CÉCILE DUROT

Asymptotique fine pour l'estimateur isotonique en régression et méthodes de jackknife

Thèses d'Orsay, 1997

<http://www.numdam.org/item?id=BJHTUP11_1997__0477__P0_0>

ORSAY
n° d'ordre

UNIVERSITE DE PARIS-SUD
CENTRE D'ORSAY

THESE

présentée
pour obtenir

LE TITRE DE DOCTEUR EN SCIENCES

Spécialité Mathématiques

par

Cécile Durot

Sujet

Asymptotique fine pour l'estimateur isotonique en régression et méthodes de jackknife. Applications à la comparaison de courbes de croissance.

Soutenue le 9 janvier 1997 devant la commission d'examen composée de

M.	Jean Bretagnolle	Président
M.	Piet Groeneboom	Rapporteur
M.	Enno Mammen	Rapporteur
M.	Lucien Birgé	Examinateur
Mme.	Sylvie Huet	Examinateur
M.	Pascal Massart	Examinateur

Remerciements

Mes premiers remerciements sont adressés à mon directeur de thèse, Pascal Massart. Ses conseils, tant en ce qui concerne la rédaction que le contenu de cette thèse, m'ont été précieux pour mener ce travail à terme. Il a su, au long de nos séances de travail, développer mon intuition mathématique et me familiariser avec diverses techniques statistiques et probabilistes. Je le remercie pour sa grande disponibilité et sa patience ! Je le remercie plus particulièrement pour ses conseils amicaux et son soutien. Sa gaité, son dynamisme et l'ambiance chaleureuse qu'il développe autour de lui ont ajouté une dimension amicale à nos relations professionnelles, aussi agréable que stimulante.

Les séances de travail auxquelles a participé Jean Bretagnolle ont été très fructueuses. Je le remercie pour l'intérêt constant qu'il a manifesté pour ma recherche, et suis très honorée qu'il ait accepté de présider mon jury de thèse.

Le dernier (et le plus conséquent) chapitre de cette thèse a pour origine un article de Piet Groeneboom. Je le remercie de l'attention toute particulière qu'il a portée à mon travail en fin de thèse, notamment en acceptant de faire partie de mon jury et de rapporter ma thèse, ceci avec beaucoup d'attention. J'ai été sensible à ses encouragements et son enthousiasme.

Je remercie Enno Mammen de l'extrême rigueur avec laquelle il a rapporté ma thèse. Je le remercie en particulier de m'avoir signalé une erreur apparaissant dans une version préliminaire du quatrième chapitre.

Sylvie Huet et Lucien Birgé ont accepté de faire partie de mon jury de thèse. Je les en remercie vivement.

Je remercie enfin Richard Tomassone pour m'avoir suggéré l'étude de l'estimateur jackknife et fourni les données m'ayant permis d'illustrer cette méthode.

Abstract

This thesis is devoted to the study of some non linear estimators of a regression function. More precisely, one observe $y_i = f(x_i) + \varepsilon_i$ $(i \in \{1, ..., n\})$, where the explanatory variables (x_i) are deterministic and the error terms (ε_i) are i.i.d. with mean zero. We consider two frameworks : the non linear parametric regression and the non parametric regression. We study the sharp asymptotic behaviour of some least squares type estimators.

For non linear parametric regression, we consider the jackknife estimator $\tilde{\theta}$ of the regression parameter θ_0. We show that, under suitable smoothness conditions, $\sqrt{n}\Gamma(\theta_0)^{-1}(\tilde{\theta} - \theta_0)$ converges in distribution to a standard gaussian vector, for some adequate positive matrix $\Gamma(\theta_0)$ and propose a consistent estimator $\tilde{s}$ of $\Gamma(\theta_0)$. Thus, $\sqrt{n}\tilde{s}^{-1}(\tilde{\theta} - \theta_0)$ converges in distribution to a standard gaussian vector. We present some applications of these results. In particular, we show how they allow the construction of confidence regions and hypothesis tests, with given asymptotic level. This valids asymptotically Tukey's heuristic, according to which the pseudo-values can be considered as i.i.d variables. We illustrate the jackknife method by the study of real data. These data result from experiments about bacteria evolution in a milky environment. We compute jackknife estimators thanks to the logiciel S.A.S. Then, we test the influence of the temperature on the bacteria evolution.

For monotone non parametric regression, we prove a central limit theorem for the $\mathbb{L}_1$-distance between the isotonic estimator f_n (that is the least squares estimator under order restriction) and the regression function f, whenever the (x_i) are uniformly spread over [0,1]. If f is smooth enough over [0,1] and if the error terms admit a variance σ^2 and moments of order greater than 12, then,

$$n^{1/6}\{n^{1/3}\int_0^1 |f_n(t) - f(t)|dt - C_f\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2 a) \qquad \text{as } n \to \infty \tag{0.1}$$

where $C_f = C\int_0^1 |\sigma^2 f'(t)|^{1/3} dt$. C and a are absolute constants : $C \approx 0.33$ et $a \approx 0.17$. Our approach consists in approximate the initial regression model by a white noise model thanks to a Komlós, Major and Tusnády type strong approximation. We have to prove (0.1) with $\tilde{f}_n$ in the place of f_n, where $\tilde{f}_n$ is the isotonic estimator defined from the approximant white noise model. The scaling properties of the Brownian motion will be widely used, which motivate the introduction of a white noise approximation. We also prove some weaker result than (0.1), but which holds under the square integrability of the errors. If f is smooth enough, then

$$\mathbb{E}\int_0^1 |f_n(t) - f(t)|dt = O(n^{-1/3}).$$

Key words: Non linear regression, Non parametric regression, Jackknife estimate, Isotonic estimate, Hypothesis test, Brownian motion

Table des matières

Chapitre 1

Introduction

Cette thèse est consacrée à l'étude de différents estimateurs non linéaires d'une fonction de régression. Nous envisageons deux cadres: la régression paramétrique non linéaire et la régression non paramétrique monotone. Nous étudions dans chacun le comportement asymptotique fin d'estimateurs du type moindres carrés. En régression paramétrique non linéaire, nous considérons l'estimateur jackknife du paramètre de régression. Nous établissons sa normalité asymptotique et proposons un estimateur consistant de sa variance asymptotique. Nous présentons quelques applications de ces résultats. Notamment, nous montrons comment ils permettent la construction de régions de confiance ou de tests, de niveau asymptotique donné. En régression non paramétrique monotone, nous démontrons un théorème de limite centrale pour la perte $\mathbb{L}_1$ de l'estimateur isotonique. Nous vérifions que la variance asymptotique est libre de la fonction de régression, et exprimons l'espérance asymptotique comme une fonctionelle explicite de la fonction de régression.

Avant d'énoncer plus précisément ces résultats, nous rappelons ce qu'est un modèle de régression, et décrivons quelques unes des méthodes d'estimation classiques.

