



UNIVERSITÉ JOSEPH FOURIER
Master 2 Recherche Mathématiques Appliqués
Parcours Statistique
28 juin 2012

ÉQUIVALENCE ASYMPTOTIQUE DANS DES MODÈLES STATISTIQUES NON-PARAMÉTRIQUES

Mémoire de stage

Étudiante

Ester MARIUCCI

Sous la direction de

Mme Sana LOUHICHI
M. Pierre ÉTORÉ

Remerciements

Je remercie très chaleureusement mes directeurs de stage, Mme Sana Louhichi et M. Pierre Étoré, qui m'ont offert l'opportunité d'effectuer un stage au sein de laboratoire LJK. Je les remercie pour leurs conseils, leur disponibilité et leur souci constant de l'avancement de mon rapport.

Je remercie également Mona, Benoît, Pierre-Olivier, Matthias, Pierre-Luc et Julien, tous les stagiaires que j'ai rencontré lors de mon stage, pour les nombreuses informations, conseils et traduction en français qu'ils ont pu m'offrir tout au long de ces quatre mois.

Un grand merci aussi à Jonathan qui m'a toujours aidé à me débrouiller dans le labyrinthe de la bureaucratie, tâche difficile pour un étranger.

Je tiens à remercier aussi tous les membres du laboratoire LJK qui se sont montrés très accueillants et qui ont toujours répondu à mes questions avec gentillesse et disponibilité.

Enfin, je remercie tout particulièrement Marco et ma famille qui ont su me soutenir et m'encourager avec amour et, surtout avec beaucoup de patience.

Résumé

On commence par étudier l'article de Brown et Low (1996) sur l'équivalence asymptotique, au sens de Le Cam, entre le modèle de régression non-paramétrique à variance connue et le modèle de signal avec bruit blanc.

On généralise un tel résultat au cas des modèles de signal perturbé par un processus à sauts. Cette dernière équivalence asymptotique permet d'estimer le drift d'un modèle de signal perturbé par un processus à sauts via des estimateurs de la fonction de régression du modèle équivalent qui lui est associé. De cette manière on peut utiliser la théorie classique, désormais bien développée, sur l'estimation dans un modèle de régression afin d'estimer la fonction de drift d'un modèle à sauts.

Quelques simulations sont présentées.

Introduction

Ce mémoire porte sur l'équivalence asymptotique au sens de Le Cam entre modèles statistiques. Un des objectifs est le développement des outils nécessaires à la compréhension du résultat de Brown et Low (1996) sur l'équivalence asymptotique entre le modèle de régression non paramétrique à variance connue :

$$Y_i = f(X_i) + \sigma(X_i)\varepsilon_i \quad i = 1, \dots, n \quad \varepsilon_i \sim \mathcal{N}(0, 1) \quad \text{i.i.d.} \quad (1)$$

et le modèle de signal avec bruit blanc :

$$dZ_t^{(n)} = f(t)dt + \frac{\sigma(t)}{\sqrt{n}}dB_t. \quad (2)$$

Le chapitre 1 est dédié à la notion de distance de Le Cam entre modèles statistiques. D'une façon naïve, la distance de Le Cam entre deux modèles statistiques qui ont le même espace de paramètres est nulle si et seulement si les deux modèles contiennent la même quantité d'information sur le paramètre inconnu à estimer.

Dire que les modèles (1) et (2) sont asymptotiquement équivalents au sens de Le Cam signifie alors que, pour n qui tend vers l'infini, estimer f dans le modèle de régression ou estimer f dans le modèle de signal avec bruit blanc donne des estimateurs qui ont, pour fonctions de perte bornées, le même risque. D'où l'intérêt de savoir contrôler la distance de Le Cam (c'est l'objectif de deux premiers chapitres).

Dans le chapitre 3, on trouve une description formelle des modèles (1) et (2).

Dans le chapitre 4, on développe en détails la preuve du résultat de Brown et Low, et dans le chapitre 5 on discute quelques applications d'un tel résultat.

Notons qu'une équivalence asymptotique entre modèles statistiques entraîne que les vitesses de convergence des estimateurs minimax coïncident dans les deux modèles. Cela est très utile lorsqu'on connaît bien des "bons estimateurs" dans un modèle mais pas dans le modèle qui lui est asymptotiquement équivalent.

Dans le chapitre 6, on généralise le résultat de Brown et Low au cas des modèles de signal perturbé par un processus à sauts. Plus précisément, on démontre que le modèle de régression à design déterministe

et à variance connue :

$$Y_i = f\left(\frac{i}{n}\right) + \sigma\left(\frac{i}{n}\right)\varepsilon_i \quad i = 1, \dots, n \quad \varepsilon_i \sim \mathcal{N}(0, 1) \text{ i.i.d.} \quad (3)$$

est asymptotiquement équivalent à un modèle de la forme :

$$dY_t^{(n)} = f(t)dt + \frac{\sigma(t)}{\sqrt{n}}dB_t + dX_t, \quad t \in [0, 1]. \quad (4)$$

où $X = (X_t, t \in [0, 1])$ est un processus purement à sauts indépendant du mouvement brownien B .

En particulier, en combinant ce dernier résultat avec le résultat de Brown et Low on déduit que le modèle de signal avec bruit blanc (2) et le modèle de signal perturbé par un processus à sauts (4) sont asymptotiquement équivalents.

Donc si on observe une trajectoire selon le modèle (4) (avec un drift f inconnu et $\sigma(\cdot)$ connu), l'estimation de f se ramène à son estimation dans le modèle (2) ou (3).

Dans le chapitre 7 on discute la qualité des estimateurs de la fonction de drift f qu'on peut construire à partir de la connaissance d'un estimateur de la fonction de régression. Les résultats sont commentés par des simulations.

Dans le chapitre 8 quelques perspectives sont discutées.

Table des matières

I	Rappels	1
1	Déficience au sens de Le Cam	3
1.1	Distance de Le Cam	3
1.2	Exemples	7
1.3	Interprétation en termes de la théorie du risque	8
1.4	Propriétés de la distance de Le Cam	9
1.5	Commentaires	16
2	Distance de Le Cam entre modèles de bruit blanc	19
2.1	Distance L_1 entre processus gaussiens	19
2.2	Distance Δ entre processus gaussiens	23
II	Étude de l'article de Brown et Low et discussion des applications	25
3	Introduction aux modèles	27
3.1	Modèle de régression non paramétrique à variance connue	27
3.2	Modèle de signal avec bruit blanc	27
3.3	Lien entre le modèle de régression non-paramétrique et le bruit blanc	28
3.4	Exemples des espaces des paramètres	29
4	Régression et bruit blanc	31
4.1	Cas déterministe : hypothèses et discussion	31
4.1.1	Équivalence asymptotique	32
4.2	Cas de design aléatoires	36
4.2.1	Hypothèses et discussions	37
4.2.2	Équivalence asymptotique	38
5	Applications et généralisations	41
5.1	Applications	41
5.1.1	Estimation du drift à partir du modèle de régression	45
5.1.2	Estimation de la fonction de régression à partir du modèle de bruit blanc	46
5.1.3	Risques minimax	47
5.1.4	Estimateur de Pinsker	50

5.2	Un peu de littérature	53
III	Généralisation, simulations et perspectives	57
6	Régression et modèles de signal perturbé par un processus à sauts	59
6.1	Distance L_1 entre processus à sauts	60
6.2	Équivalence asymptotique	62
6.3	Applications	65
6.3.1	Estimation du drift à partir du modèle de régression	66
6.3.2	Estimation de la fonction de régression à partir du modèle de signal perturbé par un processus à sauts	67
7	Simulations	69
7.1	Estimateur de Nadaraya-Watson	69
7.1.1	Implémentation de l'estimateur du drift à partir de l'estimateur de Nadaraya-Watson	70
7.2	Estimateurs par projection	74
7.2.1	Implémentation de l'estimateur du drift à partir de l'estimateur par projection : base trigonométrique	75
7.3	Comparaisons	79
8	Conclusion et perspectives	85
A	L'intégrale de Wiener	87
B	Processus gaussiens	91
C	Simulations	93

Table des figures

5.1	Galaxie originale et galaxie compressée	42
5.2	Superposition du graphe de la fonction $f(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$ et de l'estimateur de Pinsker correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$. En noir : la vraie fonction, en blue : l'estimateur de Pinsker.	52
7.1	Superposition du graphe de la fonction $f_1(x) = \sin(4\pi(x - \frac{1}{2})) x - \frac{1}{2} $ et de l'estimateur de Nadaraya-Watson correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$. En noir : la vraie fonction, en blue : l'estimateur de Nadaraya-Watson	71
7.2	Superposition du graphe de la fonction $f_2(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$ et de l'estimateur de Nadaraya-Watson correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$. En noir : la vraie fonction, en blue : l'estimateur de Nadaraya-Watson	71
7.3	Superposition du graphe de la fonction $f_3(x) = \exp(-(\frac{1/2-x}{x(x-1)})^2)$ et de l'estimateur de Nadaraya-Watson correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$. En noir : la vraie fonction, en blue : l'estimateur de Nadaraya-Watson	72
7.4	Superposition du graphe de la fonction $f_1(x) = \sin(4\pi(x - \frac{1}{2})) x - \frac{1}{2} $ et de l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur de Nadaraya-Watson de f_1 dans le modèle de régression. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$. En noir : la vraie fonction, en rouge : l'estimateur dans le modèle à saut via l'estimateur de Nadaraya-Watson	72
7.5	Superposition du graphe de la fonction $f_2(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$ et de l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur de Nadaraya-Watson de f_2 dans le modèle de régression. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$. En noir : la vraie fonction, en rouge : l'estimateur dans le modèle à saut via l'estimateur de Nadaraya-Watson	73
7.6	Superposition du graphe de la fonction $f_3(x) = \exp(-(\frac{1/2-x}{x(x-1)})^2)$ et de l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur de Nadaraya-Watson de f_3 dans le modèle de régression. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$. En noir : la vraie fonction, en rouge : l'estimateur dans le modèle à saut via l'estimateur de Nadaraya-Watson	73

- 7.7 Superposition du graphe de la fonction $f_1(x) = \sin(4\pi(x - \frac{1}{2}))|x - \frac{1}{2}|$ et de l'estimateur par projection (base trigonométrique) correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.
 En **noir** : la vraie fonction, en **blue** : l'estimateur par projection dans le modèle de régression 76
- 7.8 Superposition du graphe de la fonction $f_2(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$ et de l'estimateur par projection (base trigonométrique) correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.
 En **noir** : la vraie fonction, en **blue** : l'estimateur par projection dans le modèle de régression 76
- 7.9 Superposition du graphe de la fonction $f_3(x) = \exp(-(\frac{1/2-x}{x(x-1)})^2)$ et de l'estimateur par projection (base trigonométrique) correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.
 En **noir** : la vraie fonction, en **blue** : l'estimateur par projection dans le modèle de régression 77
- 7.10 Superposition du graphe de la fonction $f_1(x) = \sin(4\pi(x - \frac{1}{2}))|x - \frac{1}{2}|$ et de l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur par projection (base trigonométrique) de f_1 dans le modèle de régression. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.
 En **noir** : la vraie fonction, en **rouge** : l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection 77
- 7.11 Superposition du graphe de la fonction $f_2(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$ et de l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur par projection (base trigonométrique) de f_2 dans le modèle de régression. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.
 En **noir** : la vraie fonction, en **rouge** : l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection 78
- 7.12 Superposition du graphe de la fonction $f_3(x) = \exp(-(\frac{1/2-x}{x(x-1)})^2)$ et de l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur par projection (base trigonométrique) de f_3 dans le modèle de régression. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.
 En **noir** : la vraie fonction, en **rouge** : l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection 78
- 7.13 Superposition du graphe de la fonction $f_1(x) = \sin(4\pi(x - \frac{1}{2}))|x - \frac{1}{2}|$, de l'estimateur de Nadaraya-Watson de f_1 et de l'estimateur de f_1 dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur de Nadaraya-Watson de f_1 dans le modèle de régression.
 En **noir** : la vraie fonction, en **blue** : l'estimateur de Nadaraya-Watson dans le modèle de régression, en **rouge** l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur de Nadaraya-Watson. 80
- 7.14 Superposition du graphe de la fonction $f_2(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$, de l'estimateur de Nadaraya-Watson de f_2 et de l'estimateur de f_2 dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur de Nadaraya-Watson de f_2 dans le modèle de régression.
 En **noir** : la vraie fonction, en **blue** : l'estimateur de Nadaraya-Watson dans le modèle de régression, en **rouge** l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur de Nadaraya-Watson. 80

7.15 Superposition du graphe de la fonction $f_3(x) = \exp\left(-\left(\frac{1/2-x}{x(x-1)}\right)^2\right)$, de l'estimateur de Nadaraya-Watson de f_3 et de l'estimateur de f_3 dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur de Nadaraya-Watson de f_3 dans le modèle de régression.
 En **noir** : la vraie fonction, en **blue** : l'estimateur de Nadaraya-Watson dans le modèle de régression, en **rouge** l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur de Nadaraya-Watson. 81

7.16 Superposition du graphe de la fonction $f_1(x) = \sin\left(4\pi\left(x - \frac{1}{2}\right)\left|x - \frac{1}{2}\right|\right)$, de l'estimateur par projection (base trigonométrique) de f_1 et de l'estimateur de f_1 dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur par projection de f_1 dans le modèle de régression.
 En **noir** : la vraie fonction, en **blue** : l'estimateur par projection dans le modèle de régression, en **rouge** l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection. 81

7.17 Superposition du graphe de la fonction $f_2(x) = \frac{1}{4}\sin(4\pi x) + \frac{1}{3}\cos(4\pi x) - \frac{1}{2}\sin(2\pi x)$, de l'estimateur par projection (base trigonométrique) de f_2 et de l'estimateur de f_2 dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur par projection de f_2 dans le modèle de régression.
 En **noir** : la vraie fonction, en **blue** : l'estimateur par projection dans le modèle de régression, en **rouge** l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection. 82

7.18 Superposition du graphe de la fonction $f_3(x) = \exp\left(-\left(\frac{1/2-x}{x(x-1)}\right)^2\right)$, de l'estimateur par projection (base trigonométrique) de f_3 et de l'estimateur de f_3 dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur par projection de f_3 dans le modèle de régression.
 En **noir** : la vraie fonction, en **blue** : l'estimateur par projection dans le modèle de régression, en **rouge** l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection. 82

Liste des tableaux

7.1	Comparaison des risques moyens associés, respectivement, à l'estimateur par projection de f , à l'estimateur de f dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection, l'estimateur de Pinsker de f et la valeur théorique du risque associé à l'estimateur de Pinsker. On a posé $f(x) = \sin(4\pi(x - 0.5)) x - 0.5 $	84
7.2	Comparaison des risques moyens associés, respectivement, à l'estimateur par projection de f , à l'estimateur de f dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection, l'estimateur de Pinsker de f et la valeur théorique du risque associé à l'estimateur de Pinsker. On a posé $f(x) = e^{-\left(\frac{x-0.5}{x(x-1)}\right)^2}$	84
7.3	Comparaison des risques moyens associés, respectivement, à l'estimateur par projection de f , à l'estimateur de f dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection, l'estimateur de Pinsker de f et la valeur théorique du risque associé à l'estimateur de Pinsker. On a posé $f(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$.	84

Première partie

Rappels

Chapitre 1

Déficiences au sens de Le Cam

1.1 Distance de Le Cam

Une première notion en statistique mathématique est celle d'espace des échantillons. L'espace des échantillons est donné par l'ensemble de toutes les réalisations possibles \mathcal{X} d'une expérience donnée et par la tribu de ses sous-ensembles \mathcal{A} . D'habitude \mathcal{X} est un espace de suites $x = (x_1, x_2, \dots)$ où $x_i \in \mathbb{R}^k$ (comme dans le modèle de régression non paramétrique introduit dans le chapitre 3) ou bien c'est l'espace des fonctions $x = (x_t), t \in I$ (comme dans le modèle de signal avec bruit blanc introduit, lui aussi, dans le chapitre 3).

Nous supposons que sur l'espace des échantillons $(\mathcal{X}, \mathcal{A})$ est donnée une famille $\{P_\theta, \theta \in \Theta\}$ de mesures de probabilité, dépendant du paramètre θ , appartenant à un certain ensemble Θ de paramètres. Le mécanisme P_θ est la loi de probabilité qui gouverne la sortie aléatoire dans \mathcal{X} quand la "vraie" valeur du paramètre est θ .

Mathématiquement, une *expérience statistique* (ou un *modèle statistique*) est définie comme un triplet $\mathcal{E} = (\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$ où $(\mathcal{X}, \mathcal{A})$ est un espace mesurable, Θ est un ensemble et P_θ , pour tout θ dans Θ , est une mesure de probabilité sur \mathcal{A} . L'espace mesurable $(\mathcal{X}, \mathcal{A})$ est appelé *espace des échantillons de l'expérience* \mathcal{E} , tandis que l'ensemble Θ est appelé *espace des paramètres de* \mathcal{E} .

L'espace des échantillons sera souvent aboli dans la notation du modèle statistique, c'est-à-dire qu'on écrira $\mathcal{E} = (P_\theta : \theta \in \Theta)$ à la place de $\mathcal{E} = (\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$.

La question qu'on se pose dans la suite est la suivante : on observe une variable aléatoire X à valeurs dans \mathcal{X} ayant loi P_θ . La valeur du paramètre θ est inconnue et on voudrait estimer θ . Le problème est alors, en utilisant la terminologie de Le Cam, le suivant : *en général, la famille \mathcal{E} est compliquée. On voudrait alors l'approcher par une famille plus simple.*

Une méthode de choix pour passer à une famille "plus simple" repose sur la notion de statistique exhaustive. Pour cela on donne les définitions suivantes :

Définition 1.1. On dit que le modèle statistique $(\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$ est *dominé* s'il existe une mesure μ σ -finie sur $(\mathcal{X}, \mathcal{A})$ telle que, pour tout θ dans Θ , P_θ est absolument continue par rapport à μ . La mesure μ est dite *mesure dominante*.

Définition 1.2. Soit (G, \mathcal{G}) un espace mesurable. Une *statistique* est une application mesurable $S = S(x)$ de l'espace \mathcal{X} dans G , ou, autrement dit, une variable aléatoire définie dans $(\mathcal{X}, \mathcal{A})$ et à valeurs dans (G, \mathcal{G}) .

Définition 1.3. Soit $(\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$ un modèle statistique dominé par la mesure μ . Soit $S = S(x)$ une statistique. On dit que S est une *statistique exhaustive* pour θ (et la famille $\{P_\theta : \theta \in \Theta\}$) si, pour toute variable aléatoire Y définie sur \mathcal{X} et bornée, il existe une version de l'espérance conditionnelle $\mathbb{E}^\theta[Y|S]$ qui soit indépendante de θ et définie μ p.p.

Remarque 1.1.1. Considérons le cas suivant, très usuel dans la pratique :

Soit $(\Omega, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$ un modèle statistique de la forme :

- $\Omega = \mathbb{R}^n$;
- $\mathcal{A} = \sigma(X_1, \dots, X_n)$ où X_i est la i -ème projection ;
- $P_\theta = \nu^\theta \otimes \dots \otimes \nu^\theta$;

Grâce au lemme de Doob, une variable aléatoire T est mesurable par rapport à $\sigma(X_1, \dots, X_n)$ si et seulement si elle est une fonction mesurable de (X_1, \dots, X_n) , c'est-à-dire, si elle est de la forme $T = f(X_1, \dots, X_n)$ avec f \mathcal{A} -mesurable. Donc, dire que T est une statistique exhaustive équivaut à dire que, pour toute fonction f \mathcal{A} -mesurable, l'espérance conditionnelle $\mathbb{E}^\theta[f(X_1, \dots, X_n)|T]$ ne dépend pas de θ , ou, ce qui revient au même, la loi conditionnelle de (X_1, \dots, X_n) sachant T , ne dépend pas de θ .

Le théorème de factorisation suivant donne des conditions nécessaires et suffisantes pour qu'une statistique $S = S(x)$ soit exhaustive.

Théorème 1.1.2. Soit $\{P_\theta : \theta \in \Theta\}$ une famille de loi de probabilité dominée par une mesure μ σ -finie. La statistique S , définie sur $(\mathcal{X}, \mathcal{A})$ et à valeurs dans (G, \mathcal{G}) , sera exhaustive si et seulement s'il existe une fonction $g(S, \theta)$ \mathcal{G} -mesurable (pour tout $\theta \in \Theta$) telle que :

$$dP_\theta(x) = g(S(x), \theta)\mu(dx), \quad \forall x \in \mathcal{X}.$$

Pour mieux comprendre comment les statistiques exhaustives nous permettent de réduire la dimension des données par une transformation préalable, sans perte d'information, de façon à obtenir un modèle "plus simple", considérons un exemple élémentaire.

Exemple 1.1.3. Soit $(X_i)_{i=1}^n$ un vecteur d'observations indépendantes et équidistribuées de loi normale $\mathcal{N}(\theta, 1)$. Il est bien connu que la moyenne empirique

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$$

est une statistique exhaustive pour θ de loi normale $\mathcal{N}(\theta, n^{-1})$.

Dans ce cas l'expérience initiale est donnée par la famille des lois jointes du n -échantillon, θ étant inconnu (pour fixer les idées disons $\theta \in \Theta$ avec $\Theta \subset \mathbb{R}$). Si son observation se limite à la valeur de la statistique exhaustive \bar{X}_n , le statisticien se trouve devant la famille des lois $\mathcal{N}(\theta, n^{-1})$ de celle-ci, qui, de même que la famille des lois de l'échantillon tout entier, dépend du paramètre inconnu θ . Ainsi on obtient deux expériences indexées par le même θ . Si l'on admet qu'une statistique exhaustive contient toute l'information disponible sur θ , alors on est amené à la notion intuitive d'expériences équivalentes :

Proposition 1.1.4. Supposons que $\Theta \subset \mathbb{R}$. Les deux modèles statistiques donnés par les observations

$$\begin{aligned} X_i = \theta + \xi_i, \quad i = 1, \dots, n \quad \xi_i \sim \mathcal{N}(0, 1), \quad \text{indépendantes,} \quad \theta \in \Theta \\ Y = \theta + n^{-1/2}\xi, \quad \xi \sim \mathcal{N}(0, 1), \quad \theta \in \Theta \end{aligned}$$

sont équivalents.

C'est-à-dire que l'idée d'équivalence est que les deux modèles contiennent la même information sur θ . Une façon de rendre formelle cette idée est introduire la notion de la distance de Le Cam. Avant de faire cela rappelons les définitions suivantes :

Définition 1.4. Un *noyau Markovien* π de l'espace mesurable (E, \mathcal{E}) vers l'espace mesurable (F, \mathcal{F}) , $\pi : E \rightarrow F$, est une famille

$$\pi = (\pi(x, \cdot))_{x \in E} \quad (1.1)$$

de probabilités sur \mathcal{F} , ayant E comme ensemble des indices, telle que, pour toute fonction g mesurable positive sur (F, \mathcal{F}) , la fonction positive πg définie dans E par

$$\pi g(x) = \int_F \pi(x, dy)g(y), \quad (1.2)$$

est mesurable sur (E, \mathcal{E}) .

Remarque 1.1.5. Si g est la fonction indicatrice d'un élément B dans \mathcal{F} , en particulier,

$$\pi \mathbb{I}_B(x) = \pi(x, B). \quad (1.3)$$

Remarquons aussi que, pour toute suite $(g_n)_n$ de fonctions positives et mesurables sur (F, \mathcal{F}) , on a les deux propriétés suivantes :

$$g_n \uparrow g \implies \pi g_n \uparrow \pi g, \quad g = \sum_n g_n \implies \pi g = \sum_n \pi g_n. \quad (1.4)$$

La famille de probabilités (1.1) est un noyau Markovien, si et seulement si pour tout élément B dans \mathcal{F} , la fonction $\pi \mathbb{I}_B$ définie par (1.3) est mesurable sur (E, \mathcal{E}) . En fait, si une telle condition est satisfaite, on voit facilement que, pour toute fonction en escalier positive g sur (F, \mathcal{F}) , la fonction πg est mesurable sur (E, \mathcal{E}) . Puis on étend ce résultat au cas d'une fonction mesurable positive g quelconque en exprimant g comme enveloppe supérieure d'une suite croissante $(g_n)_n$ de fonction positive en escalier, et on se ramène au cas précédent grâce à la propriété (1.4).

Proposition 1.1.6. Soit π un noyau Markovien de (E, \mathcal{E}) vers (F, \mathcal{F}) et soit μ une probabilité sur (E, \mathcal{E}) . Alors :

a. La fonction $\pi \mu$ qui à tout élément B de \mathcal{F} associe le nombre $\pi \mu(B)$ défini par

$$\pi \mu(B) = \int_E \pi(x, B) \mu(dx) \quad (1.5)$$

est une probabilité sur \mathcal{F} .

b. Pour toute fonction g mesurable et positive sur (F, \mathcal{F}) on a :

$$\int_E g(x) \pi \mu(dx) = \int_E \pi g(x) \mu(dx).$$

Démonstration. Pour tout élément B de \mathcal{F} et pour toute suite $(B_n)_n$ d'éléments de \mathcal{F} , deux à deux disjoints tels que leur réunion soit B , de la relation $\mathbb{I}_B = \sum_n \mathbb{I}_{B_n}$ on déduit que $\pi \mathbb{I}_B = \sum_n \pi \mathbb{I}_{B_n}$ et donc que $\pi \mu(B) = \sum_n \pi \mu(B_n)$. Ce qui prouve que $\pi \mu$ est une probabilité sur \mathcal{F} .

Démontrons maintenant le point b. Il est vrai dans le cas particulier où $g = \mathbb{I}_B$ avec B élément de \mathcal{F} . Dans le cas général, il suffit d'exprimer g comme une enveloppe supérieure d'une suite croissante $(g_n)_n$ de fonctions positives en escalier et de se ramener au cas précédent en se servant de la propriété (1.4) et du théorème de Beppo Levi. \square

Définition 1.5. Soient P et Q deux mesures définies sur un même espace mesurable $(\mathcal{X}, \mathcal{A})$, qui admettent une densité par rapport à une mesure μ donnée respectivement par $p(x) = \frac{dP}{d\mu}(x)$ et $q(x) = \frac{dQ}{d\mu}(x)$. On définit la *distance de la variation totale* $\|\cdot\|$ des lois P et Q par

$$\|P - Q\| = \sup_{|f| \leq 1} \left| \int_{\mathcal{X}} f(x)p(x)\mu(dx) - \int_{\mathcal{X}} f(x)q(x)\mu(dx) \right|$$

Définition 1.6. (Δ -écart de Le-Cam)

Soient $\mathcal{P}^{(1)} = (\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$ et $\mathcal{P}^{(2)} = (\mathcal{Y}, \mathcal{B}, (Q_\theta)_{\theta \in \Theta})$ deux modèles statistiques, indexés par le même ensemble d'indices Θ . Supposons que les espaces d'observations $(\mathcal{X}, \mathcal{A})$ et $(\mathcal{Y}, \mathcal{B})$ soient des espaces polonais (espaces métrisables, séparables, complets).

La *déficiencia* (le défaut) de $\mathcal{P}^{(1)}$ par rapport à $\mathcal{P}^{(2)}$ est

$$\delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) = \inf_K \sup_{\theta \in \Theta} \|Q_\theta - KP_\theta\| \quad (1.6)$$

où l'inf est pris sur tous noyaux Markoviens K de $(\mathcal{X}, \mathcal{A})$ vers $(\mathcal{Y}, \mathcal{B})$ (indépendants de θ).

La *distance de Le Cam* entre $\mathcal{P}^{(1)}$ et $\mathcal{P}^{(2)}$ est définie comme

$$\Delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) = \max(\delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}), \delta(\mathcal{P}^{(2)}, \mathcal{P}^{(1)})).$$

Définition 1.7. (Modèles statistiques équivalents)

En conformité avec les notations précédentes, on dit que les modèles statistiques $\mathcal{P}^{(1)}$ et $\mathcal{P}^{(2)}$ sont *équivalents* s'ils existent un noyau Markovien $K^{(1)}$ de $(\mathcal{X}, \mathcal{A})$ vers $(\mathcal{Y}, \mathcal{B})$ tel que $K^{(1)}P_\theta = Q_\theta \quad \forall \theta \in \Theta$, et un noyau Markovien $K^{(2)}$ de $(\mathcal{Y}, \mathcal{B})$ vers $(\mathcal{X}, \mathcal{A})$ tel que $K^{(2)}Q_\theta = P_\theta \quad \forall \theta \in \Theta$.

Proposition 1.1.7. Deux modèles statistiques $\mathcal{P}^{(1)}$ et $\mathcal{P}^{(2)}$ sont équivalents si et seulement si $\Delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) = 0$.

Démonstration. Voir Le Cam [27]. □

Lemme 1.1.8. Toujours en conformité avec les notations précédentes, s'il existe une statistique exhaustive $S : \mathcal{X} \rightarrow \mathcal{Y}$ telle que Q_θ est la distribution de S sous P_θ , alors $\Delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) = 0$.

Démonstration. Il faut montrer que $\delta(\mathcal{P}^{(2)}, \mathcal{P}^{(1)}) = 0 = \delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)})$.

Définissons un noyau Markovien $K_\theta : \mathcal{Y} \rightarrow \mathcal{X}$ de la façon suivante :

$$K_\theta(y, A) = \mathbb{E}_{P_\theta}(\mathbb{1}_A | S = y), \quad \text{pour tout } A \in \mathcal{A}.$$

Alors :

$$K_\theta Q_\theta(A) = \int K_\theta(y, A) Q_\theta(dy) = \int \mathbb{E}_{P_\theta}(\mathbb{1}_A | S = y) S_\# P_\theta(dy) = P_\theta(A)$$

où on a noté par $S_\# P_\theta$ la loi image de S sous P_θ qui, par hypothèse, est égale à Q_θ .

Il reste à montrer que le noyau ainsi construit ne dépend pas de θ ce qui il est vrai parce que, par définition d'une statistique exhaustive, pour tout $A \in \mathcal{A}$, $\mathbb{E}_{P_\theta}(\mathbb{1}_A | S = y)$ ne dépend pas de θ . On a donc fabriqué un noyau K , indépendante de θ , tel que, pour tout borelien $A \in \mathcal{A}$, $P_\theta(A) = KQ_\theta(A)$. Cela implique que $\delta(\mathcal{P}^{(2)}, \mathcal{P}^{(1)}) = 0$.

Pour montrer que $\delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) = 0$ il suffit de définir un noyau Markovien $K'_\theta : \mathcal{X} \rightarrow \mathcal{Y}$ comme $K'_\theta(x, B) = \mathbb{1}_B(S(x))$, pour tout $B \in \mathcal{B}$, pour tout $x \in \mathcal{X}$. On a alors que

$$K'_\theta P_\theta(B) = \int \mathbb{1}_B(S(x)) P_\theta(dx) = \int_B Q_\theta(dx) = Q_\theta(B).$$

Par construction, il est clair que le noyau K' ne dépende pas de θ , et donc on a bien montré que $\delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) = 0$. \square

Remarque 1.1.9. Si $\mathcal{P}^{(1)} = (\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$ et $\mathcal{P}^{(2)} = (\mathcal{Y}, \mathcal{B}, (Q_\theta)_{\theta \in \Theta})$ sont deux modèles statistiques dont la distance de Le Cam est 0, alors, à partir d'une observation x dans le modèle $\mathcal{P}^{(1)}$, on peut toujours construire d'une façon artificielle une observation y dans le modèle $\mathcal{P}^{(2)}$ et vice-versa. En effet, grâce à la proposition (1.1.7) il existe un noyau Markovien $K^{(1)} : \mathcal{X} \rightarrow \mathcal{Y}$ tel que $K^{(1)} P_\theta = Q_\theta$ pour tout θ dans Θ , donc, si on observe x dans $\mathcal{P}^{(1)}$, pour obtenir une observation synthétique dans $\mathcal{P}^{(2)}$ il suffit de tirer un y selon la loi $K^{(1)}(x, \cdot)$. On fait la même construction pour générer une observation artificielle dans $\mathcal{P}^{(1)}$ à partir d'une observation dans $\mathcal{P}^{(2)}$.

Définition 1.8. (Équivalence asymptotique)

Soient $\mathcal{P}_n = (P_{n,\theta}, \theta \in \Theta)$ et $\mathcal{Q}_n = (Q_{n,\theta}, \theta \in \Theta)$ deux suites de modèles statistiques, indexées par n et telles que l'espace des observations peut aussi dépendre de n . On dit que les suites $\mathcal{P}_n, \mathcal{Q}_n$ sont *asymptotiquement équivalentes* si $\lim_{n \rightarrow \infty} \Delta(\mathcal{P}_n, \mathcal{Q}_n) = 0$.

1.2 Exemples

Soient \mathcal{E} et \mathcal{F} deux modèles statistiques qui ont le même espace des paramètres Θ ; la question naturelle qu'on se pose est de savoir quelle est l'expérience qui nous donne plus d'informations par rapport au paramètre inconnu θ .

Une façon de quantifier cela est de calculer $\delta(\mathcal{E}, \mathcal{F})$.

Si $\delta(\mathcal{E}, \mathcal{F}) = 0$, on peut dire que \mathcal{F} est *moins informatif* que \mathcal{E} , ou que \mathcal{E} est *meilleur* que \mathcal{F} ou encore que \mathcal{E} est *plus informatif* que \mathcal{F} .

Par contre, même si la définition de déficience est parfaitement raisonnable du point de vue statistique (voir aussi le paragraphe suivant), la calculer explicitement n'est pas du tout facile dans la plupart des cas.

Généralement, tout ce qu'on peut espérer est de trouver des bornes supérieures et inférieures.

En effet considérons une population à N individus, avec N suffisamment grand. Notons par p la proportion de personnes dans la population qui disent qu'ils vont voter pour le candidat X à la prochaine élection présidentielle. Pour estimer p le statisticien prélève un échantillon de 1000 individus.

On peut prendre un tel échantillon sans répétitions, ce qui nous donne l'expérience \mathcal{E}_N , ou avec répétitions, ce qui nous donne une autre expérience \mathcal{F}_N . Il va de soi que $\delta(\mathcal{E}_N, \mathcal{F}_N) = 0$ et en effet cela peut être prouvé d'une façon rigoureuse.

Au contraire, la valeur de $\delta(\mathcal{F}_N, \mathcal{E}_N)$ est difficile à calculer.

Cependant, on peut facilement argumenter qu'elle ne peut pas être beaucoup plus élevée que la probabilité d'avoir des doublons dans l'échantillon avec répétitions et que $\delta(\mathcal{F}_N, \mathcal{E}_N)$ converge vers zéro quand N tend vers l'infini.

Malgré les difficultés inhérentes au calcul de $\delta(\mathcal{E}, \mathcal{F})$, Torgersen (1972) [36] a été en mesure de donner des expressions explicites pour la déficience entre expériences répondant à certaines propriétés d'invariance. Hansen et Torgersen (1974) [22] ont étudié le cas des modèles linéaires. Dans l'un des cas examinés par Torgersen (1974) [37], l'expérience \mathcal{E} est donnée par l'observation de $n + r$ variables aléatoires indépendantes et identiquement distribuées de loi commune $\mathcal{N}(\theta, I)$ où I est la matrice identité de taille k . On peut aussi observer seulement les valeurs des n premières variables, ce qui nous donne l'expérience \mathcal{F} .

Alors, selon Torgersen,

$$\Delta(\mathcal{E}, \mathcal{F}) = \mathbb{P}\left(\log \frac{1 + \alpha}{\alpha} \leq \frac{\chi^2}{k} \leq \frac{1 + \alpha}{\alpha} \log(1 + \alpha)\right),$$

où $\alpha = \frac{r}{n}$ et χ^2 est une variable aléatoire ayant loi du χ^2 à k degré de liberté.

Un autre exemple où les calculs exacts sont difficiles à réaliser mais on peut quand même obtenir des bornes pour la distance de Le Cam, est le suivant.

Notons par $\mathcal{M}(n, \theta)$, où $\theta = (\theta_1, \dots, \theta_m)$, la loi multinomiale. Si on note par nV_θ la matrice qui a en position (i, j) le terme $n\theta_i - n\theta_i\theta_j$, alors, grâce au théorème central limite $\mathcal{M}(n, \theta)$ est "proche" de $\mathcal{N}(n\theta, nV_\theta)$ dans le sens de la convergence faible, pour m fixé et n grand.

