le biais introduit par le lissage en utilisant les ondelettes sur l'estimation de la régularité locale ?

(regarder ce que ça donne, en utilisant les différents théorèmes et bornes disponibles sur les ondelettes pour un processus Holder LORSQUE J AI LE TEMPS - certainement en Septembre)

B.4.0 □ A) Pré-lissage Spline

Le lissage spline est certainement une des méthodes de lissage les plus répandues de par sa simplicité d'implémentation. De plus la détermination des hyper-paramètres de lissage via la méthode de GCV permet de déterminer une approximation de base optimale à un coût computationnel relativement faible. Un des plus grands avantages du lissage B-Spline est l'obtention d'une base de fonctions, qui permet à coût de stockage faible de pouvoir prédire des points non observés. Une fois la base déterminée, il ne reste plus qu'à prédire les points non observés en utilisant la base de fonctions et les coefficients de la décomposition de la courbe sur cette base.

On rappelle que l'utilisation de Splines comme méthode de lissage nécessite tout de même de faire des choix : elle est sensible aux nombre de noeuds et leur emplacement. Il est donc nécessaire de les déterminer par validation croisée. Une méthode fréquemment utilisée est d'utiliser un nombre de noeuds \parallel égal au nombre d'observations, et de les placer aux points d'observations. Puis on utilise des splines pénalisées sur leur dérivée seconde ($L = L_{quad} + \lambda \int_0^1 f''(u)du$) et on détermine le paramètre de pénalisation par validation croisée afin de s'affranchir du choix du nombre de noeuds et de leur emplacement. La validation croisée sur la pénalisation est supposée compenser ce choix. Il s'agit de la méthode qui a été utilisée dans le cadre de ce stage, car très populaire et simple à mettre en place.

Il est à noter qu'une autre méthode de lissage spline est de déterminer le nombre de noeuds k par validation croisée, et de placer les points de façon uniforme sur les quantiles de la distribution des observations. Ce qui ne sera pas utilisé dans le cadre de ce stage.

En effectuant un pré-lissage de splines cubiques naturelles sur une courbe Höldérienne, on ne s'attend pas à obtenir de bonnes performances sur l'estimation de la régularité locale. En effet les courbes splines sont par construction de classe C^2 (fonctions polynomiales C^∞ avec des raccordements C^2), et la courbe lissée écrasera complètement l'information de régularité. Même si il s'agit de ce que l'on souhaite obtenir et qu'on ne connaît pas encore la régularité, il est raisonnable de penser qu'être précautionneux dans le choix de la technique de lissage de telle façon à être le plus proche de la régularité d'une fonction qui pourrait potentiellement ne même pas être dérivable est une bonne idée.

B.4.0 □ B) Pré-lissage à noyaux

Considérer un lissage non paramétrique à noyaux est une alternative au lissage spline. L'espoir est la détermination lors du pré-lissage d'utiliser une fenêtre de lissage qui permette de mieux conserver l'information irrégulière que les splines via la détermination du $h_{\text{pre}}^{*[cv]}$ optimal par validation croisée.

Pour rappel, la fenêtre de lissage retenue est une fenêtre de lissage déterminée par validation croisée, qui est un estimateur de la fenêtre de lissage optimale pour le risque quadratique qui peut s'exprimer en fonction de la régularité locale si l'on suppose les hypothèses retenues sur le processus par MPV (17). Même si le $h_{\mathcal{R}_{\text{quad}}}^*$ est techniquement une fonction de $t \in \mathcal{T}$, l'estimateur que l'on considère lui sera sélectionné pour l'ensemble du support de la courbe \mathcal{T} . On peut espérer que si la courbe change de régularité sur son support mais que celui-ci ne varie pas trop, alors la fenêtre de lissage sélectionnée sera adaptée à la régularité locale de la courbe peu importe où l'on se trouve sur le support.

B.4.0 □ C) Lisser en utilisant une base de fonction sans écraser l'information irrégulière ?