Un problème rencontré fréquemment en statistique est la détermination de la relation existant entre deux variables, dont l'une au moins est aléatoire. Il s'agit de décrire l'évolution d'une variable aléatoire, que nous appellerons variable réponse et noterons y, en fonction d'une variable x (aléatoire ou déterministe), dite variable explicative. Les modèles de régression proposent de décrire la variable réponse comme la somme d'un terme déterministe, dépendant de x et représentant la tendance générale du phénomène étudié, et d'un terme résiduel aléatoire, représentant les fluctuations autour de la tendance générale. La tendance étant entièrement décrite par le terme déterministe, la variable résiduelle est supposée centrée. Ainsi, il existe une fonction f telle que

$$y = f(x) + \varepsilon \qquad (1.1)$$

où ε est une variable aléatoire d'espérance nulle. Il s'agit alors de déterminer la fonction de régression f. Il n'est possible d'estimer f avec une précision raisonnable que si l'on dispose de plusieurs observations. Nous supposerons donc que nous disposons de n observations $(x_i,\ y_i)_{1\leq i\leq n}$ du couple (x, y), où chaque observation est régie par la relation (1.1). L'équation de régression est alors la suivante

$$y_i = f(x_i) + \varepsilon_i \qquad 1 \leq i \leq n \qquad (1.2)$$

où les (ε_i) sont d'espérance nulle. Nous supposerons en outre que les termes résiduels (ε_i) sont indépendants, de variance $\mathbb{E}\varepsilon_i^2 = \sigma_i^2$ finie. Les variables explicatives (x_i) seront supposées déterministes et connues, appartenant à $\mathbb{R}^d$, et f est une fonction à valeurs réelles.

Considérons tout d'abord un cadre paramétrique.

Estimation en régression paramétrique

Nous supposons ici qu'il existe un sous ensemble connu Θ de $\mathbb{R}^p$ tel que la fonction de régression appartienne à un ensemble de fonctions paramétré par Θ:

$$f \in \{f(\,.\,,\theta),\ \theta \in \Theta\}$$

où la fonction $f(\,.\,,\,.\,)$ est connue. Si le modèle est identifiable, il existe un unique vecteur θ_0 de Θ tel que f soit égale à $f(\,.\,,\theta_0)$, et le problème d'estimation de f se ramène au problème d'estimation du paramètre θ_0. On s'intéresse à la vitesse d'estimation du paramètre θ_0 (par rapport à la distance euclidienne).

La méthode des moindres carrés

La méthode d'estimation de θ_0 la plus couramment utilisée est la méthode des moindres carrés. Cette méthode consiste à choisir comme estimateur de θ_0 un point de Θ où $\sum_{i=1}^n (y_i - f(x_i,\theta))^2$ atteint un minimum, ce que nous résumons par

$$\hat{\theta} = \text{argmin}_{\theta\in\Theta} \sum_{i=1}^{n} (y_i - f(x_i,\theta))^2 \tag{1.3}$$

Dans le cas du modèle linéaire, c'est à dire lorsque $f(\,.\,,\theta)$ est une fonction linéaire de θ, l'équation (1.3) admet une solution, exprimable comme combinaison linéaire des variables réponse y_i. Cette solution est unique si le modèle est identifiable et l'estimateur des moindres carrés est dans ce cas consistant et asymptotiquement gaussien (voir par exemple Coursol [9]): il existe une matrice positive Γ_0 dépendant de θ_0 telle que

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_p(0,\Gamma_0) \qquad \text{quand } n \to \infty \tag{1.4}$$

Si en outre les résidus (ε_i) sont gaussiens, $\sqrt{n}(\hat{\theta} - \theta_0)$ est une variable gaussienne.

Lorsque f n'est pas linéaire en θ, la solution (si elle existe) de l'équation (1.3) ne peut généralement pas s'exprimer comme fonction explicite des observations. Seules les propriétés asymptotiques de $\hat{\theta}$ (s'il existe) sont accessibles. Jenrich [21] fût le premier à déterminer des conditions d'existence et de consistance de l'estimateur des moindres carrés en régression non linéaire. De nombreux auteurs se sont intéressés au comportement asymptotique de cet estimateur. Malinvaud [24] énonce une condition suffisante de consistance, et propose certains exemples d'estimateurs des moindres carrés non consistants en régression non linéaire. Jenrich [21] et Wu [42] établissent des conditions suffisantes de consistance et de normalité asymptotique de $\hat{\theta}$, lorsque Θ est un sous ensemble compact de $\mathbb{R}^p$ et $f(\,.\,,\,.\,)$ est une fonction régulière. Prakasa Rao [31] étudie le cas de fonctions de régression moins régulières, tandis que Shao [37] généralise les résultats de Jenrich [21] et de Wu [42] en établissant la consistance et la normalité asymptotique de $\hat{\theta}$ en levant l'hypothèse de compacité pour Θ.

Dans le cas de résidus hétéroscédastiques, de matrice de variance-covariance de la forme $\sigma^2 V$, où σ est un réel positif inconnu et V est une matrice positive connue, on pourra préférer l'estimateur des moindres carrés pondérés. Cet estimateur est obtenu par minimisation de la somme des carrés résiduels pondérés par les coefficients de la matrice inverse V^{-1} (voir par exemple Seber et Wild [35] p. 28).

Lorsque la loi du vecteur résiduel $(\varepsilon_1, ..., \varepsilon_n)$ est connue, on pourra également considérer l'estimateur du maximum de vraisemblance. En fait, toutes ces méthodes sont des exemples d'estimation dite de minimum de contraste, telle qu'introduite par Pfanzagl [27] et [28]. Le comportement asymptotique de ces estimateurs est étudié par Dacunha-Castelle et Duflo [11].

La méthode de jackknife

L'estimateur que nous avons choisi d'étudier dans le cadre de la régression non linéaire résulte de l'application de la méthode de jackknife à l'estimation par moindres carrés. Rappelons quelques définitions (pour une présentation plus complète de la méthode de jackknife, on pourra se référer à Gray et Schucany [15] et Miller [25]).

Soit θ_0 un paramètre inconnu et $\hat{\theta}$ un estimateur de θ_0, construit sur n observations. La méthode de jackknife consiste à éliminer successivement chacune des observations et à recalculer l'estimateur de θ_0 sur les $n-1$ observations restantes. Soit $\hat{\theta}_{-i}$ l'estimateur de θ_0 obtenu lorsque la $i^{ème}$ observation est éliminée. L'estimateur jackknife de θ_0 est défini par

$$\tilde{\theta} = n\hat{\theta} - \frac{n-1}{n}\sum_{i=1}^{n}\hat{\theta}_{-i} \tag{1.5}$$

La méthode de jackknife a été introduite comme technique de réduction du biais. Pour comprendre le fonctionnement de cette méthode, il est intéressant de considérer la situation élémentaire suivante: supposons que pour une constante a, le biais de $\hat{\theta}$ soit de la forme

$$E(\hat{\theta}) - \theta = a/n$$

Alors $\tilde{\theta}$ est un estimateur non biaisé de θ_0. Par exemple, considérons $Y_1, ..., Y_n$ des variables indépendantes et de même loi. L'estimateur empirique de var Y_1 est $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$. L'estimateur jackknife vaut $\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$. On retrouve donc par cette méthode l'estimateur non biaisé usuel de la variance. Plus généralement, si le biais de $\hat{\theta}$ possède un développement asymptotique du type

$$E(\hat{\theta}) - \theta = \frac{a}{n} + O(\frac{1}{n^2})$$

alors, le biais de $\tilde{\theta}$ est d'ordre $\frac{1}{n^2}$. Cette propriété de réduction du biais rend l'utilisation des estimateurs jackknife attractive pour de petits échantillons. On trouvera dans Efron [13] de nombreux exemples d'utilisation de la méthode de jackknife.