Dans un article de 2002, Carter [9] se pose le problème plus profond de trouver une majoration pour la distance de Le Cam $\Delta(\mathcal{N}, \mathcal{M})$ entre les modèles statistiques $\mathcal{M} := \{\mathcal{M}(n, \theta) : \theta \in \Theta\}$ et $\mathcal{N} := \{\mathcal{N}(n\theta, nV_\theta) : \theta \in \Theta\}$ sous des hypothèses de régularité pour Θ . Plus précisément il démontre que s'il existe une constante C_Θ telle que

$$\sup_{\theta \in \Theta} \frac{\max_{i \in \{1, \dots, m\}} \theta_i}{\min_{i \in \{1, \dots, m\}} \theta_i} \leq C_\Theta < \infty,$$

alors

$$\Delta(\mathcal{M}, \mathcal{N}) \leq C'_\Theta \frac{m \log m}{\sqrt{n}}$$

où C'_Θ est une constante qui ne dépend que de C_Θ .

1.3 Interprétation en termes de la théorie du risque

Pour éclaircir la signification statistique de la déficience, rappelons le cadre classique de la théorie de la décision.

Soit $\mathcal{E} = (\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$ un modèle statistique.

Un *problème de décision* pour le modèle \mathcal{E} est la donnée d'un espace mesurable (D, \mathcal{D}) , appelé *espace des décisions*, et d'une fonction $L : D \times \Theta \rightarrow [0, \infty)$, appelée *fonction de perte*, telle que, pour tout $\theta \in \Theta$, $L(\cdot, \theta)$ soit mesurable par rapport à (D, \mathcal{D}) .

Le statisticien observe une valeur $x \in \mathcal{X}$ obtenue à partir de la mesure de probabilité P_θ . Il ne connaît pas la valeur de θ et sa tâche est de sélectionner une décision $z \in D$. Pour l'accomplir, il choisit une mesure de probabilité $\pi(x, \cdot)$ sur D et il tire un point de D au hasard selon la loi $\pi(x, \cdot)$. S'il a tiré z quand la vraie distribution de x est P_θ , il subit une perte $L(z, \theta)$.

La fonction $\pi : x \rightarrow \pi(x, \cdot)$ est appelée *règle de décision randomisée* dans le modèle \mathcal{E} . Autrement dit, π est un noyau Markovien de $(\mathcal{X}, \mathcal{A})$ vers (D, \mathcal{D}) :

- i. $\forall A \in \mathcal{D}$ fixé, $x \mapsto \pi(x, A)$ est \mathcal{A} -mesurable;
- ii. $\forall x \in \mathcal{X}$ fixé, $A \mapsto \pi(x, A)$ est une probabilité sur (D, \mathcal{D}) .

Notons par $\mathcal{L}(D, \mathcal{D})$ l'ensemble de toutes les fonctions de perte L qui satisfont $0 \leq L(z, \theta) \leq 1, \forall z \in D, \forall \theta \in \Theta$ et par $\Pi(\mathcal{E})$ l'ensemble de toutes les règles de décision randomisée dans le modèle \mathcal{E} .

On appelle *risque* de la règle de décision randomisée $\pi \in \Pi(\mathcal{E})$ en θ associé à la fonction de perte $L \in \mathcal{L}(D, \mathcal{D})$ la quantité donnée par

$$R(\mathcal{E}, \pi, L, \theta) = \int_{\mathcal{X}} \left(\int_D L(z, \theta) \pi(x, dz) \right) P_{\theta}(dx).$$

Proposition 1.3.1. (Caractérisation de la déficience)

Soit $\varepsilon > 0$ fixé. Deux modèles statistiques \mathcal{E} et \mathcal{F} vérifient la relation $\delta(\mathcal{E}, \mathcal{F}) \leq \varepsilon$ si et seulement si : pour tout problème de décision avec fonction de perte $L \in \mathcal{L}(D, \mathcal{D})$ et pour toute fonction de décision $\pi \in \Pi(\mathcal{F})$, il existe une fonction de décision $\pi^* \in \Pi(\mathcal{E})$ telle que

$$R(\mathcal{E}, \pi^*, L, \theta) \leq R(\mathcal{F}, \pi, L, \theta) + \varepsilon, \quad \theta \in \Theta.$$

Démonstration. Voir Le Cam [27]. □

Soulignons que cette caractérisation concerne la déficience $\delta(\mathcal{E}, \mathcal{F})$ de \mathcal{E} par rapport à \mathcal{F} , et que δ n'est pas symétrique. Par conséquent, si pour l'expression symétrisée (le Δ -écart) on a $\Delta(\mathcal{E}, \mathcal{F}) \leq \varepsilon$, alors les risques disponibles dans \mathcal{F} sont aussi disponibles dans \mathcal{E} , à ε près, et vice versa, c'est-à-dire, à toute règle de décision π qui est "bonne" pour \mathcal{F} correspond une règle de décision π^* qui est presque aussi bonne pour \mathcal{E} et réciproquement (pour plus de détails voir le chapitre 5, section Risques minimax).

On peut maintenant donner une définition de la déficience en termes de fonction risque. Plus précisément on se donne la suivante :

Proposition 1.3.2. *La définition de déficience donnée par (1.6) est équivalente à la suivante :*

$$\delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) = \sup_{L \in \mathcal{L}(D, \mathcal{D})} \sup_{\pi_1 \in \Pi(\mathcal{P}^{(1)})} \inf_{\pi_2 \in \Pi(\mathcal{P}^{(2)})} \sup_{\theta \in \Theta} |R(\mathcal{P}^{(1)}, \pi_1, L, \theta) - R(\mathcal{P}^{(2)}, \pi_2, L, \theta)|,$$

où le premier sup est pris sur tous les espaces de décision (D, \mathcal{D}) .

Démonstration. Voir le théorème 2 à page 15 de Le Cam [28]. □

1.4 Propriétés de la distance de Le Cam

On rappelle que deux modèles statistiques $\mathcal{P}^{(1)}$ et $\mathcal{P}^{(2)}$ sont équivalents si et seulement si $\Delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) = 0$ (voir la proposition 1.1.7). On notera cette relation d'équivalence par le symbole \sim :

$$\mathcal{P}^{(1)} \sim \mathcal{P}^{(2)} \iff \Delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) = 0.$$

On peut alors énoncer la proposition suivante.

Proposition 1.4.1. *Le Δ -écart est une distance dans l'ensemble des classes d'équivalence déterminées par la relation d'équivalence \sim .*

Démonstration. Comme la propriété de symétrie est satisfaite par construction, il suffit de prouver l'inégalité triangulaire. Considérons les modèles statistiques suivants :

$$\mathcal{P}^{(1)} : \left(\mathcal{X}, \mathcal{A}, (P_\theta^{(1)})_{\theta \in \Theta} \right),$$

$$\mathcal{P}^{(2)} : \left(\mathcal{Y}, \mathcal{B}, (P_\theta^{(2)})_{\theta \in \Theta} \right),$$

$$\mathcal{P}^{(3)} : \left(\mathcal{Z}, \mathcal{C}, (P_\theta^{(3)})_{\theta \in \Theta} \right).$$

Supposons que $\delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) \leq \varepsilon$ et $\delta(\mathcal{P}^{(2)}, \mathcal{P}^{(3)}) \leq \varepsilon'$. Par définition de défience cela signifie qu'il existe un noyau markovien $K : \mathcal{X} \rightarrow \mathcal{Y}$ tel que $\sup_{\theta \in \Theta} \|P_\theta^{(2)} - KP_\theta^{(1)}\| \leq \varepsilon$ et qu'il existe un noyau markovien $K' : \mathcal{Y} \rightarrow \mathcal{Z}$ tel que $\sup_{\theta \in \Theta} \|P_\theta^{(3)} - K'P_\theta^{(2)}\| \leq \varepsilon'$. En particulier, si on considère le noyau $K'K : \mathcal{X} \rightarrow \mathcal{Z}$, on obtient :

$$\begin{aligned} \sup_{\theta \in \Theta} \|P_\theta^{(3)} - K'KP_\theta^{(1)}\| &\leq \sup_{\theta \in \Theta} \|P_\theta^{(3)} - K'P_\theta^{(2)}\| + \sup_{\theta \in \Theta} \|K'P_\theta^{(2)} - K'KP_\theta^{(1)}\| \\ &\leq \varepsilon' + \sup_{\theta \in \Theta} \|K'P_\theta^{(2)} - K'KP_\theta^{(1)}\|. \end{aligned}$$

Or,

$$\begin{aligned} \|K'P_\theta^{(2)} - K'KP_\theta^{(1)}\| &= \sup_A |K'P_\theta^{(2)}(A) - K'KP_\theta^{(1)}(A)| \\ &= \sup_A \left| \int K'(A, y)P_\theta^{(2)}(dy) - \int K'(A, y)KP_\theta^{(1)}(dy) \right| \\ &= \sup_A \left| \int K'(A, y)[P_\theta^{(2)} - KP_\theta^{(1)}](dy) \right| \\ &\leq \varepsilon \end{aligned}$$

On a ainsi montré que si $\varepsilon' \geq \delta(\mathcal{P}^{(1)}, \mathcal{P}^{(3)})$ et si $\forall \varepsilon \geq \delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)})$ alors on a :

$$\sup_{\theta \in \Theta} \|P_\theta^{(3)} - K'KP_\theta^{(1)}\| \leq \varepsilon + \varepsilon'$$

d'où :

$$\begin{aligned} \delta(\mathcal{P}^{(1)}, \mathcal{P}^{(3)}) &= \inf_{\varepsilon' \geq \delta(\mathcal{P}^{(2)}, \mathcal{P}^{(3)})} \inf_{\varepsilon \geq \delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)})} \delta(\mathcal{P}^{(1)}, \mathcal{P}^{(3)}) \\ &\leq \inf_{\varepsilon' \geq \delta(\mathcal{P}^{(2)}, \mathcal{P}^{(3)})} \inf_{\varepsilon \geq \delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)})} \varepsilon + \varepsilon' \\ &= \delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) + \delta(\mathcal{P}^{(2)}, \mathcal{P}^{(3)}). \end{aligned}$$

□

Notre but est de ramener un modèle complexe \mathcal{P} à un modèle "plus simple" \mathcal{Q} . On a alors besoin de savoir contrôler le Δ -écart de la distance de Le Cam. Pour cela, les définitions suivantes seront très utiles. Dans la suite on supposera fixés les deux modèles statistiques $\mathcal{P}^{(1)} = (\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$ et $\mathcal{P}^{(2)} = (\mathcal{X}, \mathcal{A}, (Q_\theta)_{\theta \in \Theta})$ (qui ont le même espace des observations!) et on supposera aussi qu'il existe une mesure μ qui domine les deux familles des lois $\{P_\theta; \theta \in \Theta\}$, $\{Q_\theta; \theta \in \Theta\}$. On posera $f_\theta := \frac{dP_\theta}{d\mu}$, $g_\theta := \frac{dQ_\theta}{d\mu}$.

Définition 1.9. (Distance L_1 entre modèles statistiques)

En conformité avec les notations qui précèdent, on appelle *distance L_1* entre les modèles $\mathcal{P}^{(1)}$, $\mathcal{P}^{(2)}$ la quantité

$$L_1(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) = \sup_{\theta \in \Theta} \int_{\mathcal{X}} |f_{\theta}(x) - g_{\theta}(x)| \mu(dx).$$

Classiquement, pour aboutir à l'équivalence asymptotique entre modèles statistiques on utilise le résultat suivant.

Théorème 1.4.2. *On a*

$$|R(\mathcal{P}^{(1)}, \pi, L, \theta) - R(\mathcal{P}^{(2)}, \pi, L, \theta)| \leq L_1(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) \|L\|. \quad (1.7)$$

où

$$\|L\| = \sup\{L(z, \theta) : \theta \in \Theta, z \in D\}.$$

Par conséquent

$$\Delta(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}) \leq L_1(\mathcal{P}^{(1)}, \mathcal{P}^{(2)}). \quad (1.8)$$

À partir de (1.8) on a donc

$$\Delta(\mathcal{P}_n^{(1)}, \mathcal{P}_n^{(2)}) \rightarrow 0 \quad \text{si} \quad L_1(\mathcal{P}_n^{(1)}, \mathcal{P}_n^{(2)}) \rightarrow 0. \quad (1.9)$$

Démonstration. Il suffit de remarquer que

$$\begin{aligned} & \left| \int_{\mathcal{X}} \int_D L(z, \theta) \pi(x, dz) f_{\theta}(x) \mu(dx) - \int_{\mathcal{X}} \int_D L(z, \theta) \pi(x, dz) g_{\theta}(x) \mu(dx) \right| \\ & \leq \left| \sup_{z \in D} L(z, \theta) \right| \int_{\mathcal{X}} \int_D |\pi(x, dz)| |f_{\theta}(x) - g_{\theta}(x)| \mu(dx) \\ & = \left| \sup_{z \in D} L(z, \theta) \right| \int_D \pi(x, dz) \int_{\mathcal{X}} |f_{\theta}(x) - g_{\theta}(x)| \mu(dx) \end{aligned} \quad (1.10)$$

$$\begin{aligned} & = \left| \sup_{z \in D} L(z, \theta) \right| \int_{\mathcal{X}} |f_{\theta}(x) - g_{\theta}(x)| \mu(dx) \\ & \leq \|L\| \sup_{\theta \in \Theta} \int_{\mathcal{X}} |f_{\theta}(x) - g_{\theta}(x)| \mu(dx). \end{aligned} \quad (1.11)$$

où on a obtenu l'équation (1.10) en appliquant le théorème de Fubini, et l'équation (1.11) en utilisant que $\pi(x, \cdot)$ est une mesure de probabilité sur (D, \mathcal{D}) . \square

Parfois, pour contrôler la distance L_1 , on utilise la distance de Hellinger qui est très utile surtout quand on a des lois produit.

Définition 1.10. Soient P et Q deux mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$ absolument continues par rapport à une mesure μ . Soient $p(x) := \frac{dP}{d\mu}(x)$, $q(x) := \frac{dQ}{d\mu}(x)$. La *distance de Hellinger* entre P et Q est définie par

$$H^2(P, Q) = \int_{\mathcal{X}} \left(p(x)^{1/2} - q(x)^{1/2} \right)^2 \mu(dx)$$

Propriété 1.4.3. On a

$$L_1(P, Q) \leq 2H(P, Q), \quad (1.12)$$

L_1 est la distance entre les mesures P et Q ainsi définies :

$$L_1(P, Q) = \int_{\mathcal{X}} |p(x) - q(x)| \mu(dx).$$

Démonstration. Montrer la propriété (1.12) équivaut à montrer que

$$\left(\int_{\mathcal{X}} |p(x) - q(x)| \mu(dx) \right)^2 \leq 4 \int_{\mathcal{X}} (p(x)^{1/2} - q(x)^{1/2})^2 \mu(dx).$$

Or,

$$\begin{aligned} & \left(\int_{\mathcal{X}} |p(x) - q(x)| \mu(dx) \right)^2 \\ &= \left(\int_{\mathcal{X}} |p(x)^{1/2} - q(x)^{1/2}| |p(x)^{1/2} + q(x)^{1/2}| \mu(dx) \right)^2 \\ &\leq \int_{\mathcal{X}} |p(x)^{1/2} - q(x)^{1/2}|^2 \mu(dx) \int_{\mathcal{X}} |p(x)^{1/2} + q(x)^{1/2}|^2 \mu(dx). \end{aligned}$$

Comme

$$\begin{aligned} \int_{\mathcal{X}} |p(x)^{1/2} + q(x)^{1/2}|^2 &= \int_{\mathcal{X}} (p(x) + q(x) + 2\sqrt{p(x)q(x)}) \mu(dx) \\ &\leq 2 \left(1 + \int_{\mathcal{X}} \sqrt{p(x)q(x)} \mu(dx) \right) \\ &\leq 2 \left(1 + \|\sqrt{p}\|_2 \|\sqrt{q}\|_2 \right) \\ &= 4, \end{aligned}$$

on a bien trouvé

$$\left(\int_{\mathcal{X}} |p(x) - q(x)| \mu(dx) \right)^2 \leq 4 \int_{\mathcal{X}} (p(x)^{1/2} - q(x)^{1/2})^2 \mu(dx),$$

ce qui conclut la preuve. \square

Corollaire 1.4.4. Soient $\mathcal{P}_n = (P_{n,\theta}, \theta \in \Theta)$ et $\mathcal{Q}_n = (Q_{n,\theta}, \theta \in \Theta)$ deux suites de modèles statistiques définis sur un même espace. Si

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} H(P_{n,\theta}, Q_{n,\theta}) = 0,$$

alors

$$\lim_{n \rightarrow \infty} \Delta(\mathcal{P}_n, \mathcal{Q}_n) = 0.$$

Démonstration. La démonstration découle des relations (1.12) et (1.9). \square

Propriété 1.4.5.

$$H^2(\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)) = 2 \left[1 - \left[\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2} \right]^{1/2} \exp \left[-\frac{(\mu_1 - \mu_2)^2}{4(\sigma_1^2 + \sigma_2^2)} \right] \right] \quad (1.13)$$

Démonstration. D'une façon générale,

$$\int_{\mathcal{X}} \left(p(x)^{1/2} - q(x)^{1/2} \right)^2 \mu(dx) = 2 \left[1 - \int_{\mathcal{X}} \sqrt{p(x)q(x)} \mu(dx) \right],$$

dans ce cas on a :

- i. $p(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left\{ -\frac{1}{2} \left(\frac{x-\mu_1}{\sigma_1} \right)^2 \right\}$
- ii. $q(x) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left\{ -\frac{1}{2} \left(\frac{x-\mu_2}{\sigma_2} \right)^2 \right\},$

donc

$$\begin{aligned} \int_{\mathbb{R}} \sqrt{p(x)q(x)} dx &= \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2} \int_{\mathbb{R}} \exp \left\{ -\frac{1}{4\sigma_1^2\sigma_2^2} \left[\left(x\sqrt{\sigma_1^2 + \sigma_2^2} - \frac{\mu_2\sigma_1^2 + \mu_1\sigma_2^2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right)^2 \right. \right. \\ &\quad \left. \left. + \mu_2^2\sigma_1^2 + \mu_1^2\sigma_2^2 - \left(\frac{\mu_2^2\sigma_1^4 + \mu_1^2\sigma_2^4 + 2\mu_2\mu_1\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right) \right] \right\} \\ &= \left[\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2} \right]^{1/2} \exp \left[-\frac{(\mu_1 - \mu_2)^2}{4(\sigma_1^2 + \sigma_2^2)} \right]. \end{aligned}$$

□

Propriété 1.4.6. Si P et Q sont deux mesures produit, $P = \otimes_{j=1}^m P_j$ et $Q = \otimes_{j=1}^m Q_j$, absolument continues par rapport à la mesure de Lebesgue avec $p_j = \frac{dP_j}{dx}$, $q_j = \frac{dQ_j}{dx}$, alors

$$H^2(P, Q) = 2 \left[1 - \prod_{j=1}^m \left[1 - \frac{H^2(P_j, Q_j)}{2} \right] \right]. \quad (1.14)$$

Démonstration.

$$\begin{aligned} H^2(P, Q) &= 2 \left[1 - \int_{\mathbb{R}^m} \sqrt{p(x)q(x)} dx \right] \\ &= 2 \left[1 - \int_{\mathbb{R}^m} \prod_{j=1}^m \sqrt{p_j(x_j)q_j(x_j)} dx_j \right] \\ &= 2 \left[1 - \prod_{j=1}^m \int_{\mathbb{R}} \sqrt{p_j(x_j)q_j(x_j)} dx_j \right] \\ &= 2 \left[1 - \prod_{j=1}^m \left(1 - \frac{H^2(P_j, Q_j)}{2} \right) \right]. \end{aligned}$$

□

Une autre façon de démontrer que deux suites de modèles statistiques $\mathcal{P}_n = (P_{n,\theta}, \theta \in \Theta)$ et $\mathcal{Q}_n = (Q_{n,\theta}, \theta \in \Theta)$ sont asymptotiquement équivalentes est d'utiliser la définition 1.6 et les résultats suivants :

Lemme 1.4.7. (*Théorème de Scheffé*) Soient P et Q deux mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$ absolument continues par rapport à une mesure μ . Soient $p(x) := \frac{dP}{d\mu}(x)$, $q(x) := \frac{dQ}{d\mu}(x)$. On a

$$\|P - Q\| = \frac{1}{2} \int_{\mathcal{X}} |p(x) - q(x)| \mu(dx) = 1 - \int_{\mathcal{X}} \min(p(x), q(x)) \mu(dx)$$

Démonstration. Voir Tsybakov [38], lemma 2.1, page 84. \square

Propriété 1.4.8. On a

$$\frac{H^2(P, Q)}{2} \leq \|P - Q\| \leq H(P, Q) \sqrt{1 - \frac{H^2(P, Q)}{4}}.$$

Démonstration. Voir Tsybakov [38], lemma 2.3, page 86. \square

Corollaire 1.4.9. Soient $\mathcal{P}_n = (P_{n,\theta}, \theta \in \Theta)$ et $\mathcal{Q}_n = (Q_{n,\theta}, \theta \in \Theta)$ deux modèles statistiques qui ont le même espace des échantillons. Si

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} H(P_{n,\theta}, Q_{n,\theta}) = 0,$$

alors les deux modèles sont asymptotiquement équivalents.

Démonstration.

$$\begin{aligned} \delta(\mathcal{P}_n, \mathcal{Q}_n) &= \inf_K \sup_{\theta \in \Theta} \|Q_{\theta,n} - KP_{\theta,n}\| \\ &\leq \sup_{\theta \in \Theta} H(P_{\theta,n}, Q_{\theta,n}) \sqrt{1 - \frac{H^2(P_{\theta,n}, Q_{\theta,n})}{4}} \\ &\rightarrow 0 \quad \text{quand } n \rightarrow \infty. \end{aligned}$$

De même, en utilisant la symétrie de la distance d'Hellinger, on montre que $\delta(\mathcal{Q}_n, \mathcal{P}_n) \rightarrow 0$ quand $n \rightarrow \infty$. On conclut ainsi, par définition de distance de Le Cam, que $\Delta(\mathcal{P}_n, \mathcal{Q}_n) \rightarrow 0$ quand $n \rightarrow \infty$, ce qui termine la preuve. \square

Pour majorer la distance de Le Cam on peut également se servir des distances suivantes :

Définition 1.11. (Distance κ)

Soient P et Q deux mesures positives sur l'espace mesurable $(\mathcal{X}, \mathcal{A})$, absolument continues par rapport à une mesure μ et de densité, respectivement, $p(x) = \frac{dP}{d\mu}(x)$, $q(x) = \frac{dQ}{d\mu}(x)$. Posons $g(x) = \frac{d(P+Q)}{d\mu}(x)$. La distance κ entre P et Q est définie par

$$\kappa^2(P, Q) = \frac{1}{2} \int \frac{(p(x) - q(x))^2}{g(x)} \mu(dx)$$

Propriété 1.4.10. En conformité avec les notations qui précèdent, on a

$$\frac{1}{2} H^2(P, Q) \leq \kappa^2(P, Q) \leq H^2(P, Q)$$

et

$$\kappa^2(P, Q) \leq 2\|P - Q\|.$$

Démonstration. Pour démontrer la première inégalité il suffit de remarquer qu'on peut écrire

$$\kappa^2(P, Q) = \frac{1}{2} \int_{\mathcal{X}} (\sqrt{p(x)} - \sqrt{q(x)})^2 \frac{(\sqrt{p(x)} + \sqrt{q(x)})^2}{g(x)} \mu(dx),$$

et utiliser le fait que

$$g(x) \leq (\sqrt{p(x)} + \sqrt{q(x)})^2 \leq 2g(x).$$

Pour la deuxième inégalité il suffit d'utiliser la propriété 1.4.8. \square

Corollaire 1.4.11. Soient $\mathcal{P}_n = (P_{n,\theta}, \theta \in \Theta)$ et $\mathcal{Q}_n = (Q_{n,\theta}, \theta \in \Theta)$ deux modèles statistiques qui ont le même espace des échantillons. Si

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} \kappa^2(P_{n,\theta}, Q_{n,\theta}) = 0,$$

alors les deux modèles sont asymptotiquement équivalents.

Démonstration. Il découle immédiatement de la propriété (1.4.10). \square

Définition 1.12. (Distance de Pearson)

Soient P et Q deux mesures positives dominées par une mesure μ avec $p(x) = \frac{dP}{d\mu}(x)$ et $q(x) = \frac{dQ}{d\mu}(x)$. La distance de Pearson entre P et Q est définie par

$$\chi^2(P, Q) = \int_{\mathcal{X}} \frac{(q(x) - p(x))^2}{p(x)} \mu(dx)$$

Propriété 1.4.12. On a

$$\chi^2(P, Q) \leq 2\kappa^2(P, Q)$$

Démonstration. La démonstration découle du fait que $p(x) \leq \frac{d(P+Q)}{d\mu}(x)$. \square

Corollaire 1.4.13. Soient $\mathcal{P}_n = (P_{n,\theta}, \theta \in \Theta)$ et $\mathcal{Q}_n = (Q_{n,\theta}, \theta \in \Theta)$ deux modèles statistiques qui ont le même espace des échantillons. Si

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} \chi^2(P_{n,\theta}, Q_{n,\theta}) = 0,$$

alors les deux modèles sont asymptotiquement équivalents.

Démonstration. La démonstration découle des propriétés 1.4.12 et 1.4.10. \square

Une autre façon de majorer la distance de Le Cam est d'utiliser la *divergence de Kullback*.

Définition 1.13. (Divergence de Kullback)

Soient P et Q deux mesures de probabilité définies sur l'espace mesurable $(\mathcal{X}, \mathcal{A})$.

La *divergence de Kullback* entre P et Q est définie comme

$$K(P, Q) = \begin{cases} \int \log \left(\frac{dP}{dQ} \right) dP & \text{si } P \ll Q, \\ +\infty & \text{sinon.} \end{cases}$$

Remarque 1.4.14. La quantité $K(P, Q)$ est bien définie, c'est-à-dire, si $P \ll Q$ alors l'intégrale $\int \log \left(\frac{dP}{dQ} \right) dP$ est bien définie.

En effet on peut démontrer que si $P \ll Q$ alors

$$\int \left(\log \frac{dP}{dQ} \right)_- dP \leq \|P - Q\|$$

où $a_- = \max\{0, -a\}$. (Pour plus de détails sur la preuve voir Tsybakov [38], Lemma 2.2 page 85.)

Remarque 1.4.15. La divergence de Kullback ne définit pas une distance (en particulier elle n'est pas symétrique).

Propriété 1.4.16. Si P et Q sont deux mesures produit, $P = \otimes_{i=1}^n P_i$, $Q = \otimes_{i=1}^n Q_i$, alors

$$K(P, Q) = \sum_{i=1}^n K(P_i, Q_i).$$

Propriété 1.4.17. On a

$$H^2(P, Q) \leq K(P, Q).$$

Démonstration. Il suffit de supposer que $K(P, Q) < \infty$. Dans ce cas il existe une mesure μ et deux fonctions, p et q , tels que

$$K(P, Q) = \int_{\{x:p(x)q(x)>0\}} p(x) \left(\log \frac{p(x)}{q(x)} \right) \mu(dx).$$

Or, comme $-\log(x+1) \geq -x$ si $x > -1$, on a

$$\begin{aligned} K(P, Q) &= \int_{\{x:p(x)q(x)>0\}} p(x) \left(\log \frac{p(x)}{q(x)} \right) \mu(dx) \\ &= 2 \int_{\{x:p(x)q(x)>0\}} p \left(\log \sqrt{\frac{p(x)}{q(x)}} \right) \mu(dx) \\ &= -2 \int_{\{x:p(x)q(x)>0\}} p \log \left(\left[\sqrt{\frac{q(x)}{p(x)}} - 1 \right] + 1 \right) \mu(dx) \\ &\geq -2 \int_{\{x:p(x)q(x)>0\}} p(x) \left[\sqrt{\frac{q(x)}{p(x)}} - 1 \right] \mu(dx) \\ &= -2 \left(\int_{\{x:p(x)q(x)>0\}} \sqrt{p(x)q(x)} - 1 \right) \mu(dx) \\ &= H^2(P, Q). \end{aligned}$$

□

Corollaire 1.4.18. Soient $\mathcal{P}_n = (P_{n,\theta}, \theta \in \Theta)$ et $\mathcal{Q}_n = (Q_{n,\theta}, \theta \in \Theta)$ deux modèles statistiques qui ont le même espace des échantillons. Si

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} K(P_{n,\theta}, Q_{n,\theta}) = 0,$$

alors les deux modèles sont asymptotiquement équivalents.

Démonstration. La démonstration découle de la propriété 1.4.17 et du corollaire 1.4.9

□

1.5 Commentaires

Le point clé dans la théorie de Le Cam est le fait de caractériser l'équivalence entre modèles statistiques en utilisant des "randomizations" entre espace probabilisés (qu'on peut voir comme noyau Markovien dans le cas où les espaces des observations sont des espaces polonais. Pour une définition rigoureuse de "randomization" dans le cas général voir Le Cam(1986) [27]).

Supposons avoir à disposition deux modèles statistiques $\mathcal{P}^{(1)}$ et $\mathcal{P}^{(2)}$ qui consistent, respectivement, en l'ensemble des lois de probabilité $(P_\theta : \theta \in \Theta)$ définies sur l'ensemble mesurable $(\mathcal{X}, \mathcal{A})$ et des lois de probabilité $(Q_\theta : \theta \in \Theta)$ définies sur $(\mathcal{Y}, \mathcal{B})$.

D'une façon naïve, les deux modèles sont équivalents s'ils apportent la même quantité d'informations pour $\theta \in \Theta$.

D'une façon formelle, ils sont équivalents si leur Δ -écart est zéro. La première étape pour majorer le Δ -écart est de proposer un candidat noyau Markovien de \mathcal{X} à \mathcal{Y} . Alors la borne est établie en donnant une borne pour la distance entre les observations transformées via le noyau Markovien et les observations qui proviennent du modèle duquel on souhaite se rapprocher.

Les transformations (noyaux Markoviens) explicites entre modèles sont très utiles car elles génèrent une correspondance entre les estimateurs dans les deux modèles, c'est-à-dire, si la distribution de $T(\mathcal{X})$ est proche à celle de \mathcal{Y} , alors l'estimateur $\hat{\theta}(T(\mathcal{X}))$ a presque le même risque que $\hat{\theta}(\mathcal{Y})$.

Un raisonnement similaire peut être fait dans le cas où on gère deux suites de modèles statistiques qui ont le même espace de paramètres. Considérons à ce propos les suites :

$$\mathcal{P}_n^{(1)} = \left(\mathcal{X}_n, \mathcal{A}_n, \left(P_{\theta,n}^{(1)} \right)_{\theta \in \Theta} \right),$$

$$\mathcal{P}_n^{(2)} = \left(\mathcal{Y}_n, \mathcal{B}_n, \left(P_{\theta,n}^{(2)} \right)_{\theta \in \Theta} \right).$$

Une "randomization", $T(\mathcal{X}_n)$, qui envoie les observations \mathcal{X}_n dans les observations \mathcal{Y}_n , envoie aussi les lois du modèle $\mathcal{P}_n^{(1)}$ aux lois qui rapprochent les distributions $\left\{ P_{\theta,n}^{(2)} : \theta \in \Theta \right\}$ via $P_{\theta,n}^{(1)} T \approx P_{\theta,n}^{(2)}$.

On dira que les expériences sont asymptotiquement équivalents si la distance de la variation totale entre la loi $P_{\theta,n}^{(2)}$ et la distribution de T sous $P_{\theta,n}^{(1)}$ converge vers 0 uniformément en θ . L'effet d'une telle équivalence asymptotique est que toute investigation dans un modèle génère automatiquement un résultat asymptotiquement analogue dans l'autre modèle.

Chapitre 2

Distance de Le Cam entre modèles de bruit blanc

Dans ce chapitre on se propose de majorer la distance de Le Cam entre modèles de bruit blanc. Le cadre qu'on va considérer est le suivant. Soit $I = [a, b]$ un intervalle qui contient 0 et \mathcal{F} une classe fonctionnelle fixée. Notons par P_f la loi du processus défini par l'équation stochastique :

$$dY_t = f(t)dt + \sigma(t)dB_t$$

où σ est une fonction connue et $B = (B_t, t \in [a, b])$ est un mouvement brownien.

On se propose alors de majorer $\Delta(\mathcal{M}, \mathcal{N})$, où

$$\mathcal{M} := (\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I, (P_\mu)_{\mu \in \mathcal{F}}) \quad (2.1)$$

et

$$\mathcal{N} := (\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I, (P_\nu)_{\nu \in \mathcal{F}}) \quad (2.2)$$

avec $\mathcal{C}(I, \mathbb{R})$ l'espace des fonctions continues de I dans \mathbb{R} et \mathcal{C}^I la sigma-algèbre engendrée par les ouverts pour la norme sup.

Grâce au théorème 1.4.2, une façon d'estimer $\Delta(\mathcal{M}, \mathcal{N})$ est de calculer $L_1(\mathcal{M}, \mathcal{N})$. Ce calcul fera l'objet du paragraphe suivant.

2.1 Distance L_1 entre processus gaussiens

Soit $I = [a, b] \subset \mathbb{R}$ un intervalle qui contient 0 et soient $Y = (Y_t, t \in I)$, $Z = (Z_t, t \in I)$ deux processus gaussiens qui ont, respectivement, $\mu(t)$ et $\nu(t)$ comme fonction de moyenne, $\sigma^2(t)$ comme fonction de variance (la même pour tous les deux), et vérifient

$$dY_t = \mu(t)dt + \sigma(t)dB_t \quad (2.3)$$

$$dZ_t = \nu(t)dt + \sigma(t)dB_t \quad (2.4)$$

avec $B = (B_t, t \in I)$ un mouvement brownien, c'est-à-dire une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $(\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I)$. On notera par $\mathbb{W}_{f, \sigma}$ la loi d'un processus X de la forme $dX_t = f(t)dt + \sigma(t)dB_t$. On utilisera l'abréviation \mathbb{W}_σ pour $\mathbb{W}_{0, \sigma}$ et on posera $\mathcal{A}_I = \sigma(B_t : t \in I)$.

Le lemme suivant nous assure qu'il existe une mesure, \mathbb{W}_σ , qui domine à la fois la loi de Y et celle de Z .

Lemme 2.1.1. *Soit $X = (X_t, t \in I)$ un processus gaussien défini par l'équation :*

$$dX_t = f(t)dt + \sigma(t)dB_t.$$

avec f et σ qui vérifient :

$$\int_I \frac{f(s)^2}{\sigma(s)^2} ds < \infty. \quad (2.5)$$

Alors, la mesure de probabilité $\mathbb{W}_{f,\sigma}$ admet une densité par rapport à \mathbb{W}_σ :

$$\frac{d\mathbb{W}_{f,\sigma}}{d\mathbb{W}_\sigma}(\omega) = \exp\left(\int_I \frac{f(s)}{\sigma^2(s)} d\omega_s - \frac{1}{2} \int_I \frac{f^2(s)}{\sigma^2(s)} ds\right) \quad (2.6)$$

Remarque 2.1.2. L'expression $\int_I \frac{f(s)}{\sigma^2(s)} d\omega_s$ doit être comprise comme la limite sur toutes les partitions $a = t_0 < \dots < t_k = b$ de l'intervalle $I = [a, b]$

$$\int_I \frac{f(s)}{\sigma^2(s)} d\omega_s := \lim_{t_0 < \dots < t_k} \sum_{i=1}^k \frac{f(t_{i-1})}{\sigma^2(t_{i-1})} (\omega_{t_i} - \omega_{t_{i-1}}),$$

voir la remarque A.0.6 de l'annexe A pour plus de détails.