Le lissage spline donne une fonction de classe C^2 , ce qui est un désavantage dans le cadre du prélassage qui sert à déterminer les paramètres de régularité de courbes issues d'un processus que l'on ne suppose pas plus régulier que continu. Toutefois, le fait d'utiliser une base de fonctions pour effectuer le lissage a de nombreux avantages par rapport au lissage à noyaux qui peuvent éventuellement s'avérer utiles dans certaines situations spécifiques pour la mise en production de modèles.

En effet, une fois que l'on a déterminé les composantes de la décomposition de notre signal sur la base de fonctions, on n'a plus besoin de se référer aux données pour prédire une valeur. Il s'agit d'une méthode très économique en mémoire, ce qui peut être très avantageux dans le cadre de la mise en production de modèles lorsqu'il y a de nombreuses courbes observées.

B.4.1 Ondelettes

B.4.1 □ A) Une brève introduction aux ondelettes

Les ondelettes proviennent du monde du traitement du signal. Elles répondent à un problème de représentation des données à la fois dans le domaine temporel et dans le domaine fréquentiel. En effet, la transformée de Fourier nous donne accès aux fréquences présentes dans un signal mais ne nous permet pas de localiser à quel moment sont intervenues les fréquences spécifiques. Le théorème d'indétermination de Heisenberg stipule que l'on ne peut avoir une résolution parfaite à la fois dans le domaine fréquentiel et le domaine temporel, il y a un compromis qui doit être fait. La question devient alors :



Comment représenter une fonction dans le domaine temporel et dans le domaine fréquentiel de façon optimale ? En d'autres termes, quelle résolution temporelle et quelle résolution fréquentielle choisir ?

Une première approche proposée en 1946 par Denis Gabor est la transformée de Fourier à court terme (STFT). Celle-ci consiste à regarder la transformée de Fourier d'une fonction sur une fenêtre de taille fixe et à faire glisser cette fenêtre sur la fonction. On obtient ainsi la représentation fréquentielle de la fonction sur un intervalle de temps centré en un point que l'on peut faire varier.

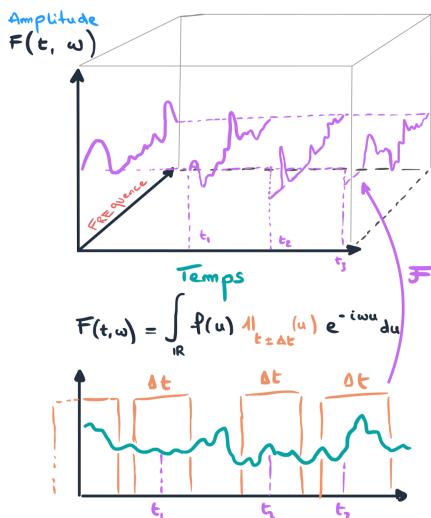


FIGURE B.8 – Transformée de Fourier à court terme d'une fonction

Cependant contrairement à ce que peut suggérer le dessin présenté ici, la résolution fréquentielle n'est pas parfaite. Elle est d'ailleurs dans le cadre de la Transformée de Fourier à court terme constante, que ce soit sur le domaine temporel ou le domaine fréquentiel. La résolution fréquentielle est donc constante quelque soit la fréquence considérée.



Quel est le problème avec cette approche ?

le problème ne vient pas du monde mathématique mais plutôt du monde réel : les signaux que l'on observe présentent la caractéristique suivante : Les signaux de basse fréquence ont tendance à s'étendre sur la durée, et les signaux de hautes fréquences ont tendance à être très localisés, sous forme d'impulsion. Il devient alors clair que pour correctement identifier et localiser les fréquences présentes dans un signal, il est judicieux (voire parfois nécessaire) de varier la résolution fréquentielle et temporelle (limitées par le théorème d'indétermination de Heisenberg) en fonction de ce qui est le plus difficile à distinguer. C'est ce que proposent les ondelettes.