Nous nous intéressons à présent à l'utilisation de la méthode de jackknife en régression.

Estimateur jackknife des moindres carrés

Revenons au modèle de régression paramétrique

$$y_i = f(x_i, \theta_0) + \varepsilon_i \tag{1.6}$$

où les (x_i) appartiennent à un compact $\mathcal{X}$ de $\mathbb{R}^d$, θ_0 appartient à un sous ensemble Θ de $\mathbb{R}^p$, et les résidus (ε_i) sont centrés. Nous supposons dans cette introduction (bien que nos résultats restent valables sous des hypothèses moins restrictives) que Θ est compact et que les résidus (ε_i) sont indépendants et équi-distribués.

Appliquons la méthode de jackknife à l'estimation par moindres carrés. Rappelons que l'estimateur des moindres carrés de θ_0 est un p-vecteur $\hat{\theta}$ de Θ qui minimise $\sum_{i=1}^n (y_i - f(x_i, \theta))^2$ sur Θ. On définit, pour tout $i \in \{1, ..., n\}$, $\hat{\theta}_{-i}$ comme estimateur des moindres carrés obtenu lorsque la i$^{\text{ème}}$ observation est éliminée, c'est à dire que $\hat{\theta}_{-i}$ est un p-vecteur de Θ minimisant $\sum_{j \neq i}(y_j - f(x_j, \theta))^2$ sur Θ. Comme nous l'avons mentionné plus haut, l'estimateur jackknife défini par (1.5) possède des propriétés de réduction du biais de l'estimateur des moindres carrés, qui justifient son utilisation pour de petits échantillons. Les propriétés asymptotiques de l'estimateur jackknife en régression (consistance, normalité asymptotique) ont en outre été largement étudiées pour le modèle linéaire (voir Miller[26], Shao[36], Hinkley[20], Weber et Welsh[40]) ainsi que pour certains modèles de régression linéaire généralisée (Shao[38]). Pour ce qui concerne la régression non linéaire, la normalité asymptotique de l'estimateur jackknife a été conjecturée par plusieurs auteurs (Duncan [12], Fox et al[14], Miller[26]) sans démonstration. C'est précisément l'objet du second chapitre de cette thèse que de démontrer ce résultat. Nous montrons que $\sqrt{n}\Gamma(\theta_0)^{-1}(\tilde{\theta} - \theta_0)$ converge en distribution vers un vecteur gaussien standard, pour une matrice positive $\Gamma(\theta_0)$ adequate. Pour appliquer un tel résultat de convergence en loi, il est nécessaire de construire un estimateur consistant de $\Gamma(\theta_0)$. Introduisons pour ce faire les pseudo-valeurs $\tilde{\theta}_i$ suivant la définition de Tukey [39]:

$$\forall i \in \{1, ..., n\} \qquad \tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{-i}$$

Il se trouve que $\tilde{\theta}$ est la moyenne empirique des pseudo-valeurs $\tilde{\theta}_i$, c'est à dire que $\tilde{\theta} = \frac{1}{n}\sum_{i=1}^n \tilde{\theta}_i$. Il est donc naturel (et utile pour les applications que nous détaillons plus loin) de considérer comme estimateur de $\Gamma(\theta_0)$ la racine carrée de la matrice positive

$$\tilde{s}^2 = \frac{1}{n-1}\sum_{i=1}^n (\tilde{\theta}_i - \tilde{\theta})\,{}^t(\tilde{\theta}_i - \tilde{\theta}) \tag{1.7}$$

Nous établissons des conditions suffisantes de consistance de cet estimateur, et en déduisons que $\sqrt{n}\tilde{s}^{-1}(\tilde{\theta} - \theta_0)$ converge en distribution vers une gaussienne standard quand $n \to \infty$. Ces résultats sont obtenus sous des conditions de moment et des hypothèses convenables de régularité.

Shao [37] a établi des conditions suffisantes de normalité asymptotique de l'estimateur des moindres carrés en régression non linéaire. Notre travail consiste donc essentiellement à démontrer l'équivalence asymptotique de l'estimateur jackknife $\tilde{\theta}$ et de l'estimateur des moindres carrés $\hat{\theta}$, à la vitesse $n^{-1/2}$. Ceci implique en particulier l'égalité des variances asymptotiques de $\sqrt{n}(\tilde{\theta} - \theta_0)$ et $\sqrt{n}(\hat{\theta} - \theta_0)$. Nous démontrons alors la consistance de notre estimateur de variance $\tilde{s}^2$ en établissant l'équivalence asymptotique de $\tilde{s}^2$ et d'un autre estimateur

$\hat{s}^2 = (n-1)\sum_{i=1}^{n}(\hat{\theta}_{-i} - \hat{\theta})\ {}^t(\hat{\theta}_{-i} - \hat{\theta})$ dont Shao établit la consistance dans [37]. Plus précisément, nous démontrons le fait suivant: s'il existe un entier $m \geq 2$ tel que $\mathbb{E}|\varepsilon_1|^m$ soit fini, si f est régulière au voisinage de θ_0 et si les estimateurs des moindres carrés obtenus dans le modèle complet et dans les modèles diminués convergent en probabilité et uniformément vers θ_0 (i.e. $\max_{1\leq i\leq n} ||\hat{\theta}_{-i} - \theta_0|| = o_{\mathbb{P}}(1)$ et $||\hat{\theta} - \theta_0|| = o_{\mathbb{P}}(1)$), alors

- $||\hat{\theta} - \tilde{\theta}|| = o_{\mathbb{P}}(n^{3/m-1})$
- $||\hat{s}^2 - \tilde{s}^2|| = o_{\mathbb{P}}(n^{6/m-2})$

(nous utilisons ici la même notation $||\ .\ ||$ pour représenter la norme euclidienne dans $\mathbb{R}^p$ ou la norme associée pour les matrices $p \times p$ à coefficients réels). Ainsi, si $\mathbb{E}|\varepsilon_1|^6$ est fini,

$$\sqrt{n}\tilde{s}^{-1}(\tilde{\theta} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_p(0, I_p) \qquad \text{quand } n \to \infty$$

Pour en venir à la motivation principale de ce chapitre, précisons que ce résultat permet de valider asymptotiquement des pratiques de construction de régions de confiance et de tests couramment utilisées et qu'on peut situer comme suit. Elles se fondent sur l'heuristique de Tukey [39] qui, définissant les pseudo-valeurs $(\tilde{\theta}_i)$, postule qu'elles se comportent comme des variables indépendantes et de même loi normale. En s'appuyant sur ce postulat, on peut dès lors réaliser une analyse de variance comme dans un modèle linéaire gaussien, à partir des pseudo-valeurs. Citons à présent deux applications explicites de notre résultat de convergence qui confortent le point de vue développé ci-dessus.

- *Construction de régions de confiance pour θ_0.*
  Soit χ_p^2 une variable aléatoire de loi du chi deux à p degrés de liberté et $C_{p,\alpha}$ le quantile défini par $\mathbb{P}(\chi_p^2 > C_{p,\alpha}) = 1 - \alpha$. Alors, sous des hypothèses de régularité convenables, $\mathcal{R} = \left\{\theta \in \Theta,\ n\ {}^t(\tilde{\theta} - \theta)\tilde{s}^{-2}(\tilde{\theta} - \theta) \leq C_{p,\alpha}\right\}$ est une région de confiance pour θ_0 de niveau asymptotique α.