Démonstration. Écrivons :

$$dY_t = f(t)dt + \sigma(t) \left[dB_t - \frac{f(t)}{\sigma(t)} dt \right] = \sigma(t)dB_t, \quad (2.7)$$

et définissons un nouveau processus sur (Ω, \mathcal{A}) , $\tilde{B} = (\tilde{B}_t, t \in I)$, de la manière suivante :

$$\tilde{B}_t = B_t - \int_a^t \frac{f(s)}{\sigma(s)} ds.$$

Grâce à l'hypothèse (2.5), le processus $\exp\left(\int_a^t \frac{f(s)}{\sigma(s)} dB_s - \frac{1}{2} \int_a^t \frac{f^2(s)}{\sigma^2(s)} ds\right)$ est une martingale et donc, grâce au théorème de Girsanov, le processus \tilde{B} est un mouvement brownien sous $\tilde{\mathbb{P}}$, où $\tilde{\mathbb{P}}$ est une mesure de probabilité sur (Ω, \mathcal{A}) qui a comme densité par rapport à \mathbb{P} la fonction donnée par :

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} \Big|_{\mathcal{A}_t} = \exp\left(\int_I \frac{f(t)}{\sigma(t)} dB_t - \frac{1}{2} \int_I \frac{f^2(t)}{\sigma^2(t)} dt\right).$$

D'un côté on a

$$dX_t = f(t)dt + \sigma(t)dB_t,$$

de l'autre

$$dY_t = f(t)dt + \sigma(t)d\tilde{B}_t,$$

avec B et \tilde{B} mouvements browniens respectivement sur $(\Omega, \mathcal{A}, \mathbb{P})$ et $(\Omega, \mathcal{A}, \tilde{\mathbb{P}})$.

En particulier, la loi de X sous \mathbb{P} est la même que la loi de Y sous $\tilde{\mathbb{P}}$.

Donc, pour toute fonction $F : \mathcal{C}(I, \mathbb{R}) \rightarrow \mathbb{R}$ mesurable bornée, on a

$$\begin{aligned} \int F(\omega) \mathbb{W}_{f, \sigma}(d\omega) &= \mathbb{E}_{\mathbb{P}}[F(X)] = \mathbb{E}_{\tilde{\mathbb{P}}}[F(Y)] = \mathbb{E}_{\mathbb{P}} \left[F(Y) \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} \right] \\ &= \mathbb{E}_{\mathbb{P}} \left[F(Y) \exp \left(\int_I \frac{f(t)}{\sigma(t)} dB_t - \frac{1}{2} \int_I \frac{f^2(t)}{\sigma^2(t)} dt \right) \right] \\ &= \mathbb{E}_{\mathbb{P}} \left[F(Y) \exp \left(\int_I \frac{f(t)}{\sigma(t)} \frac{dY_t}{\sigma(t)} - \frac{1}{2} \int_I \frac{f^2(t)}{\sigma^2(t)} dt \right) \right] \\ &= \int F(\omega) \exp \left(\int_I \frac{f(t)}{\sigma^2(t)} d\omega_t - \frac{1}{2} \int_I \frac{f^2(t)}{\sigma^2(t)} dt \right) \mathbb{W}_{\sigma}(d\omega), \end{aligned}$$

où on a utilisé (2.7) dans l'avant-dernière égalité. Ceci conclut la preuve. \square

On se propose maintenant de calculer la distance L_1 entre les processus Y et Z définis par les équations (2.3), (2.4). Pour cela notons par g_Y et g_Z les densités de probabilité de Y et de Z par rapport à \mathbb{W}_{σ} , c'est-à-dire :

$$\forall A \in \mathcal{C}^I, \quad \mathbb{P}(Y \in A) = \int_A g_Y(\omega) \mathbb{W}_{\sigma}(d\omega)$$

et

$$\forall A \in \mathcal{C}^I, \quad \mathbb{P}(Z \in A) = \int_A g_Z(\omega) \mathbb{W}_{\sigma}(d\omega).$$

En conformité avec la définition 1.10 et les notations (2.1) et (2.2), la distance L_1 entre les modèles \mathcal{M}, \mathcal{N} est

$$L_1(\mathcal{M}, \mathcal{N}) = \int |g_Y(\omega) - g_Z(\omega)| \mathbb{W}_{\sigma}(d\omega).$$

Dans ce cas, on notera la distance L_1 entre les deux modèles par $L_1(Y, Z)$, ou encore par $L_1(g_Y, g_Z)$.

Proposition 2.1.3. *Soient f et g , les densités respectives d'une variable aléatoire gaussienne $\mathcal{N}(\mu, \sigma^2)$ et de $\mathcal{N}(\nu, \sigma^2)$. Alors*

$$L_1(f, g) = \mathbb{E} \left| \exp \left(X - \frac{(\nu - \mu)^2}{2\sigma^2} \right) - 1 \right| = 2 \left[1 - 2\phi \left(\frac{\nu - \mu}{2\sigma} \right) \right] \quad (2.8)$$

où $X \sim \mathcal{N} \left(0, \frac{(\mu - \nu)^2}{2\sigma^2} \right)$ et ϕ est la fonction de répartition de $\mathcal{N}(0, 1)$.

Démonstration. Sans perte de généralité supposons que $\mu < \nu$. Par définition de la distance L_1 on a

$$\begin{aligned} L_1(f, g) &= \int_{\mathbb{R}} |f(x) - g(x)| dx \\ &= \int_{-\infty}^{\frac{\mu+\nu}{2}} (f(x) - g(x)) dx + \int_{\frac{\mu+\nu}{2}}^{\infty} (g(x) - f(x)) dx. \end{aligned}$$

Or,

$$\begin{aligned} \int_{-\infty}^{\frac{\mu+\nu}{2}} f(x) dx &= \mathbb{P} \left(\mathcal{N}(\mu, \sigma^2) \leq \frac{\mu + \nu}{2} \right) = \mathbb{P} \left(\mathcal{N}(0, 1) \leq \frac{\nu - \mu}{2\sigma} \right) \\ &= \phi \left(\frac{\nu - \mu}{2\sigma} \right) \end{aligned}$$

De même on a

$$\begin{aligned}
\int_{\mathbb{R}} |f(x) - g(x)| dx &= \phi\left(\frac{\nu - \mu}{2\sigma}\right) - \phi\left(\frac{\mu - \nu}{2\sigma}\right) + \\
&+ \left(1 - \phi\left(\frac{\nu - \mu}{2\sigma}\right)\right) - \left(1 - \phi\left(\frac{\mu - \nu}{2\sigma}\right)\right) \\
&= 2\left[\phi\left(\frac{\mu - \nu}{2\sigma}\right) - \phi\left(\frac{\nu - \mu}{2\sigma}\right)\right] \\
&= 2\left[1 - 2\phi\left(\frac{\nu - \mu}{2\sigma}\right)\right]
\end{aligned}$$

Ce qui montre que

$$L_1(f, g) = 2\left[1 - 2\phi\left(\frac{\nu - \mu}{2\sigma}\right)\right]. \quad (2.9)$$

D'autre part on peut aussi exprimer L_1 comme

$$\begin{aligned}
L_1(f, g) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \left| \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) - \exp\left(-\frac{(x - \nu)^2}{2\sigma^2}\right) \right| dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \left| \exp\left(-\frac{(x - \nu)^2 - 2(\mu - \nu)(x - \nu) + (\mu - \nu)^2}{2\sigma^2}\right) - \exp\left(-\frac{(x - \nu)^2}{2\sigma^2}\right) \right| dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \left| \exp\left(\frac{2(\mu - \nu)(x - \nu) - (\mu - \nu)^2}{2\sigma^2}\right) - 1 \right| \exp\left(-\frac{(x - \nu)^2}{2\sigma^2}\right) dx \\
&= \mathbb{E} \left| \exp\left(\frac{2(\mu - \nu)(Y - \nu) - (\mu - \nu)^2}{2\sigma^2}\right) - 1 \right| \\
&= \mathbb{E} \left| \exp\left(\frac{(\mu - \nu)Z}{\sigma} - \frac{(\mu - \nu)^2}{2\sigma^2}\right) - 1 \right| \\
&= \mathbb{E} \left| \exp\left(X - \frac{(\mu - \nu)^2}{2\sigma^2}\right) - 1 \right|,
\end{aligned}$$

où $Y \sim \mathcal{N}(\nu, \sigma^2)$ et $Z \sim \mathcal{N}(0, 1)$. Ce dernier résultat et l'équation (2.9) concluent la preuve. \square

Proposition 2.1.4. *Soient Y et Z deux processus gaussiens définis respectivement par les équations (2.3), (2.4). En conformité avec les notations qui précèdent, on a*

$$L_1(Y, Z) = 2\left(1 - 2\phi\left(-\frac{D}{2}\right)\right), \quad (2.10)$$

où

$$D^2 = \int_I \frac{(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt.$$

En particulier, $L_1(Y, Z) = O(D)$.

Démonstration. Grâce au lemme 2.1.1, on sait que les lois des processus Y et Z satisfont les relations :

$$\begin{aligned}
\mathcal{L}_Y(d\omega) &= \exp\left(\int_I \frac{\mu(t)}{\sigma^2(t)} d\omega_t - \frac{1}{2} \int_I \frac{\mu^2(t)}{\sigma^2(t)} dt\right) \mathbb{W}_\sigma(d\omega) \\
\mathcal{L}_Z(d\omega) &= \exp\left(\int_I \frac{\nu(t)}{\sigma^2(t)} d\omega_t - \frac{1}{2} \int_I \frac{\nu^2(t)}{\sigma^2(t)} dt\right) \mathbb{W}_\sigma(d\omega)
\end{aligned}$$

Donc, en particulier, la loi de Y admet une densité par rapport à la loi de Z donnée par la formule :

$$\begin{aligned} g_{Y/Z} &= \exp \left(\int_I \frac{\mu(t) - \nu(t)}{\sigma^2(t)} d\omega_t - \frac{1}{2} \int_I \frac{\mu^2(t) - \nu^2(t)}{\sigma^2(t)} dt \right) \\ &= \exp \left(\int_I \frac{(\mu(t) - \nu(t))}{\sigma^2(t)} (d\omega_t - \nu(t)dt) - \frac{1}{2} \int_I \frac{(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt \right); \end{aligned} \quad (2.11)$$

d'où,

$$\begin{aligned} L_1(Y, Z) &= \int \left| g_Y(\omega) - g_Z \right| \mathbb{W}_\sigma(d\omega) \\ &= \int \left| \exp \left(\int \frac{\mu(t) - \nu(t)}{\sigma^2(t)} (d\omega_t - \nu(t)dt) - \frac{1}{2} \int \frac{(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt \right) - 1 \right| g_Z(\omega) \mathbb{W}_\sigma(d\omega) \\ &= \mathbb{E}_\mathbb{P} \left| \exp \left(\int \frac{\mu(t) - \nu(t)}{\sigma^2(t)} (dZ_t - \nu(t)dt) - \frac{1}{2} \int \frac{(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt \right) - 1 \right| \\ &= \mathbb{E}_\mathbb{P} \left| \exp \left(\int \frac{(\mu(t) - \nu(t))}{\sigma^2(t)} \sigma(t) dB_t - \frac{1}{2} \int \frac{(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt \right) - 1 \right|; \end{aligned}$$

Or, la variable aléatoire $\int_I \frac{(\mu(t) - \nu(t))}{\sigma(t)} dB_t$ est de loi $\mathcal{N}(0, \int_I \frac{(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt)$, donc, grâce à la proposition 2.1.3, on a

$$L_1(Y, Z) = 2 \left[1 - 2\phi \left(\frac{1}{2} \sqrt{\int_I \frac{(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt} \right) \right],$$

ce qui conclut la preuve. \square

2.2 Distance Δ entre processus gaussiens

Grâce aux résultats des paragraphes précédents (et en gardant les mêmes notations) on peut finalement donner une majoration de la quantité $\Delta(\mathcal{M}, \mathcal{N})$. Il suffit en effet de combiner le théorème 1.4.2, qui nous dit que $\Delta(\mathcal{M}, \mathcal{N}) \leq L_1(\mathcal{M}, \mathcal{N})$, avec la proposition 2.1.4, qui nous assure que $L_1(\mathcal{M}, \mathcal{N}) = 2 \left(1 - 2\phi \left(-\frac{D}{2} \right) \right)$, pour aboutir au résultat suivant :

Proposition 2.2.1. *Sous les notations précédentes, on a*

$$\Delta(\mathcal{M}, \mathcal{N}) \leq 2 \left(1 - 2\phi \left(-\frac{D}{2} \right) \right)$$

où

$$D^2 = \int_I \frac{(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt$$

et ϕ est la fonction de répartition de $\mathcal{N}(0, 1)$.

L'intérêt de la proposition précédente sera plus clair si on se place dans la situation suivante : soient deux suites de modèles statistiques \mathcal{M}_n et \mathcal{N}_n définis par :

$$\mathcal{M}_n := (\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I, (P_{\mu,n})_{\mu \in \mathcal{F}})$$

et

$$\mathcal{N}_n := (\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I, (P_{\nu,n})_{\nu \in \mathcal{F}})$$

étant $(P_{f,n})_{f \in \mathcal{F}}$ la loi du processus

$$dY_t^{(n)} = f(t, n)dt + \sigma(t, n)dB_t.$$

De la proposition 2.2.1 on déduit alors le corollaire :

Corollaire 2.2.2. *En conformité avec les notations qui précèdent, si*

$$\lim_{n \rightarrow \infty} \int_I \frac{(\mu(t, n) - \nu(t, n))^2}{\sigma^2(t, n)} dt = 0, \quad (2.12)$$

alors

$$\lim_{n \rightarrow \infty} \Delta(\mathcal{M}_n, \mathcal{N}_n) = 0.$$

Démonstration. La démonstration découle immédiatement du théorème 1.4.2 et du fait que $L_1(\mathcal{M}_n, \mathcal{N}_n) = O(D)$ avec $D = \sqrt{\int_I \frac{(\mu(t, n) - \nu(t, n))^2}{\sigma^2(t, n)} dt}$. \square

Exemple 2.2.3. Soit $\mathcal{F} = L_2(I)$ et considérons

$$\text{i. } \mu(t, n)(x) = \mu(t)(x) = \sum_{j \in \mathbb{Z}} c_j(\mu) e^{ijx}$$

$$\text{ii. } \nu(t, n)(x) = \sum_{j=-n}^n c_j(\mu) e^{ijx}$$

avec $(c_j(\mu))_{j \in \mathbb{Z}}$ coefficients de Fourier de μ .

Alors il est clair que la condition (2.12) est satisfaite. En effet il suffit d'utiliser le fait que, pour tout $f \in L_2(I)$, $\lim_{n \rightarrow \infty} \|f - S_n(f)\|_2 = 0$ où $S_n(f)(x) = \sum_{j=-n}^n c_j(f) e^{ijx}$.

Deuxième partie

Étude de l'article de Brown et Low et discussion des applications

Chapitre 3

Introduction aux modèles

3.1 Modèle de régression non paramétrique à variance connue

On suppose avoir observé n couples de variables aléatoires (X_{ni}, Y_{ni}) , $i = 1, \dots, n$ indépendantes et identiquement distribuées, liées par la relation suivante :

$$Y_{ni} = f(X_{ni}) + \sigma(X_{ni})\varepsilon_{ni} \quad (3.1)$$

où :

- les X_{n1}, \dots, X_{nn} sont les “design” du modèle, ou variables explicatives. Deux cas peuvent se produire : soit les v.a. X_{ni} sont déterministes, soit elles sont aléatoires. Dans le cas déterministe on supposera que

$$X_{ni} = x_{ni} = H^{-1}(i/n), \quad i = 1, \dots, n; \quad (3.2)$$

où $H : \mathbb{R} \rightarrow [0, 1]$ est une fonction de répartition strictement croissante. Dans le cas où le design est aléatoire on supposera que les X_{ni} sont indépendantes et de fonction de répartition commune H .

- Les ε_{ni} , $i = 1, \dots, n$ représentent le “bruit”. Elles sont i.i.d. de loi normale centrée réduite.
- Les Y_{ni} sont les variables à expliquer.
- f est une fonction réelle inconnue dite *fonction de régression*.
- σ est une fonction connue.

On peut alors construire une suite de modèles statistiques en définissant, pour tout n , le modèle statistique

$$\mathcal{P}_n^Y = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_{n,f}^Y)_{f \in \mathcal{F}}) \quad (3.3)$$

où \mathcal{F} est une classe fonctionnelle fixée à l’avance qui contient la fonction de régression f et $P_{n,f}^Y$ est la loi du vecteur $(Y_{n,1}, \dots, Y_{n,n})$.

3.2 Modèle de signal avec bruit blanc

Soit $I \subset \mathbb{R}$ un intervalle, éventuellement infini, et supposons, sans perte de généralité, que $0 \in I$. On suppose avoir observé, pour tout $n \in \mathbb{N}$, un processus gaussien $Z^{(n)} = (Z_t^{(n)}, t \in I)$ (pour la définition et les propriétés d’un processus gaussien voir l’Annexe B) qui satisfait l’équation

$$dZ_t^{(n)} = f(t)dt + \frac{\lambda(t)}{\sqrt{n}}dB_t. \quad (3.4)$$

où $B = (B_t : t \in I)$ est un mouvement brownien standard, $f : I \rightarrow \mathbb{R}$ est une fonction inconnue qui appartient à un certain espace fonctionnel \mathcal{F} fixé à l'avance, et $\lambda : I \rightarrow \mathbb{R}$ est une fonction connue. Cela nous permet de définir une suite de modèles statistiques :

$$\mathcal{P}_n^Z = (\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I, (P_{n,f}^Z)_{f \in \mathcal{F}}) \quad (3.5)$$

avec $\mathcal{C}(I, \mathbb{R})$ est l'espace des fonctions continues de I dans \mathbb{R} , \mathcal{C}^I est la sigma-algèbre de Borel de $\mathcal{C}(I, \mathbb{R})$ et $P_{n,f}^Z$ est la distribution du processus $Z^{(n)}$ qu'on vient de définir.

Remarque 3.2.1. A priori, l'espace des observations devrait être l'ensemble des toutes les possibles fonctions de I dans \mathbb{R} , mais, comme l'intégrale de Wiener est continue, on considérera plutôt $\mathcal{C}(I, \mathbb{R})$.

3.3 Lien entre le modèle de régression non-paramétrique et le bruit blanc

Dans le chapitre 4 on donnera une démonstration rigoureuse de l'équivalence asymptotique entre les modèles statistiques (3.3) et (3.5). Avant de faire cela, on voudrait ici montrer, d'une façon heuristique, que le modèle de bruit blanc est "proche" du modèle de régression. Juste pour faciliter l'intuition, modifions un peu le modèle (3.5), et supposons avoir observé un processus Y qui satisfait l'équation stochastique

$$dY_t = f(t)dt + \varepsilon dB_t, \quad t \in [0, 1]; \quad (3.6)$$

où $0 < \varepsilon < 1$, $f : [0, 1] \rightarrow \mathbb{R}$ et $B = (B_t : t \in [0, 1])$ est le mouvement brownien sur $[0, 1]$. Discrétisons (3.6) de la façon suivante :

1. intégrons le processus Y sur $[t, t + \Delta]$, $\Delta > 0$:

$$\frac{Y(t + \Delta) - Y(t)}{\Delta} = \frac{1}{\Delta} \int_t^{t+\Delta} f(s)ds + \frac{\varepsilon}{\Delta} (B(t + \Delta) - B(t)).$$

2. Définissons

$$y(t) := \frac{Y(t + \Delta) - Y(t)}{\Delta}, \quad \xi(t) := \frac{\varepsilon}{\Delta} (B(t + \Delta) - B(t)).$$

Pour tout $t \in [0, 1]$, par définition du mouvement brownien, la variable aléatoire $\xi(t)$ est gaussienne de moyenne zéro et variance donnée par :

$$\mathbb{E}(\xi^2(t)) = \frac{\varepsilon^2}{\Delta^2} \mathbb{E}[(B(t + \Delta) - B(t))^2] = \frac{\varepsilon^2}{\Delta}.$$

3. Posons $\varepsilon = 1/\sqrt{n}$ et $\Delta = 1/n$ de façon à ce que, pour tout t , $\xi(t) \sim \mathcal{N}(0, 1)$ et

$$y(t) = f(t) + \xi(t) + \left(\frac{1}{\Delta} \int_t^{t+\frac{1}{n}} f(s)ds - f(t) \right).$$

Pour n grand et f suffisamment régulière, la quantité $\frac{1}{\Delta} \int_t^{t+\frac{1}{n}} f(s)ds - f(t)$ est très petite, donc $y(t) \approx f(t) + \xi(t)$.

4. En particulier, pour $Y_i = y(i/n)$ et $\xi_i = \xi(i/n)$, nous obtenons

$$Y_i \approx f(i/n) + \xi_i,$$

qui est le modèle de régression non paramétrique à design déterministe et avec les bruits i.i.d. de loi normale centrée réduite.

De plus, remarquons qu'à partir de l'observation d'un processus Y dans le modèle de bruit blanc (donc qui satisfait l'équation (3.6)) le statisticien a accès à une suite d'observations gaussiennes de la forme :

$$y_j = \theta_j + \varepsilon \xi_j, \quad j = 1, 2, \dots, \quad (3.7)$$

où les variables aléatoires ξ_j , $j = 1, 2, \dots$, sont i.i.d. de loi normale réduite centrée. En effet, si on considère une base orthonormale dans $L_2[0, 1]$, $\{\phi_j\}_{j=1}^\infty$, alors, en utilisant (3.6) on aboutit à :

$$\int_0^1 \phi_j(t) dY_t = \int_0^1 f(t) \phi_j(t) dt + \varepsilon \int_0^1 \phi_j(t) dB_t.$$

Posons

$$\theta_j := \int_0^1 f(t) \phi_j(t) dt, \quad \xi_j := \int_0^1 \phi_j(t) dB_t, \quad j = 1, 2, \dots$$

Alors, si on définit y_j par :

$$y_j := \int_0^1 \phi_j(t) dY_t, \quad j = 1, 2, \dots,$$

on aboutit au modèle (3.7).

En effet, comme les fonctions ϕ_j sont orthonormales dans $L_2[0, 1]$, les variables ξ_j sont i.i.d. de loi $\mathcal{N}(0, 1)$.

3.4 Exemples des espaces des paramètres

En conformité avec les notations qui précèdent, l'espace des paramètres \mathcal{F} des modèles (3.3) et (3.5) représente l'ensemble de tous les choix possibles permis pour la fonction inconnue f à estimer. Des exemples de tels espaces, qui nous intéressent par la suite, sont les suivants.

Exemple 3.4.1. La classe \mathcal{F} est donnée en termes d'une condition de Lipschitz :

Si $\alpha \geq 1$

$$\mathcal{F}_{\alpha, B}^{(1)} = \left\{ f : \left| f(x + \Delta) - f(x) - \sum_{i=1}^{[\alpha]-1} \frac{f^{(i)}(x)}{i!} \Delta^i \right| \leq B |\Delta|^\alpha \text{ et } \sup_{x \in I} |f(x)| \leq B \right\},$$

si $0 < \alpha < 1$,

$$\mathcal{F}_{\alpha, B}^{(1)} = \left\{ f : |f(x + \Delta) - f(x)| \leq B |\Delta|^\alpha \right\}$$

où f est une fonction $[\alpha] - 1$ -fois dérivable, $x, x + \Delta$ appartiennent à I .

Exemple 3.4.2. La classe \mathcal{F} est donnée en termes d'une condition de Sobolev :

$$\mathcal{F}_{\alpha, B}^{(2)} = \left\{ f : \int \left(f^{(\alpha)}(t) \right)^2 dt \leq B \text{ et } \sup_{x \in I} |f(x)| \leq B, \quad \alpha \in \mathbb{N} \right\}$$

où $f^{(\alpha)}$ indique la α -ième dérivée de f qu'on suppose exister, on suppose aussi que $f^{(\alpha-1)}$ est absolument continue.

Chapitre 4

Équivalence asymptotique entre le modèle de régression non paramétrique et le bruit blanc

Le but de ce chapitre est de montrer d'une façon rigoureuse l'équivalence asymptotique entre le modèle (3.3) de régression non paramétrique à variance connue et le modèle (3.5) de signal avec bruit blanc. L'intérêt d'une telle équivalence sera précisé dans le chapitre suivant. On étudie d'abord le cas des design déterministes et ensuite aléatoires dans le modèle de régression non-paramétrique.

4.1 Cas déterministe : hypothèses et discussion

Soit $I = [\alpha, \beta]$, $-\infty < \alpha < \beta < \infty$.

– On suppose que $\sigma^2(\cdot) > 0$ est une fonction absolument continue sur I pour laquelle il existe une constante $C_1 < \infty$ telle que

$$\left| \frac{d}{dt} \ln \sigma(t) \right| \leq C_1, \quad t \in I \quad (4.1)$$

– On suppose qu'il existe une constante $B < \infty$ telle que

$$\sup_{t \in I} \{|f(t)| : f \in \mathcal{F}\} = B \quad (4.2)$$

– Soit H une fonction absolument continue sur I tel que

$$H'(t) = h(t) > 0 \quad \text{p.p. sur } I \quad (4.3)$$

– Soient x_{ni} comme définis dans (3.2). Définissons une fonction en escalier de la façon suivante :

$$\bar{f}_n(t) = \begin{cases} f(x_{ni}) & \text{si } \xi_{i-1} \leq t < \xi_i, \quad i = 1, \dots, n; \\ f(x_{nn}) & \text{si } t = \beta; \end{cases} \quad (4.4)$$

où $\xi_i = H^{-1}(\frac{i}{n})$ (on a aboli la dépendance en n par soucis de simplification). On suppose que

$$\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}} n \int_{\alpha}^{\beta} (f(t) - \bar{f}_n(t))^2 h(t) dt = 0. \quad (4.5)$$

Remarque 4.1.1. La condition donnée par l'équation (4.5) est une condition sur la régularité de f . En particulier, elle est satisfaite dans le cas où f appartient à un des espaces de paramètres présentés dans les exemples 3.4.1 et 3.4.2 avec $\alpha > \frac{1}{2}$.

En effet, vérifions que si on choisit $\mathcal{F} = \mathcal{F}_{\alpha,B}^{(1)}$ (défini dans l'exemple 3.4.1), avec $\alpha > \frac{1}{2}$, alors la condition donnée par l'équation (4.5) est satisfaite :

Par soucis de simplicité, supposons que $I = [0,1]$, que H est la fonction de répartition d'une loi uniforme sur $[0,1]$ et que $\sigma(\cdot) = 1$. Soit $\frac{i-1}{n} \leq t < \frac{i}{n}$. Par définition de $\bar{f}_n(t)$ et de $\mathcal{F}_{\alpha,B}^{(1)}$,

$$|f(t) - \bar{f}_n(t)| = |f(t) - f(i/n)| \leq B|t - 1/n|^\alpha.$$

Donc

$$\begin{aligned} n \int_0^1 (f(t) - \bar{f}_n(t))^2 dt &\leq nB^2 \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left| t - \frac{i}{n} \right|^{2\alpha} dt \\ &\leq n^2 B^2 \int_0^{\frac{1}{n}} t^{2\alpha} dt \\ &= \frac{n^2 B^2}{2\alpha + 1} \left(\frac{1}{n} \right)^{2\alpha+1} \end{aligned}$$

et cette dernière quantité converge vers zéro quand n tend à l'infini, vu que, par hypothèse, $2\alpha + 1 > 2$. Donc la condition (4.5) est bien satisfaite.

4.1.1 Équivalence asymptotique

Théorème 4.1.2. *Sous les hypothèses (4.1)-(4.5), le modèle de régression non paramétrique à design déterministes \mathcal{P}_n^Y introduit dans le chapitre 3 (modèle (3.1)) et défini par l'équation :*

$$\mathcal{P}_n^Y : Y_{ni} = f(x_{ni}) + \sigma(x_{ni})\epsilon_{ni} \quad (4.6)$$

est asymptotiquement équivalent au modèle de signal avec bruit blanc \mathcal{P}_n^Z introduit dans le chapitre 3 (modèle (3.5)) et défini par l'équation :

$$\mathcal{P}_n^Z : dZ_t^{(n)} = f(t)dt + \frac{\lambda(t)}{\sqrt{n}}dB_t \quad (4.7)$$

avec $\lambda^2(t) = \frac{\sigma^2(t)}{h(t)}$.

Démonstration. Soit $\bar{Z}^{(n)} = (\bar{Z}_t^{(n)} : t \in I)$ un processus gaussien qui satisfait l'équation stochastique :

$$d\bar{Z}_t^{(n)} = \bar{f}_n(t)dt + \frac{\lambda(t)}{\sqrt{n}}dB_t. \quad (4.8)$$

Exactement comme on a déjà fait pour la construction du modèle (3.5), et avec les mêmes notations, définissons le modèle statistique qui lui est associé de la manière suivante :

$$\mathcal{P}_n^{\bar{Z}} = (\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I, (P_{n,f}^{\bar{Z}})_{f \in \mathcal{F}}),$$

où $P_{n,f}^{\bar{Z}}$ est la distribution du processus $\bar{Z}^{(n)}$.

La démonstration se déroule en quatre étapes.

– **ÉTAPE 1** : Montrer que $\lim_{n \rightarrow \infty} \Delta(\mathcal{P}_n^Z, \mathcal{P}_n^{\bar{Z}}) = 0$.

Grâce au théorème 1.4.2, vu que les espaces des observations sont les mêmes pour les deux modèles, il suffit de montrer que

$$\lim_{n \rightarrow \infty} L_1(Z^{(n)}, \bar{Z}^{(n)}) = 0.$$

Or, grâce à l'hypothèse (4.2) et au fait que la fonction $\sigma^2(\cdot)$ est absolument continue sur I et toujours positive, l'hypothèse $\int_I \frac{f(s)^2}{\sigma(s)^2} ds < \infty$ est satisfaite. On peut alors appliquer la proposition 2.1.4 pour aboutir à

$$L_1(Z^{(n)}, \bar{Z}^{(n)}) = O(D)$$

$$\text{avec } D^2 = \int_{\alpha}^{\beta} \frac{n(f(t) - \bar{f}_n(t))^2}{\lambda^2(t)} dt = \int_{\alpha}^{\beta} \frac{n(f(t) - \bar{f}_n(t))^2 h(t)}{\sigma^2(t)} dt.$$

Donc $D^2 \leq \max_{t \in [\alpha, \beta]} \frac{1}{\sigma^2(t)} \int_{\alpha}^{\beta} n(f(t) - \bar{f}_n(t))^2 h(t) dt$ qui tend vers zéro quand n tend vers l'infini grâce à l'hypothèse (4.5), d'où la convergence vers zéro en distance L_1 des processus $Z^{(n)}, \bar{Z}^{(n)}$. D'où l'ÉTAPE 1 est démontrée.

Avant de démontrer l'ÉTAPE 2 introduisons une suite de variables aléatoires définies de la façon suivante

$$K_i^2 = \frac{n\sigma^2(x_{ni})}{\int_{\xi_{i-1}}^{\xi_i} \frac{dt}{\lambda^2(t)}} \quad \text{et} \quad S_i^{(n)} = K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\bar{Z}_t^{(n)}}{\lambda^2(t)}, \quad (4.9)$$

et notons par \mathcal{P}_n^S le modèle statistique associé aux $(S_i^{(n)})_{1 \leq i \leq n}$ défini comme :

$$\mathcal{P}_n^S = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_{n,f}^S)_{f \in \mathcal{F}}), \quad (4.10)$$

avec $P_{n,f}^S$ la loi du vecteur $(S_1^{(n)}, \dots, S_n^{(n)})$.

– **ÉTAPE 2** : Montrer que $\Delta(\mathcal{P}_n^{\bar{Z}}, \mathcal{P}_n^S) = 0$.

Cela découle immédiatement du lemme 1.1.8, car, pour tout n , $\{S_i^{(n)} : i = 1, \dots, n\}$ est une statistique exhaustive pour $\bar{Z}^{(n)}$.

Pour voir qu'il s'agit bien d'une statistique exhaustive, il suffit de considérer la statistique $S : (\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ ainsi définie :

$$\forall \omega \in \mathcal{C}(I, \mathbb{R}) \quad S(\omega) = (S_1(\omega), \dots, S_n(\omega)),$$

avec $S_i(\omega) = K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\omega_t}{\lambda^2(t)}$. Or, grâce à la proposition 2.1.1,

$$\begin{aligned} \frac{dP_{n,f}^{\bar{Z}}}{d\mathbb{W}_{\lambda/\sqrt{n}}}(\omega) &= \exp\left(n \int_I \frac{\bar{f}(s)}{\lambda^2(s)} d\omega_s - \frac{n}{2} \int_I \frac{\bar{f}^2(s)}{\lambda^2(s)} ds\right) \\ &= \exp\left(n \sum_{i=1}^n f(x_{ni}) \int_{\xi_{i-1}}^{\xi_i} \frac{d\omega_s}{\lambda^2(s)} - \frac{n}{2} \sum_{i=1}^n f^2(x_{ni}) \int_{\xi_{i-1}}^{\xi_i} \frac{ds}{\lambda^2(s)}\right) \\ &= \exp\left(n \sum_{i=1}^n f(x_{ni}) \frac{S_i(\omega)}{K_i} - \frac{n}{2} \sum_{i=1}^n f^2(x_{ni}) \int_{\xi_{i-1}}^{\xi_i} \frac{ds}{\lambda^2(s)}\right), \end{aligned}$$

ce qui conclut l'ÉTAPE 2 en utilisant le théorème de factorisation 1.1.2.

– **ÉTAPE 3** : Montrer que $\lim_{n \rightarrow \infty} \Delta(\mathcal{P}_n^S, \mathcal{P}_n^Y) = 0$.

L'idée pour démontrer cela est de remarquer que les variables aléatoires $S_i^{(n)}$, $i = 1 \dots n$, sont indépendantes et donc d'utiliser la propriété (1.4.6) pour montrer que $\lim_{n \rightarrow \infty} H(P_{n,f}^S, P_{n,f}^Y) = 0$ et conclure grâce à la propriété (1.12). Voyons tout cela en détails.

Montrons d'abord que les variables $S_i^{(n)}$, $i = 1 \dots n$, sont indépendantes. Pour cela il est utile de remarquer que, grâce aux propriétés de l'intégrale de Wiener (voir Annexe A), que $(S_1^{(n)}, \dots, S_n^{(n)})$ est un vecteur gaussien. Pour montrer l'indépendance il suffit alors de montrer que, pour tout $i \neq j$, $\mathbb{E}[S_i^{(n)} S_j^{(n)}] = \mathbb{E}[S_i^{(n)}] \mathbb{E}[S_j^{(n)}]$.

Notons que, grâce au théorème A.0.3,

$$\mathbb{E}\left[\int_{\xi_{i-1}}^{\xi_i} \frac{dB_t}{\sqrt{n\lambda(t)}}\right] = 0 = \mathbb{E}\left[\int_{\xi_{j-1}}^{\xi_j} \frac{dB_t}{\sqrt{n\lambda(t)}}\right]$$

et que, grâce à la propriété donnée par l'équation (A.4),

$$\mathbb{E}\left[\int_{\xi_{i-1}}^{\xi_i} \frac{dB_t}{\sqrt{n\lambda(t)}} \int_{\xi_{j-1}}^{\xi_j} \frac{dB_t}{\sqrt{n\lambda(t)}}\right] = \int_I \mathbb{I}_{[\xi_{i-1}, \xi_i] \cap [\xi_{j-1}, \xi_j]}(t) \frac{dt}{n\lambda^2(t)} = 0,$$

car pour tous $i \neq j$, $\mathbb{I}_{[\xi_{i-1}, \xi_i] \cap [\xi_{j-1}, \xi_j]} \equiv 0$.