B.4.1 □ B) Motivation dans le cadre de l'analyse de données fonctionnelles

La capacité de capturer de façon efficiente les irrégularités³ de la fonction lissée est une motivation pour l'utilisation de la base d'ondelettes pour effectuer le pré-lissage de données, dont on espère qu'il n'écrase pas la majorité de l'information irrégulière de nos données. Si une des méthodes possibles, comme mentionnée précédemment, est d'utiliser un lissage non paramétrique à noyaux, les bases de fonctions ont de nombreux avantages. Un des avantages est le fait qu'une fois les projections sur la base déterminées, il n'y a plus besoin de se référer de nouveau aux données originales par la suite. Cela donne une représentation très parcimonieuse des données. Alors pour déterminer la valeur de $\hat{X}(t)$ en un point t non observé, il suffit d'évaluer l'expression $\sum_k \langle X | \psi_k \rangle \psi_k(t)$ avec $(\psi_k)_{k \in [1, K]}$ la base d'ondelettes tronquée déterminée par validation croisée.

3. on pourra se référer pour la justification technique de cette affirmation l'annexe A.7

Annexe C

Un peu d'Histoire

C.1 Histoire des séries temporelles



une grande partie des informations présentées dans cette section histoire provient de la référence (26)

Parmi les étapes importantes du développement des séries temporelles, on peut noter l'article *Time Series Analysis : Forecasting and Control* de Box et Jenkins (1970) qui introduit le modèle ARIMA et une approche aujourd’hui standardisée d’évaluation du modèle à utiliser ainsi que son estimation. Ce développement est dû en grande partie à l’utilisation de telles données dans les secteurs économiques et des affaires afin de suivre l’évolution et la dynamique de différentes métriques

L’étude des séries temporelles a été divisée en l’étude du domaine fréquentiel, qui étudie le spectre des processus pour le décomposer en signaux principaux, et du domaine temporel, qui étudie les dépendances des indices temporels. L’utilisation de chacune des approches était sujet à débats mouvementés jusqu’aux alentours de l’an 2000.

Le développement des capacités de calcul a été une révolution notamment pour l’identification des modèles (le critère AIC, l’estimation par vraisemblance dans les années 1980, ...).

À partir des années 1980, les modèles non linéaires émergent (ARCH par Engle, modèles à seuil ...) et trouvent application en économie notamment. Enfin l’étude multivariée (modèle VAR) fait surface dans les années 1980 par Christopher Sims (28, lien de l’article)

Une large partie de la théorie s’appuie notamment sur l’étude des racines de l’unité, en considérant un polynôme d’opérateur $P(B) = (I + \sum_k a_k B^k)$ à partir duquel les relations d’autocorrelations peuvent se ré-écrire.

C.2 Histoire des données fonctionnelles



Pour une description plus complète de l'histoire du développement de l'analyse fonctionnelle, on pourra se référer à [cet article de Wang, Chiou et Müller \(27\)](#)

Bien que l'histoire du développement de l'Analyse de Données Fonctionnelles (FDA) puisse être retracée jusqu'aux travaux de Grenander et Karhunen (14) dans les années 1940 et 1950, où l'outil a été utilisé pour étudier les courbes de croissance en biométrie, ce sous-domaine de la statistique a été étudié de manière systématique à partir des années 1980.

En effet, c'est J.O. Ramsay qui a introduit l'appellation de "données fonctionnelles" en 1982 (20) et qui contribuera en partie à sa popularisation. La thèse de Dauxois et Pousse en 1976 sur l'analyse factorielle dans le cadre des données fonctionnelles(5) a ouvert la voie à l'analyse par composante principale fonctionnelle (FPCA), un outil clé pour l'étude des données fonctionnelles. La FPCA permet d'étudier des objets fonctionnels qui sont de dimension infinie, difficiles à manipuler et impossibles à observer empiriquement, en dimension finie et surtout sur \mathbb{R}^d que l'on connaît bien.