- *Test d'homogénéité de fonctions de régression.*
  Considérons k modèles de régression de la forme

$$y_i^{(j)} = f(x_i^{(j)}, \theta_0^{(j)}) + \varepsilon_i^{(j)} \qquad 1 \leq i \leq n_j \qquad 1 \leq j \leq k$$

où les résidus sont indépendants et de même loi centrée, et pour tout $j \in \{1, ..., k\}$, $\theta_0^{(j)}$ est un paramètre inconnu. Notons $(\tilde{\theta}_i^{(j)})$ les pseudo-valeurs obtenues dans le $j^{ème}$ modèle et $\tilde{\theta}^{(j)}$ l'estimateur jackknife de $\theta_0^{(j)}$. Soit $\tilde{\theta}_c$ la moyenne empirique des estimateurs jackknife (c'est à dire $\tilde{\theta}_c = \frac{1}{k}\sum_{j=1}^{k}\tilde{\theta}^{(j)}$) et $\tilde{s}_c^2$ la matrice

$$\tilde{s}_c^2 = \frac{1}{n-k}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\tilde{\theta}_i^{(j)} - \tilde{\theta}^{(j)})\ {}^t(\tilde{\theta}_i^{(j)} - \tilde{\theta}^{(j)})$$

où $n = \sum_{j=1}^{k} n_j$. Sous des hypothèses de régularité convenables, le test défini par la région de rejet

$$\left\{\sum_{j=1}^{k} n_j\ {}^t(\tilde{\theta}^{(j)} - \tilde{\theta}_c)\tilde{s}_c^{-2}(\tilde{\theta}^{(j)} - \tilde{\theta}_c) > C_{p(k-1),\alpha}\right\}$$

où $C_{p(k-1),\alpha}$ est le quantile défini par $\mathbb{P}(\chi^2_{p(k-1)} > C_{p(k-1),\alpha}) = 1 - \alpha$ est un test de l'hypothèse $\theta_0^{(1)} = ... = \theta_0^{(k)}$ contre l'alternative $\exists j, j' :\ \theta_0^{(j)} \neq \theta_0^{(j')}$ de niveau asymptotique $1 - \alpha$.

Afin d'illustrer la méthode de jackknife présentée ci-dessus, nous avons étudié des données réelles et effectué quelques tests d'homogénéité. Ces données sont les résultats d'expériences portant sur l'évolution en milieu lacté d'un taux de bactéries. Nous disposons de quatre ensembles de données, chacun étant constitué de mesures, à différentes dates, du taux de bactéries d'un milieu lacté placé à une température donnée. Nous avons déterminé les estimations jackknife pour chacune des courbes de croissance au moyen d'un programme réalisé à l'aide du logiciel S.A.S. Nous avons alors pu tester l'influence de la température sur l'évolution du taux de bactéries.

Considérons à présent un cadre non paramétrique.

Estimation en régression non paramétrique

Nous considérons à nouveau le modèle de régression (1.2). On vise ici à estimer f en imposant des contraintes beaucoup moins restrictives que dans le paragraphe précédent. Typiquement, c'est une contrainte a priori sur la régularité ou la monotonie de f qui, dans l'approche non paramétrique, se substitue à la donnée d'un ensemble $\{f(\,.\,,\theta),\ \theta \in \Theta\}$ dans l'approche paramétrique.

De façon à mieux situer la méthode d'estimation que nous avons choisi d'étudier ainsi que les résultats de convergence que nous avons en vue de démontrer, nous présentons tout d'abord quelques méthodes d'estimation non paramétrique classiques.

Les méthodes linéaires: noyau et projection

Les méthodes les plus élémentaires sont linéaires; ce sont les méthodes de noyau ou de projection. Dans les deux cas, on estime une approximation de f de moindre complexité, définie soit comme la convolée de f avec une fonction régulière (dite noyau) soit comme projection de f sur un espace de fonctions de dimension finie.

La méthode du noyau fut tout d'abord introduite pour l'estimation non paramétrique d'une densité (Rosenblatt [32]), puis adaptée par Priestley et Chao [29] à l'estimation d'une fonction de régression. Elle consiste à estimer la convolée de f avec un noyau $K_{h_n}(\,.\,) = h_n^{-1}K(\,.\,h_n^{-1})$, où K est une fonction régulière et $\{h_n\}_{n\in\mathbb{N}}$ est une suite de réels positifs convergeant vers 0 quand $n \to \infty$. Priestley et Chao [29] ont étudié la consistance ponctuelle d'estimateurs à noyau lorsque la fonction de régression et le noyau sont supposés Lipschitziens. Benedetti [2] a généralisé ce résultat et établi la normalité asymptotique ponctuelle d'estimateurs à noyau. Cheng et Lin [8] ont ensuite évalué la vitesse de convergence uniforme de ces estimateurs.

Nous énoncons ici un résultat, dû à Csörgö et Horváth [10], de convergence pour l'estimateur à noyau d'une densité. Bien que ce résultat ait été démontré dans un cadre différent du notre (l'estimation d'une densité, et non pas l'estimation d'une fonction de régession), il permettra de mieux situer notre propre résultat. Il s'agit en effet d'un théorème de limite centrale pour la distance $\mathbb{L}_p$ entre une densité et son estimateur à noyau f_n. Nous nous contentons de l'énoncer pour $p = 1$, lorsque le support de la densité à estimer est compact. Soit f une densité à support compact. Si f est deux fois dérivable, le noyau K est suffisamment régulier et si $h_n = n^{-\alpha}$ avec

$\frac{1}{4} < \alpha < \frac{1}{2}$, alors il existe des constantes m et σ^2 dépendant de K telles que

$$h_n^{-1/2}\left\{\sqrt{nh_n}\int |f_n(t)-f(t)|dt - m\int f^{1/2}(t)dt\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2) \qquad \text{quand } n\to\infty \tag{1.8}$$

De même que les estimateurs à noyau, les estimateurs par projection ont initialement été introduits afin de résoudre des problèmes d'estimation non paramétrique d'une densité par rapport à une mesure donnée μ (Cencov [7]) avant d'être adaptés au problème de l'estimation d'une fonction de régression. La méthode de projection consiste à estimer la projection de f sur un sous-espace vectoriel de $\mathbb{L}_2(\mu)$ donné de dimension N_n, où la suite $\{N_n\}_{n\in\mathbb{N}}$ vérifie $\lim_{n\to\infty} N_n = \infty$. Plus précisément, soit $\{\phi_j\}_{1\leq j\leq N_n}$ une famille orthonormée connue de fonctions (les fonctions ϕ_j dépendant éventuellement de n) telles qu'il existe des coefficients β_j $(1\leq j\leq N_n)$ pour lesquels f soit convenablement approchée par la fonction $\sum_{j=1}^{N_n}\beta_j\phi_j$. Alors, l'estimateur par projection de f est obtenu par

$$\hat{f}_n = \sum_{j=1}^{N_n}\hat{\beta}_j\phi_j \tag{1.9}$$

où pour tout $j\in\{1,...,N_n\}$, $\hat{\beta}_j$ représente (en régression) l'estimateur de β_j ajusté par moindres carrés dans le modèle de régression linéaire

$$y_i = \sum_{i=1}^{N_n}\beta_j\phi_j(x_i)+\varepsilon_i \qquad 1\leq i\leq n$$

Ainsi, en projetant f sur l'espace de fonctions engendré par la famille $\{\phi_j\}_{1\leq j\leq N_n}$, on ramène le problème d'estimation de f au problème d'estimation d'un nombre fini de paramètres. Par exemple, si f admet une représentation en série

$$f = \sum_{j=1}^{\infty}\beta_j\phi_j$$

où $\{\phi_j\}_{j\in\mathbb{N}}$ est une base connue de fonctions et $\{\beta_j\}_{j\in\mathbb{N}}$ sont des coefficients inconnus, on considère la projection de f sur l'espace engendré par les N_n premiers termes de la base $\{\phi_j\}_{j\in\mathbb{N}}$ pour obtenir l'estimateur (1.9).