Donc, d'un côté,

$$\begin{aligned} \mathbb{E}[S_i^{(n)} S_j^{(n)}] &= \mathbb{E}\left[\left(K_i \int_{\xi_{i-1}}^{\xi_i} \frac{f(x_{ni})}{\lambda^2(t)} dt + K_i \int_{\xi_{i-1}}^{\xi_i} \frac{dB_t}{\sqrt{n\lambda(t)}}\right) \left(K_j \int_{\xi_{j-1}}^{\xi_j} \frac{f(x_{nj})}{\lambda^2(t)} dt + K_j \int_{\xi_{j-1}}^{\xi_j} \frac{dB_t}{\sqrt{n\lambda(t)}}\right)\right] \\ &= K_i \int_{\xi_{i-1}}^{\xi_i} \frac{f(x_{ni})}{\lambda^2(t)} dt \cdot K_j \int_{\xi_{j-1}}^{\xi_j} \frac{f(x_{nj})}{\lambda^2(t)} dt, \end{aligned}$$

et de l'autre côté, toujours grâce au théorème A.0.3 on a :

$$\mathbb{E}[S_i^{(n)}] \mathbb{E}[S_j^{(n)}] = K_i \int_{\xi_{i-1}}^{\xi_i} \frac{f(x_{ni})}{\lambda^2(t)} dt \cdot K_j \int_{\xi_{j-1}}^{\xi_j} \frac{f(x_{nj})}{\lambda^2(t)} dt,$$

d'où,

$$\mathbb{E}[S_i^{(n)} S_j^{(n)}] = \mathbb{E}[S_i^{(n)}] \mathbb{E}[S_j^{(n)}] \quad \forall i \neq j.$$

On a donc montré que les variables $(S_i^{(n)})_{1 \leq i \leq n}$ sont indépendantes et de loi $\mathcal{N}(\mathbb{E}[S_i^{(n)}], V[S_i^{(n)}])$ avec

$$V[S_i^{(n)}] = \frac{K_i^2}{n} \int_{\xi_{i-1}}^{\xi_i} \frac{dt}{\lambda^2(t)} = \sigma^2(x_{ni})$$

$$\begin{aligned} \mathbb{E}[S_i^{(n)}] &= K_i \int_{\xi_{i-1}}^{\xi_i} \frac{\bar{f}_n(t)}{\lambda^2(t)} dt \\ &= \sqrt{n} \sigma(x_{ni}) f(x_{ni}) \left[\int_{\xi_{i-1}}^{\xi_i} \frac{h(t)}{\sigma^2(t)} dt \right]^{\frac{1}{2}} \\ &= f(x_{ni}) \left[\frac{\sigma(x_{ni})}{\sigma(\xi_{ni})} \right], \end{aligned}$$

où $\xi_{i-1} \leq \bar{\xi}_{ni} \leq \xi_i$. Notons que l'existence de $\bar{\xi}_{ni}$ est garantie par l'argument suivant : par définition des $(\xi_i)_i$ on a $\int_{\xi_{i-1}}^{\xi_i} h(t) dt = \frac{1}{n}$,

$$\int_{\xi_{i-1}}^{\xi_i} \frac{h(t)}{\sigma^2(t)} dt - \frac{1}{\sigma^2(s)n} = \int_{\xi_{i-1}}^{\xi_i} h(t) \left[\frac{1}{\sigma^2(t)} - \frac{1}{\sigma^2(s)} \right] dt$$

et

$$\begin{aligned} \int_{\xi_{i-1}}^{\xi_i} h(t) \min_{s \in [\xi_{i-1}, \xi_i]} \left[\frac{1}{\sigma^2(t)} - \frac{1}{\sigma^2(s)} \right] dt &\leq \int_{\xi_{i-1}}^{\xi_i} h(t) \left[\frac{1}{\sigma^2(t)} - \frac{1}{\sigma^2(s)} \right] dt \\ &\leq \int_{\xi_{i-1}}^{\xi_i} h(t) \max_{s \in [\xi_{i-1}, \xi_i]} \left[\frac{1}{\sigma^2(t)} - \frac{1}{\sigma^2(s)} \right] dt. \end{aligned}$$

Or, comme la première quantité est négative et la dernière est positive, il doit exister un point $\bar{\xi}_{ni}$ dans $[\xi_{i-1}, \xi_i]$, qui annule la fonction continue $\int_{\xi_{i-1}}^{\xi_i} h(t) \left[\frac{1}{\sigma^2(t)} - \frac{1}{\sigma^2(s)} \right] dt$, c'est-à-dire, $\int_{\xi_{i-1}}^{\xi_i} \frac{h(t)}{\sigma^2(t)} dt = \frac{1}{\sigma^2(\bar{\xi}_{ni})n}$.

Remarquons aussi, que grâce à l'hypothèse (4.1) on a

$$\mathbb{E}\left(S_i^{(n)}\right) = f(x_{ni}) \frac{\sigma(x_{ni})}{\sigma(\bar{\xi}_{ni})},$$

ceci nous amène à la relation

$$\mathbb{E}\left(S_i^{(n)}\right) = f(x_{ni})(1 + O(\xi_i - \xi_{i-1})),$$

uniformément en $f \in \mathcal{F}$, i et n . En effet, en faisant un développement limité au deuxième ordre de $\frac{\sigma(x_{ni})}{\sigma(\bar{\xi}_{ni})}$, on obtient

$$\frac{\sigma(x_{ni})}{\sigma(\bar{\xi}_{ni})} = 1 + \frac{\dot{\sigma}(\bar{\xi}_{ni})}{\sigma(\bar{\xi}_{ni})}(x_{ni} - \bar{\xi}_{ni}) + O(|x_{ni} - \bar{\xi}_{ni}|^2),$$

et, en utilisant l'hypothèse (4.1), on aboutit à

$$\begin{aligned} \frac{\sigma(x_{ni})}{\sigma(\bar{\xi}_{ni})} &\leq 1 + C_1 |x_{ni} - \bar{\xi}_{ni}| + O(|x_{ni} - \bar{\xi}_{ni}|^2) \\ &= 1 + O(\xi_i - \xi_{i-1}). \end{aligned}$$

Montrons finalement que $\lim_{n \rightarrow \infty} H^2(P_{n,f}^S, P_{n,f}^Y) = 0$.

Pour cela on peut utiliser les propriétés 1.4.6 et 1.4.5 vu que $(x_{ni}, Y_{ni})_{1 \leq i \leq n}$ est une suite de variables aléatoires indépendantes de loi $\mathcal{N}(f(x_{ni}), \sigma^2(x_{ni}))$ et $(S_i^{(n)})_{1 \leq i \leq n}$ est une suite de variables aléatoires indépendantes de loi $\mathcal{N}(f(x_{ni})(1 + O(\xi_i - \xi_{i-1})), \sigma^2(x_{ni}))$. Si on note par $\mathcal{L}(S_i^{(n)})$ et $\mathcal{L}(Y_{ni})$ les lois, respectivement, de la variable aléatoire $S_i^{(n)}$ et celle de Y_{ni} , on trouve

$$\begin{aligned} H^2(P_{n,f}^S, P_{n,f}^Y) &= 2 \left[1 - \prod_{i=1}^n \left[1 - \frac{H^2(\mathcal{L}(S_i^{(n)}), \mathcal{L}(Y_{ni}))}{2} \right] \right] \\ &= 2 \left[1 - \prod_{i=1}^n \left[1 - \left[1 - \exp\left(f^2(x_{ni})O(\xi_i - \xi_{i-1})^2\right) \right] \right] \right] \\ &= O\left(\sum_{i=1}^n f^2(x_{ni})(\xi_i - \xi_{i-1})^2\right). \end{aligned}$$

Pour contrôler la quantité $H^2(\mathcal{L}(S_i^{(n)}), \mathcal{L}(Y_{ni}))$ on utilise la propriété 1.4.5,

$$\begin{aligned} H^2(\mathcal{L}(S_i^{(n)}), \mathcal{L}(Y_{ni})) &= 2 \left[1 - \left(\frac{2\sigma^2(x_{ni})}{2\sigma^2(x_{ni})} \right)^{\frac{1}{2}} \exp \left(- \frac{(f(x_{ni})(1 + O(\xi_i - \xi_{i-1})) - f(x_{ni}))^2}{8\sigma^2(x_{ni})} \right) \right] \\ &= 2 \left[1 - \exp(f^2(x_{ni})O(\xi_i - \xi_{i-1})^2) \right], \end{aligned}$$

et, grâce aux hypothèses (4.2), (4.3)

$$\begin{aligned} O \left(\sum_{i=1}^n f^2(x_{ni})(\xi_i - \xi_{i-1})^2 \right) &\leq B^2 \sum_{i=1}^n (\xi_i - \xi_{i-1}) O \left(\sup_{1 \leq i \leq n} (\xi_i - \xi_{i-1}) \right) \\ &= B^2(\beta - \alpha) O \left(\sup_{1 \leq i \leq n} \left(H^{-1} \left(\frac{i}{n} \right) - H^{-1} \left(\frac{i-1}{n} \right) \right) \right) \\ &\leq B^2(\beta - \alpha) O \left(\frac{1}{n} \frac{1}{\min_{t \in [\alpha, \beta]} H'(t)} \right) \\ &= o(1) \quad \text{uniformément pour } f \in \mathcal{F}, \end{aligned}$$

ce qui montre, en utilisant la propriété 1.12 et le théorème 1.4.2, que $\lim_{n \rightarrow \infty} \Delta(\mathcal{P}_n^S, \mathcal{P}_n^Y) = 0$

– **ÉTAPE 4** : Conclusion.

Il faut montrer que $\lim_{n \rightarrow \infty} \Delta(\mathcal{P}_n^Z, \mathcal{P}_n^Y) = 0$. Cela découle immédiatement des étapes précédentes et de l'inégalité triangulaire satisfaite par Δ (voir proposition 1.4.1). □

Remarque 4.1.3. Les hypothèses (4.1)-(4.2) peuvent être améliorées. On les a posées pour pouvoir montrer que $H^2(P_{n,f}^Y, P_{n,f}^S) = O(1)$. Cependant, si par exemple on travaille avec une fonction σ constante, les v.a. $S_i^{(n)}$ ont une moyenne donnée par $f(x_{ni})$ et donc, dans ce cas, la distance d'Hellinger $H^2(P_{n,f}^Y, P_{n,f}^S)$ est exactement zéro, sans avoir besoin des hypothèses (4.1)-(4.2). L'hypothèse (4.5) a été introduite pour pouvoir définir les points $\xi_i, i = 1, \dots, n$ et l'hypothèse (4.5) sert à conclure que la distance $L_1(Z^{(n)}, \bar{Z}^{(n)})$ converge vers 0 (voir la proposition 2.1.4).

4.2 Cas de design aléatoires

Des résultats analogues à ceux de la section précédente peuvent être obtenus même dans le cas où les design du modèle de régression non-paramétrique sont aléatoires. On rappelle que le modèle de régression considéré est le suivant :

$$Y_{ni} = f(X_{ni}) + \sigma(X_{ni})\varepsilon_{ni}, \quad i = 1, \dots, n, \quad (4.11)$$

où :

- les X_{n1}, \dots, X_{nn} sont indépendantes et de fonction de répartition commune H , avec $H : \mathbb{R} \rightarrow [0, 1]$ fonction de répartition strictement croissante.
- $(\varepsilon_{ni})_{1 \leq i \leq n}$ sont des v.a. i.i.d. de loi normale centrée réduite.
- σ est une fonction connue.
- f est une fonction inconnue appartenant à une certaine classe fonctionnelle \mathcal{F} .

Le modèle statistique associé à (4.11) est

$$\mathcal{P}_n^{Y,X} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left(P_{n,f}^Y \right)_{f \in \mathcal{F}} \right). \quad (4.12)$$

On verra que le modèle de régression non-paramétrique (4.12), qu'on a noté par $\mathcal{P}^{1,n,X}$ (pour souligner la dépendance par rapport à la partie aléatoire) est asymptotiquement équivalent à un modèle de signal avec bruit blanc dont le drift dépend des observations aléatoires (X_{ni}) ainsi que du paramètre inconnu f .

Pour pouvoir décrire un tel modèle de bruit blanc, qu'on supposera défini sur l'intervalle $I = [\alpha, \beta]$ avec $-\infty < \alpha < \beta < \infty$, on aura besoin d'introduire certaines notations.

Notons par $x_{(1)}, \dots, x_{(n)}$ les valeurs ordonnées par ordre croissant des observations X_{n1}, \dots, X_{nn} et posons $x_{(0)} = \alpha$ et $x_{(n+1)} = \beta$. Pour tous $x_{(i-1)} \leq t \leq x_{(i)}$, $i = 1, \dots, n+1$, posons

$$\tilde{H}_n(t) = \frac{i-1}{n+1} + \frac{t - x_{(i-1)}}{(n+1)(x_{(i)} - x_{(i-1)})}. \quad (4.13)$$

Définissons $\tilde{\xi}_i = \tilde{H}_n^{-1}(i/n)$ et définissons \tilde{f}_n par l'équation

$$\tilde{f}_n(t) = \begin{cases} f(x_{ni}) & \text{si } \tilde{\xi}_{i-1} \leq t < \tilde{\xi}_i, \quad i = 1, \dots, n; \\ f(x_{nn}) & \text{si } t = \beta; \end{cases}$$

Notons avec $\tilde{h}_n(t)$ la dérivée à gauche de \tilde{H}_n au temps t et posons $\tilde{\lambda}_n^2(t) = \frac{\sigma^2(t)}{\tilde{h}_n(t)}$.

Définissons finalement une suite de processus gaussiens qui satisfont l'équation stochastique :

$$dZ_t^{(n,X)} = \tilde{f}_n(t)dt + \frac{\tilde{\lambda}_n(t)}{\sqrt{n}} dB_t \quad (4.14)$$

où on a indexé par X pour souligner le fait que la distribution du processus $Z^{(n,X)} = (Z_t^{(n,X)}, t \in I)$ dépend des observations aléatoires $(X_{ni})_i$ via les fonctions $\tilde{\lambda}_n$ et \tilde{f}_n .

Cela nous permet d'introduire un nouveau modèle statistique :

$$\mathcal{P}_n^{(Z,X)} = \left(\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I, \left(P_{n,f}^Z \right)_{f \in \mathcal{F}} \right) \quad (4.15)$$

avec $\mathcal{C}(I, \mathbb{R})$ l'espace des fonctions continues de I dans \mathbb{R} , \mathcal{C}^I la sigma-algèbre de Borel de $\mathcal{C}(I, \mathbb{R})$ et $P_{n,f}^{(2)}$ la distribution du processus $Z^{(n,X)}$ qu'on vient de définir.

4.2.1 Hypothèses et discussions

Soit $I = [\alpha, \beta]$, $-\infty < \alpha < \beta < \infty$.

– On suppose que $\sigma^2(\cdot) > 0$ est une fonction absolument continue sur I pour laquelle il existe une constante $C_1 < \infty$ telle que

$$\left| \frac{d}{dt} \ln \sigma(t) \right| \leq C_1, \quad t \in I \quad (4.16)$$

– On suppose qu'il existe une constante $B < \infty$ telle que

$$\sup\{|f(t)| : t \in I, f \in \mathcal{F}\} = B \quad (4.17)$$

– On suppose que \tilde{h}_n soit une dérivée à gauche de \tilde{H}_n définie par l'équation (4.13).

– Finalement on suppose que

$$\lim_{n \rightarrow \infty} \text{Prob} \left\{ \sup_{f \in \mathcal{F}} n \int_{\alpha}^{\beta} (f(t) - \tilde{f}_n(t))^2 \left(\frac{\tilde{h}_n(t)}{\sigma^2(t)} \right) dt \right\} = 0. \quad (4.18)$$

Remarque 4.2.1. L'hypothèse (4.18) n'est pas, dans la pratique, particulièrement plus difficile à vérifier que l'hypothèse (4.5). Par exemple, si on pose $I = [0, 1]$, H fonction de répartition d'une loi uniforme sur $[0, 1]$, $\sigma(t) \equiv 1$ et on considère l'espace fonctionnel défini dans l'exemple 3.4.1 avec $\alpha > \frac{1}{2}$, alors la condition (4.18) est satisfaite. En effet :

$$\int (f(t) - \tilde{f}_n(t))^2 \tilde{h}_n(t) dt = \int \left(f(\tilde{H}_n^{-1}(v)) - \tilde{f}_n(\tilde{H}_n^{-1}(v)) \right)^2 dv$$

et on a

$$\mathbb{E} \left\{ n \left(f(\tilde{H}_n^{-1}(v)) - \tilde{f}_n(\tilde{H}_n^{-1}(v)) \right) \right\} \leq n B^2 \mathbb{E} [x_{(j)} - x_{(j-1)}]^{2\alpha} \rightarrow 0$$

uniformément en v , n où $j = \min\{i : x_{(i)} > \tilde{H}_n^{-1}(v)\}$.

4.2.2 Équivalence asymptotique

Théorème 4.2.2. *Sous les hypothèses (4.16)-(4.18), le modèle de régression non paramétrique à design aléatoires $\mathcal{P}_n^{(Y,X)}$ précédemment introduit (modèle (4.12)) et défini par l'équation :*

$$\mathcal{P}_n^{(Y,X)} : Y_{ni} = f(X_{ni}) + \sigma(X_{ni})\varepsilon_{ni}, \quad i = 1 \dots, n \quad (4.19)$$

est asymptotiquement équivalent au modèle de signal avec bruit blanc (modèle (4.15)) défini par l'équation :

$$\mathcal{P}_n^{(Z,X)} : dZ_t^{n,X} = \tilde{f}(t)dt + \frac{\tilde{\lambda}(t)}{\sqrt{n}} dB_t, \quad (4.20)$$

avec la relation $\tilde{\lambda}_n^2(t) = \frac{\sigma^2(t)}{\tilde{h}_n(t)}$.

Démonstration. La preuve se déroule en quatre étapes et elle suit exactement le même schéma que la preuve du théorème 4.1.2. On commence donc par définir une suite de processus gaussiens :

$$d\tilde{Z}_t^{(n,X)} = \tilde{f}_n(t)dt + \frac{\tilde{\lambda}_n(t)}{\sqrt{n}} dB_t$$

et on note par $\mathcal{P}_n^{(\tilde{Z},X)}$ la suite de modèles statistiques qui lui est associée :

$$\mathcal{P}_n^{(\tilde{Z},X)} = \left(\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I; P_{n,f}^{(\tilde{Z},X)} : f \in \mathcal{F} \right),$$

où $P_{n,f}^{(\tilde{Z},X)}$ est la loi du processus $\tilde{Z}^{(n,X)} = (\tilde{Z}_t^{(n)} : t \in I)$.

– **ÉTAPE 1** : Grâce à l'hypothèse (4.18), $L_1(\mathcal{P}_n^{Z,X}, \mathcal{P}_n^{\tilde{Z},X})$ converge vers 0 en probabilité et donc le Δ -écart entre les deux suites de modèles stochastiques, grâce au théorème 1.4.2, converge vers zéro en probabilité.

– **ÉTAPE 2** : Comme dans le cas déterministe on définit :

$$K_i^2 = \frac{n\sigma^2(X_{ni})}{\int_{\tilde{\xi}_{i-1}}^{\tilde{\xi}_i} dt / (\tilde{\lambda}_n^2(t))} \quad \text{et} \quad S_i^{(n,X)} = K_i \int_{\tilde{\xi}_{i-1}}^{\tilde{\xi}_i} \frac{d\tilde{Z}_t^{(n)}}{\tilde{\lambda}_n^2(t)}. \quad (4.21)$$

et on note par $\mathcal{P}_n^{(S,X)}$ le modèle statistique associé aux $(S_i^{(n,X)})_{1 \leq i \leq n}$:

$$\mathcal{P}_n^{(S,X)} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_{n,f}^{(S,X)})_{f \in \mathcal{F}} \right),$$

où $P_{n,f}^{(S,X)}$ est la loi du vecteur $(S_1^{(n)}, \dots, S_n^{(n)})$. Grâce au lemme 1.1.8 $\Delta(\mathcal{P}_n^{(\tilde{Z},X)}, \mathcal{P}_n^{(S,X)})=0$.

– **ÉTAPE 3** : Avec les mêmes arguments que ceux utilisés dans l'ÉTAPE 3 de la preuve du théorème 4.1.2 on montre que les variables $S_i^{(n,X)}$, $i = 1, \dots, n$, sont indépendantes de loi $\mathcal{N}\left(\mathbb{E}\left[S_i^{(n,X)}\right], V\left[S_i^{(n,X)}\right]\right)$ avec

$$V\left[S_i^{(n,X)}\right] = \mathbb{E}[\sigma^2(X_{ni})]$$

$$\mathbb{E}\left[S_i^{(n,X)}\right] = \mathbb{E}\left[f(X_{ni})(1 + O(\tilde{\xi}_i - \tilde{\xi}_{i-1}))\right]$$

où $\xi_{i-1} \leq \bar{\xi}_{ni} \leq \xi$. De la même façon on trouve :

$$H^2(P_n^{(S,X)}, P_n^{(Y,X)}) = O\left(\mathbb{E}\left[\sum_{i=1}^n f^2(X_{n,i})(\tilde{\xi}_i - \tilde{\xi}_{i-1})^2\right]\right)$$

qui converge vers zéro lorsque n qui tend vers l'infini. Cela montre que $\lim_{n \rightarrow \infty} \Delta(\mathcal{P}_n^{S,X}, \mathcal{P}_n^{Y,X})=0$.

– **ÉTAPE 4** On conclut en utilisant les étapes précédentes et le fait que le Δ -écart définit une pseudo-métrique. □

Chapitre 5

Applications et généralisations

Étant donnée une expérience, on peut prendre une décision dont on mesure le risque. On peut définir qu'une expérience est meilleure ou moins bonne qu'une autre du point de vue du risque. Prenons l'exemple d'une classe dans laquelle les élèves seraient identifiés par leurs noms et prénoms. On aimerait pouvoir les identifier par leurs seules initiales, ce qui représente une véritable économie en termes de stockage de données. Cependant cela conduit à une perte d'information trop importante, puisque plusieurs élèves peuvent avoir les mêmes initiales. Cette expérience sera jugée moins bonne (plus restreinte) du point de vue du risque. De manière générale, une sous-expérience d'une expérience donnée est moins bonne du point de vue du risque. Mais elle peut aussi être équivalente. C'est le cas de la proposition 1.1.4. L'expérience \mathcal{G} est équivalente à l'expérience \mathcal{E} si à partir de \mathcal{G} on peut reconstruire une expérience, qui n'est pas \mathcal{E} , mais qui aura le même comportement que \mathcal{E} du point de vue des décisions et de leurs risques (une formulation mathématique est donnée par la proposition 1.3.1). On peut parler d'exhaustivité lorsque l'équivalence est vérifiée, qu'il n'y a aucune perte d'information (mathématiquement $\Delta(\mathcal{G}, \mathcal{E}) = 0$). Le Cam introduit aussi l'idée d'exhaustivité approchée, d'équivalence asymptotique, dont on peut avoir l'intuition en regardant une photo et sa version compressée : l'image compressée ne représente que 10% de l'originale en termes de place, de quantité d'information. Cependant, cette information est "suffisante" puisque la photo compressée est tout à fait lisible et exploitable.

Le résultat de Le Cam sur l'équivalence entre expériences, ou modèles statistiques, a pour conséquences directes ou indirectes le codage, la compression de données, la théorie de l'information, l'estimation fonctionnelle, la notion de parcimonie (à savoir que si on s'y prend bien, beaucoup de phénomènes peuvent s'exprimer avec peu de paramètres), le bootstrap (rééchantillonnage). Les statistiques en médecine, en traitement du signal et de l'image, les logiciels de notation musicale en sont aussi des champs d'application parmi d'autres.

5.1 Applications issues de l'équivalence asymptotique entre le modèle de régression et celui du bruit blanc

Dans les paragraphes qui suivent on va focaliser notre attention sur des applications issues de l'équivalence asymptotique entre le modèle de régression non-paramétrique et le modèle de signal avec bruit blanc démontrée dans le chapitre 4.

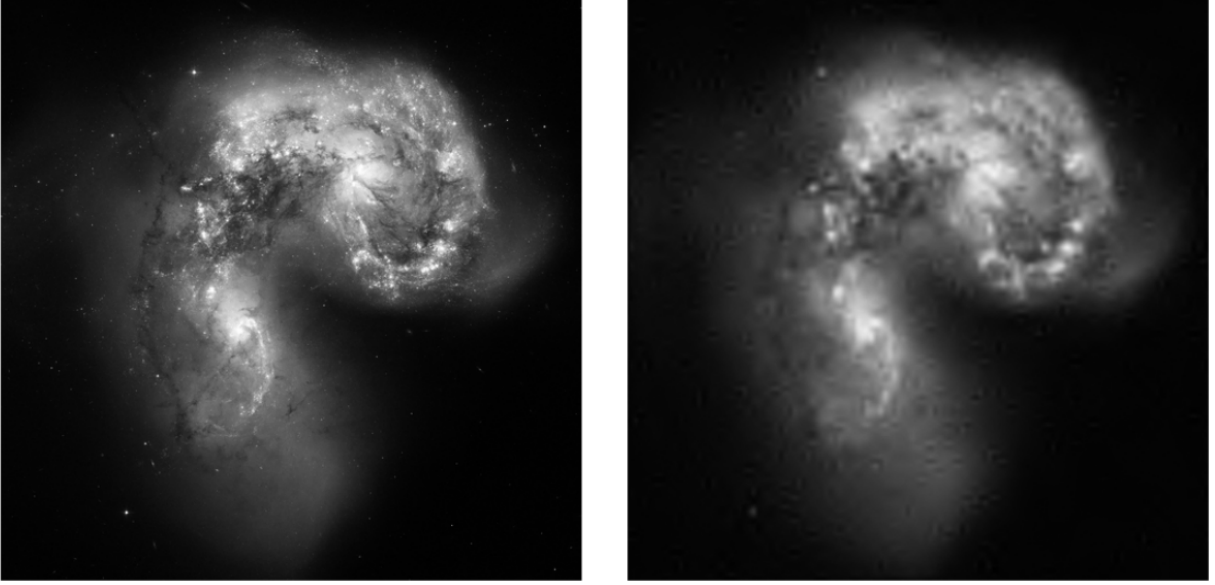


FIGURE 5.1 – Galaxie originale et galaxie compressée

Pour cela remarquons encore une fois que, grâce à la proposition 1.3.1, si $\mathcal{E}^{(n)}$, $\mathcal{G}^{(n)}$ sont deux modèles statistiques asymptotiquement équivalents au sens de la définition 1.8, et dans la suite de modèles $\mathcal{E}^{(n)}$ il existe une suite de règles de décision $\pi_1^{(n)}$ telle que les risques associés $R(\mathcal{E}^{(n)}, \pi_1^{(n)}, L_n, \theta)$ convergent vers la quantité $\rho(\theta)$, pour toute suite de fonctions de perte L_n uniformément bornées, alors il existe une suite de règles de décision $\pi_2^{(n)}$ dans le modèle $\mathcal{G}^{(n)}$ telle que les risques $R(\mathcal{G}^{(n)}, \pi_2^{(n)}, L_n, \theta)$ convergent vers la même quantité $\rho(\theta)$, uniformément pour θ dans Θ .

Définition 5.1. On dira que la suite de règles de décision $\pi_1^{(n)}$ définie dans le modèle $\mathcal{E}^{(n)}$ est *asymptotiquement équivalente* à la suite de règles de décision $\pi_2^{(n)}$ définie dans le modèle $\mathcal{G}^{(n)}$ si, pour toute fonction de perte L_n uniformément bornée, on a :

$$\lim_{n \rightarrow \infty} \left| R(\mathcal{E}^{(n)}, \pi_1^{(n)}, L_n, \theta) - R(\mathcal{G}^{(n)}, \pi_2^{(n)}, L_n, \theta) \right| = 0.$$

uniformément dans l'espace des paramètres Θ .

Voyons maintenant comment traduire l'équivalence asymptotique entre le modèle de régression et le modèle de bruit blanc démontrée dans le chapitre 4 (voir le théorème 4.1.2) en termes d'équivalence asymptotique entre suites de règles de décision.

On rappelle ici, brièvement, le contexte.

D'un côté on a le modèle de régression :

$$Y_{ni} = f(x_{ni}) + \lambda(x_{ni})\sqrt{h(x_{ni})}\varepsilon_i \quad (5.1)$$

auquel on associe le modèle statistique :

$$\mathcal{P}_n^Y = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left(P_{n,f}^Y \right)_{f \in \mathcal{F}} \right), \quad (5.2)$$

de l'autre côté, le modèle de bruit blanc :

$$dZ_t^{(n)} = f(t)dt + \frac{\lambda(t)}{\sqrt{n}}dB_t \quad (5.3)$$

auquel on associe le modèle statistique :

$$\mathcal{P}_n^Z = \left(\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I, \left(P_{n,f}^Z \right)_{f \in \mathcal{F}} \right), \quad (5.4)$$

Le but est de transférer les informations disponibles sur les règles de décision d'un modèle à l'autre.

Pour fixer les idées, supposons donnée une suite d'espaces de décisions (D_n, \mathcal{D}_n) .

De plus, supposons que nous disposons du modèle statistique (5.2) et d'une suite de règles de décision (non randomisée) $\pi_1^{(n)}$ qui lui est associé. On a donc une fonction $\pi_1^{(n)} : \mathbb{R}^n \rightarrow D_n$ qui associe à chaque point de \mathbb{R}^n (et donc à toutes les possibles réalisations du modèle de régression (5.1)) la "bonne" décision.

La question qu'on se pose est alors la suivante :

est-ce qu'à partir de la suite de règles de décision $\pi_1^{(n)}$ associée au modèle de régression (5.2) on peut construire une suite de règles de décision $\pi_2^{(n)}$ associée au modèle de bruit blanc (5.3) qui lui est asymptotiquement équivalente ?

De toute évidence on peut aussi se poser le problème inverse, c'est-à-dire en supposant qu'on dispose du modèle de bruit blanc (5.4) et d'une suite de règles de décision $\pi_2^{(n)}$ qui lui est associé. Supposons, de plus, que dans ce cas la suite est randomisée : donc ce dont on dispose concrètement est un noyau Markovien $\pi_2^{(n)}$ de $(\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I)$ vers (D_n, \mathcal{D}_n) . La question est alors toujours la même :

est-ce qu'à partir de la suite de règles de décision (randomisée) $\pi_2^{(n)}$ on peut construire une suite de règles de décision (randomisée) $\pi_1^{(n)}$ dans le modèle de régression (5.2) qui lui est asymptotiquement équivalente ?

Le résultat suivant, corollaire du théorème 4.1.2, répond exactement aux questions posées.

Corollaire 5.1.1. *Soit $\{\pi_1^{(n)}\}_n$ une suite de règles de décision définie dans le modèle de régression (5.2). Définissons une suite de règles de décision $\{\pi_2^{(n)}\}_n$ dans le modèle de bruit blanc (5.4) par la relation :*

$$\pi_2^{(n)}(\omega) := \pi_1^{(n)}(y_1(\omega), \dots, y_n(\omega)) \quad \text{où} \quad y_i(\omega) = K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\omega_t}{\lambda^2(t)} \quad \forall i = 1, \dots, n.$$

avec $K_i^2 = \frac{nh(x_{ni})\lambda^2(x_{ni})}{\int_{\xi_{i-1}}^{\xi_i} \frac{dt}{\lambda^2(t)}}$ et $\omega : I \rightarrow \mathbb{R}$ fonction continue.

Alors la suite de règles de décision $\{\pi_2^{(n)}\}_n$ est asymptotiquement équivalente à $\{\pi_1^{(n)}\}_n$.

Inversement, supposons connue une suite de règles de décision $\{\pi_2^{(n)}\}_n$ dans le modèle de bruit blanc (5.4). Alors $\{\pi_1^{(n)}\}_n$ est une suite de règles de décisions randomisées dans le modèle de régression (5.2) asymptotiquement équivalente à $\{\pi_1^{(n)}\}_n$, où, pour tout $A \in \mathcal{D}_n$, on a défini

$$\pi_1^{(n)}((y_1, \dots, y_n), A) := \mathbb{E} \left[\pi_2^{(n)}(\bar{Z}^{(n)}, A) \middle| K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\bar{Z}_t^n}{\lambda^2(t)} = y_i, i = 1, \dots, n \right] \quad (5.5)$$

avec $\bar{Z}^{(n)}$ qui satisfait

$$d\bar{Z}_t = \bar{f}_n(t)dt + \frac{\lambda(t)}{\sqrt{n}},$$

où \bar{f}_n est définie comme dans le chapitre 4 (voir (6.11)).

Démonstration. La preuve du corollaire se déduit de celle du théorème 4.1.2. Commençons par démontrer la première affirmation. Comme dans ce cas la suite de règles de décision est non randomisée, on a que

$$\begin{aligned} & \left| R(\mathcal{P}_n^Y, \pi_1^{(n)}, L_n, f) - R(\mathcal{P}_n^Z, \pi_2^{(n)}, L_n, f) \right| = \\ & \left| \int_{\mathbb{R}^n} L_n(f, \pi_1^{(n)}(y)) P_{n,f}^Y(dy) - \int_{\mathcal{C}(I, \mathbb{R})} L_n(f, \pi_2^{(n)}(\omega)) P_{n,f}^Z(d\omega) \right| = \\ & \left| \int_{\mathbb{R}^n} L_n(f, \pi_1^{(n)}(y)) P_{n,f}^Y(dy) - \int_{\mathcal{C}(I, \mathbb{R})} L_n(f, \pi_1^{(n)}(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^Z(d\omega) \right|. \end{aligned}$$

Or, si on considère les variables aléatoires S_1, \dots, S_n ainsi définies :

$$S_i = K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\bar{Z}_t}{\lambda(t)^2}, \quad i = 1, \dots, n$$

et on note par

$$\mathcal{P}_n^S = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_{n,f}^S)_{f \in \mathcal{F}})$$

le modèle statistique qui lui est associé, et avec

$$\mathcal{P}_n^{\bar{Z}} = (\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I, (P_{n,f}^{\bar{Z}})_{f \in \mathcal{F}})$$

le modèle statistique associé au processus $\bar{Z}^{(n)}$. On a

$$\begin{aligned} & \left| \int_{\mathbb{R}^n} L_n(f, \pi_1^{(n)}(y)) P_{n,f}^Y(dy) - \int_{\mathcal{C}(I, \mathbb{R})} L_n(f, \pi_1^{(n)}(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^Z(d\omega) \right| \\ & \leq \left| \int_{\mathbb{R}^n} L_n(f, \pi_1^{(n)}(y)) P_{n,f}^Y(dy) - \int_{\mathcal{C}(I, \mathbb{R})} L_n(f, \pi_1^{(n)}(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^{\bar{Z}}(d\omega) \right| \\ & + \left| \int_{\mathcal{C}(I, \mathbb{R})} L_n(f, \pi_1^{(n)}(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^{\bar{Z}}(d\omega) - \int_{\mathcal{C}(I, \mathbb{R})} L_n(f, \pi_1^{(n)}(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^Z(d\omega) \right| \\ & = \left| \int_{\mathbb{R}^n} L_n(f, \pi_1^{(n)}(y)) P_{n,f}^Y(dy) - \int_{\mathbb{R}^n} L_n(f, \pi_1^{(n)}(y)) P_{n,f}^S(dy) \right| \\ & + \left| \int_{\mathcal{C}(I, \mathbb{R})} L_n(f, \pi_1^{(n)}(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^{\bar{Z}}(d\omega) - \int_{\mathcal{C}(I, \mathbb{R})} L_n(f, \pi_1^{(n)}(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^Z(d\omega) \right| \\ & \leq \sup_{\substack{f \in \mathcal{F} \\ z \in D^n}} |L_n(f, z)| \left[L_1(\mathcal{P}_n^Y, \mathcal{P}_n^S) + L_1(\mathcal{P}_n^Z, \mathcal{P}_n^{\bar{Z}}) \right] \\ & \leq \sup_{\substack{f \in \mathcal{F} \\ z \in D^n}} |L_n(f, z)| \left[H(P_{n,f}^S, P_{n,f}^Y) + L_1(\mathcal{P}_n^Z, \mathcal{P}_n^{\bar{Z}}) \right], \end{aligned}$$

où on a utilisé dans le dernier passage le fait que la distance L_1 est majorée par la distance d'Hellinger (propriété 1.4.3).

Or, grâce à l'ÉTAPE 1 de la preuve du théorème 4.1.2, on sait que $\lim_{n \rightarrow \infty} L_1(\mathcal{P}_n^Z, \mathcal{P}_n^{\bar{Z}}) = 0$ et, grâce à l'ÉTAPE 3 de la même preuve, on sait que $\lim_{n \rightarrow \infty} H(P_{n,f}^S, P_{n,f}^Y) = 0$, ce qui nous fait conclure que

$$\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}} \left| R(\mathcal{P}_n^Y, \pi_1^{(n)}, L_n, f) - R(\mathcal{P}_n^Z, \pi_2^{(n)}, L_n, f) \right| = 0.$$

Démontrons maintenant la deuxième affirmation. Remarquons tout d'abord que la suite de règles de décisions $\pi_1^{(n)}$ donnée par (5.5) est bien définie. En effet elle ne dépend pas du drift $\bar{f}_n(t)$ du processus $\bar{Z}^{(n)}$ car, pour tout n , $\{S_i^{(n)} : i = 1, \dots, n\}$ est une statistique exhaustive pour $\bar{Z}^{(n)}$.