Au cours des années 2000, de nombreux outils statistiques déjà développés pour des données à valeurs dans \mathbb{R}^d depuis un siècle, tels que la régression linéaire (éventuellement généralisée), les séries temporelles ou encore les modèles additifs, ont été adaptés aux données fonctionnelles. Par exemple, les modèles de régression linéaire fonctionnelle ont été développés avec une réponse fonctionnelle (21) ou scalaire (3) en 1999. Les modèles linéaires généralisés ont également été étudiés (13, 19), avec l'estimation de la fonction de lien par méthode non paramétrique à direction révélatrice (*Single Index Model*) récemment étudiée en 2011 (4). Cette méthode avait déjà été utilisée en économétrie pour des données de \mathbb{R}^d depuis 1963 (23), et leur estimation directe a été étudiée une décennie auparavant par M.Hristache, Juditsky et Spokoiny (12). De même, les modèles additifs ont été étendus aux données fonctionnelles en 1999 par Lin et Zhang (16). Enfin, le livre de Bosq, *Linear Processes in Function Spaces : Theory and Applications* (1), publié en 2000, a contribué au développement des séries temporelles pour les données fonctionnelles.

Depuis lors, des ressources telles que l'ouvrage de Kokoszka et Reimherr, *Introduction to Functional Data Analysis (2017)* (15), rendent la théorie et la mise en production des méthodes d'analyse et de prédiction de données fonctionnelles plus accessibles.

Annexe D

Algorithmes & Implementations

D.1 Algorithmes de simulation

Algorithm 1: Get_single_mc_sim : Génération de FAR pour chaque valeur sur la grille

Data:

Input: \vec{t} : endroits où évaluer la régularité

Régularité de X_n sur $[0, 1]$: $H : t \mapsto H_t$

fonction moyenne : μ

Output: $L = \left\{ L[N, \lambda] : N \in \vec{N}, \lambda \in [30, 45, \dots, 480], n \in [1, N] \right\}$

Result: $\left\{ (T_n^{[\lambda]}[m], X_n(T_n^{[\lambda]}[m]))_{m \in [1, M_n]} : N \in \vec{N}, \lambda \in [30, 45, \dots, 480], n \in [1, N] \right\}$

** ---- Paramètres utilisés ---- **

```
1  $G \leftarrow 100$  # Grille d'approximation de  $l'$ 
2
3  $B \leftarrow 100$  # Burn-in de la relation FAR(1)
4
5  $\vec{\Delta} \leftarrow [0.01 \dots 0.2]_{30}$  # diamètre du voisinage  $J_\Delta$ 
6
7  $\vec{N} \leftarrow [100 200 300 400]_4$  # nombre de courbes
8
9  $\Lambda \leftarrow [30 45 \dots 480]$  # nombre moyen de points observés sur une courbe
10
# ** -----
11  $L = []$ 
12 for  $N \in \vec{N}$  do #  $\vec{N} = [100, 200, 300, 400]$ 
|   for  $\lambda \in \Lambda$  do
|        $(T_n^{[\lambda]}[m], X_n(T_n^{[\lambda]}[m]))_{m \in [1, M_n]} \leftarrow \text{far\_sim}(N, \lambda, H, \mu, B, G, \vec{t}, \vec{\Delta})$  # génération
|   end
|    $L[N, \lambda] \leftarrow (T_n^{[\lambda]}[m], X_n(T_n^{[\lambda]}[m]))_{m \in [1, M_n]}$ 
17 end
18 return  $L$ 
```

Algorithm 2: Simulation de Monte Carlo

```
1 for  $mc \leftarrow 1$  to 200 do
2   |  $L_{mc} \leftarrow \text{Get\_single\_mc\_sim}(H, \mu)$ 
3 end
4 return  $\{L_{mc} : mc \in [1, 200]\}$ 
```

Algorithm 3: far_sim : Simulation d'un FAR(1)