Un cas particulier de l'estimation par projection est l'estimation par histogramme. Supposons f définie sur $[0,1]$. Soit $I_1,...,I_{N_n}$ des intervalles formant une partition de $[0,1]$, et pour tout $j\in\{1,...,N_n\}$, soit $\phi_j = \frac{1}{\sqrt{|I_j|}}\mathbb{I}_{I_j}$, où $|I_j|$ désigne la longueur de l'intervalle I_j. On considère alors la projection de f sur le sous-espace de $L^2[0,1]$ engendré par la famille orthonormée $\{\phi_j\}_{1\leq j\leq N_n}$.

C'est encore un résultat obtenu dans le cadre de l'estimation d'une densité que nous avons choisi d'énoncer afin d'introduire notre propre résultat, la littérature dans ce cadre étant bien plus fournie que celle relative à l'estimation d'une fonction de régression. Ce résultat, dû à Kogure [23], fournit l'ordre de grandeur du risque minimum pour un estimateur par histogramme. Si f est une densité à support compact,

$$\limsup_n\left[\inf_{\hat{f}_n} n^{1/3}\mathbb{E}\int|\hat{f}_n(t)-f(t)|dt\right] = c\int|f(t)f'(t)|^{1/3}dt \tag{1.10}$$

où l'infimum est pris sur l'ensemble des estimateurs par histogramme et c est une constante absolue. Ainsi, le risque $\mathbb{L}_1$ du meilleur histogramme est d'ordre $n^{-1/3}$. Cependant, ce résultat ne permet pas de construire le meilleur histogramme, c'est à dire de déterminer la partition optimale du support de f. En effet, la partition optimale dépend étroitement de la fonction inconnue f. En l'absence d'informations a priori sur f, on choisira une partition régulière, vraisemblablement non optimale.

Les estimateurs isotoniques (en estimation de fonction de régression) et de Grenander (en estimation de densité) sont des estimateurs par histogramme construits automatiquement à partir des observations, c'est à dire sans connaissance a priori (autre que la monotonie) sur la fonction à estimer. C'est la répartition même des observations qui guide le choix d'une bonne partition.

Estimation isotonique

Grenander [16] a introduit un estimateur non paramétrique pour une densité monotone ou unimodale, à présent connu sous le nom d'estimateur de Grenander, pour lequel aucun choix arbitraire de paramètre n'est nécessaire. Cet estimateur est comparable aux meilleurs estimateurs par histogrammes. En effet, Birgé [4] a démontré que le risque $\mathbb{L}_1$ de l'estimateur de Grenander (c'est à dire l'espérance de la distance $\mathbb{L}_1$ entre l'estimateur de Grenander et la fonction de densité à estimer) est majoré, à une constante multitplicative près, par l'infimum des risques $\mathbb{L}_1$ des estimateurs par histogrammes. Cela signifie que l'estimateur de Grenander est un estimateur par histogramme qui s'adapte automatiquement à la forme de la densité à estimer pour choisir la subdivision optimale de $[0,1]$. Les propriétés non asymptotiques de l'estimateur de Grenander ont été étudiées par Birgé [4], tandis que ses propriétés asymptotiques ponctuelles ont été étudiées par Prakasa Rao [30]. Une présentation complète de l'estimation isotonique (c'est à dire sous contrainte d'ordre) peut être trouvée dans Barlow et al [1].

Groeneboom s'est interessé au comportement asymptotique global de l'estimateur de Grenander, et plus particulièrement à la normalité asymptotique de la distance $\mathbb{L}_1$ entre la densité à estimer et son estimateur de Grenander. Dans [17], il annonce le résultat suivant. Soit f une densité décroissante de support [0,1] et f_n l'estimateur de Grenander de f. Si f est suffisament régulière, alors

$$n^{1/6}\left\{n^{1/3}\int_0^1 |f_n(t)-f(t)|dt - D_f\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2) \qquad \text{quand } n\to\infty \tag{1.11}$$

avec $D_f = C\int_0^1 |f'(t)f(t)|^{1/3}dt$. C et σ^2 sont des constantes absolues: $C \approx 0.33$ et $\sigma^2 \approx 0.17$. Groeneboom propose dans [17] une ligne de preuve pour ce résultat. Une démonstration complète en a été établie très récemment par Groeneboom, Hooghiemstra et Lopuhaä [19].

L'attrait de l'estimateur de Grenander, comparé à un estimateur à noyau, réside dans ses propriétés d'adaptivité. En particulier, un inconvénient du théorème de normalité asymptotique (1.8) de Csörgö et Horváth est qu'il concerne un estimateur à noyau dont la fenêtre est à choisir par le statisticien, alors que la description de l'estimateur de Grenander ne dépend que des données. Un prix à payer est l'hypothèse de monotonie, qui est fondamentale pour un bon fonctionnement de l'estimateur de Grenander. Une autre difficulté, technique celle la, liée à l'estimateur de Grenander est qu'il est non linéaire, ce qui rend la démonstration d'un résultat comme (1.11) plus délicate que celle de (1.8).

Brunk [5] a défini un estimateur pour une fonction de régression monotone analogue à l'estimateur de Grenander pour une fonction de densité. C'est cet estimateur, généralement appelé estimateur isotonique, que nous avons choisi d'étudier dans le cadre de la régression non paramétrique monotone. L'estimateur isotonique f_n de f est défini comme l'estimateur non paramétrique des moindres carrés sous la contrainte d'ordre $f_n(x_1) \geq ... \geq f_n(x_n)$. f_n se trouve être la pente du plus petit majorant concave de la fonction de répartition empirique F_n définie par

$$\forall t \in [0,1] \qquad F_n(t) = \frac{1}{n}\sum_{i=1}^{n} y_i \mathbb{I}_{x_i \leq t}$$

Brunk [6] a démontré la convergence ponctuelle de l'estimateur isotonique et déterminé la distribution limite de $n^{1/3}(f_n(t) - f(t))$ en un point t fixé tel que $f'(t) \neq 0$. Wright [41] a généralisé ce résultat au cas de fonctions de régression moins régulières. Dans son théorème 1, il suppose que f satisfait $|f(t) - f(x)| = A|x - t|^\alpha(1 + o(1))$ quand $x \to t$, où α et A sont des constantes positives, et montre que dans ce cas, la vitesse de convergence est $n^{-\alpha/2\alpha+1}$. La démonstration de Brunk est inspirée de la démonstration d'un résultat analogue de convergence ponctuelle pour l'estimateur de Grenander d'une densité monotone (Prakasa Rao [30]).