Écrivons :

$$\begin{aligned} & \left| R(\mathcal{P}_n^Y, \pi_1^{(n)}, L_n, f) - R(\mathcal{P}_n^Z, \pi_2^{(n)}, L_n, f) \right| = \\ & \left| \int_{\mathbb{R}^n} \int_{D_n} L_n(f, z) \pi_1^{(n)}(y, dz) P_{n,f}^Y(dy) - \int_{\mathcal{C}(I, \mathbb{R})} \int_{D_n} L_n(f, z) \pi_2^{(n)}(\omega, dz) P_{n,f}^Z(d\omega) \right| = \\ & \left| \int_{\mathbb{R}^n} \int_{D_n} L_n(f, z) \mathbb{E}[\pi_2^{(n)}(\bar{Z}, dz) | S_i = y_i, i = 1 \dots, n] P_{n,f}^Y(dy) - \int_{\mathcal{C}(I, \mathbb{R})} \int_{D_n} L_n(f, z) \pi_2^{(n)}(\omega, dz) P_{n,f}^Z(d\omega) \right|. \end{aligned} \quad (5.6)$$

Grâce à l'inégalité triangulaire, on peut majorer (5.6) par la somme de deux morceaux :

$$\left| \int_{\mathbb{R}^n} \int_{D_n} L_n(f, z) \mathbb{E}[\pi_2^{(n)}(\bar{Z}, dz) | S_i = y_i, i = 1 \dots, n] P_{n,f}^S(dy) - \int_{\mathcal{C}(I, \mathbb{R})} \int_{D_n} L_n(f, z) \pi_2^{(n)}(\omega, dz) P_{n,f}^Z(d\omega) \right| \quad (5.7)$$

et

$$\begin{aligned} & \left| \int_{\mathbb{R}^n} \int_{D_n} L_n(f, z) \mathbb{E}[\pi_2^{(n)}(\bar{Z}, dz) | S_i = y_i, i = 1 \dots, n] P_{n,f}^Y(dy) \right. \\ & \left. - \int_{\mathbb{R}^n} \int_{D_n} L_n(f, z) \mathbb{E}[\pi_2^{(n)}(\bar{Z}, dz) | S_i = y_i, i = 1 \dots, n] P_{n,f}^S(dy) \right|. \end{aligned} \quad (5.8)$$

Or, remarquons qu'on peut manipuler le premier morceau (5.7) de la manière suivante :

$$\begin{aligned} & \left| \int_{\mathbb{R}^n} \int_{D_n} L_n(f, z) \mathbb{E}[\pi_2^{(n)}(\bar{Z}, dz) | S_i = y_i, i = 1 \dots, n] P_{n,f}^S(dy) - \int_{\mathcal{C}(I, \mathbb{R})} \int_{D_n} L_n(f, z) \pi_2^{(n)}(\omega, dz) P_{n,f}^Z(d\omega) \right| \\ & = \left| \int_{D_n} L_n(f, z) \mathbb{E}[\pi_2^{(n)}(\bar{Z}, dz)] - \int_{\mathcal{C}(I, \mathbb{R})} \int_{D_n} L_n(f, z) \pi_2^{(n)}(\omega, dz) P_{n,f}^Z(d\omega) \right| \\ & = \left| \int_{D_n} L_n(f, z) \int_{\mathcal{C}(I, \mathbb{R})} \pi_2^{(n)}(\omega, dz) P_{n,f}^{\bar{Z}}(d\omega) - \int_{\mathcal{C}(I, \mathbb{R})} \int_{D_n} L_n(f, z) \pi_2^{(n)}(\omega, dz) P_{n,f}^Z(d\omega) \right| \\ & \leq \sup_{\substack{f \in \mathcal{F} \\ z \in D^n}} |L_n(f, z)| \left[L_1(\mathcal{P}_n^Z, \mathcal{P}_n^{\bar{Z}}) \right]. \end{aligned}$$

Grâce au théorème 1.4.2, (5.7) est majoré par la distance $L_1(\mathcal{P}_n^Y, \mathcal{P}_n^S)$ qui converge vers zéro, uniformément dans l'espace des paramètres, comme c'est déjà montré dans l'ÉTAPE 1 de la preuve du théorème 4.1.2.

Pour conclure alors que les suites de règles de décision $\pi_1^{(n)}$ et $\pi_2^{(n)}$ sont asymptotiquement équivalentes, il suffit de montrer que (5.8) converge vers zéro, uniformément dans l'espace des paramètres \mathcal{F} .

Or, le théorème 4.1.2 nous assure que (5.8) est majoré par la distance $L_1(\mathcal{P}_n^Y, \mathcal{P}_n^S)$ qui est, à son tour (grâce à la propriété 1.4.3), majorée par la distance d'Hellinger $H(P_{n,f}^S, P_{n,f}^Y)$ dont on sait déjà qu'elle converge uniformément vers zéro (voir l'ÉTAPE 3 de la preuve du théorème 4.1.2). \square

5.1.1 Estimation du drift à partir du modèle de régression

Considérons tout d'abord le modèle de régression non paramétrique défini dans le chapitre 3 :

$$Y_i = f(X_i) + \lambda(X_i) \sqrt{h(X_i)} \varepsilon_i, \quad i = 1, \dots, n, \quad (\varepsilon_i)_i \text{ i.i.d } \mathcal{N}(0, 1) \quad (5.9)$$

où la fonction de régression f est inconnue et appartient à une certaine classe fonctionnelle \mathcal{F} .

On suppose avoir construit, pour tout n , un estimateur de f , T_n , qui nous satisfait. Le problème qu'on se pose est alors le suivant :

comment peut-on estimer le drift du modèle de signal avec bruit blanc :

$$dZ_t = f(t)dt + \frac{\lambda(t)}{\sqrt{n}}dB_t, \quad t \in I \quad (5.10)$$

à partir des estimateurs $(T_n)_n$, de manière à atteindre asymptotiquement le même risque que celui associé aux T_n ?

La réponse apparaît implicitement dans le corollaire 5.1.1. En effet on peut voir une règle de décision non randomisée tout simplement comme un estimateur de f . On propose alors comme estimateur du drift f l'estimateur F_n ainsi défini :

$$F_n(\omega) := T_n(S_1^{(n)}(\omega), \dots, S_n^{(n)}(\omega))$$

où les $(S_i^{(n)})_{1 \leq i \leq n}$ sont définies comme :

$$K_i^2 = \frac{n\sigma^2(x_{ni})}{\int_{\xi_{i-1}}^{\xi_i} \frac{dt}{\lambda^2(t)}} \quad S_i^{(n)}(\omega) = K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\omega_t}{\lambda^2(t)}, \quad i = 1, \dots, n,$$

et $\omega : t \rightarrow Z_t^{(n)}$ est une trajectoire du processus $Z^{(n)} = (Z_t^n, t \in I)$

5.1.2 Estimation de la fonction de régression à partir du modèle de bruit blanc

Malheureusement, estimer la fonction de régression à partir d'un estimateur du drift dans le modèle de bruit blanc n'est pas aussi facile qu'estimer le drift à partir du modèle de régression.

Plus précisément, supposons que nous disposons d'une suite d'estimateurs F_n du drift du modèle de bruit blanc :

$$dZ_t^{(n)} = f(t)dt + \frac{\lambda(t)}{\sqrt{n}}dB_t$$

et de n observations, Y_1, \dots, Y_n , issues du modèle de régression :

$$Y_i = f(X_i) + \lambda(X_i)\sqrt{h(X_i)}\varepsilon_i, \quad i = 1, \dots, n, \quad \varepsilon_i \text{ i.i.d } \mathcal{N}(0, 1). \quad (5.11)$$

La question qu'on se pose est : comment trouver une suite d'estimateur $(T_n)_n$ de la fonction de régression f qui ait asymptotiquement le même risque de la suite d'estimateurs $(F_n)_n$?

La réponse n'est plus claire. En effet, pour pouvoir utiliser le corollaire 5.1.1, on aurait besoin de disposer d'une suite d'estimateurs $(F_n)_n$ qui soient des noyaux Markoviens de $(\mathcal{C}(I, \mathbb{R}), \mathcal{C}^I)$ vers un espace des décisions fixé (D_n, \mathcal{D}_n) . De plus, même si on dispose d'une telle suite d'estimateurs, il y a encore d'autres difficultés. En effet, la construction donnée par le corollaire 5.1.1 nous amène à définir la suite $(T_n)_n$ de la manière suivante : pour tout $A \in \mathcal{D}_n$,

$$T_n((y_1, \dots, y_n), A) := \mathbb{E} \left[F_n(\bar{Z}^{(n)}, A) \middle| K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\bar{Z}_t^{(n)}}{\lambda^2(t)} = y_i, i = 1, \dots, n \right] \quad (5.12)$$

avec $\bar{Z}^{(n)}$ qui satisfait

$$d\bar{Z}_t = \bar{f}_n(t)dt + \frac{\lambda(t)}{\sqrt{n}}dB_t,$$

où \bar{f}_n est définie comme dans le chapitre 4 (voir (6.11)). Or, on voit bien que le calcul de l'espérance conditionnelle peut être difficile dans la pratique, ce qui rend pratiquement inaccessible l'expression explicite des T_n , qui sont, en tout cas, des estimateurs "randomisés".

Un cas où il est possible d'améliorer la situation, c'est-à-dire d'obtenir des estimateurs classiques, non-randomisés, est celui où l'espace D_n des décisions est un sous-ensemble fermé et convexe d'un espace séparable de Banach et toutes les fonctions de perte $L_n(f, \cdot)$ sont convexes et ont une limite infinie à l'infini :

$$\lim_{\|z\| \rightarrow \infty} L_n(\theta, z) = \infty.$$

Dans ce cas, en définissant une suite d'estimateurs T'_n de la fonction de régression par :

$$T'_n(y_1, \dots, y_n) := \int_{D_n} z \mathbb{E} \left[F_n(\bar{Z}^{(n)}, dz) \middle| K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\bar{Z}_t^{(n)}}{\lambda^2(t)} = y_i, i = 1, \dots, n \right], \quad (5.13)$$

grâce à l'inégalité de Jensen, on obtient que la suite d'estimateurs définie par la relation (5.13), est (du point de vue du risque) au moins aussi bonne que la suite d'estimateurs (5.12). En effet, si on note par \mathcal{P}_n^Y le modèle statistique associé au modèle de régression (5.11), on a :

$$\begin{aligned} & R(\mathcal{P}_n^Y, T'_n, L_n, f) - R(\mathcal{P}_n^Y, T_n, L_n, f) \\ &= \int_{\mathbb{R}^n} L_n(f, T'_n(y)) P_{n,f}^Y(dy) - \int_{\mathbb{R}^n} \int_{D_n} L_n(f, z) \mathbb{E} [F_n(\bar{Z}^{(n)}, dz) \middle| K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\bar{Z}_t^{(n)}}{\lambda^2(t)} = y_i, i = 1, \dots, n] P_{n,f}^Y(dy) \\ &= \int_{\mathbb{R}^n} L_n \left(f, \int_{D_n} z \mathbb{E} \left[F_n(\bar{Z}^{(n)}, dz) \middle| K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\bar{Z}_t^{(n)}}{\lambda^2(t)} = y_i, i = 1, \dots, n \right] \right) P_{n,f}^Y(dy) \\ &\quad - \int_{\mathbb{R}^n} \int_{D_n} L_n(f, z) \mathbb{E} [F_n(\bar{Z}^{(n)}, dz) \middle| K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\bar{Z}_t^{(n)}}{\lambda^2(t)} = y_i, i = 1, \dots, n] P_{n,f}^Y(dy) \end{aligned}$$

Or, l'inégalité de Jensen nous dit que pour toute variable aléatoire V à valeurs dans un sous-ensemble fermé et convexe d'un espace de Banach séparable, et fonction convexe Φ définie sur cet espace et telle que $\lim_{\|a\| \rightarrow \infty} \Phi(a) = \infty$, on a

$$\Phi(\mathbb{E}[V]) = \Phi \left(\int z dV_{\#}(z) \right) \leq \mathbb{E}[\Phi(V)] = \int \Phi(z) V_{\#}(dz).$$

On pose $\Phi = L_n(f, \cdot)$ et V une variable aléatoire de loi image $V_{\#}(dz) = \mathbb{E} [F_n(\bar{Z}^{(n)}, dz) \middle| K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\bar{Z}_t^{(n)}}{\lambda^2(t)} = y_i, i = 1, \dots, n]$. Alors grâce à l'inégalité précédente on déduit que :

$$R(\mathcal{P}_n^Y, T'_n, L_n, f) - R(\mathcal{P}_n^Y, T_n, L_n, f) \leq 0,$$

comme souhaité.

5.1.3 Risques minimax

Une autre propriété très utile est lorsqu'on dispose des deux modèles qui sont asymptotiquement équivalents, alors chaque risque asymptotiquement minimax (par rapport à des fonctions de perte bornées) dans un modèle peut être transféré dans l'autre modèle.

Afin de rendre complètement claire l'affirmation précédente, on introduit les définitions suivantes.

Définition 5.2. Le *risque minimax* associé au modèle statistique \mathcal{E} , à la fonction de perte bornée L et au paramètre $f \in \mathcal{F}$, est défini comme suit :

$$R(\mathcal{F}) := \inf_{\hat{f}} \sup_{f \in \mathcal{F}} R(\mathcal{E}, \hat{f}, L, f),$$

où l'infimum est pris sur tous les possibles estimateurs \hat{f} de f .

Définition 5.3. Un estimateur T est dit *estimateur minimax* du paramètre $f \in \mathcal{F}$ relatif au modèle statistique \mathcal{E} et à la fonction de perte bornée L , si

$$\sup_{f \in \mathcal{F}} R(\mathcal{E}, T, L, f) = \inf_{\hat{f}} \sup_{f \in \mathcal{F}} R(\mathcal{E}, \hat{f}, L, f).$$

Définition 5.4. Une suite d'estimateurs $(T_n)_n$ est dite *suite d'estimateurs asymptotiquement minimax* du paramètre $f \in \mathcal{F}$ relative à la suite des modèles statistiques $\mathcal{E}^{(n)}$ et aux fonctions de perte bornée L_n , si

$$\lim_{n \rightarrow \infty} \left| \sup_{f \in \mathcal{F}} R(\mathcal{E}^{(n)}, T_n, L_n, f) - \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} R(\mathcal{E}^{(n)}, \hat{f}_n, L_n, f) \right| = 0$$

Dans la suite, soient $\mathcal{E}^{(n)}$ et $\mathcal{F}^{(n)}$ deux suites de modèles statistiques qui ont le même espace des paramètres \mathcal{F} .

Lemme 5.1.2. Si $\mathcal{E}^{(n)}$ et $\mathcal{F}^{(n)}$ sont asymptotiquement équivalents, alors ils ont, asymptotiquement, le même risque minimax :

$$\lim_{n \rightarrow \infty} \left| \inf_{T_n} \sup_{f \in \mathcal{F}} R(\mathcal{E}^{(n)}, T_n, L_n, f) - \inf_{F_n} \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n, L_n, f) \right| = 0;$$

où l'inf est pris, respectivement, sur tous les possibles estimateurs T_n de f dans le modèle \mathcal{E}_n et sur tous les possibles estimateurs F_n de f dans le modèle \mathcal{F}_n .

Démonstration. Montrons tout d'abord que, pour tout $\varepsilon > 0$ il existe n_0 tel que pour tout $n \geq n_0$ on a :

$$\inf_{T_n} \sup_{f \in \mathcal{F}} R(\mathcal{E}^{(n)}, T_n, L_n, f) - \inf_{F_n} \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n, L_n, f) \leq \varepsilon. \quad (5.14)$$

Soit $\varepsilon > 0$ fixé. Clairement, il suffit de montrer qu'il existe une suite $(T_n^*)_n$ tel que, pour tout $n \geq n_0$ on a :

$$\sup_{f \in \mathcal{F}} R(\mathcal{E}^{(n)}, T_n^*, L_n, f) - \inf_{F_n} \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n, L_n, f) \leq \varepsilon.$$

Soit $(F_n^{\min})_n$ une suite d'estimateurs asymptotiquement minimax du paramètre f relative à la suite $\mathcal{F}^{(n)}$. Alors, par définition,

$$\lim_{n \rightarrow \infty} \left| \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n^{\min}, L_n, f) - \inf_{F_n} \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n, L_n, f) \right| = 0,$$

donc il existe n_1 tel que, pour tout $n \geq n_1$,

$$\sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n^{\min}, L_n, f) - \inf_{F_n} \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n, L_n, f) \leq \frac{\varepsilon}{2}.$$

De plus, comme les suites de modèles $\mathcal{E}^{(n)}$, $\mathcal{F}^{(n)}$ sont asymptotiquement équivalentes, grâce à la proposition 1.3.1, il existe une suite T_n^* d'estimateurs (potentiellement randomisés) associée à la suite $\mathcal{E}^{(n)}$ et un n_2 tels que, pour tout $n \geq n_2$,

$$\sup_{f \in \mathcal{F}} R(\mathcal{E}^{(n)}, T_n^*, L_n, f) - \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n^{\min}, L_n, f) \leq \frac{\varepsilon}{2}. \quad (5.15)$$

Si on pose $n_0 = \max(n_1, n_2)$, l'inégalité (5.14) suit alors par inégalité triangulaire.

Pour démontrer l'autre inégalité on suit exactement le même raisonnement. \square

Proposition 5.1.3. *Soient $\mathcal{E}^{(n)}$ et $\mathcal{F}^{(n)}$ deux modèles asymptotiquement équivalents. Si $(T_n)_n$ est une suite d'estimateurs asymptotiquement minimax du paramètre f relative à la suite de modèles statistiques $\mathcal{E}^{(n)}$ et $(F_n)_n$ est une suite d'estimateurs de f relative à la suite de modèles $\mathcal{F}^{(n)}$ asymptotiquement équivalente à $(T_n)_n$, alors $(F_n)_n$ est une suite d'estimateurs asymptotiquement minimax du paramètre f .*

Démonstration. Par l'inégalité triangulaire, la quantité

$$\lim_{n \rightarrow \infty} \left| \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n, L_n, f) - \inf_{F_n^*} \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n^*, L_n, f) \right|$$

est majorée par les deux quantités suivantes :

$$\lim_{n \rightarrow \infty} \left| \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n, L_n, f) - \inf_{T_n^*} \sup_{f \in \mathcal{F}} R(\mathcal{E}^{(n)}, T_n^*, L_n, f) \right| \quad (5.16)$$

et

$$\lim_{n \rightarrow \infty} \left| \inf_{T_n^*} \sup_{f \in \mathcal{F}} R(\mathcal{E}^{(n)}, T_n^*, L_n, f) - \inf_{F_n^*} \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n^*, L_n, f) \right| \quad (5.17)$$

Or, comme $(F_n)_n$ et $(T_n)_n$ sont asymptotiquement équivalentes et la suite $(T_n)_n$ est asymptotiquement minimax pour la suite de modèles $\mathcal{E}^{(n)}$, la quantité (5.16) est nulle. De même, grâce au lemme 5.1.2, la quantité (5.17) est nulle, ce qui montre bien que

$$\lim_{n \rightarrow \infty} \left| \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n, L_n, f) - \inf_{F_n^*} \sup_{f \in \mathcal{F}} R(\mathcal{F}^{(n)}, F_n^*, L_n, f) \right| = 0.$$

\square

Corollaire 5.1.4. *Soit $(T_n)_n$ une suite d'estimateurs asymptotiquement minimax de la fonction de régression f relative à la suite de modèles statistiques $\mathcal{P}^{(1,n)}$ définie par (5.2). Alors la suite d'estimateurs $(F_n)_n$ du drift associée à la suite de modèles statistiques $\mathcal{P}^{(2,n)}$ ((5.4)) ainsi construit :*

$$F_n(\omega) := T_n(y_1(\omega), \dots, y_n(\omega)) \quad \text{où} \quad y_i(\omega) = K_i \int_{\xi_{i-1}}^{\xi_i} \frac{d\omega_t}{\lambda^2(t)} \quad \forall i = 1, \dots, n.$$

avec $K_i^2 = \frac{nh(x_{ni})\lambda^2(x_{ni})}{\int_{\xi_{i-1}}^{\xi_i} \frac{dt}{\lambda^2(t)}}$ et $\omega : I \rightarrow \mathbb{R}$ fonction continue, est asymptotiquement minimax.

Démonstration. Grâce à la proposition précédente, il suffit de démontrer que les suites $(T_n)_n$ et $(F_n)_n$ sont asymptotiquement équivalentes. Pour cela, appliquons le corollaire 5.1.1 en posant $\pi_1^{(n)} \equiv T_n$ et en remarquant que, par construction, $\pi_2^{(n)} \equiv F_n$. L'équivalence asymptotique entre les suites de règles de décision $(\pi_1^{(n)})_n$ et $(\pi_2^{(n)})_n$ s'écrit alors sous la forme

$$\lim_{n \rightarrow \infty} \left| \sup_{f \in \mathcal{F}} R(\mathcal{P}_n^Y, T_n, L_n, f) - \sup_{f \in \mathcal{F}} R(\mathcal{P}_n^Z, F_n, L_n, f) \right| = 0,$$

ce qui conclut la preuve. \square

Comme on a déjà souligné précédemment, associer à une suite d'estimateur minimax dans le modèle de bruit blanc une suite d'estimateur dans le modèle de régression, est souvent difficile dans la pratique.

Cependant, si on connaît un estimateur minimax dans le modèle de bruit blanc, même si on ne peut pas lui associer directement un estimateur dans le modèle de régression, on peut transférer les vitesses de convergence dans ce dernier modèle (voir proposition 5.1.3). Une construction explicite d'un tel estimateur, pour une certaine classe fonctionnelle \mathcal{F} , sera l'objet du paragraphe suivant.

5.1.4 Estimateur de Pinsker

Considérons le modèle suivant de signal avec bruit blanc :

$$dZ(t) = f(t)dt + \varepsilon dB(t), \quad t \in [0, 1], \quad 0 < \varepsilon < 1. \quad (5.18)$$

avec $f : [0, 1] \rightarrow \mathbb{R}$ qui appartient à la classe de Sobolev $W(\beta, L)$ ainsi définie :

$$W(\beta, L) = \{f \in L_2[0, 1] : \theta = \{\theta_j\} \in \Theta(\beta, Q)\}, \quad Q = \frac{L^2}{\pi^{2\beta}},$$

où

- $\theta_j = \int_0^1 f(s)\phi_j(s)ds$
- $\{\phi_j\}_{j=1}^\infty$ est la base trigonométrique définie par :

$$\begin{aligned} \phi_1(x) &\equiv 1 \\ \phi_{2k}(x) &= \sqrt{2} \cos(2\pi kx) \\ \phi_{2k+1}(x) &= \sqrt{2} \sin(2\pi kx) \end{aligned}$$

où $x \in [0, 1]$ et $k = 1, 2, \dots$

- $\Theta(\beta, Q)$ est l'ellipsoïde

$$\Theta(\beta, Q) = \left\{ \theta = \{\theta_j\} \in \ell^2(\mathbb{N}) : \sum_{j=1}^{\infty} a_j^2 \theta_j^2 \leq Q \right\}$$

avec

$$a_j = \begin{cases} j^\beta & \text{si } j \text{ est pair,} \\ (j-1)^\beta & \text{sinon.} \end{cases} \quad (5.19)$$

Comme s'est montré dans le chapitre 3, la séquence suivante d'observations est à la disposition du statisticien une fois qu'on se donne le modèle (5.18) :

$$z_j = \int_0^1 \phi_j(t) dZ(t) = \theta_j + \varepsilon \xi_j, \quad j = 1, 2, \dots,$$

où les $\xi_j = \int_0^1 \phi_j(x) dB(x)$ sont variables aléatoires i.i.d. de loi commune normale centrée et réduite.

Définissons l'estimateur suivant de f :

$$\tilde{f}_\varepsilon(x) = \sum_{j=1}^{\infty} l_j^* z_j \phi_j(x) \quad (5.20)$$

où

$$l_j^* = (1 - \kappa^* a_j)_+ \quad (5.21)$$

avec

$$\kappa^* = \left(\frac{\beta}{(2\beta + 1)(\beta + 1)Q} \right)^{\frac{\beta}{2\beta+1}} \varepsilon^{\frac{2\beta}{2\beta+1}}. \quad (5.22)$$

Le nombre de termes non nuls $N = \max\{j : l_j^* > 0\}$ dans la somme (5.20) est fini, de sorte qu'on peut écrire

$$\tilde{f}_\varepsilon(x) = \sum_{j=1}^N l_j^* z_j \phi_j(x).$$

On peut facilement démontrer que $N = N_\varepsilon$ converge vers l'infini quand ε tend vers zéro avec vitesse $\varepsilon^{-2/(2\beta+1)}$.

Pour une interprétation graphique, en figure 5.2 on a superposé au graphe de la fonction $f(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$ le graphe de l'estimateur de Pinsker qui lui est associé.

Supposons avoir observé une trajectoire $\omega : t \rightarrow Z_t$ du processus Z défini par (5.18).

Théorème 5.1.5. *Soient $\beta > 0$, $L > 0$. Alors*

$$\lim_{\varepsilon \rightarrow 0} \sup_{f \in W(\beta, L)} \varepsilon^{-\frac{4\beta}{2\beta+1}} \mathbb{E}_f \|\tilde{f}_\varepsilon - f\|_2^2 = \liminf_{\varepsilon \rightarrow 0} \sup_{T_\varepsilon} \sup_{f \in W(\beta, L)} \varepsilon^{-\frac{4\beta}{2\beta+1}} \mathbb{E}_f \|T_\varepsilon - f\|_2^2 = C^*,$$

où $\inf T_\varepsilon$ désigne la borne inférieure sur tous les estimateurs, \mathbb{E}_f indique l'esperance par rapport à la loi de l'observation ω sous le modèle (5.18), $\|\cdot\|_2$ est la norme $L_2[0, 1]$, et

$$\begin{aligned} C^* &= L^{\frac{2}{2\beta+1}} (2\beta + 1)^{\frac{1}{2\beta+1}} \left(\frac{\beta}{\pi(\beta + 1)} \right)^{\frac{2\beta}{2\beta+1}} \\ &= [Q(2\beta + 1)]^{\frac{1}{2\beta+1}} \left(\frac{\beta}{\beta + 1} \right)^{\frac{2\beta}{2\beta+1}} \end{aligned} \quad (5.23)$$

avec $Q = \frac{L^2}{\pi^{2\beta}}$.

Démonstration. Voir Pinsker (1980) [34]. □

Remarque 5.1.6. Le théorème 5.1.5 entraîne que l'estimateur (5.20) est *asymptotiquement efficace* dans le sens que :

$$\lim_{\varepsilon \rightarrow 0} \sup_{f \in W(\beta, L)} \frac{\mathbb{E}_f \|\tilde{f}_\varepsilon - f\|_2^2}{\mathcal{R}_\varepsilon^*} = 1,$$

où $\mathcal{R}_\varepsilon^*$ est le *risque minimax* défini comme

$$\mathcal{R}_\varepsilon^* := \inf_{T_\varepsilon} \sup_{f \in W(\beta, L)} \mathbb{E}_f \|T_\varepsilon - f\|_2^2.$$

Considérons maintenant le modèle de régression non paramétrique donné par

$$Y_i = f(i/n) + \xi_i, \quad i = 1, \dots, n, \quad (5.24)$$

où les ξ_i sont des variables aléatoires i.i.d de loi $\mathcal{N}(0, \sigma^2)$, avec $\sigma^2 > 0$.

Si on pose $\varepsilon = \sigma/\sqrt{n}$, grâce au théorème 4.1.2 on sait que les modèles (5.18) et (5.24) sont asymptotiquement équivalents, donc (voir commentaires chapitre 1 et proposition 1.3.1) on peut énoncer le théorème suivant :

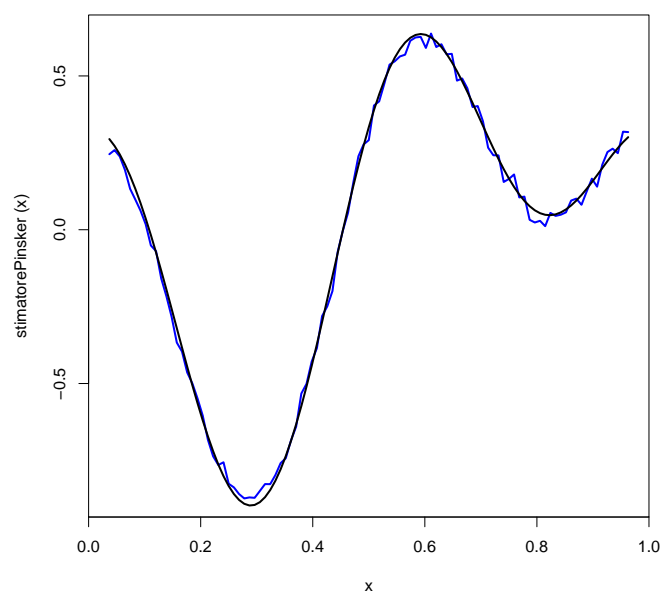


FIGURE 5.2 – Superposition du graphe de la fonction $f(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$ et de l'estimateur de Pinsker correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.
En **noir** : la vraie fonction, en **blue** : l'estimateur de Pinsker.

Théorème 5.1.7. *Il existe un estimateur \tilde{f}_n de f dans le modèle (5.24) tel que*

$$\begin{aligned} \lim_{n \rightarrow \infty} \sup_{f \in W(\beta, L)} \mathbb{E}_f(n^{\frac{2\beta}{2\beta+1}} \|\tilde{f}_n - f\|_2^2) &= \lim_{n \rightarrow \infty} \inf_{T_n} \sup_{f \in W(\beta, L)} \mathbb{E}_f(n^{\frac{2\beta}{2\beta+1}} \|T_n - f\|_2^2) \\ &= C^* \sigma^{\frac{4\beta}{2\beta+1}}, \end{aligned}$$

où $\inf T_\varepsilon$ désigne la borne inférieure sur tous les estimateurs, \mathbb{E}_f indique l'espérance par rapport à la loi des observations (Y_1, \dots, Y_n) sous le modèle (5.24), $\|\cdot\|_2$ est la norme $L_2[0, 1]$, $\beta \geq 1$, $L > 0$ et C^* est définie par (5.23).

Remarque 5.1.8. Une autre application de l'équivalence asymptotique entre les modèles (3.3), (3.5), par exemple, est le fait de pouvoir transférer dans le modèle de régression le résultat de Fan (1991) [16] sur la vitesse de convergence des estimateurs de fonctionnelles quadratiques.

5.2 Un peu de littérature

L'idée d'approcher des expériences compliquées par une expérience plus simple remonte au papier de Wald (1943) [39]. Dans les années 50, David Blackwell et Stein trouvèrent des conditions nécessaires et suffisantes pour que une expérience puisse être considérée "meilleure" ou "plus instructive" par rapport à une autre, en s'inspirant d'un papier non publié de Shapley et Sherman (1949). Aujourd'hui, ce résultat est connu sous le nom de théorème de Blackwell-Sherman-Stein.

En 1964 Le Cam, motivé par le théorème de Blackwell-Sherman-Stein, introduit la notion de déficience entre deux modèles statistiques. La déficience peut être interprétée comme une tentative de rendre quantitatif le théorème de Blackwell-Sherman-Stein, c'est-à-dire de quantifier la "déficience" entre deux expériences. Généralement, les articles de Le Cam ne sont pas aisés à manipuler, et son travail ne sera appréhendé dans son ensemble qu'à partir des années 90. Une véritable avancée dans la théorie de l'équivalence asymptotique entre expériences est due aux résultats de Brown et Low (1996) [4] et Nussbaum (1996) [33].

Le premier établit l'équivalence entre le modèle de régression non paramétrique et le modèle de signal avec bruit blanc présentés dans le chapitre 3 (pour une preuve d'une telle équivalence voir le chapitre 4), le deuxième montre qu'un échantillon de taille n de variables aléatoires identiquement distribuées et indépendantes de densité commune f sur $[0, 1]$, avec f appartenant à une certaine classe d'Hölder, est asymptotiquement équivalente à un modèle de bruit blanc avec drift $f^{\frac{1}{2}}$ et variance $\frac{1}{4}n^{-1}$.

Or, le modèle de bruit blanc est bien étudié. On dispose de beaucoup de résultats sur l'estimation minimax dans le cas gaussien, comme par exemple, les vitesses de convergence optimales (voir Donoho, Johnstone, Kerkyacharian et Picard (1995)[15] et Tsybakov (1997) [38]).

En utilisant le résultat de Nussbaum, on peut facilement transférer tous ces résultats relatifs au modèle de bruit blanc, dans le modèle de régression pour résoudre le problème de l'estimation de la fonction de régression.

Ces dernières années, plusieurs développements ont été publiés.

Grama et Nussbaum (1998) [20] ont démontré une équivalence asymptotique pour modèles linéaires généralisés qui a été étendu à une plus grande classe de fonctions dans Grama et Nussbaum (2002) [21], Jähmish et Nussbaum (2003) [23].

Plus précisément, le cadre dans lequel se pose l'article de Grama et Nussbaum (1998) est le suivant :

On dispose d'un échantillon de n variables aléatoires indépendantes X_i dont la loi appartient à une famille exponentielle \mathcal{P} de paramètres $\theta_i = f(\frac{i}{n}) \in \Theta$ où $f : [0, 1] \rightarrow \Theta$ est une fonction inconnue à estimer, appartenant à une certaine classe fonctionnelle Σ .

Le résultat principal de l'article est que, sous certaines hypothèses sur Σ , un tel modèle est asymptotiquement équivalent, au sens de Le Cam, à une expérience gaussienne en forme homoscedastique :

$$dY_t^n = \Gamma(f(t))dt + \frac{1}{\sqrt{n}}dW_t, \quad t \in [0, 1],$$

où $f \in \Sigma$, et Γ est une fonction $\Gamma : \Theta \rightarrow \mathbb{R}$ telle que $\Gamma'(\theta) = \sqrt{I(\theta)}$ où $I(\theta)$ est l'information de Fisher dans la famille exponentielle \mathcal{P} et avec W un mouvement brownien standard.

Dans le 2002, Brown, Cai, Low et Zhang [3] reprennent l'article de Brown et Low (1996) [4] en généralisant au cas du design aléatoires et, surtout, en construisant d'une façon explicite les noyaux Markoviens qui réalisent l'équivalence asymptotique.

Une autre généralisation de l'article de Brown et Low (1996) [4] est due à Carter (2007) [7]. Il montre l'équivalence asymptotique entre un modèle de régression non paramétrique à variance inconnue et un modèle de bruit blanc. Dans ce cas aussi, l'équivalence est démontrée en exhibant une construction explicite qui permet d'obtenir un processus gaussien à partir du modèle de régression et vice versa.

Ils existent aussi des résultats d'équivalence asymptotique pour des modèles de diffusion. Le premier résultat à cet effet est attribuable à Milstein et Nussbaum (1998) [31] qui ont démontré comment certains processus de diffusion peuvent être approchés de leurs versions discrètes données par des modèles d'autoregression.

Plus concrètement, ils considèrent un modèle de diffusion à temps continu de la forme

$$dY(t) = f(Y(t))dt + \varepsilon dW(t) \quad t \in [0, 1] \quad Y(0) = 0, \quad (5.25)$$

où $dW(t)$ est un bruit blanc, ε un paramètre petit et f une fonction inconnue appartenant à une certaine classe fonctionnelle Σ .

Dans ce cas, la discrétisation à considérer est suggérée par le schéma d'Euler pour la résolution d'équations différentielles stochastiques et le modèle qu'on construit est le suivant :

$$Y_{i+1} = Y_i + n^{-1}f(Y_i) + \varepsilon n^{-\frac{1}{2}}\xi_i, \quad i = 1, \dots, n, \quad Y_1 = 0 \quad (5.26)$$

où les ξ_i sont variables aléatoires i.i.d normales réduites et Y_i est une solution approchée de (5.25) définie seulement aux points $t_i, i = 1, \dots, n$ d'une grille qui est une partition de $[0, 1]$.

Le résultat principal du papier de Milstein et Nussbaum est que, sous certaines hypothèses sur la classe fonctionnelle Σ , et si $n = n_\varepsilon$ est tel que $\varepsilon n_\varepsilon \rightarrow \infty$ quand $\varepsilon \rightarrow 0$, les modèles donnés par les relations (5.25) et (8.6) sont asymptotiquement équivalents quand $\varepsilon \rightarrow 0$.

D'autres résultats qui concernent l'équivalence asymptotique pour des modèles de diffusion peuvent être trouvés dans Genon-Catalot, Laredo et Nussbaum (2002) [17], Delattre et Hoffman (2002) [13], Dalalyan et Reiss (2004) [11], Grama et Neumann (2006) [19] et Konakov, Mammen, Woerner (2012) [25].