Input: $N \in \mathbb{N}^*$: nb X_n
 $\lambda \in \mathbb{N}^*$: nb moy pts observés
 $H : t \mapsto H_t$: régularité
 $\mu : [0, 1] \rightarrow \mathbb{R}$: moyenne
 $B \in \mathbb{N}^*$: Burn-in du FAR(1)
 $G \in \mathbb{N}^*$: grille méthode des rectangles
 $\overrightarrow{\Delta}$: tous les diamètres testés
 \vec{t} : endroits où évaluer la régularité

Result: $\left\{ (T_n^{[\lambda]}[m], X_n(T_n^{[\lambda]}[m]))_{m \in \llbracket 1, M_n \rrbracket} : n \in \llbracket 1, N \rrbracket, \lambda \& N \text{ fixés} \right\}$
avec $X_{n+1}(t) = \int \beta(u, t) X_n(u) du + \eta_{n+1}$

```

1  $N_t \leftarrow \text{len}(\vec{t})$ 
2  $N_{calc} \leftarrow N + B$ 
3 # points de comparaison avec le lissage
    $T(\Delta) = \begin{bmatrix} t_1[1] = \vec{t}[1] - \frac{\Delta}{2} & t_2[1] = \vec{t}[1] & t_3[1] = \vec{t}[1] + \frac{\Delta}{2} \\ \vdots & \vdots & \vdots \\ t_1[N_t] = \vec{t}[N_t] - \frac{\Delta}{2} & t_2[N_t] = \vec{t}[N_t] & t_3[N_t] = \vec{t}[N_t] + \frac{\Delta}{2} \end{bmatrix}$ 
4 # Points observés
5  $T_{obs} \leftarrow \mathcal{U}([0, 1])^{\otimes \lambda}$ 
6 # Grille de la méthode des rectangles médians
7  $\text{grid}_f \leftarrow \left( \frac{(k-1)+(k)}{2} \right)_{k \in \llbracket 1, G \rrbracket}$ 
8 # Ensemble des points
9  $T \leftarrow T_{obs} \cup T(\Delta) \cup \text{grid}_f$ 
10  $T \leftarrow \text{order}(T)$ 

# Simulation de mouvement brownien multi fractionnaire de régularité  $(H, L)$ 
11  $L = []$ 
12 for  $k \in \llbracket 1, N_{calc} \rrbracket$  do
13    $\varepsilon_k \leftarrow \text{mfBm\_sim}(T, H, L)$ 
14    $\mu_k \leftarrow \mu(T)$ 
15    $L[k] \leftarrow [T, \varepsilon_k, \mu_k]$  #  $\in \mathcal{M}_{\text{len } T, 3}(\mathbb{R})$ 
16 end

# Simulation de la relation FAR(1)
17  $# X_{n+1}(t) = \mu(t) + \int_0^1 \beta(u, t) X_n(u) du + \varepsilon_{n+1}$ 
18 for  $k \in \llbracket 1, N_{calc} \rrbracket$  do
19    $K_\beta = \begin{bmatrix} \ddots & \vdots & \ddots \\ \dots & \beta(s, t) & \dots \\ \ddots & \vdots & \ddots \end{bmatrix}$ 
20    $I(\beta, \mathbb{X}_{k-1}) = \frac{1}{G} K_\beta \mathbb{X}_{k-1}$ 
21    $\mathbb{X}_k \leftarrow \mu_k + I(\beta, \mathbb{X}_{k-1}) + \varepsilon_k$ 
22 end
23 return  $(\mathbb{X}_n)_{n \in \llbracket B+1, N_{calc} \rrbracket}$ 