Nous étudions dans cette thèse le comportement asymptotique de l'estimateur isotonique en régression, dans le cas où la fonction de régression f est définie sur $[0,1]$, les résidus (ε_i) sont indépendants et de même loi centrée et les variables explicatives (x_i) sont uniformément réparties sur $[0,1]$: $\forall i \in \{1,...,n\}$, $x_i = i/n$. Nous montrons que, sous des hypothèses convenables, la distance $\mathbb{L}_1$ entre l'estimateur isotonique et f converge à vitesse $n^{-1/3}$ vers une fonctionnelle de f. De plus, cette distance, convenablement recentrée, converge à vitesse $n^{-1/2}$ vers une loi gaussienne centrée dont la variance est libre de f. Plus précisément, si f est suffisament régulière et si ε_1 admet un moment d'ordre au moins 12, alors

$$n^{1/6}\{n^{1/3}\int_0^1 |f_n(t) - f(t)|dt - C_f\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2) \qquad \text{quand } n \to \infty \tag{1.12}$$

avec $C_f = C\int_0^1 |f'(t)|^{1/3}dt$. C et σ^2 sont des constantes absolues: $C \approx 0.33$ et $\sigma^2 \approx 0.17$. Nous obtenons donc pour l'estimateur isotonique en régression un résultat de normalité asymptotique analogue à celui démontré par Groeneboom et al [19] pour l'estimateur de Grenander. La preuve de Groeneboom et al (qui a été établie indépendemment de la notre) repose essentiellement sur des travaux antérieurs de Groeneboom [18] portant sur les mouvements Browniens avec dérive quadratique, tandis que notre propre preuve s'inspire davantage des heuristiques de Groeneboom [17].

Notre démarche consiste à approcher le modèle de régression initial par un modèle de bruit blanc grâce à une approximation forte (nous employons un raffinement, dû à Sakhanenko [34], de la construction de Komlós, Major et Tusnády). Ceci revient à approcher $F_n(t)$ par $\frac{1}{\sqrt{n}}W(t) + F(t)$, où F est la primitive de f nulle en zéro et W est un mouvement Brownien standard. Il s'agit alors de démontrer (1.12) avec $\tilde{f}_n$ en lieu et place de f_n, où $\tilde{f}_n$ est l'estimateur isotonique défini comme étant la pente du plus petit majorant concave de $\frac{1}{\sqrt{n}}W + F$. Les propriétés de changement d'échelle ou d'origine sur le mouvement Brownien seront abondamment exploitées, ce qui motive l'introduction d'une approximation bruit blanc. Nous montrons en particulier que f_n et $\tilde{f}_n$ sont équivalents vis à vis du type de convergence que nous visons, c'est à dire que

$$\mathbb{E}\int |\tilde{f}_n(t) - f_n(t)|dt = o(n^{-1/2}) \tag{1.13}$$

La preuve du résultat de convergence (1.12) en modèle de bruit blanc est assez longue. Nous en donnons ici les principaux arguments. Notons tout d'abord que la fonction inverse (généralisée) U_n de l'estimateur isotonique $\tilde{f}_n$ admet une expression plus explicite que $\tilde{f}_n$. En effet, U_n peut encore s'écrire

$$\forall a \in \mathbb{R} \qquad U_n(a) = \operatorname*{argmax}_{u\in[0,1]}\{W(u) + \sqrt{n}(F(u) - au)\}$$

En outre, si g désigne la fonction inverse de f définie sur $\mathbb{R}$,

$$\int_0^1 |\tilde{f}_n(t) - f(t)|dt = \int_{\mathbb{R}} |U_n(a) - g(a)|da \tag{1.14}$$

Il nous suffit donc d'établir un théorème de limite centrale pour la distance $\mathbb{L}_1$ entre U_n et g, et de calculer le biais asymptotique de cette distance. Dans la suite, f est supposée continue et $[m, M]$ désigne l'image de $[0, 1]$ par f.

Nous établissons l'inégalité exponentielle suivante: il existe une constante positive C_0 ne dépendant que de f telle que

$$\forall a \in \mathbb{R}\ \forall x > 0, \quad \mathbb{P}(n^{1/3}|U_n(a) - g(a)| > x) \leq 2\exp(-C_0x^3)$$

Cette inégalité constitue un outil essentiel utilisé à plusieurs endroits de la preuve. Elle contribue en particulier à justifier le fait que

$$\mathbb{E}\int_{\mathbb{R}} |U_n(a) - g(a)|da = \mathbb{E}\int_m^M |U_n(a) - g(a)|da + o(n^{-1/2})$$

Aussi nous suffit-il d'étudier le comportement asymptotique de $\int_m^M |U_n(a) - g(a)|da$. Nous démontrons tout d'abord un théorème de limite centrale pour $\int_m^M |U_n(a) - g(a)|da$ en introduisant un processus B_n^* à trajectoires dans $C([m, M])$, dont nous prouvons la convergence fonctionnelle vers un processus gaussien à la vitesse $n^{-1/2}$. Soit en effet

$$\forall t \in [m, M] \qquad B_n^*(t) = \int_m^t |U_n(a) - g(a)|da - \mathbb{E}\int_m^t |U_n(a) - g(a)|da$$

La normalité asymptotique de $\sqrt{n}B_n^*(M)$ découle a fortiori de la convergence fonctionnelle et établit le résultat cherché. La convergence fonctionnelle est établie en s'appuyant sur un théorème de Billingsley [3]. Il s'agit de garantir la convergence en variance, l'indépendance asymptotique des accroissements et la tension. Une conséquence immédiate de l'inégalité exponentielle ci-dessus est que pour tout s et t dans $[m, M]$, pour tout entier p,

$$\mathbb{E}||B_n^*(t) - B_n^*(s)||_p = |t - s|O(n^{-1/3})$$

Les points délicats ici sont évidemment de "gagner" la vitesse $n^{-1/2}$ depuis cette vitesse $n^{-1/3}$, et de calculer explicitement la variance asymptotique d'un accroissement de $\sqrt{n}B_n^*$. Nous utilisons pour cela la faible dépendance des acrroissements de $\sqrt{n}B_n^*$. Nous montrons en effet que les "petits" accroissements de $\sqrt{n}B_n^*$ peuvent être approchés par des variables indépendantes convenablement construites, auxquelles il est possible d'appliquer l'inégalité de Rosenthal. Nous obtenons ainsi un contrôle des moments de tout ordre des accroissements de $\sqrt{n}B_n^*$. L'intégrabilité uniforme se prouve alors par des arguments standards. La tension de la suite de processus $\{\sqrt{n}B_n^*\}_{n\in\mathbb{N}}$ est quant à elle obtenue par un argument de chainage grâce aux inégalités de moment.

Une des difficultés qu'il faut surmonter aussi bien lors de l'évaluation asymptotique de $n^{1/3}\mathbb{E}\int_m^M |U_n(a) - g(a)|da$ que pour justifer l'approximation bruit blanc, est qu'il convient d'établir que si un processus est uniformément proche d'un mouvement Brownien avec dérive dominée par une parabole, alors les lieux où les maximums sont atteints sont proches. Ceci est garanti par une étude de la vitesse avec laquelle le mouvement Brownien avec dérive minorée par une parabole "fuit" après avoir atteint son maximum.