Les résultats sur l'équivalence asymptotique dans modèle multidimensionnel sont rares. On peut citer le papier de Carter (2006) [6] (qui a montré l'équivalence asymptotique entre un problème de régression gaussienne bidimensionnelle et le problème de l'estimation de densité), Reiss (2006)[35], Brown et Zhang (1998) [5], Dalalyan et Reiss (2008) [12] et Klemelä et Tsybakov (2001) [24].

Sont également rares les résultats d'équivalence asymptotique pour données dépendantes. Des exemples récents dans cette direction sont donnés par Carter (2009) [8], Meister (2011) [30].

D'autres applications à la théorie de l'équivalence asymptotique entre modèles statistiques sont dues à Brown, Carter, Low, Zhang (2004) [3] qui ont montré l'équivalence asymptotique entre un processus de Poisson avec intensité variable et bruit blanc avec drift, à Golubev, Nussbaum, Zhou (2009) [18] qui ont établi une approximation discrète gaussienne au problème de l'estimation de la densité spectrale d'une processus stationnaire, à Dette, Nagel et Neumeyer (2004) [14] qui ont utilisé le résultat de Brown et Low [4] pour enquêter sur la performance du bootstrap non paramétrique appliqué au modèle de régression non paramétrique.

Troisième partie

Généralisation, simulations et
perspectives

Chapitre 6

Régression et modèles de signal perturbé par un processus à sauts

L'objectif de ce chapitre est de généraliser le théorème 4.1.2 du chapitre 4 sur l'équivalence asymptotique entre le modèle de signal avec bruit blanc et le modèle de régression. En particulier, on modifiera le modèle de bruit blanc en incluant des trajectoires discontinues et on démontrera, en ce qui concerne l'estimation du drift, que ce dernier modèle est encore une fois équivalent au modèle de régression habituelle, rencontré à plusieurs reprises au cours du mémoire.

Commençons par décrire ce contexte.

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé fixé et (D_n, \mathcal{D}_n) une suite d'espaces de décision fixée. Dans la suite on notera toujours par $X = (X_t : t \in [0, 1])$ un processus purement de saut (c'est-à-dire tel que sa partie continue soit nulle presque sûrement) et $B = (B_t : t \in [0, 1])$ un mouvement brownien défini sur l'espace $(\Omega, \mathcal{A}, \mathbb{P})$. On supposera que X et B sont définis sur $(\Omega, \mathcal{A}, \mathbb{P})$ et sont indépendants.

On notera (D, \mathcal{D}) l'espace des fonctions de $[0, 1]$ dans \mathbb{R} continues à droite et qui admettent une limite à gauche en chaque point (càdlàg) muni de la topologie de Skorohod. (Voir [1]).

On considère une suite de processus stochastiques $Z^{(n)} = (Z_t^{(n)} : t \in [0, 1])$ définie par :

$$dZ_t^{(n)} = f(t)dt + \frac{\sigma(t)}{\sqrt{n}}dB_t + dX_t \quad (6.1)$$

où f est une fonction inconnue appartenant à une certaine classe \mathcal{F} .

On note par \mathcal{P}_n^Z le modèle statistique associé au processus $Z^{(n)}$:

$$\mathcal{P}_n^Z := \left(D, \mathcal{D}, \left(P_{f,n}^Z \right)_{f \in \mathcal{F}} \right), \quad (6.2)$$

où $P_{f,n}^Z$ est la loi du processus $Z^{(n)}$.

On se pose alors la question suivante : quel est un équivalent asymptotique discret, au sens de Le Cam, de la suite de modèles statistiques (6.2) ?

L'objectif étant de construire une suite de modèles statistiques qui, asymptotiquement, apporte la même information sur la fonction f que la suite (6.2), l'idée est que, d'un point de vue de l'inférence sur f , les modèles statistiques associés aux suites de processus :

$$dZ_t^{(n)} = f(t)dt + \frac{\sigma(t)}{\sqrt{n}}dB_t + dX_t,$$

$$d\tilde{Z}_t^{(n)} = f(t)dt + \frac{\sigma(t)}{\sqrt{n}}dB_t$$

sont équivalents (car toute l'information sur f est contenue dans la partie continue du processus $Z^{(n)}$, qui est justement donnée par le processus $\tilde{Z}^{(n)}$.)

Comme on sait déjà que la suite de modèles de signal avec bruit blanc est asymptotiquement équivalente à un modèle de régression, on s'attend alors à obtenir le même résultat d'équivalence asymptotique dans le cas d'un modèle de signal discontinu donné par (6.1).

Pour pouvoir énoncer rigoureusement un tel résultat il faut étudier tout d'abord le comportement, en termes de distance L_1 , de processus à sauts de la forme (6.1).

6.1 Distance L_1 entre processus à sauts

Lemme 6.1.1. *Soit $Z = (Z_t : t \in [0, 1])$ un processus stochastique défini par l'équation :*

$$dZ_t = f(t)dt + \sigma(t)dB_t + dX_t,$$

avec f et σ qui vérifient :

$$\int_0^1 \frac{f^2(s)}{\sigma^2(s)} ds < \infty. \quad (6.3)$$

Soit ξ la loi du processus $Y = (Y_t : t \in [0, 1])$ défini par l'équation

$$dY_t = \sigma(t)dB_t + dX_t.$$

Alors la mesure ξ domine la loi de Z et la densité est donnée par :

$$\frac{d\mathcal{L}_Z}{d\xi}(\omega) = \exp\left(\int_0^1 \frac{f(s)}{\sigma^2(s)} d\omega_s^c - \frac{1}{2} \int_0^1 \frac{f^2(s)}{\sigma^2(s)} ds\right), \quad (6.4)$$

où on écrit $\omega = \omega^c + \omega^d$ pour indiquer la décomposition en partie continue et discontinue, respectivement, d'une trajectoire càdlàg.

Démonstration. Écrivons :

$$dY_t = f(t)dt + \sigma(t)\left[dB_t - \frac{f(t)}{\sigma(t)}dt \right] + dX_t,$$

et définissons un nouveau processus sur (Ω, \mathcal{A}) , $\tilde{B} = (\tilde{B}_t, t \in [0, 1])$, de la manière suivante :

$$\tilde{B}_t = B_t - \int_0^t \frac{f(s)}{\sigma(s)} ds.$$

Grâce à l'hypothèse (6.3), le processus $\exp\left(\int_0^t \frac{f(s)}{\sigma(s)} dB_s - \frac{1}{2} \int_0^t \frac{f^2(s)}{\sigma^2(s)} ds\right)$ est une martingale et donc, grâce au théorème de Girsanov, le processus \tilde{B}_t est un mouvement brownien sous $\tilde{\mathbb{P}}$, où $\tilde{\mathbb{P}}$ est une mesure de probabilité sur (Ω, \mathcal{A}) qui a comme densité par rapport à \mathbb{P} la fonction donnée par :

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} \Big|_{\mathcal{A}_{[0,1]}} = \exp\left(\int_0^1 \frac{f(t)}{\sigma(t)} dB_t - \frac{1}{2} \int_0^1 \frac{f^2(t)}{\sigma^2(t)} dt\right).$$

D'une part, on a :

$$dX_t = f(t)dt + \sigma(t)dB_t + dX_t,$$

de l'autre part :

$$dY_t = f(t)dt + \sigma(t)d\tilde{B}_t + dX_t,$$

avec B et \tilde{B} mouvements browniens respectivement sur $(\Omega, \mathcal{A}, \mathbb{P})$ et $(\Omega, \mathcal{A}, \tilde{\mathbb{P}})$.

En particulier, comme X est indépendant de B , la loi de Z sous \mathbb{P} est la même que la loi de Y sous $\tilde{\mathbb{P}}$.

Remarquons aussi que, par définition de Y , on peut écrire $dB_s = \frac{dY_s^c}{\sigma(s)}$. Donc, pour toute fonction $F : D \rightarrow \mathbb{R}$ mesurable bornée, on a

$$\begin{aligned} \int F(\omega) \mathcal{L}_Z(d\omega) &= \mathbb{E}_{\mathbb{P}}[F(Z)] = \mathbb{E}_{\tilde{\mathbb{P}}}[F(Y)] = \mathbb{E}_{\mathbb{P}} \left[F(Y) \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} \right] \\ &= \mathbb{E}_{\mathbb{P}} \left[F(Y) \exp \left(\int_0^1 \frac{f(t)}{\sigma(t)} dB_t - \frac{1}{2} \int_0^1 \frac{f^2(t)}{\sigma^2(t)} dt \right) \right] \\ &= \int F(\omega) \exp \left(\int_0^1 \frac{f(t)}{\sigma^2(t)} d\omega_t^c - \frac{1}{2} \int_0^1 \frac{f^2(t)}{\sigma^2(t)} dt \right) \xi(d\omega), \end{aligned}$$

ce qui conclut la preuve. □

Proposition 6.1.2. *Soit Y un processus stochastique défini par l'équation :*

$$dY_t = \mu(t)dt + \frac{\sigma(t)}{\sqrt{n}} dB_t + dX_t$$

et soit Z un processus stochastique défini par l'équation :

$$dZ_t = \nu(t)dt + \frac{\sigma(t)}{\sqrt{n}} dB_t + dX_t$$

avec μ, ν, σ tels que :

$$\int_0^1 \frac{\mu^2(s)}{\sigma^2(s)} ds < \infty, \quad \int_0^1 \frac{\nu^2(s)}{\sigma^2(s)} ds < \infty. \quad (6.5)$$

Alors

$$L_1(Y, Z) = 2 \left(1 - 2\phi \left(-\frac{D}{2} \right) \right), \quad (6.6)$$

où

$$D^2 = \int_I n \frac{(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt.$$

et $\phi(x) = \int_{-\infty}^x h(s)ds$ avec h densité d'une variable normale réduite.

En particulier, $L_1(Y, Z) = O(D)$.

Démonstration. Grâce à l'hypothèse (6.5) et au lemme (6.1.1), il existe une mesure ξ qui domine à la fois la loi de Y et la loi de Z . Plus précisément, les densités respectives, g_Y et g_Z , sont données par les formules :

$$\begin{aligned} g_Y(\omega) &= \exp \left(n \int_0^1 \frac{\mu(s)}{\sigma^2(s)} d\omega_s^c - \frac{n}{2} \int_0^1 \frac{\mu^2(s)}{\sigma^2(s)} ds \right). \\ g_Z(\omega) &= \exp \left(n \int_0^1 \frac{\nu(s)}{\sigma^2(s)} d\omega_s^c - \frac{n}{2} \int_0^1 \frac{\nu^2(s)}{\sigma^2(s)} ds \right). \end{aligned}$$

En particulier, la loi de Y admet une densité par rapport à la loi de Z donnée par :

$$\begin{aligned} g_{Y/Z}(\omega) &= \exp\left(n \int_0^1 \frac{\mu(s) - \nu(s)}{\sigma^2(s)} d\omega_s^c - \frac{n}{2} \int_0^1 \frac{\mu^2(s) - \nu^2(s)}{\sigma^2(s)} ds\right) \\ &= \exp\left(n \int_0^1 \frac{\mu(s) - \nu(s)}{\sigma^2(s)} (d\omega_s^c - \nu(s) ds) - \frac{n}{2} \int_0^1 \frac{(\mu(s) - \nu(s))^2}{\sigma^2(s)} ds\right). \end{aligned}$$

Finalement,

$$\begin{aligned} L_1(Y, Z) &= \int_0^1 \left| g_Y(\omega) - g_Z \right| \xi(d\omega) \\ &= \int_0^1 \left| \exp\left(n \int_0^1 \frac{\mu(t) - \nu(t)}{\sigma^2(t)} (d\omega_t^c - \nu(t) dt) - \frac{n}{2} \int_0^1 \frac{(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt\right) - 1 \right| g_Z(\omega) \xi(d\omega) \\ &= \mathbb{E}_{\mathbb{P}} \left| \exp\left(n \int_0^1 \frac{\mu(t) - \nu(t)}{\sigma^2(t)} (dZ_t^c - \nu(t) dt) - \frac{n}{2} \int_0^1 \frac{(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt\right) - 1 \right| \\ &= \mathbb{E}_{\mathbb{P}} \left| \exp\left(\sqrt{n} \int_0^1 \frac{\mu(t) - \nu(t)}{\sigma(t)} dB_t - \frac{n}{2} \int_0^1 \frac{(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt\right) - 1 \right|. \end{aligned}$$

Or, la variable aléatoire $\int_0^1 \frac{\sqrt{n}(\mu(t) - \nu(t))}{\sigma(t)} dB_t$ suit la loi $\mathcal{N}(0, \int_0^1 \frac{n(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt)$, donc, grâce à la proposition 2.1.3, on a

$$L_1(Y, Z) = 2 \left[1 - 2\phi\left(\frac{1}{2} \sqrt{\int_0^1 \frac{n(\mu(t) - \nu(t))^2}{\sigma^2(t)} dt}\right) \right],$$

ce qui conclut la preuve. \square

6.2 Équivalence asymptotique

Comme on l'a annoncé, l'objectif de cette section est d'énoncer un résultat d'équivalence asymptotique entre la suite de modèles statistiques de signal à trajectoires discontinues \mathcal{P}_n^Z définie par (6.2) et le modèle de régression habituelle. Pour cela redonnons explicitement la définition du modèle de régression qu'on va utiliser.

On notera \mathcal{P}_n^Y le modèle statistique de régression donné par :

$$\mathcal{P}_{n,f}^Y = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_{n,f}^Y \right) \quad (6.7)$$

où $P_{n,f}^Y$ est la loi du vecteur (Y_{n1}, \dots, Y_{nn}) ainsi défini :

$$Y_{ni} = f\left(\frac{i}{n}\right) + \sigma\left(\frac{i}{n}\right) \varepsilon_{ni}, \quad (6.8)$$

où :

- $(\varepsilon_{ni})_{1 \leq i \leq n}$ sont v.a. i.i.d. de loi normale centrée réduite;
- σ une fonction connue;
- f une fonction inconnue appartenant à la classe \mathcal{F} .

De plus, supposons satisfaites les hypothèses suivantes :

- $\sigma^2(\cdot) > 0$ est une fonction connue absolument continue sur $[0, 1]$ pour laquelle il existe une constante $C_1 < \infty$ telle que

$$\left| \frac{d}{dt} \ln \sigma(t) \right| \leq C_1, \quad t \in [0, 1] \quad (6.9)$$

- Il existe une constante $B < \infty$ telle que

$$\sup_{t \in [0, 1]} \{|f(t)| : f \in \mathcal{F}\} = B \quad (6.10)$$

- On définit une fonction en escalier de la façon suivante :

$$\bar{f}_n(t) = \begin{cases} f\left(\frac{i}{n}\right) & \text{si } \frac{i-1}{n} \leq t < \frac{i}{n}, \quad i = 1, \dots, n; \\ f(1) & \text{si } t = 1; \end{cases} \quad (6.11)$$

- On suppose que

$$\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}} n \int_0^1 (f(t) - \bar{f}_n(t))^2 dt = 0. \quad (6.12)$$

On peut finalement énoncer le résultat suivant :

Théorème 6.2.1. *Sous les hypothèses (6.9)-(6.12), les modèles statistiques (6.2), (6.7) sont asymptotiquement équivalentes au sens de Le Cam.*

Démonstration. La démonstration se déroule en suivant les lignes directrices de la preuve du théorème 4.1.2. En analogie avec celle preuve là, commençons par définir une suite de processus $\bar{Z}^{(n)} = (\bar{Z}_t^{(n)} : t \in [0, 1])$ qui satisfont les équations stochastiques :

$$d\bar{Z}_t^{(n)} = \bar{f}_n(t)dt + \frac{\sigma(t)}{\sqrt{n}}dB_t + dX_t. \quad (6.13)$$

Définissons ensuite le modèle statistique associé de la façon suivante :

$$\mathcal{P}_n^{\bar{Z}} = (D, \mathcal{D}, \{P_{n,f}^{\bar{Z}} : f \in \mathcal{F}\}) \quad (6.14)$$

où $P_{n,f}^{\bar{Z}}$ est la loi du processus $\bar{Z}^{(n)}$.

Comme dans le cas de la preuve du théorème 4.1.2, divisons la démonstrations en quatre étapes.

- **ÉTAPE 1** : Montrer que $\lim_{n \rightarrow \infty} \Delta(\mathcal{P}_n^Z, \mathcal{P}_n^{\bar{Z}}) = 0$.

Grâce au théorème 1.4.2, étant donné que les espaces des observations sont les mêmes pour les deux modèles, il suffit de montrer que $\lim_{n \rightarrow \infty} L_1(Z^{(n)}, \bar{Z}^{(n)}) = 0$.

Or, grâce à la proposition 6.1.2, on sait que

$$L_1(Z^{(n)}, \bar{Z}^{(n)}) = O(D)$$

en étant $D^2 = n \int_0^1 (f(t) - \bar{f}_n(t))^2 dt$.

Donc D^2 tend vers zéro quand n tend vers l'infini grâce à l'hypothèse (6.12), d'où la convergence vers zéro en distance L_1 des suites de processus $Z^{(n)}$, $\bar{Z}^{(n)}$.

- Introduisons une suite de variables aléatoires définies de la façon suivante

$$S_i^{(n)} = K_i \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{d\bar{Z}_t^c}{\sigma^2(t)} = \frac{K_i}{n} f\left(\frac{i}{n}\right) \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{dt}{\sigma^2(t)} + K_i \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{dB_t}{\sqrt{n}\sigma(t)}$$

Notons avec \mathcal{P}_n^S le modèle statistique associé au vecteur aléatoire $(S_i^{(n)})_{1 \leq i \leq n}$ défini comme :

$$\mathcal{P}_n^S = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{P_{n,f}^S : f \in \mathcal{F}\}),$$

en étant $P_{n,f}^S$ la loi du vecteur $(S_1^{(n)}, \dots, S_n^{(n)})$.

– **ÉTAPE 2** : Montrer que $\Delta(\mathcal{P}_n^{\bar{Z}}, \mathcal{P}_n^S) = 0$.

Cela découle immédiatement du lemme 1.1.8, puisque, pour tout n , $\{S_i^{(n)} : i = 1, \dots, n\}$ une statistique exhaustive pour $\bar{Z}^{(n)}$.

Pour voir qu'il s'agit bien d'une statistique exhaustive, il suffit de considérer la statistique $S : (D, \mathcal{D}) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ ainsi définie :

$$\forall \omega \in D \quad S(\omega) = (S_1(\omega), \dots, S_n(\omega)),$$

avec $S_i(\omega) = K_i \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{d\omega_t^c}{\sigma^2(t)}$.

Or, grâce au lemme 6.1.1, on peut écrire

$$\begin{aligned} \frac{dP_n^{\bar{Z}}}{d\xi}(\omega) &= \exp\left(n \int_0^1 \frac{\bar{f}_n(t)}{\sigma^2(t)} d\omega_t^c - \frac{n}{2} \int_0^1 \frac{\bar{f}_n^2(t)}{\sigma^2(t)} dt\right) \\ &= \exp\left(n \sum_{i=1}^n f\left(\frac{i}{n}\right) \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{d\omega_t^c}{\sigma^2(t)} - \frac{n}{2} \int_0^1 \frac{\bar{f}_n^2(t)}{\sigma^2(t)} dt\right) \\ &= \exp\left(n \sum_{i=1}^n f\left(\frac{i}{n}\right) \frac{S_i(\omega)}{K_i} - \frac{n}{2} \sum_{i=1}^n f^2\left(\frac{i}{n}\right) \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{dt}{\sigma^2(t)}\right), \end{aligned}$$

ce qui conclut, grâce au théorème de factorisation 1.1.2.

– **ÉTAPE 3** : Montrer que $\lim_{n \rightarrow \infty} \Delta(\mathcal{P}_n^S, \mathcal{P}_n^Y) = 0$.

On a déjà fait le calcul dans le chapitre 4. En effet les modèles \mathcal{P}_n^S et \mathcal{P}_n^Y coïncident respectivement avec les modèles (4.10) et (4.6). On conclut alors en utilisant l'ÉTAPE 3 de la preuve du théorème 4.1.2.

– **ÉTAPE 4** : Montrer que $\lim_{n \rightarrow \infty} \Delta(\mathcal{P}_n^Y, \mathcal{P}_n^{\bar{Z}}) = 0$.

Cela découle immédiatement des étapes précédentes et de l'inégalité triangulaire satisfaite par Δ (voir proposition 1.4.1). □

Remarquons que le théorème précédent comprend le cas où le modèle (6.2) est un modèle de signal perturbé par un processus de Lévy. En effet, si on suppose que le processus X est somme de deux processus de saut indépendants, $X^{(1)}$ (qui représente les sauts de grande amplitude) et $X^{(2)}$ (qui représente les sauts d'amplitude plus petite que 1) :

$$X_t = X_t^{(1)} + X_t^{(2)},$$

où

$$- X_t^{(1)} = \int_{[0,t] \times]1, +\infty[} x J(ds, dx),$$

$$- X_t^{(2)} = \lim_{\varepsilon \downarrow 0} \int_{[0,t] \times]\varepsilon, 1]} x [J(ds, dx) - \nu(dx)],$$

– ν est une mesure positive ne chargeant pas $\{0\}$ et telle que $\int (1 \wedge x^2) < \infty$,

– J est une mesure aléatoire de Poisson sur $\mathbb{R}^+ \times \mathbb{R}$ d'intensité ρ telle que $\rho([0, t] \times B) = t\nu(B)$, définie par :

$$J(D) = \#\{s : (s, X_s - X_{s-})\} \quad \forall D \in \mathcal{B}(\mathbb{R}^+ \times \mathbb{R}),$$

alors, grâce à la formule de décomposition de Lévy-Itô, le processus $L^{(n)}$:

$$L_t^{(n)} = \frac{\sigma(t)}{\sqrt{n}} dB_t + dX_t, \quad (6.15)$$

est un processus de Lévy de triplet inhomogène $(0, \frac{\sigma^2(\cdot)}{n}, \nu)$ (pour plus des détails sur le processus de Lévy à triplet inhomogène voir [10], partie IV).

Donc, si on note avec \mathcal{P}_n^L le modèle statistique associé au processus (6.15) :

$$\mathcal{P}_n^L = \left(D, \mathcal{D}, \left(P_{n,f}^L \right)_{f \in \mathcal{F}} \right), \quad (6.16)$$

comme corollaire immédiat du théorème 6.2.1, on obtient le résultat qui suit.

Corollaire 6.2.2. *Sous les hypothèses (6.9)-(6.12) les modèles statistiques (6.16) et (6.7), sont asymptotiquement équivalents au sens de Le Cam.*

6.3 Applications

Au cours de la dernière décennie, les processus de Lévy et d'autres processus stochastiques à sauts, sont devenus de plus en plus populaires.

Il existe plusieurs domaines d'applications ; citons par exemple : le contrôle stochastique, les séries temporelles à temps continu, les processus de Lévy fractionnels, l'études du risque dans les compagnies d'assurance, la fiabilité, la théorie de la fragmentation, les équations différentielles stochastiques perturbées par processus de Lévy et encore beaucoup d'autres (voir le livre de Kyprianou (2006) [26] et la bibliographie ci-jointe). Cependant, le domaine dans lequel la plus grande utilisation en a été faite est la finance mathématique. Les processus de Lévy sont entrés dans l'économétrie financière en 1963 lorsque Mandelbrot [29] proposa des processus de Lévy α -stable en tant que modèles pour le prix du coton, mais c'est seulement au cours des dernières années qu'ils ont été exploités pour leur flexibilité dans la modélisation de la distribution des rendements du sous-jacent, à un horizon de temps donné.

Actuellement on sait que le modèle de Black-Scholes [2], défini sous forme de rendement instantané par

$$\frac{dS_t}{S_t} = \left(\mu + \frac{\sigma^2}{2} \right) dt + \sigma dW_t, \quad (6.17)$$

ou sous forme intégrée par

$$S_t = S_0 \exp(\mu t + \sigma W_t),$$

est inconsistant du point de vue empirique. En effet, la modélisation des titres par un mouvement brownien géométrique ne permet pas de rendre compte de différents problèmes tels que la présence empirique des discontinuités dans la trajectoire du prix, le phénomène du smile (ou skew de volatilité implicite) ou encore le fait que les queues de la distribution empirique du log-prix sont lourdes. Une solution possible consiste alors à introduire des sauts dans la dynamique du prix, en considérant à la place de l'équation (6.17), l'équation suivante :

$$\frac{dS_t}{S_t} = \left(\mu + \frac{\sigma^2}{2} \right) dt + \sigma dL_t, \quad (6.18)$$

où $L = (L_t, t \in [0, 1])$ est un processus de Lévy.

D'une façon encore plus générale, on peut modéliser le log-prix par équations de la forme

$$dZ_t = f(t)dt + dL_t, \quad (6.19)$$

où L est un processus de Lévy de triplet inhomogène $(0, \sigma^2(\cdot), \nu)$.

Le corollaire 6.2.2 nous aide alors à résoudre, au moins partiellement, le problème de la calibration. En effet, comme on spécifiera dans le paragraphe qui suit, si on suppose connue la fonction de volatilité $\sigma(\cdot)$, grâce à l'équivalence asymptotique entre le modèle statistique associé à (6.19) et le modèle de régression habituelle, dès qu'on est capables de trouver un estimateur minimax dans le modèle de régression on est également capables à fabriquer un estimateur minimax pour le drift $f(\cdot)$ de (6.19).

6.3.1 Estimation du drift à partir du modèle de régression

Supposons que nous disposons de n observations issues du modèle de régression non-paramétrique \mathcal{P}_n^Y défini par :

$$Y_i = f\left(\frac{i}{n}\right) + \sigma\left(\frac{i}{n}\right)\varepsilon_i, \quad i = 1, \dots, n, \quad (\varepsilon_i)_{1 \leq i \leq n} \text{ i.i.d } \mathcal{N}(0, 1) \quad (6.20)$$

où la fonction de régression f est la seule inconnue et appartient à la classe fonctionnelle \mathcal{F} .

De plus, supposons que nous connaissons une suite $(T_n)_n$ d'estimateurs asymptotiquement minimax du modèle (6.20) pour la classe \mathcal{F} donnée.

Comment exploiter la connaissance de la suite $(T_n)_n$ pour construire une suite $(F_n)_n$ d'estimateurs asymptotiquement minimax (pour la même classe fonctionnelle \mathcal{F}) du modèle \mathcal{P}_n^Z défini par

$$dZ_t^{(n)} = f(t)dt + \frac{\sigma(t)}{\sqrt{n}}dB_t + dX_t, \quad t \in [0, 1]? \quad (6.21)$$

La réponse se situe dans le résultat suivant, corollaire du théorème 6.2.1.

Corollaire 6.3.1. *Soit $(T_n)_n$ une suite d'estimateurs asymptotiquement minimax de la fonction de régression f du modèle (6.20), pour la classe \mathcal{F} .*

Soit $F_n : (D, \mathcal{D}) \rightarrow \mathbb{R}$ une suite de fonction ainsi définie : pour tout ω fonction càdlàg, $\omega = \omega^c + \omega^d$,

$$F_n(\omega) := T_n(y_1(\omega), \dots, y_n(\omega)), \quad i = 1, \dots, n,$$

où

$$K_i^2 = \frac{n\sigma^2\left(\frac{i}{n}\right)}{\int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{dt}{\sigma^2(t)}}, \quad y_i(\omega) = K_i \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{d\omega_t^c}{\sigma^2(t)}, \quad i = 1, \dots, n, \quad \omega^c(t) = \omega(t) - \sum_{s \leq t} \left(\omega(s) - \lim_{r \uparrow s} \omega(r) \right).$$

Alors $(F_n(\mathbf{Z}))_n$, où \mathbf{Z} est une trajectoire du processus (6.21), est une suite d'estimateurs asymptotiquement minimax du drift dans la classe \mathcal{F} qui est asymptotiquement équivalente à $(T_n)_n$.

Démonstration. Grâce à la proposition 5.1.3 il suffit de démontrer que les suites $(T_n)_n$ et $(F_n)_n$ sont asymptotiquement équivalentes.

Comme d'habitude notons avec $\mathcal{P}_n^{\bar{Z}}$ le modèle statistique associé au processus

$$d\bar{Z}_t = \bar{f}_n(t)dt + \sigma(t)dB_t + dX_t,$$

et avec \mathcal{P}_n^S celui associé aux variables aléatoires S_i , $i = 1, \dots, n$, ainsi définies :

$$S_i = K_i \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{d\bar{Z}_t^c}{\sigma^2(t)}, \quad i = 1, \dots, n.$$

Il faut démontrer que, pour toute suite de fonctions de perte uniformément bornées, $(L_n)_n$, $L_n : D_n \rightarrow [0, \infty)$, on a :

$$\lim_{n \rightarrow \infty} \left| R(\mathcal{P}_n^Y, T_n, L_n, f) - R(\mathcal{P}_n^Z, F_n, L_n, f) \right| = 0,$$

uniformément en $f \in \mathcal{F}$.

Pour cela écrivons la chaîne suivante de définitions.

$$\begin{aligned} & \left| R(\mathcal{P}_n^Y, T_n, L_n, f) - R(\mathcal{P}_n^Z, F_n, L_n, f) \right| = \\ & \left| \int_{\mathbb{R}^n} L_n(f, T_n(y)) P_{n,f}^Y(dy) - \int_D L_n(f, F_n(\omega)) P_{n,f}^Z(d\omega) \right| = \\ & \left| \int_{\mathbb{R}^n} L_n(f, T_n(y)) P_{n,f}^Y(dy) - \int_D L_n(f, T_n(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^Z(d\omega) \right|. \end{aligned}$$

Or, pour pouvoir démontrer que cette dernière quantité tend vers zéro, utilisons la preuve du théorème 6.2.1 et notamment le fait que $\lim_{n \rightarrow \infty} L_1(\mathcal{P}_n^Z, \mathcal{P}_n^{\bar{Z}}) = 0$ et que $\lim_{n \rightarrow \infty} H(P_{n,f}^Y, P_{n,f}^S) = 0$.

Écrivons :

$$\begin{aligned} & \left| \int_{\mathbb{R}^n} L_n(f, T_n(y)) P_{n,f}^Y(dy) - \int_D L_n(f, T_n(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^Z(d\omega) \right| \\ & \leq \left| \int_{\mathbb{R}^n} L_n(f, T_n(y)) P_{n,f}^Y(dy) - \int_D L_n(f, T_n(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^{\bar{Z}}(d\omega) \right| \\ & \quad + \left| \int_D L_n(f, T_n(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^{\bar{Z}}(d\omega) - \int_D L_n(f, T_n(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^Z(d\omega) \right| \\ & = \left| \int_{\mathbb{R}^n} L_n(f, T_n(y)) P_{n,f}^Y(dy) - \int_{\mathbb{R}^n} L_n(f, T_n(y)) P_{n,f}^S(dy) \right| \\ & \quad + \left| \int_D L_n(f, T_n(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^{\bar{Z}}(d\omega) - \int_D L_n(f, T_n(y_1(\omega), \dots, y_n(\omega))) P_{n,f}^Z(d\omega) \right| \\ & \leq \sup_{\substack{f \in \mathcal{F} \\ z \in D^n}} |L_n(f, z)| \left[L_1(\mathcal{P}^Y, \mathcal{P}^S) + L_1(\mathcal{P}^{\bar{Z}}, \mathcal{P}^Z) \right] \\ & \leq \sup_{\substack{f \in \mathcal{F} \\ z \in D^n}} |L_n(f, z)| \left[H(P_{n,f}^Y, P_{n,f}^S) + L_1(\mathcal{P}^{\bar{Z}}, \mathcal{P}^Z) \right], \end{aligned}$$

où on a utilisé dans le dernier passage le fait que la distance L_1 est majorée par la distance d'Hellinger (propriété 1.4.3). \square

6.3.2 Estimation de la fonction de régression à partir du modèle de signal perturbé par un processus à sauts

Le problème de l'estimation de la fonction de régression est désormais classique et bien connu. Il n'est donc pas pratique de passer par le modèle de signal perturbé par un processus à sauts (qui n'est pas encore trop étudié du point de vue statistique) pour parvenir à un estimateur minimax dans le modèle de régression. Au moins au niveau théorique, on peut énoncer le résultat suivant, corollaire du théorème 6.2.1.

Corollaire 6.3.2. *En conformité aux notations du paragraphe précédent, si $(F_n)_n$ est une suite d'estimateurs randomisés asymptotiquement minimax dans le modèle à saut \mathcal{P}_n^Z , alors la suite d'estimateurs randomisés $(T_n)_n$ ainsi définie : pour tout $A \in \mathcal{D}_n$,*

$$T_n((y_1, \dots, y_n), A) := \mathbb{E} \left[F_n(\bar{Z}^{(n)}, A) \middle| K_i \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{d\bar{Z}_t^c}{\sigma^2(t)} = y_i, i = 1, \dots, n \right]$$

est asymptotiquement minimax dans le modèles de régression \mathcal{P}_n^Y .

Démonstration. Grâce à la proposition 5.1.3 il suffit de démontrer que les suites $(T_n)_n, (F_n)_n$ sont asymptotiquement équivalentes.

On suit alors exactement la même preuve que celle de la deuxième partie du corollaire 5.1.1, à la seule différence qu'ici il faut définir les variables aléatoires $(S_i)_{1 \leq i \leq n}$ de la manière suivante :

$$S_i = K_i \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{d\bar{Z}_t^c}{\sigma^2(t)}, \quad i = 1, \dots, n,$$

pour pouvoir utiliser l'ÉTAPE 3 du théorème 6.2.1. □

Chapitre 7

Simulations

L'objectif de cette partie est de construire des estimateurs de la fonction de drift f dans le modèle de signal perturbé par un processus à sauts :

$$dZ_t = f(t)dt + \frac{\sigma(t)}{\sqrt{n}}dB_t + dX_t. \quad (7.1)$$

Dans le chapitre 6 (voir théorème 6.2.1) nous avons démontré que le modèle ci-dessus est équivalent, au sens de Le Cam, au modèle de régression non-paramétrique :

$$Y_i = f\left(\frac{i}{n}\right) + \sigma\left(\frac{i}{n}\right)\varepsilon_i, \quad i = 1, \dots, n, \quad \varepsilon_1, \dots, \varepsilon_n \text{ i.i.d} \sim \mathcal{N}(0, 1). \quad (7.2)$$

Par conséquent, à partir de l'observation d'une trajectoire $t \rightarrow Z_t$ issue du modèle (7.1), on peut construire un estimateur du drift f en fabriquant un estimateur de la fonction de régression f dans le modèle de régression (7.2).

Plus précisément, on considérera, comme estimateurs de la fonction de régression, l'estimateur de Nadaraya-Watson et l'estimateur par projection construit en prenant la base trigonométrique de $L^2[0, 1]$.

Pour des raisons de simplicité on supposera $\sigma(\cdot) = \sigma$ constant.

7.1 Estimateur de Nadaraya-Watson

On dispose d'un échantillon composé de n variables aléatoires indépendantes Y_1, \dots, Y_n issues du modèle de régression non paramétrique à design déterministes donné, pour $i = 1, \dots, n$, par

$$Y_i = f\left(\frac{i}{n}\right) + \sigma\varepsilon_i.$$

Une estimateur classique de la fonction inconnue f est l'estimateur non-paramétrique suivant, dit estimateur de Nadaraya-Watson et noté **NW**. (Pour plus de détails voir [32]).

Il est construit à partir d'une fonction noyau K (fonction intégrable, d'intégrale égale à 1) et d'une fenêtre h_n . Plus précisément, l'estimateur **NW** s'écrit sous la forme d'une moyenne pondérée des valeurs (Y_1, \dots, Y_n) :

$$f_n^{NW}(x) = \frac{1}{nh_n} \frac{\sum_{i=1}^n Y_i K\left(\frac{\frac{i}{n} - x}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\frac{i}{n} - x}{h_n}\right)} \mathbb{I}\left(\sum_{i=1}^n K\left(\frac{\frac{i}{n} - x}{h_n}\right) \neq 0\right).$$

Dans les figures 7.1, 7.2, 7.3 on a tracé l'estimateur de Nadaraya-Watson relatif, respectivement, aux fonctions :

$$\begin{aligned} f_1(x) &= \sin\left(4\pi\left(x - \frac{1}{2}\right)\right)\left|x - \frac{1}{2}\right|, \\ f_2(x) &= \frac{1}{4}\sin(4\pi x) + \frac{1}{3}\cos(4\pi x) - \frac{1}{2}\sin(2\pi x), \\ f_3(x) &= \exp\left(-\left(\frac{1/2 - x}{x(x-1)}\right)^2\right) \end{aligned}$$

en choisissant la fenêtre h_n de façon à minimiser le risque quadratique moyen.