```

Bibliographie

- (1) Denis Bosq. *Linear processes in function spaces : theory and applications*, volume 149. Springer Science & Business Media, 2000.
- (2) V. Capasso and D. Bakstein. *An Introduction to Continuous-Time Stochastic Processes : Theory, Models, and Applications to Finance, Biology, and Medicine*. Modeling and Simulation in Science, Engineering and Technology. Springer New York, 2015.
- (3) Hervé Cardot, Frédéric Ferraty, and Pascal Sarda. Functional linear model. *Statistics & Probability Letters*, 45(1) :11–22, 1999.
- (4) Dong Chen, Peter Hall, and Hans-Georg Müller. Single and multiple index functional regression models with nonparametric link. 2011.
- (5) Jacques Dauxois and Alain Pousse. *Les analyses factorielles en calcul des probabilités et en statistique : Essai d'étude synthétique*. PhD thesis, Éditeur inconnu, 1976.
- (6) Ton Dieker. *Simulation of fractional Brownian motion*. PhD thesis, Masters Thesis, Department of Mathematical Sciences, University of Twente . . . , 2004.
- (7) ENGIE. Quelle est la consommation d'électricité par personne dans un foyer ? ([lire en ligne](#)) , 2022.
- (8) Steven Golovkine, Nicolas Klutchnikoff, and Valentin Patilea. Adaptive estimation of irregular mean and covariance functions, 2021.
- (9) Steven Golovkine, Nicolas Klutchnikoff, and Valentin Patilea. Learning the smoothness of noisy curves with application to online curve estimation. *Electronic Journal of Statistics*, 16(1), Jan 2022.
- (10) X. Gourdon. *Les maths en tête. Analyse - 3e édition*. Editions Ellipses, 2020. Théorème et Applications : densité des fonctions dérivables nulle part - pages : 398—399 , ex4 : 401.
- (11) Siegfried Hörmann and Piotr Kokoszka. Weakly dependent functional data. *The Annals of Statistics*, 38(3) :1845 – 1884, 2010.
- (12) Marian Hristache, Anatoli Juditsky, and Vladimir Spokoiny. Direct estimation of the index coefficient in a single-index model. *Annals of Statistics*, pages 595–623, 2001.
- (13) Gareth M James. Generalized linear models with functional predictors. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 64(3) :411–432, 2002.
- (14) Kari Karhunen. Zur spektraltheorie stochastischer prozesse. *Ann. Acad. Sci. Fennicae, A1*, 34, 1946.
- (15) Piotr Kokoszka and Matthew Reimherr. *Introduction to functional data analysis*. CRC press, 2017.
- (16) Xihong Lin and Daowen Zhang. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society Series B : Statistical Methodology*, 61(2) :381–400, 1999.

- (17) Patilea Maissoro, Vimond. Learning smoothness of functional times series under weak dependency assumption. 2023. not available yet.
- (18) Stéphane Mallat. Wavelet tour — part vi — wavelet zoom. ([lire en ligne](#)) .
- (19) Hans-Georg Müller and Ulrich Stadtmüller. Generalized functional linear models. 2005.
- (20) James O Ramsay. When the data are functions. *Psychometrika*, 47 :379–396, 1982.
- (21) James O Ramsay and CJ1125714 Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society : Series B (Methodological)*, 53(3) :539–561, 1991.
- (22) Brian Ripley. Mass r package documentation | support functions and datasets for venables and ripley's mass. ([lire en ligne](#)) .
- (23) William F Sharpe. A simplified model for portfolio analysis. *Management science*, 9(2) :277–293, 1963.
- (24) Yousri Slaoui. Recursive nonparametric regression estimation for dependent strong mixing functional data. *Statistical Inference for Stochastic Processes*, 23(3) :665–697, 2020.
- (25) Stilian A. Stoev and Murad S. Taqqu. How rich is the class of multifractional brownian motions ? *Stochastic Processes and their Applications*, 116(2) :200–221, 2006.
- (26) Ruey S. Tsay. Time series and forecasting : Brief history and future research. *Journal of the American Statistical Association*, 95(450), 2000. DOI : <https://doi.org/10.2307/2669408>.
- (27) Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. *Annual Review of Statistics and its application*, 3 :257–295, 2016. télécharger.
- (28) James H. Stock & Mark W. Watson. Vector autoregressions. *Journal of the American Statistical Association / Journal of Economic Perspective*, 15(4), 2001. page 101 - DOI : <https://doi.org/10.1257/jep.15.4.101> - télécharger.