Nous prouvons également un résultat plus faible que (1.12), mais valable dès que les termes résiduels ε_i admettent un moment d'ordre 2. Si pour tout $i \in \{1, ..., n\}$, $x_i = i/n$ et si la fonction de régression f est suffisament régulière, alors

$$\mathbb{E}\int_0^1 |f_n(t) - f(t)|dt = O(n^{-1/3}).$$

Plan de cette thèse

Le second chapitre est consacré à l'étude du comportement asymptotique de l'estimateur jackknife en régression paramétrique, et plus précisément à la démonstration de la convergence en loi de cet estimateur convenablement normalisé vers une gaussienne standard. En outre, nous présentons plus en détails dans ce chapitre les applications de ces résultats à la construction de régions de confiance et de tests d'hypothèse.

Dans le chapitre 3, nous illustrons par une étude de données réelles la méthode d'estimation et de tests d'hypothèse présentée au second chapitre. Nous montrons ainsi sur un exemple comment déterminer pratiquement les estimations jackknife, et proposons un programme de calcul réalisé au moyen du logiciel S.A.S.

Enfin, le chapitre 4 est dévolu à l'étude du comportement asymptotique de l'estimateur isotonique en régression non paramétrique.

Deux annexes viennent ensuite compléter cette thèse. Nous y démontrons l'invariance de la distance $\mathbb{L}_1$ entre deux fonctions monotones par passage à l'inverse de ces fonctions (soit l'égalite (1.14)) et un raffinement du théorème de convergence fonctionnelle de Billingsley mentionné plus haut.

Les chapitres 2 et 4 correspondent chacun à un article.

Partie I

Méthode de jackknife en régression non linéaire

Chapitre 2

Asymptotic normality of the jackknife least-squares estimator in non linear regression

The purpose of this paper is to establish the asymptotic normality of the jackknife least-squares estimator for the regression parameter in non linear regression. The main motivation is to provide asymptotic justifications for empirical methods consisting in using the so called jackknife pseudo-values in the usual analysis of variance inference. These methods were originally founded on heuristics suggested by Tukey about the weak asymptotic dependency of these pseudo-values. Our results allow us to valid the analysis of variance on the pseudo-values asymptotically.

2.1 Introduction

Jackknife techniques have been introduced as a method for bias reduction (see Efron [13] for examples and Miller [25] for a review). So, jackknife estimation is widely used for estimating an unknown parameter or the variance matrix of some asymptotically efficient estimator, in the hope to get a good small sample performance. Asymptotic normality of the jackknife estimator for a parameter and consistency of the jackknife estimator for the variance matrix have been studied in the past in several contexts. The problem of jackknifing linear regression models has been extensively studied (see for example Miller [26], Shao [36], Hinkley [20], Weber and Welsh [40]). Shao [38] examined the case of generalized linear models. Only a few authors considered the context of non linear regression. Duncan [12] and Fox et al [14] stated empirical results for the jackknife in non linear regression, while Shao [37] established consistency of a jackknifed estimator for the least squares estimator's variance matrix.

The aim of this paper is to prove asymptotic efficiency results for jackknifed estimators in nonlinear regression. Our main motivation is to provide asymptotic justifications for the empirical methods of analysis of variance, based on the jackknife pseudo-values introduced by Tukey [39]. Throughout this paper the following nonlinear regression model is considered

$$y_i = f(x_i, \theta_0) + \varepsilon_i \quad i = 1, ..., n \tag{2.1}$$

where the x_i's are deterministic vectors, y_i is the observation at point x_i and θ_0 is an unknown parameter. The regression function f is defined over a subset $\mathcal{X} \times \Theta$ of $\mathbb{R}^d \times \mathbb{R}^p$ where $\mathcal{X}$ is compact. The ε_i's are unobservable independent random variables with zero mean and finite unknown standard deviation. The errors are allowed to be heteroscedastic, more precisely we assume that ε_i/σ_i are i.i.d. for some unknown constants σ_i.

The jackknife estimators are constructed as linear combinations of least-squares estimators. The least-squares estimator of θ_0 based on the n data points (x_i, y_i) is a p-vector $\hat{\theta}$ of Θ which minimizes $\Sigma_n(\theta) = \sum_{i=1}^n (y_i - f(x_i, \theta))^2$ over Θ. The jackknife method consists in sequentially deleting data points (x_i, y_i), and recomputing the least-squares estimator. Let $\hat{\theta}_{-i}$ be the least-squares estimator when the ith data point is deleted, which means that $\hat{\theta}_{-i}$ is a p-vector of Θ which minimizes $\Sigma_{-i}(\theta) = \sum_{j\neq i}(y_j - f(x_j, \theta))^2$ over Θ. Then the pseudo-values as defined by Tukey [39] are the p-vectors $\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{-i}$. Their empirical mean

$$\tilde{\theta} = n\hat{\theta} - \frac{n-1}{n}\sum_{i=1}^{n}\hat{\theta}_{-i} \tag{2.2}$$

is the jackknife estimator of θ_0. We shall show in this paper that (under some appropriate assumptions) $\sqrt{n}\Gamma^{-1}(\theta_0)(\tilde{\theta} - \theta_0)$ converges to a standard multivariate gaussian distribution for some adequate positive matrix $\Gamma(\theta_0)$. Moreover, we shall see that this convergence holds true if one substitute to $\Gamma(\theta_0)$ some jackknife estimator $\tilde{s}$. The estimator that we propose is defined as the square root of the nonnegative matrix

$$\tilde{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\tilde{\theta}_i - \tilde{\theta})\,{}^t(\tilde{\theta}_i - \tilde{\theta}) \tag{2.3}$$

This generalizes on asymptotic results for the jackknife estimators for multiple linear models which have been proved by Miller [26] and Weber and Welsh [40]. Moreover, it provides a proof to Miller's claim "although the proofs have not been worked out, the jackknife should extend to the case of nonlinear least-squares".

Let us describe now some applications of such asymptotic results to the construction of confidence regions and tests of comparison.

Tukey [39] suggested that in most cases, the pseudo-values $\tilde{\theta}_1, ..., \tilde{\theta}_n$ can be treated as n approximatively independent identically distributed variables for constructing a confidence interval for instance. Miller [25] shows that, in the context of linear regression models, this "rule of thumb" can be rigourously justified. In the same spirit, we have in view to describe various questions in the context of non linear regression, where our results provide asymptotic justifications to methods of confidence region constructing or hypothesis testing which are currently used because of their simplicity and which are based on Tukey's heuristics. For confidence regions, one can more precisely derive from the weak convergence of $\sqrt{n}\tilde{s}^{-1}(\tilde{\theta} - \theta_0)$ to a standard normal, that $n\,{}^t(\tilde{\theta} - \theta_0)\tilde{s}^{-2}(\tilde{\theta} - \theta_0)$ converges in distribution towards a random variable χ^2_p distributed as a chi-square with p degrees of freedom. Let $C_{p,\alpha}$ be some quantile of the chi-square distribution with p degrees of freedom, that is $\mathbb{P}(\chi^2_p > C_{p,\alpha}) = 1 - \alpha$. Then,

$$\mathcal{R} = \left\{\theta \in \Theta,\ n\,{}^t(\tilde{\theta} - \theta)\tilde{s}^{-2}(\tilde{\theta} - \theta) \leq C_{p,\alpha}\right\}$$

is a confidence region for θ_0 with asymptotic level equal to α. If the parameter of interest is $\beta_0 = h(\theta_0)$, where h is a smooth enough function defined from Θ to $\mathbb{R}^q$ $(q \leq p)$, our convergence results prove that under suitable regularity conditions, we can construct confidence regions in the same way. A confidence region with asymptotic level α for β_0 is

$$\mathcal{R}_h = \{\beta \in Im(h), n\, {}^t(\tilde\beta - \beta)\tilde s_h^{-2}(\tilde\beta - \beta) \le C_{p,\alpha}\}$$

where $\tilde\beta = nh(\hat\theta) - \frac{n-1}{n}\sum_{i=1}^n h(\hat\theta_{-i})$ and $\tilde s_h^2$ is the empirical variance of the pseudo-values $\tilde\beta_i = nh(\hat\theta) - (n-1)h(\hat\theta_{-i})$.