Remarquons que les fonctions ci-dessus proposées appartiennent à l'espace $\mathcal{F}_{1,10}^{(2)}$ introduit dans le paragraphe 3.4 et donc, en particulier, ils satisfont les hypothèses des théorèmes 4.1.2 et 6.2.1.

(Pour le code, voir l'annexe C)

7.1.1 Implémentation de l'estimateur du drift à partir de l'estimateur de Nadaraya-Watson

Supposons avoir à disposition une trajectoire $\omega : t \rightarrow Z_t$ tirée selon la loi du processus :

$$dZ_t = f(t)dt + \frac{\sigma}{\sqrt{n}}dB_t + dX_t,$$

et l'estimateur de Nadaraya-Watson qu'on a défini dans le paragraphe précédent.

Grâce au corollaire 6.3.1, ces données nous suffisent pour construire un estimateur de f asymptotiquement équivalent à l'estimateur de Nadaraya-Watson.

En effet, l'estimateur de f à proposer n'est rien d'autre que :

$$F_n(\omega)(x) := \frac{1}{h_n} \frac{\sum_{i=1}^n \left(\left(\omega^c\left(\frac{i}{n}\right) - \omega^c\left(\frac{i-1}{n}\right) \right) K\left(\frac{(i-1)/n - x}{h_n}\right) \right)}{\sum_{i=1}^n K\left(\frac{\frac{i}{n} - x}{h_n}\right)} \mathbb{I}\left(\sum_{i=1}^n K\left(\frac{\frac{i}{n} - x}{h_n}\right) \neq 0\right),$$

où ω^c est la partie continue de la trajectoire càdlàg ω , c'est-à-dire, $\omega^c(t) := \omega(t) - \sum_{s \leq t} \left(\omega(s) - \lim_{r \uparrow s} \omega(r) \right)$.

En analogie avec ce qu'on a fait dans le paragraphe précédent, on trouvera (voir figures 7.4, 7.5, 7.6) l'estimateur F_n du drift f correspondant, respectivement, aux fonctions :

$$\begin{aligned} f_1(x) &= \sin\left(4\pi\left(x - \frac{1}{2}\right)\right)\left|x - \frac{1}{2}\right|, \\ f_2(x) &= \frac{1}{4}\sin(4\pi x) + \frac{1}{3}\cos(4\pi x) - \frac{1}{2}\sin(2\pi x), \\ f_3(x) &= \exp\left(-\left(\frac{1/2 - x}{x(x-1)}\right)^2\right). \end{aligned}$$

(Pour le code, implémenté en R, voir l'annexe C)

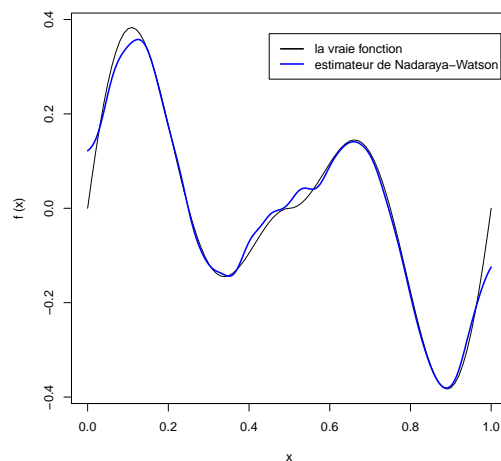


FIGURE 7.1 – Superposition du graphe de la fonction $f_1(x) = \sin\left(4\pi\left(x - \frac{1}{2}\right)\right)\left|x - \frac{1}{2}\right|$ et de l'estimateur de Nadaraya-Watson correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.
En **noir** : la vraie fonction, en **bleu** : l'estimateur de Nadaraya-Watson

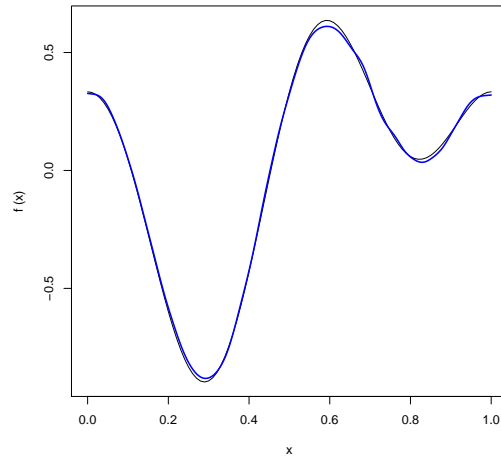


FIGURE 7.2 – Superposition du graphe de la fonction $f_2(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$ et de l'estimateur de Nadaraya-Watson correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.
En **noir** : la vraie fonction, en **bleu** : l'estimateur de Nadaraya-Watson

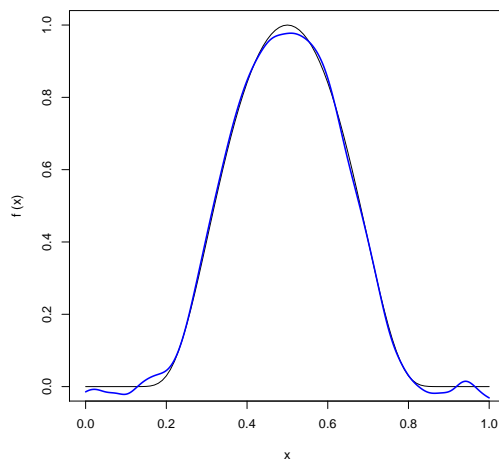


FIGURE 7.3 – Superposition du graphe de la fonction $f_3(x) = \exp\left(-\left(\frac{1/2-x}{x(x-1)}\right)^2\right)$ et de l'estimateur de Nadaraya-Watson correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.

En **noir** : la vraie fonction, en **bleu** : l'estimateur de Nadaraya-Watson

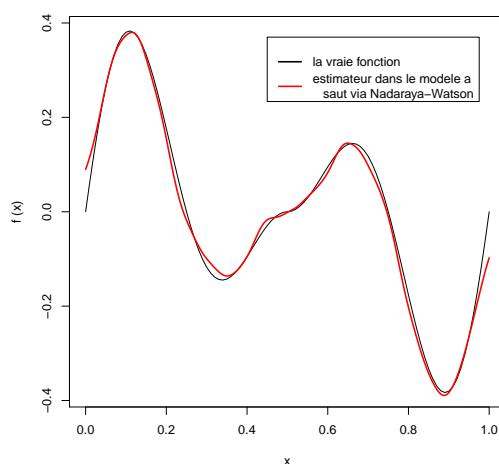


FIGURE 7.4 – Superposition du graphe de la fonction $f_1(x) = \sin\left(4\pi\left(x - \frac{1}{2}\right)\right)\left|x - \frac{1}{2}\right|$ et de l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur de Nadaraya-Watson de f_1 dans le modèle de régression. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.

En **noir** : la vraie fonction, en **rouge** : l'estimateur dans le modèle à saut via l'estimateur de Nadaraya-Watson

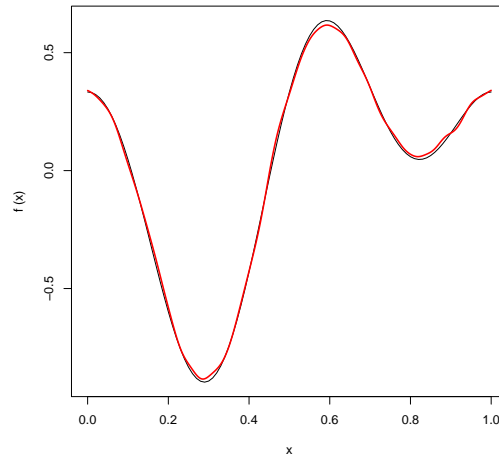


FIGURE 7.5 – Superposition du graphe de la fonction $f_2(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$ et de l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur de Nadaraya-Watson de f_2 dans le modèle de régression. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$. En **noir** : la vraie fonction, en **rouge** : l'estimateur dans le modèle à saut via l'estimateur de Nadaraya-Watson

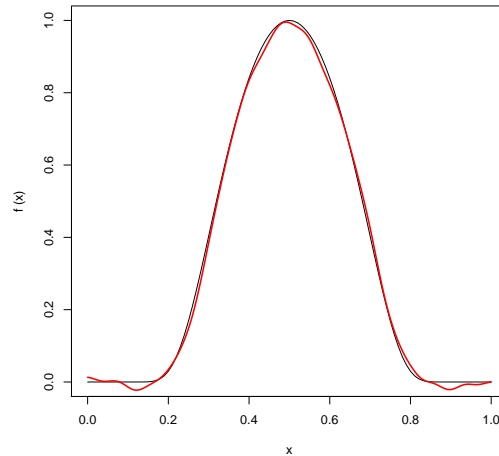


FIGURE 7.6 – Superposition du graphe de la fonction $f_3(x) = \exp\left(-\left(\frac{1/2-x}{x(x-1)}\right)^2\right)$ et de l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur de Nadaraya-Watson de f_3 dans le modèle de régression. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.

En **noir** : la vraie fonction, en **rouge** : l'estimateur dans le modèle à saut via l'estimateur de Nadaraya-Watson

7.2 Estimateurs par projection

Soit $(\phi_j)_j$ une base orthonormale donnée de $L_2[0, 1]$.

On suppose ici que la fonction de régression f appartient à la classe de *Sobolev périodique* suivante :

$$W_{\beta, L}^{\text{per}} = \{f : [0, 1] \rightarrow \mathbb{R}, f \in \mathcal{C}^{(\beta)}([0, 1]), \|f^{(\beta)}\|_2 \leq L, f^{(j)}(0) = f^{(j)}(1) \quad \forall j = 0, \dots, \beta - 1\}.$$

Donc, en particulier, f admet une écriture de la forme :

$$f(x) = \sum_{j=1}^{\infty} \theta_j \phi_j(x), \quad \sum_{j=1}^{\infty} \theta_j^2 < \infty,$$

où

$$\theta_j = \int_0^1 f(x) \phi_j(x) dx.$$

Pour estimer f , il suffit alors de fournir une estimation des coefficients de Fourier θ_j , que l'on construit à partir des Y_i .

L'idée est la suivante : on prend l'approximation de f par sa projection sur ϕ_1, \dots, ϕ_N et on remplace les $(\theta_j)_{1 \leq j \leq N}$ par leurs estimateurs.

Or, comme

$$\frac{1}{n} \sum_{i=1}^n f\left(\frac{i}{n}\right) \phi_j\left(\frac{i}{n}\right) \xrightarrow{n \rightarrow \infty} \int_0^1 f(x) \phi_j(x) dx$$

l'approche naturelle à suivre est de remplacer les inconnues $f\left(\frac{i}{n}\right)$ par les observations Y_i , vu que l'espérance de Y_i est bien $f\left(\frac{i}{n}\right)$ pour tout $i = 1, \dots, n$.

On obtient les estimateurs $\hat{\theta}_j$ des θ_j :

$$\hat{\theta}_j = \frac{1}{n} \sum_{i=1}^n Y_i \phi_j\left(\frac{i}{n}\right)$$

et donc on peut finalement définir la statistique

$$\hat{f}_{nN}(x) = \sum_{j=1}^N \hat{\theta}_j \phi_j(x)$$

qui est dite *estimateur par projection de f* .

Naturellement N joue un rôle important, il s'agit en effet d'un paramètre de lissage dont le choix doit être fait de façon à établir l'équilibre entre le biais et la variance.

Si on se concentre sur le risque quadratique de l'estimateur \hat{f}_{nN} défini par

$$R_n(\hat{f}_{nN}) = \mathbb{E} \left[\int_0^1 (\hat{f}_{nN}(t) - f(t))^2 dt \right] = \mathbb{E} \left[\sum_{j=1}^N (\hat{\theta}_j - \theta_j)^2 \right] + \sum_{j=N+1}^{\infty} \theta_j^2,$$

en faisant les calculs, on trouve que (voir [38]) :

$$R_n(\hat{f}_{nN}) = O\left(\frac{N}{n} + N^{-2\beta}\right).$$

Un estimateur \hat{f}_{nN} qui minimise le risque quadratique s'obtient alors en prenant $N = N_{\beta, n}$, où $N_{\beta, n}$ est la partie entière de $n^{\frac{1}{2\beta+1}}$.

On peut aussi démontrer les résultats suivants :

Théorème 7.2.1. *Pour tout $\beta \in \mathbb{N}^*$ et pour tout $L > 0$,*

$$\limsup_{n \rightarrow \infty} \sup_{f \in W_{\beta, L}^{\text{per}}} n^{\frac{2\beta}{2\beta+1}} R_n(\hat{f}_{nN}) \leq C_{\beta, L, \sigma}$$

où $C_{\beta, L, \sigma}$ est une constante.

Théorème 7.2.2. *Pour tout $\beta \in \mathbb{N}^*$ et pour tout $L > 0$,*

$$\liminf_{n \rightarrow \infty} \inf_{T_n} \sup_{f \in W_{\beta, L}^{\text{per}}} n^{\frac{2\beta}{2\beta+1}} R_n(T_n) \geq D_{\beta, L, \sigma},$$

où l'infimum est pris sur toutes les procédures d'estimation $(T_n)_n$ et $D_{\beta, L, \sigma}$ désigne une constante indépendante de n .

Les résultats ci-dessus nous disent alors que l'estimateur de f dans le modèle discret \hat{f}_{nN} est minimax (en prenant comme fonction de perte la norme L_2) dans la classe fonctionnelle $W_{\beta, L}^{\text{per}}$ avec une vitesse de convergence égale à $n^{-\frac{2\beta}{2\beta+1}}$.

(Pour plus de détails sur les estimateurs par projection voir [38].)

7.2.1 Implémentation de l'estimateur du drift à partir de l'estimateur par projection : base trigonométrique

Soit, pour tout $x \in [0, 1]$,

$$\begin{aligned} \phi_1(x) &= 1 \\ \phi_{2j}(x) &= \sqrt{2} \cos(2\pi jx) \\ \phi_{2j+1}(x) &= \sqrt{2} \sin(2\pi jx), \end{aligned}$$

la base trigonométrique de $L_2[0, 1]$.

Par analogie avec les paragraphes précédents, (voir figures 7.7, 7.8, 7.9) on a tracé les estimateurs par projection, en ayant pris comme base la base trigonométrique et comme fonctions les éléments de la classe de Sobolev périodique $W_{1,10}^{\text{per}}$ suivants :

$$\begin{aligned} f_1(x) &= \sin\left(4\pi\left(x - \frac{1}{2}\right)\right) \left|x - \frac{1}{2}\right|, \\ f_2(x) &= \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x), \\ f_3(x) &= \exp\left(-\left(\frac{1/2 - x}{x(x-1)}\right)^2\right). \end{aligned}$$

(Pour le code, implémenté en R, voir l'annexe C.)

Encore une fois, supposons avoir à disposition une trajectoire $\omega : t \rightarrow Z_t$ tirée selon la loi du processus :

$$dZ_t = f(t)dt + \frac{\sigma}{\sqrt{n}} dB_t + dX_t,$$

et l'estimateur par projection sur la base trigonométrique qu'on a défini dans le paragraphe précédent.

Grâce au corollaire 6.3.1, ces données nous suffisent pour construire un estimateur de f asymptotiquement équivalent à l'estimateur par projection.

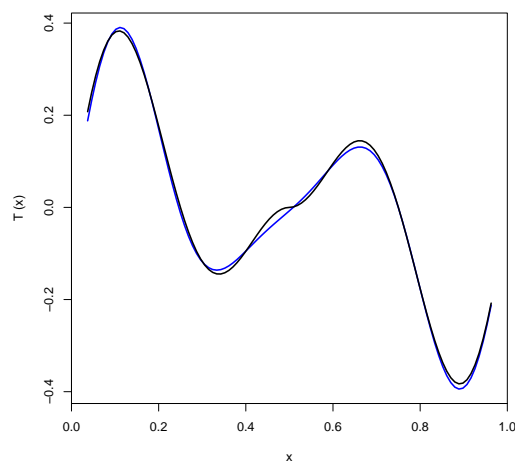


FIGURE 7.7 – Superposition du graphe de la fonction $f_1(x) = \sin\left(4\pi\left(x - \frac{1}{2}\right)\right)\left|x - \frac{1}{2}\right|$ et de l'estimateur par projection (base trigonométrique) correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$. En **noir** : la vraie fonction, en **bleu** : l'estimateur par projection dans le modèle de régression

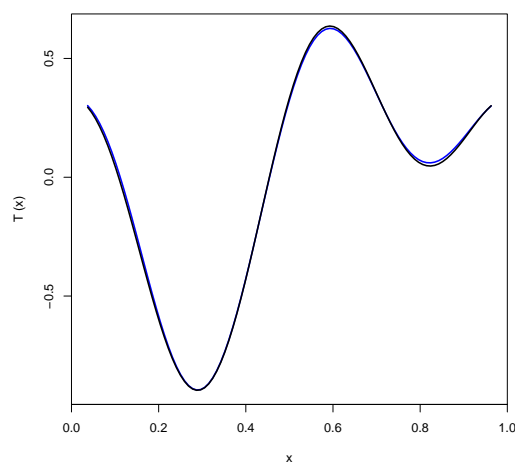


FIGURE 7.8 – Superposition du graphe de la fonction $f_2(x) = \frac{1}{4}\sin(4\pi x) + \frac{1}{3}\cos(4\pi x) - \frac{1}{2}\sin(2\pi x)$ et de l'estimateur par projection (base trigonométrique) correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.

En **noir** : la vraie fonction, en **bleu** : l'estimateur par projection dans le modèle de régression

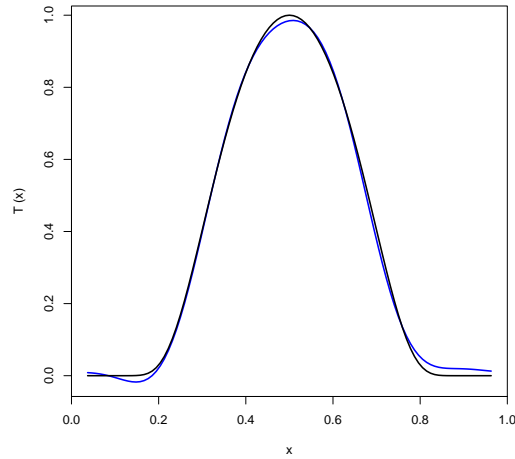


FIGURE 7.9 – Superposition du graphe de la fonction $f_3(x) = \exp\left(-\left(\frac{1/2-x}{x(x-1)}\right)^2\right)$ et de l'estimateur par projection (base trigonométrique) correspondant. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$. En **noir** : la vraie fonction, en **bleu** : l'estimateur par projection dans le modèle de régression

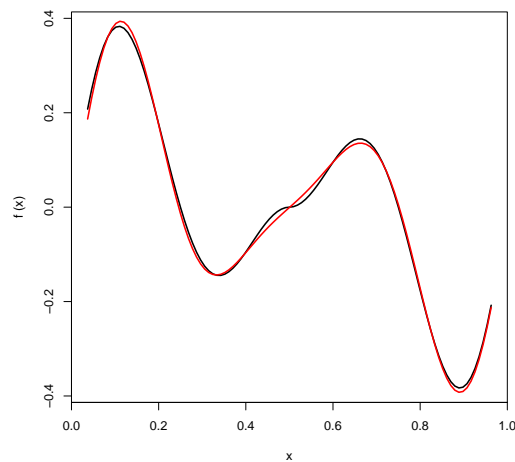


FIGURE 7.10 – Superposition du graphe de la fonction $f_1(x) = \sin\left(4\pi\left(x - \frac{1}{2}\right)\right)\left|x - \frac{1}{2}\right|$ et de l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur par projection (base trigonométrique) de f_1 dans le modèle de régression. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$. En **noir** : la vraie fonction, en **rouge** : l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection

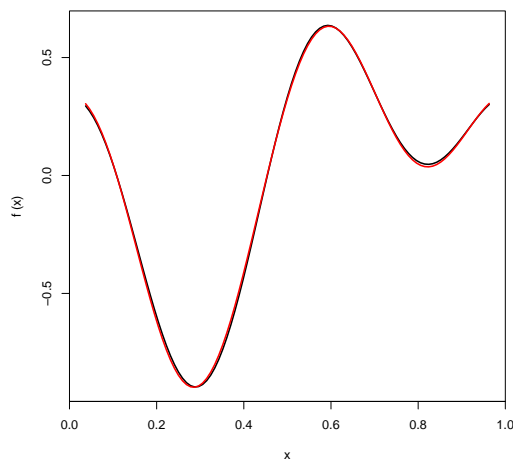


FIGURE 7.11 – Superposition du graphe de la fonction $f_2(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$ et de l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur par projection (base trigonométrique) de f_2 dans le modèle de régression. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.

En **noir** : la vraie fonction, en **rouge** : l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection

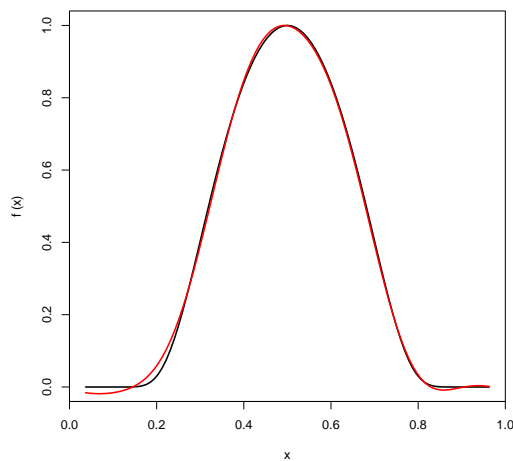


FIGURE 7.12 – Superposition du graphe de la fonction $f_3(x) = \exp\left(-\left(\frac{1/2-x}{x(x-1)}\right)^2\right)$ et de l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur par projection (base trigonométrique) de f_3 dans le modèle de régression. Simulation faite en posant $n = 1000$ et $\sigma = 0.1$.

En **noir** : la vraie fonction, en **rouge** : l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection

Plus en détails, l'estimateur à proposer n'est rien d'autre que :

$$F_n(\omega)(x) := \sum_{j=1}^N \sum_{i=1}^n \left(\omega^c\left(\frac{i}{n}\right) - \omega^c\left(\frac{i-1}{n}\right) \right) \phi_j\left(\frac{i}{n}\right) \phi_j(x).$$

Par analogie avec ce qu'on a fait dans le paragraphe précédent, on a tracé (voir figures 7.10, 7.11, 7.12) l'estimateur F_n du drift f construit à partir de l'estimateur par projection sur la base trigonométrique T_n du modèle de régression.

7.3 Comparaisons

Pour faciliter une interprétation graphique des résultats des paragraphes précédents, on a tracé (voir figures 7.13, 7.14, 7.15, 7.16, 7.17 et 7.18) les graphes des différentes fonctions f à estimer, auxquels on a superposé les graphes des estimateurs correspondants dans le modèle de régression et dans le modèle de signal perturbé par un processus à sauts.

De plus, remarquons qu'en combinant le résultat de Brown et Low (voir le théorème 4.1.2) et l'équivalence asymptotique démontrée dans le chapitre 6 entre le modèle de régression non paramétrique et le modèle de signal perturbé par un processus à sauts (voir le théorème 6.2.1), on en déduit que, du point de vue de l'estimation du drift f , le modèle de signal avec bruit blanc :

$$dZ_t = f(t)dt + \frac{\sigma(t)}{\sqrt{n}}dB_t, \quad t \in [0, 1] \quad (7.3)$$

est asymptotiquement équivalent au modèle de signal perturbé par un processus à sauts :

$$dY_t = f(t)dt + \frac{\sigma(t)}{\sqrt{n}}dB_t + dX_t, \quad t \in [0, 1]. \quad (7.4)$$

Une question naturelle à se poser est alors la suivante : si on observe une trajectoire issue du modèle (7.4) et si on souhaite estimer la fonction de drift f , quelle est la stratégie la plus convenable à suivre ? Est-il préférable d'utiliser un estimateur qui provient du modèle de régression ou plutôt du modèle de signal avec bruit blanc ?

Les expériences n'ont pas montré une réponse claire. En effet, nous avons relevé des cas où l'estimateur de Pinsker est préférable (au niveau du risque) à l'estimateur de f dans le modèle (7.4) construit à partir de l'estimateur par projection du modèle de régression et vice versa.

Dans les tableaux 7.1, 7.2 et 7.3 on a catalogué les risques moyens associés aux estimateurs suivants : dans la première colonne, il s'agit des résultats relatifs à l'estimateur par projection sur la base trigonométrique du modèle de régression ; dans la deuxième, les résultats concernent l'estimateur dans le modèle de signal perturbé par un processus à sauts obtenu grâce au corollaire 6.3.1 via l'estimateur par projection dans le modèle de régression ; dans la troisième on trouve, dans le modèle de bruit blanc, ceux de l'estimateur de Pinsker. Finalement, on rapporte dans la dernière colonne la valeur théorique asymptotique prévue par le théorème de Pinsker 5.1.5. Dans tous les cas considérés, on a calculé le risque par rapport à la fonction de perte donnée par la distance $L_2([0, 1])$ en utilisant une méthode Monte Carlo.

Les résultats confirment que, parmi les estimateurs proposés, il n'y a pas un qui donne toujours les meilleures estimations (en termes du risque) : même si des tendances émergent, les risques moyens sont toujours très proches entre eux. Par contre, en règle générale, on trouve une différence sensible entre les

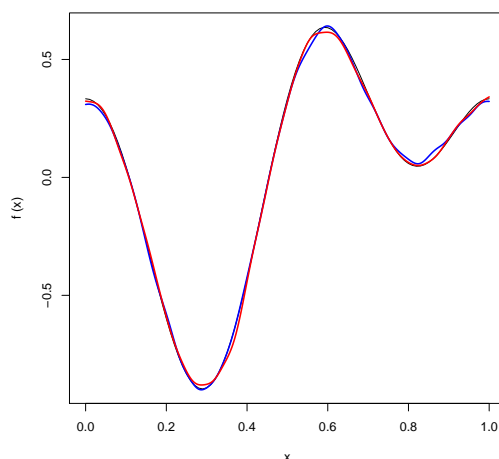


FIGURE 7.13 – Superposition du graphe de la fonction $f_1(x) = \sin\left(4\pi\left(x - \frac{1}{2}\right)\right)\left|x - \frac{1}{2}\right|$, de l'estimateur de Nadaraya-Watson de f_1 et de l'estimateur de f_1 dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur de Nadaraya-Watson de f_1 dans le modèle de régression. En **noir** : la vraie fonction, en **bleu** : l'estimateur de Nadaraya-Watson dans le modèle de régression, en **rouge** l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur de Nadaraya-Watson.

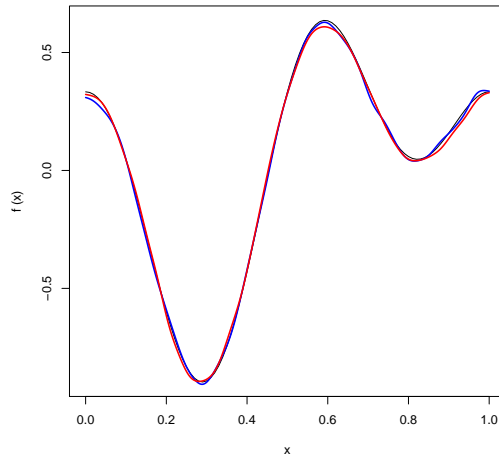


FIGURE 7.14 – Superposition du graphe de la fonction $f_2(x) = \frac{1}{4}\sin(4\pi x) + \frac{1}{3}\cos(4\pi x) - \frac{1}{2}\sin(2\pi x)$, de l'estimateur de Nadaraya-Watson de f_2 et de l'estimateur de f_2 dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur de Nadaraya-Watson de f_2 dans le modèle de régression. En **noir** : la vraie fonction, en **bleu** : l'estimateur de Nadaraya-Watson dans le modèle de régression, en **rouge** l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur de Nadaraya-Watson.

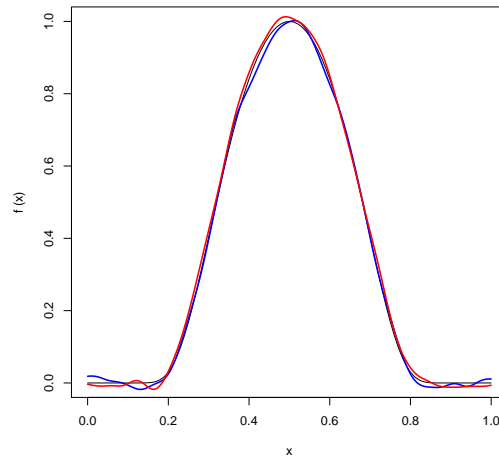


FIGURE 7.15 – Superposition du graphe de la fonction $f_3(x) = \exp\left(-\left(\frac{1/2-x}{x(x-1)}\right)^2\right)$, de l'estimateur de Nadaraya-Watson de f_3 et de l'estimateur de f_3 dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur de Nadaraya-Watson de f_3 dans le modèle de régression.

En **noir** : la vraie fonction, en **bleu** : l'estimateur de Nadaraya-Watson dans le modèle de régression, en **rouge** l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur de Nadaraya-Watson.

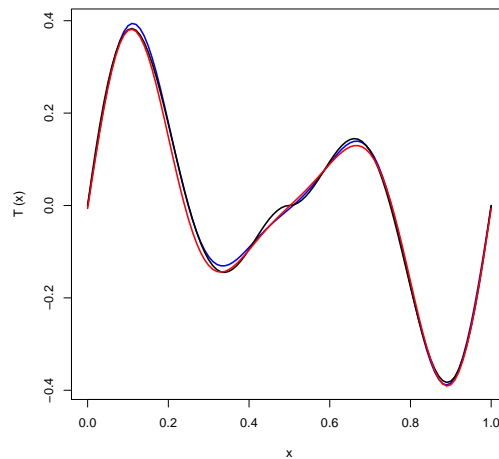


FIGURE 7.16 – Superposition du graphe de la fonction $f_1(x) = \sin\left(4\pi\left(x - \frac{1}{2}\right)\right)\left|x - \frac{1}{2}\right|$, de l'estimateur par projection (base trigonométrique) de f_1 et de l'estimateur de f_1 dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur par projection de f_1 dans le modèle de régression.

En **noir** : la vraie fonction, en **bleu** : l'estimateur par projection dans le modèle de régression, en **rouge** l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection.

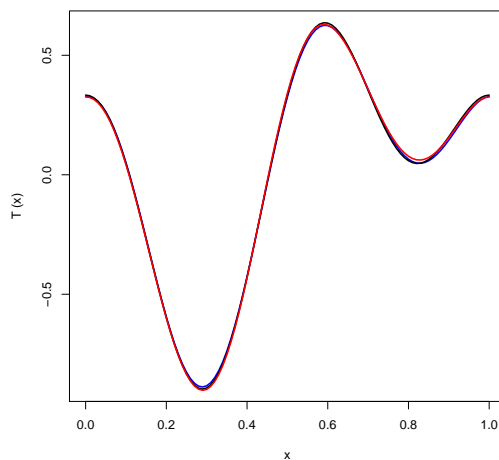


FIGURE 7.17 – Superposition du graphe de la fonction $f_2(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$, de l'estimateur par projection (base trigonométrique) de f_2 et de l'estimateur de f_2 dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur par projection de f_2 dans le modèle de régression.

En **noir** : la vraie fonction, en **bleu** : l'estimateur par projection dans le modèle de régression, en **rouge** l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection.

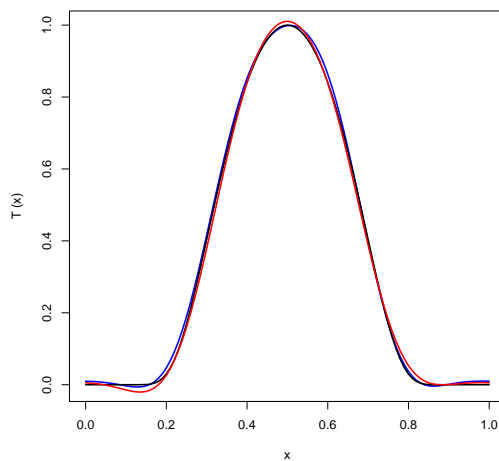


FIGURE 7.18 – Superposition du graphe de la fonction $f_3(x) = \exp\left(-\left(\frac{1/2-x}{x(x-1)}\right)^2\right)$, de l'estimateur par projection (base trigonométrique) de f_3 et de l'estimateur de f_3 dans le modèle de signal perturbé par un processus à sauts obtenu à partir de l'estimateur par projection de f_3 dans le modèle de régression.

En **noir** : la vraie fonction, en **bleu** : l'estimateur par projection dans le modèle de régression, en **rouge** l'estimateur dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection.

résultats donnés par les estimateurs et les valeurs idéales prévues par le théorème de Pinsker, toujours d'un facteur 10 environ. Il y a deux explications possibles pour un tel phénomène : d'un côté, le théorème de Pinsker donne une valeur asymptotique, qui ne pourrait être atteinte que pour des valeurs de n extrêmement grandes par rapport au bruit σ . De l'autre côté, les différences entre la fonction à estimer et l'estimateur sont très petites dans chaque point, fait qui rend les calculs particulièrement sensibles aux erreurs numériques des logiciels de calcul. C'est exactement pour cette raison qu'on n'a pas pu faire des simulations pour des petites valeurs de $\varepsilon = \frac{\sigma}{n}$.

	Régression	Est. transporté M.à S.	Pinsker	Valeur théorique
$n = 100, \sigma = 0.1$	0.1909541	0.1912585	0.05314375	0.004235654
$n = 500, \sigma = 0.1$	0.01431202	0.01265321	0.03187722	0.001448573
$n = 500, \sigma = 0.5$	0.06127406	0.05791142	0.09397223	0.01238513
$n = 1000, \sigma = 0.5$	0.04594394	0.04855009	0.07242672	0.007802142
$n = 5000, \sigma = 0.1$	0.005464376	0.005873015	0.013926	0.0003120857
$n = 5000, \sigma = 1$	0.05266939	0.05562003	0.06813567	0.006723682

TABLE 7.1 – Comparaison des risques moyens associés, respectivement, à l'estimateur par projection de f , à l'estimateur de f dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection, l'estimateur de Pinsker de f et la valeur théorique du risque associé à l'estimateur de Pinsker. On a posé $f(x) = \sin(4\pi(x - 0.5))|x - 0.5|$

	Régression	Est. transporté M.à S.	Pinsker	Valeur théorique
$n = 100, \sigma = 0.1$	0.02977428	0.03081126	0.05372695	0.004235654
$n = 500, \sigma = 0.1$	0.01944834	0.019056	0.03172858	0.001448573
$n = 500, \sigma = 0.5$	0.05690559	0.06568017	0.09753825	0.01238513
$n = 1000, \sigma = 0.5$	0.04500288	0.04852682	0.06976673	0.007802142
$n = 5000, \sigma = 0.1$	0.005602151	0.006408627	0.01488441	0.0003120857
$n = 5000, \sigma = 1$	0.05620099	0.05670829	0.06881296	0.006723682

TABLE 7.2 – Comparaison des risques moyens associés, respectivement, à l'estimateur par projection de f , à l'estimateur de f dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection, l'estimateur de Pinsker de f et la valeur théorique du risque associé à l'estimateur de Pinsker. On a posé $f(x) = e^{-\left(\frac{x-0.5}{x(x-1)}\right)^2}$

	Régression	Est. transporté M.à S.	Pinsker	Valeur théorique
$n = 100, \sigma = 0.1$	0.1780349	0.1784109	0.05407487	0.004235654
$n = 500, \sigma = 0.1$	0.01052737	0.0112597	0.03370397	0.001448573
$n = 500, \sigma = 0.5$	0.05636669	0.05717137	0.09320534	0.01238513
$n = 1000, \sigma = 0.5$	0.04294714	0.04463352	0.07612717	0.007802142
$n = 5000, \sigma = 0.1$	0.005850944	0.004763002	0.0148907	0.0003120857
$n = 5000, \sigma = 1$	0.05189352	0.05633176	0.06905875	0.006723682

TABLE 7.3 – Comparaison des risques moyens associés, respectivement, à l'estimateur par projection de f , à l'estimateur de f dans le modèle de signal perturbé par un processus à sauts via l'estimateur par projection, l'estimateur de Pinsker de f et la valeur théorique du risque associé à l'estimateur de Pinsker. On a posé $f(x) = \frac{1}{4} \sin(4\pi x) + \frac{1}{3} \cos(4\pi x) - \frac{1}{2} \sin(2\pi x)$

Chapitre 8

Conclusion et perspectives

Dans ce mémoire nous avons étudié l'article de Brown et Low [4] portant sur l'équivalence asymptotique au sens de Le Cam entre le modèle de signal avec bruit blanc :

$$dZ(t) = f(t)dt + \frac{\sigma(t)}{\sqrt{n}}dB_t, \quad t \in [0, 1] \quad (8.1)$$

et le modèle de régression non-paramétrique à variance connue :

$$Y_i = f(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, \dots, n \quad \varepsilon_i \sim \mathcal{N}(0, 1) \quad \text{i.i.d.} \quad (8.2)$$

Nous avons généralisé le résultat précédent : nous avons montré une équivalence asymptotique entre le modèle de signal perturbé par un processus de Lévy L de triplet inhomogène $(0, \frac{\sigma(\cdot)}{\sqrt{n}}, \nu)$, $\sigma(\cdot) > 0$:

$$dX_t = f(t)dt + dL_t, \quad t \in [0, 1] \quad (8.3)$$

et le même modèle de régression non-paramétrique que (8.2). Plus précisément, ce qu'on a démontré est que les modèles sont équivalents pour ce qui concerne l'estimation de la fonction inconnue f .

Nombreuses sont les possibles généralisations. En particulier, celles qui nous apparaissent les plus naturelles et qui seront l'objet d'un travail futur sont les suivantes :

1. Pour résoudre un problème de calibration (estimation de la fonction du drift et de la mesure de Lévy), on souhaite chercher un équivalent asymptotique au modèle :

$$dX_t = f(t)dt + dL_t, \quad t \in [0, 1] \quad L \sim \text{Lévy}\left(0, \frac{\sigma(\cdot)}{\sqrt{n}}, \nu\right). \quad (8.4)$$

2. Une autre question intéressante concerne les modèles récurrents et les processus de diffusion. Dans l'objectif d'estimer une fonction f dans un modèle de diffusion, Milstein et Nussbaum [31] ont montré l'équivalence asymptotique entre le modèle de diffusion :

$$dy(t) = f(y(t))dt + \varepsilon dB_t, \quad t \in [0, 1], \quad y(0) = 0 \quad (8.5)$$

et le modèle discret donné par :

$$Y_{i+1} = Y_i + n^{-1}f(Y_i) + \varepsilon n^{-\frac{1}{2}}\xi_i, \quad i = 1, \dots, n, \quad Y_1 = 0, \quad (8.6)$$

où les ξ_i sont variables aléatoires i.i.d normales réduites et centrées et Y_i est une solution approchée de (8.5) définie seulement aux points t_i , $i = 1, \dots, n$ d'une grille qui est une partition de $[0, 1]$.

Ce qui nous apparaît raisonnable est alors l'existence d'une équivalence asymptotique entre les modèles de la forme :

$$dy(t) = f(y(t))dt + dL_t, \quad t \in [0, 1], \quad y(0) = 0, \quad L \sim \text{Lévy}(0, \varepsilon, \nu) \quad (8.7)$$

et

$$Y_{i+1} = Y_i + n^{-1}f(Y_i) + \varepsilon n^{-\frac{1}{2}}\xi_i, \quad i = 1, \dots, n, \quad Y_1 = 0, \quad (8.8)$$

car la seule inconnue est la fonction f qui appartient à la partie continue du processus défini par (8.7).

Il serait alors intéressant d'étudier quelques applications d'une telle équivalence : qu'entraîne-t-elle au niveau de l'existence et unicité de la solution ? Que peut-on déduire sur le comportement limite des solutions ?

3. Finalement, un projet de travail plus ardu est l'étude de l'existence d'un modèle discret asymptotiquement équivalent à un modèle de signal perturbé par un processus à accroissements dépendants M :

$$dJ_t = f(t)dt + dM_t, \quad t \in [0, 1]. \quad (8.9)$$

En suivant la même approche utilisée dans le cas d'accroissements indépendants, on s'attendra à ce que le modèle (8.9) soit asymptotiquement équivalent à un modèle de régression non-paramétrique qui, cette fois, devrait avoir des innovations dépendantes.

Des applications à des problèmes de calibration en finance seront envisageables.

Annexe A

L'intégrale de Wiener

Le but de cette annexe est de rappeler brièvement la définition et les propriétés de l'intégrale de Wiener. Pour cela commençons par rappeler la définition du mouvement brownien.

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

Définition A.1. Un *processus stochastique* est une fonction mesurable $X(t, \omega)$ définie sur l'espace produit $I \times \Omega$, $I \subset \mathbb{R}$ intervalle. En particulier,

- i. pour tout t , $X(t, \cdot)$ est une variable aléatoire,
- ii. pour tout ω , $X(\cdot, \omega)$ est une fonction mesurable.

Pour des raisons de commodité, on écrira la variable aléatoire $X(t, \cdot)$ comme $X(t)$ ou encore X_t . Dans le même esprit, on pourra exprimer le processus stochastique $X(t, \omega)$ comme $X(t)(\omega)$ ou, plus simplement, $X(t)$ ou encore $X = (X_t, t \in I)$.

Définition A.2. Un processus stochastique $B(t, \omega)$ est appelé *mouvement brownien* s'il satisfait les conditions suivantes

1. $\mathbb{P}\{\omega : B(0, \omega) = 0\} = 1$.
2. Pour tous $0 \leq s < t$, la variable aléatoire $B(t) - B(s)$ suit une loi normale $\mathcal{N}(0, t - s)$.
3. $B(t, \omega)$ est à incréments indépendantes, i.e. pour tous $0 \leq t_1 < t_2 < \dots < t_n$, les variables aléatoires

$$B(t_1), B(t_2) - B(t_1), \dots, B(t_n) - B(t_{n-1})$$

sont indépendantes.

4. $\mathbb{P}\{\omega : B(\cdot, \omega) \text{ est continue}\} = 1$.

On souhaite maintenant donner un sens aux expressions de la forme

$$\int_a^b f(t) dB(t, \omega),$$

où f est une fonction déterministe (i.e. qui ne dépend pas de ω) et $B(t, \omega)$ est le mouvement brownien. Pour cela, on va définir l'intégrale de Wiener en deux étapes.

Étape 1 Supposons que f soit une fonction en escalier de la forme $f = \sum_{i=1}^n a_i \mathbb{I}_{[t_{i-1}, t_i)}$ avec $t_0 = a$, $t_n = b$. Définissons :

$$I(f) = \sum_{i=1}^n a_i (B(t_i) - B(t_{i-1})). \quad (\text{A.1})$$

Il est clair que $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$ pour tous $\alpha, \beta \in \mathbb{R}$, pour toutes fonctions en escalier f, g . De plus on a la propriété suivante :

Lemme A.0.1. *Pour toutes fonctions en escalier f , la variable aléatoire $I(f)$ suit une loi gaussienne de moyenne nulle et de variance donnée par*

$$\mathbb{E}(I(f)^2) = \int_a^b f(t)^2 dt \quad (\text{A.2})$$

Démonstration. Il est bien connu qu'une combinaison linéaire de variables aléatoires gaussiennes indépendantes est encore une variable aléatoire gaussienne, donc, grâce aux conditions (2) et (3) de la définition du mouvement brownien, la variable aléatoire définie par l'équation (A.1) suit une loi gaussienne de moyenne zéro. Calculons la variance :

$$\begin{aligned} \mathbb{E}(I(f)^2) &= \mathbb{E} \left[\sum_{i,j=1}^n a_i a_j (B(t_i) - B(t_{i-1}))(B(t_j) - B(t_{j-1})) \right] \\ &= \mathbb{E} \left[\sum_{i \neq j} a_i a_j (B(t_i) - B(t_{i-1}))(B(t_j) - B(t_{j-1})) \right. \\ &\quad \left. + \sum_{i=1}^n a_i^2 (B(t_i) - B(t_{i-1}))^2 \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n a_i^2 (B(t_i) - B(t_{i-1}))^2 \right] \\ &= \sum_{i=1}^n a_i^2 (t_i - t_{i-1}) \\ &= \int_a^b f(t)^2 dt; \end{aligned}$$

où on a utilisé, grâce aux propriétés (2) et (3) du mouvement brownien : pour tous $i \neq j$

$$\mathbb{E}[(B(t_i) - B(t_{i-1}))(B(t_j) - B(t_{j-1}))] = \mathbb{E}[B(t_i) - B(t_{i-1})] \mathbb{E}[B(t_j) - B(t_{j-1})] = 0$$

et

$$\mathbb{E}[B(t_i) - B(t_{i-1})]^2 = t_i - t_{i-1}.$$

□

Étape 2 Soit $L^2(\Omega)$ l'espace de Hilbert composé par toutes les variables aléatoires $X : \Omega \rightarrow \mathbb{R}$, de carré intégrables, muni du produit scalaire $\langle X, Y \rangle = \mathbb{E}[XY]$. Soit $f \in L^2[a, b]$ et choisissons une suite de fonctions en escalier, $(f_n)_n$, telle que f_n converge vers f en $L^2[a, b]$ quand n tend vers l'infini. Le lemme A.0.1 nous dit que la suite $(I(f_n))_n$ est de Cauchy dans $L^2(\Omega)$, et donc, vu que $L^2(\Omega)$ est un espace de Hilbert, $(I(f_n))_n$ converge dans $L^2(\Omega)$. Appelons une telle limite $I(f)$, c'est-à-dire, posons :

$$I(f) := \lim_{n \rightarrow \infty} I(f_n) \quad \text{dans } L^2(\Omega) \quad (\text{A.3})$$

Question A.0.2. $I(f)$ est-il bien défini ?

Pour montrer que $I(f)$ est bien défini il faut montrer que la limite dans l'équation (A.3) ne dépend pas du choix de la suite $(f_n)_n$.

Soit $(g_m)_m$ une autre suite de fonctions en escalier qui converge vers f dans $L^2[a, b]$. Par linéarité de I et grâce à (A.1),

$$\begin{aligned}\mathbb{E}[|I(f_n) - I(g_m)|^2] &= \mathbb{E}[|I(f_n - g_m)|^2] \\ &= \int_a^b (f_n(t) - g_m(t))^2 dt.\end{aligned}$$

En écrivant $f_n(t) - g_m(t) = [f_n(t) - f(t)] - [g_m(t) - f(t)]$ et en utilisant l'inégalité $(x - y)^2 \leq 2(x^2 + y^2)$, on aboutit à

$$\begin{aligned}\int_a^b (f_n(t) - g_m(t))^2 dt &\leq 2 \int_a^b [(f_n(t) - f(t))^2 + (g_m(t) - f(t))^2] dt \\ &\rightarrow 0 \quad \text{quand } n, m \rightarrow \infty.\end{aligned}$$

Il s'en suit que

$$\lim_{n \rightarrow \infty} I(f_n) = \lim_{m \rightarrow \infty} I(g_m) \quad \text{dans } L^2(\Omega)$$

et donc $I(f)$ est bien défini.

Définition A.3. Soit $f \in L^2[a, b]$. La limite $I(f)$ définie par l'équation (A.3) est appelée *intégrale de Wiener de f* .

L'intégrale de Wiener $I(f)$ de f sera indiqué par

$$I(f)(\omega) = \left(\int_a^b f(t) dB(t) \right) (\omega), \quad \omega \in \Omega, \quad \text{presque sûrement}$$

où, plus simplement, par

$$I(f) = \int_a^b f(t) dB(t).$$

Théorème A.0.3. Pour toute fonction $f \in L^2[a, b]$, l'intégrale de Wiener $\int_a^b f(t) dB(t)$ est une variable aléatoire gaussienne de moyenne 0 et variance $\|f\|^2 = \int_a^b f(t)^2 dt$.

Démonstration. Grâce au lemme A.0.1 l'assertion est vraie pour toute fonction f à escalier dans $L^2[a, b]$. Pour une fonction f quelconque dans $L^2[a, b]$, l'assertion découle du bien connu fait suivant : si (X_n) suit une loi gaussienne $\mathcal{N}(\mu_n, \sigma_n^2)$ et (X_n) converge vers X dans $L^2(\Omega)$, alors X suit une loi gaussienne $\mathcal{N}(\mu, \sigma^2)$ avec $\mu := \lim_{n \rightarrow \infty} \mu_n$ et $\sigma^2 := \lim_{n \rightarrow \infty} \sigma_n^2$. \square

Corollaire A.0.4. Si $f, g \in L^2[a, b]$, alors

$$\mathbb{E}[I(f)I(g)] = \int_a^b f(t)g(t)dt. \quad (\text{A.4})$$

En particulier, si f et g sont orthogonaux, les variables aléatoires gaussiennes $I(f)$ et $I(g)$ sont indépendantes.

Démonstration. Par linéarité de I et grâce au théorème précédent, on a

$$\begin{aligned}\mathbb{E}[(I(f) + I(g))^2] &= \mathbb{E}[(I(f + g))^2] \\ &= \int_a^b (f(t) + g(t))^2 dt \\ &= \int_a^b f^2(t) dt + 2 \int_a^b f(t)g(t) dt + \int_a^b g^2(t) dt\end{aligned}\quad (\text{A.5})$$

Par contre, toujours en appliquant le théorème précédente, on trouve que

$$\begin{aligned}\mathbb{E}[(I(f) + I(g))^2] &= \mathbb{E}[I(f)^2 + 2I(f)I(g) + I(g)^2] \\ &= \int_a^b f^2(t) dt + 2\mathbb{E}[I(f)I(g)] + \int_a^b g^2(t) dt.\end{aligned}\quad (\text{A.6})$$

Manifestement, l'équation (A.4) découle des équations (A.5) et (A.6). \square

Remarque A.0.5. Le corollaire A.0.4 montre que l'intégrale de Wiener $I : L^2[a, b] \rightarrow L^2(\Omega)$ est une isométrie (i.e. préserve le produit scalaire).

Remarque A.0.6. D'une façon générale, un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et un processus stochastique $Z = (Z_t, t \in [a, b])$ défini sur $(\Omega, \mathcal{A}, \mathbb{P})$, permettent de définir l'intégrale d'une fonction f contre Z comme la limite sur toutes les partitions $a = t_0 < \dots < t_k = b$ de l'intervalle $[a, b]$

$$\lim_{t_0 < \dots < t_k} \sum_{i=1}^k f(t_{i-1})(Z_{t_i} - Z_{t_{i-1}}) \quad (\text{A.7})$$

lorsque cette limite existe.

On notera l'intégrale d'une fonction f contre Z de la façon suivante :

$$\int_a^b f(t) dZ_t. \quad (\text{A.8})$$

En particulier, si on choisit comme espace probabilisé $(\mathcal{C}[a, b], \mathcal{C}^{[a, b]}, \mu)$, l'espace de toutes les fonctions continues de $[a, b]$ dans \mathbb{R} (muni de la tribu borélienne $\mathcal{C}^{[a, b]}$ induite par la norme du sup et d'une mesure μ) on peut considérer le processus stochastique identité Id :

$$\forall \gamma \in \mathcal{C}[a, b] \quad \text{Id}(\gamma)_t = \gamma(t).$$

Le lien avec la définition A.3 est alors le suivant. Si on note par \mathbb{W} la mesure de Wiener (i.e. la loi image du processus $(B_t, t \in [a, b])$), alors la loi de la variable aléatoire $\int_a^b f(s) d\text{Id}(s)$ sous \mathbb{W} est donnée par $\int_a^b f(s) dB_s$, ce qui justifie la notation introduite ci-dessus (A.8).

Annexe B

Processus gaussiens

Définition B.1. Une famille \mathcal{S} de variables aléatoires d -dimensionnelles définies sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ est dite une *famille gaussienne* si, pour tout choix de $X_1, \dots, X_m \in \mathcal{S}$ et $\gamma_1, \dots, \gamma_m \in \mathbb{R}^d$, la variable aléatoire $\langle \gamma_1, X_1 \rangle + \dots + \langle \gamma_m, X_m \rangle$ est gaussienne.

Définition B.2. Un processus $X = (X_t, t \in T)$ est dit *gaussien* si toutes ses lois finies dimensionnelles $\mathcal{L}(X_{t_1}, \dots, X_{t_n})$ sont gaussiennes ($\forall n, \forall (t_1, \dots, t_n) \in T^n$). Autrement dit, X est *gaussien* si $(X_t)_{t \in T}$ est une famille gaussienne.

Il est connu que la loi d'un vecteur gaussien $(X_{t_1}, \dots, X_{t_n})$ est déterminée (via sa fonction caractéristique) par le vecteur moyenne $(\mathbb{E}[X_{t_1}], \dots, \mathbb{E}[X_{t_n}])$ et la matrice de covariance $(Cov(X_{t_i}, X_{t_j})_{i,j=1}^n)$. On comprend dès lors que toute la loi d'un processus gaussien est connue dès qu'on se donne la fonction moyenne $a(t) = \mathbb{E}[X_t]$ et l'opérateur de covariance $K(s, t) = Cov(X_s, X_t)$. En effet, la loi fini-dimensionnelle de $(X_{t_1}, \dots, X_{t_n})$ est alors la loi normale de dimension n $\mathcal{N}(a_n, K_n)$ avec $a_n = (a(t_1), \dots, a(t_n))$ et $K_n = (K(t_i, t_j)_{i,j=1}^n)$. Les fonctions a et K définissent donc toutes les lois fini-dimensionnelles de X et donc aussi sa loi.

Proposition B.0.7. Soit $I = [a, b]$ un intervalle de \mathbb{R} qui contient 0. Soit f une fonction dans $L^2(I)$. Le processus stochastique $X = (X_t, t \in I)$ défini par

$$X_t = \int_a^b f(t) dB_t$$

où B est un mouvement brownien, est un processus gaussien de fonction moyenne $\mathbb{E}[X_t] \equiv 0$ et d'opérateur de covariance donné par $K(t_i, t_j) = \int_s^t f^2(u) \mathbb{1}_{[s, t_i] \cap [s, t_j]}(u) du$.

Démonstration. Pour montrer que le processus X est gaussien il faut montrer que pour tous $a \leq t_1 < \dots < t_n \leq b$, pour tous nombre réels $\gamma_1, \dots, \gamma_n$, la variable aléatoire $\gamma_1 X_{t_1} + \dots + \gamma_n X_{t_n}$ est gaussienne. On utilise alors exactement le même argument utilisé dans la preuve du théorème A.0.3. Le calcul de la fonction moyenne et de l'opérateur de covariance découlent du théorème A.0.3 et de l'équation (A.4). \square

Corollaire B.0.8. Soit $I = [a, b]$ un intervalle de \mathbb{R} qui contient 0. Soit f une fonction dans $L^2(I)$, μ une fonction dans $L^1(I)$. Le processus stochastique $X = (X_t, t \in I)$ défini par

$$X_t = \int_a^b \mu(u) du + \int_a^b f(u) dB_u$$

où B est un mouvement brownien, est un processus gaussien de fonction moyenne

$$\mathbb{E}[X_t] = \int_a^b \mu(u) du$$

et d'opérateur de covariance donné par

$$K(t_i, t_j) = \int_a^b f^2(u) \mathbb{I}_{[a, t_i] \cap [a, t_j]}(u) du.$$

Démonstration. Grâce à la proposition précédente on sait que pour tous $a \leq t_1 < \dots < t_n \leq b$, pour tous nombre réels $\gamma_1, \dots, \gamma_n$, la variable aléatoire $\gamma_1 \int_a^{t_1} f(u) dB_u + \dots + \gamma_n \int_a^{t_n} f(u) dB_u$ est gaussienne centrée. Il est clair alors que pour tous $a \leq t_1 < \dots < t_n \leq b$, pour tous nombre réels $\gamma_1, \dots, \gamma_n$, la variable aléatoire $\gamma_1 X_{t_1} + \dots + \gamma_n X_{t_n}$ est encore gaussienne centrée. Pour obtenir la formule souhaitée de la fonction moyenne, il suffit de remarquer que si X suit une loi normale $\mathcal{N}(\nu, \sigma^2)$, alors $c + X$ suit une loi normale $\mathcal{N}(c + \nu, \sigma^2)$, le calcul de l'opérateur de covariance découle de la proposition précédente. \square

Annexe C

Simulations

```
1 #####
2 #
3 # NADARAYA-WATSON #
4 # #
5 #####

7 library(KernSmooth)

9 ## On teste differentes fonctions de regression ,
10 # appartenantes a la classe de Sobolev periodiques et donc ,
11 # en particulier , satisfaisantes les hypotheses du theoreme
12 # sur l'equivalence asymptotique entre le modele de regression et celui a saut .
13
14 f<-function(t){
15 f=0.25*sin(4*pi*t)+1/3*cos(4*pi*t)-1/2*sin(2*pi*t)

17 # r=(-t+0.5)/(t*(t-1))
18 # f=exp(-r^2)
19
20 # f=sin(4*pi*(t-0.5))*abs(t-0.5)
21 }

23 # On simule les donnees

25 n=1000
26 sigma = 0.1
27 X<-c(0:n)/n
28 Y<-f(X)+rnorm(n+1, sd=sigma)
29
30 # On superpose le graphe de f au graphe de l'estimateur de Nadaraya-Watson.
31 # La fonction dpill() nous permet de choisir la fenetre optimale.

33 h<-dpill(X,Y)
```

```

plot(f)
35 lines(locpoly(X,Y,degree=0,bandwidth=h),col="blue",lwd=2)
legend(0.5,0.37,c("la vraie fonction", "estimation via regression"),
37       col=c("black","blue"), lty=1, lwd=2)
dev.print(pdf, file="NW.pdf")
39
  ## On utilise l'estimateur de Nadaraya-Watson pour simuler f dans
41 #le modele a saut.
  ## On simule la partie continue de la trajectoire issue du modele:
43 #  $dZ_t=f(t)dt+\sigma/\sqrt{n}dB_t+dX_t$ .

45 Z = c(0:(n+1))

47 # Simulons un mouvement brownien en se souvenant de toutes les valeurs
  #dans les pointes de la forme  $i/(n+1)$ .
49
  brown = c(0,cumsum(rnorm(n+1,sd=1/sqrt(n+1))))
51
  # Finalement, definissons la partie continue de la trajectoire Z, qui,
53 # au temps t, est l'integrale de f entre 0 et t +  $\sigma/\sqrt{n+1} * B_t$ .

55 for (i in 0:(n+1)) {
  Z[i+1] = integrate(f, 0, i/(n+1))$value +  $\sigma/\sqrt{n+1} * \text{brown}[i+1]$ 
57 }

59 # On superpose le graphe de f au graphe de l'estimateur du drift
  #construit a partir du modele de regression.
61
  S=c(1:(n+1))
63 for (i in 1:(n+1)) {
  S[i]=(n+1)*(Z[i+1]-Z[i])
65 }

67 plot(f)
  lines(locpoly(X,S,degree=0,bandwidth=h),col="red",lwd=2)
69 legend(0.5,0.37,c("la vraie fonction", "estimation via modele a saut"),
        col=c("black","red"), lty=1, lwd=2)
71 dev.print(pdf, file="NWsaut.pdf")

73
#####
75 # #
  # ESTIMATEUR PAR PROJECTION #
77 # #
#####
79

81 ##Simulation des donnees

```

```

83 n = 1000
   N = floor(n^(1/3))
85 sigma = 0.1
   Y<-f(c(0:n)/n)+rnorm(n+1, sd=sigma)
87
   ##fonctions a estimer
89
   f<-function(t){
91 # f=0.25*sin(4*pi*t)+1/3*cos(4*pi*t)-1/2*sin(2*pi*t)

93 r=(-t+0.5)/(t*(t-1))
   f=exp(-r^2)
95
   # f=sin(4*pi*(t-0.5))*abs(t-0.5)
97 }

99 ## Estimateur de f ##
   phi<-function(k,t) {
101 if (k == 1) {
   phi <- 1
103 }
   else if ((k %% 2)==0) {
105 phi <- sqrt(2)*cos(k*t*pi)
   }
107 else if ((k %% 2)==1) {
   phi <- sqrt(2)*sin((k-1)*t*pi)
109 }
   }
111 coeff<-function(y){
   bla=c(1:N)*0
113 for (j in 1:N) {
   for (k in 0:n) {
115 bla[j]<-bla[j] + 1/n * y[k+1] * phi(j,k/n)
   }
117 }
   bla
119 }
   stimatore<-function(t,co) {
121 v<-0
   for (j in 1:N) {
123 v<-v + co[j] * phi(j,t)
   }
125 v
   }
127 T<-function(t){
   theta<-coeff(Y)
129 T<-stimatore(t,theta)

```

```

}
131  ## Generons la trajectoire continue Z ##
133 F<-function(t){
      Z = c(0:(n+1))
135
      # Simulons un mouvement brownien en se souvenant
137 #de toutes les valeurs dans les pointes de la forme i/(n+1)

139 brown = c(0,cumsum(rnorm(n+1,sd=1/sqrt(n+1))))

141 # Finalement, definissons la partie continue de la trajectoire Z,
      # qui, au temps t, est l'integrale de f entre 0 et t + sigma/sqrt(n+1) * B_t.
143 for (i in 0:(n+1)) {
      Z[i+1] = integrate(f, 0, i/(n+1))$value + sigma/sqrt(n+1)*brown[i+1]
145 }
      S=c(1:(n+1))
147 for (i in 1:(n+1)) {
      S[i]=(n+1)*(Z[i+1]-Z[i])
149 }

151 ## Definissons l'estimateur du drift construit a partir
      # du modele de regression
153 eta<-coeff(S)
      F<-stimatore(t, eta)
155 }

157

159

161 # Plot des graphes avec estimation par base trigonometrique
      curve(T, col="blue", lwd=2)
163 curve(f, col="black", lwd=2, add=TRUE)
      legend(0.03, 0.67, c("la vraie fonction", "estimation via regression"),
165           col=c("black", "blue"), lty=1, lwd=2)
      #dev.print(pdf, file="trigo3.pdf")
167 curve(f, col="black", lwd=2)
      curve(F, col="red", lwd=2, add=TRUE)
169 legend(0.03, 0.67, c("la vraie fonction", "estimation via modele a saut"),
           col=c("black", "red"), lty=1, lwd=2)
171 #dev.print(pdf, file="trigosaut3.pdf")

173 #####
175 # #
      # ESTIMATEUR DE PINSKER #
177 # #

```

```

#####
179  ## Dans nos notations, l'estimateur de Pinsker coincide formellement
181  #avec l'estimateur trigonometrique, avec les coeffisants multiplies
    # par des constantes l_j et avec un choix different de N.
183
    f<-function(t){
185 f=0.25*sin(4*pi*t)+1/3*cos(4*pi*t)-1/2*sin(2*pi*t)

187 # r=(-t+0.5)/(t*(t-1))
    # f=exp(-r^2)
189
    # f=sin(4*pi*(t-0.5))*abs(t-0.5)
191 }

193 beta = 1
    gamma = (2*beta+1)
195 L <- 10

197 Q = L^2/(pi^(2*beta))
    epsilon <- sigma / sqrt(n)
199
    k <- (beta/(gamma*(beta+1)*Q))^(beta/gamma) * epsilon^(2*beta/gamma)
201
    lPinsker <- function(j) {
203 a <- ((j-1)^beta) * (j %% 2) + (j^beta) * ((j+1) %% 2)
    l <- max(0, (1 - k * a))
205 }

207 # Le suivant ne marche que pour beta=1, sinon des logarithms sont a prévoir...
    NPinsker <- ceiling(1/k)
209
    integrando <- function(t) {
211 integrando <- f(t) * phi(j,t)
    }
213
    theta <- function() {
215 theta <- integrate(integrando, 0, 1, subdivisions=10000)$value + epsilon * rnorm(1)
    }
217
    coeffPinsker <- c(1:NPinsker)
219
    for (j in 1:NPinsker) {
221 coeffPinsker[[j]] <- theta()
    }
223
    # La fonction stimatorePinsker differe de la fonction
225 #stimatore par les constantes l_j

```

```

227 stimatorePinsker<-function(t) {
      v<-0
229   for (j in 1:NPinsker) {
      v<-v + lPinsker(j) * coeffPinsker[j] * phi(j,t)
231   }
      v
233 }

235 curve(stimatorePinsker, col="red", lwd=2)
      curve(f, col="blue", lwd=2,add=TRUE)
237

239 #####
      #
241 #          CALCUL DES RISQUES          #
      #
243 #####

245 ## Voila les fonctions necessaires pour le calcul du risque
      #dans le modele de regression avec Montecarlo
247
      diffquadreg<-function(t) {
249   diffquadreg<-((f(t)-T(t))^2)
      }
251
      # Fonction de perte L: distance L2
253 Lreg<-function() {
      res <- integrate(diffquadreg,0,1)$value
255   res <- sqrt(res)
      res
257 }

259
      ## Voila les fonctions necessaires pour le calcul du risque
261 #dans le modele white noise avec Montecarlo

263 diffquadwhite<-function(t) {
      diffquadwhite<-((f(t)-F(t))^2)
265 }

267 # Fonction de perte L: distance L2
      Lwhite<-function() {
269   res <- integrate(diffquadwhite,0,1)$value
      res <- sqrt(res)
271   res
      }
273

```

```

275 # Calcul des risques: necessite de beaucoup de operations,
    #deja pour n=K=100 il faut attendre une dizaine de secondes.
277
    K = 100
279 rischioregr = 0
    for (i in 1:K) {
281 # Pour chaque i on tire un echantillon Y de taille n et on calcule,
    #par Lregr, la fonction de perte entre f et son estimateur T calcule par Y.
283 #On calcul le risque par Montecarlo calcule sur les K (suppose grand)
    #possibilites pour Y.
285
        Y<-f(c(0:n)/n)+rnorm(n+1, sd=sigma)
287 rischioregr = rischioregr + Lregr()/K
    }
289 rischioregr

291 rischiowhite <- 0
    for (i in 1:K) {
293 # De meme, on calcule le risque de l'estimateur F par Montecarlo.
    #A noter que la fonction F calcule deja une trajectoire a chaque iteration,
295 #donc il n'y a pas de calcul prealable d'une trajectoire Z.

297 rischiowhite = rischiowhite + Lwhite()/K
    }
299 rischiowhite

301 diffquadPinsker<-function(t) {
    diffquadPinsker<-((f(t)-stimatorePinsker(t))^2)
303 }

305 # Fonction de perte L: distance L2
    LPinsker<-function() {
307 res <- integrate(diffquadPinsker,0,1)$value
    res <- sqrt(res)
309 res
    }
311
    K=100
313 rischioPinsker=0
    for (i in 1:K) {
315 for (j in 1:NPinsker) {
        coeffPinsker[[j]] <- theta()
317 }
        rischioPinsker = rischioPinsker + LPinsker()/K
319 }
    print(rischioPinsker)
321

```

```
# Calcul du risque de l'estimateur de Pinsker par la formule
323 #donne dans le chapitre 5

325 epsilon <- sigma / sqrt(n)
    gamma <- 2*beta+1
327 C <- L^(2/gamma) * gamma^(1/gamma) * (beta/(pi * (beta+1)))^(2*beta/gamma)
    vitesse <- epsilon^(4*beta/gamma)
329 risque <- C*vitesse
    risque
```

Bibliographie

- [1] P. Billingsley. Convergence of probability measures. In *John Wiley & Sons, Inc., New York-London-Sydney*. 1968.
- [2] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 3, 1973.
- [3] Cai T.T. Low M.G. Brown, L.D. and C. Zhang. Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.*, 30 :688–707, 2002.
- [4] L.D. Brown and M. Low. Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, 24 :2384–2398, 1996.
- [5] L.D. Brown and C. Zhang. Asymptotic nonequivalence of nonparametric experiments when the smoothness index is $1/2$. *Ann. Statist.*, 26 :279–287, 1998.
- [6] A.V. Carter. A continuous gaussian process approximation to a nonparametric regression in two dimensions. *Bernoulli*, 12 :143–156, 2006.
- [7] A.V. Carter. Asymptotic approximation of nonparametric regression experiments with unknown variances. *Ann. Statist.*, 35 :1644–1673, 2007.
- [8] A.V. Carter. Asymptotically sufficient statistics in nonparametric regression experiments with correlated noise. *Journal of Probability and Statistics*, 2009.
- [9] V. Carter, A. Deficiency distance between multinomial and multivariate normal experiments. *Ann. Statist.*, 30 :708–730, 2002.
- [10] R. Cont and Tankov P. Financial modelling with jump processes. In *Chapman & Hall/CRC, Boca Raton, FL*. 2004.
- [11] A. Dalalyan and M. Reiss. Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab. Theory Rel. Fields*, 134 :248–282, 2006.
- [12] A. Dalalyan and M. Reiss. Asymptotic statistical equivalence for ergodic diffusions : the multidimensional case. 2008.
- [13] S. Delattre and M. Hoffmann. Asymptotic equivalence for a null recurrent diffusion. *Bernoulli*, 8 :139–174, 2002.
- [14] Nagel E. R. Dette, H. and N. Neumeyer. Bootstrap tests for the error distribution in linear and non parametric regression models. 2004.
- [15] I. M. Kerkycharian G. Donoho, D. L. Johnstone and D. Picard. Wavelet shrinkage : asymptopia ? (with discussion). *Roy. Statist. Soc.*, B57, 1995.
- [16] J. Fan. On the estimation of quadratic functions. *Ann. Statist.*, 19 :1273–1294, 1991.

- [17] Laredo C. Geno-Catalot, V. and M. Nussbaum. Asymptotic equivalence of estimating a poisson intensity and a positive diffusion drift. *Ann. Statist.*, 30 :731–753, 2002.
- [18] K. Nussbaum M. Golubev, G. and H. Zhou H. Asymptotic equivalence of spectral density estimation and gaussian white noise. 2009.
- [19] G. Grama, I. and H. Neumann, M. Asymptotic equivalence of nonparametric autoregression and nonparametric regression. *Ann. Statist.*, 34 :1701–1732, 2006.
- [20] I. Grama and M. Nussbaum. Asymptotic equivalence for nonparametric generalized linear models. *Probab.Theor.Related Fields*, 111 :493–533, 1998.
- [21] I. Grama and M. Nussbaum. Asymptotic equivalence for nonparametric regression. 11(1) :1–36, 2002.
- [22] O. Hansen and N. Torgersen, E. Comparison of linear normal experiments. *Ann. Statist.*, 2 :367–373, 1974.
- [23] M. Janisch and M. Nussbaum. Asymptotic equivalence for a model of independent non identically distributed observations. *Statist. Decisions*, 21(3) :197–218, 2003.
- [24] J. Klemela and B. Tsybakov, A. Sharp adaptive estimation of linear functional. *Ann. Statist.*, 29 :1567–1600, 2001.
- [25] Mammen E. Konakov, V. and J. Woerner. Statistical convergence of markov experiments to diffusion limits.
- [26] E. Kyprianou, A. Introductory lectures on fluctuations of lévy processes with applications. In *Springer, Berlin*. 2006.
- [27] L. Le Cam. Asymptotic methods in statistical decision theory. In *Springer-Verlag, New York*. 1986.
- [28] L. Le Cam and Yang G.L. Asymptotics in statistics. some basic concepts. In *Springer, New York*. 2000.
- [29] B. Mandelbrot, B. The variation of certain speculative prices. *Journal of Business*, XXXVI :392–417, 1963.
- [30] A. Meister. Asymptotic equivalence of functional linear regression and a white noise inverse problem. *Ann. Statist.*, 39 :1471–1495, 2011.
- [31] G. Milstein and M. Nussbaum. Diffusion approximation for non parametric autoregression. *Probab. Theory Related Fields*, 112 :535–543, 1998.
- [32] E. Nadaraya. On estimating regression. *Theory of Probability and its Application.*, 9 :141–142, 1964.
- [33] M. Nussbaum. Asymptotic equivalence of density estimation and white noise. *Ann. Statist.*, 24 :2399–2430, 1996.
- [34] M.S. Pinsker. Optimal filtering of square integrable signals in gaussian white noise. *Problems Inform.Transmission*, 16 :120–133, 1980.
- [35] M. Reiss. On le cam’s distance between gaussian nonparametric regression and white noise. *Manuscript*, 2006.
- [36] N. Torgersen, E. Comparison of translation experiments. *Ann. Statist.*, 43 :1383–1399, 1972.
- [37] N. Torgersen, E. Comparison of experiments by factorization. *Stat.Res.Report*, 1974.
- [38] A.B. Tsybakov. Asymptotically efficient signal estimation in l_2 under general loss functions. *Problems of Information Transmission*, 33.

- [39] A. Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.*, 54 :426–482, 1943.