Another interesting problem where our results can turn to be useful is the question of comparing different regression parameters. Let us consider for instance the problem of testing homogeneity between different regression functions. This problem occurs when one wants for example to study a phenomenon in several contexts and one wishes to test the influence of the context. Suppose that in each context, the phenomenon can be modelized by a nonlinear regression involving the same parametric regression function f defined from $\mathcal{X} \times \Theta$ to $\mathbb{R}$. Consider the experiment in the jth context ($j \in \{1, ..., k\}$, where k is the number of contexts considered).

$$y_i^{(j)} = f(x_i^{(j)}, \theta_0^{(j)}) + \varepsilon_i^{(j)}, \quad i = 1, ..., n_j$$

where $\theta_0^{(j)}$ denotes the true unknown parameter in the jth model. Then, comparison of the k regression models reduces to comparison of the k parameters $\theta_0^{(j)}$. This comparison can be made using a (multidimensional if $p > 1$) analysis of variance on pseudo-values, whenever assumptions of theorem 2.1 hold true for each model and whenever the experiments have been made independently of each other. Let $\tilde\theta_i^{(j)}$ be the ith pseudo-value obained in the jth model and $\tilde\theta^{(j)}$ the jackknife estimator of $\theta_0^{(j)}$. Let $\tilde\theta_c$ denotes the empirical mean of the jackknife estimators in the k models

$$\tilde\theta_c = \frac{1}{k}\sum_{j=1}^{k}\tilde\theta^{(j)} \tag{2.4}$$

and $\tilde s_c^2$ denotes the matrix

$$\tilde s_c^2 = \frac{1}{n-k}\sum_{j=1}^{k}\sum_{i=1}^{n}(\tilde\theta_i^{(j)} - \tilde\theta^{(j)})\, {}^t(\tilde\theta_i^{(j)} - \tilde\theta^{(j)}) \tag{2.5}$$

where $n = \sum_{j=1}^k n_j$. Consider the following $\mathcal{U}_p$ statistic of analysis of variance

$$\mathcal{U}_p = \sum_{j=1}^{k} n_j\, {}^t(\tilde\theta^{(j)} - \tilde\theta_c)\tilde s_c^{-2}(\tilde\theta^{(j)} - \tilde\theta_c).$$

$\mathcal{U}_p/(n-k)$ is the Hotelling-Lawley statistic of the following model of analysis of variance over the pseudo-values

$$\begin{pmatrix} {}^t\tilde\theta_1^{(1)} \\ \vdots \\ {}^t\tilde\theta_{n_1}^{(1)} \\ \vdots \\ {}^t\tilde\theta_1^{(k)} \\ \vdots \\ {}^t\tilde\theta_{n_k}^{(k)} \end{pmatrix} = \begin{pmatrix} 1_{n_1} & & 0 \\ & \ddots & \\ 0 & & 1_{n_k} \end{pmatrix} \begin{pmatrix} {}^t\theta_0^{(1)} \\ \vdots \\ {}^t\theta_0^{(k)} \end{pmatrix} + \begin{pmatrix} {}^t\gamma_1^{(1)} \\ \vdots \\ {}^t\gamma_{n_1}^{(1)} \\ \vdots \\ {}^t\gamma_1^{(k)} \\ \vdots \\ {}^t\gamma_{n_k}^{(k)} \end{pmatrix}$$

where the $\gamma_i^{(j)}$ are random p-vectors. The Hotelling-Lawley test is asymptotically valid iif $\mathcal{U}_p$ is asymptotically distributed under the true model as a chi square with $p(k-1)$ degrees of freedom. Provided that the $(\varepsilon_i^{(j)})_{i,j}$ are i.i.d. this in fact is a consequence of our asymptotic normality results (namely theorem 2.1 below). So, the following region $\mathcal{T}$ is an available reject region for the test of homogeneity of the k parameters, with asymptotic level equal to α

$$\mathcal{T} = \{\mathcal{U}_p > C_{p(k-1),\alpha}\}$$

where $C_{p(k-1),\alpha}$ is the quantile defined by $\mathbb{P}(\chi^2_{p(k-1)} > C_{p(k-1),\alpha}) = 1 - \alpha$.

Note that all statistics involved in this paper (for example $\hat{\theta}$, $\tilde{\theta}$, $\tilde{s}^2$) depend on n, but for notational simplicity, this subscript will be omitted throughout the paper.

2.2 Main results

In this section, we state some asymptotic results for the jackknife estimator of the unknown parameter θ_0 and the variance matrix. We shall prove that under suitable smoothness conditions on f and moments conditions on the errors ε_i, the jackknife estimator $\tilde{\theta}$ of θ_0 normalized by $\sqrt{n}$ is asymptotically normally distributed and the sample variance $\tilde{s}^2$ of the associated pseudo-values provides a consistent estimator for the asymptotic variance of $\sqrt{n}(\tilde{\theta} - \theta_0)$. Shao [37] gave sufficient conditions for the least-squares estimator $\hat{\theta}$ to be asymptotically efficient. Our main task will be to show that $\tilde{\theta}$ is asymptotically equivalent to $\hat{\theta}$ at the rate $n^{-1/2}$. This will in particular implies that $\sqrt{n}(\tilde{\theta} - \theta_0)$ and $\sqrt{n}(\hat{\theta} - \theta_0)$ have the same asymptotic variance matrix. This will allow us to derive the consistency of our estimator $\tilde{s}^2$ of the variance matrix from the consistency of the estimator $\hat{s}^2$ proposed by Shao in his analysis of the asymptotic behaviour of $\hat{\theta}$.

We shall assume some regularity conditions for the regression function f and the errors ε_i, and some topological properties for Θ and $\mathcal{X}$, in such a way that the least-squares estimator is asymptotically gaussian and the jackknife estimator of variance proposed by Shao is consistant. Shao [37] proved that under the following assumptions $\mathcal{S}_n$, the least-squares estimator of θ_0 is asymptotically gaussian, and under assumptions $\mathcal{S}_n$ and $\mathcal{S}_c$, his jackknife estimator of variance is consistent.

$\mathcal{S}_n$: For each $i \in \{1, ..., n\}$, let f_i denote the function defined over Θ which takes θ into $f(x_i, \theta)$.

1. For all i, there is some centered random variable δ_i with unit variance and some scalar σ_i such that $\varepsilon_i = \sigma_i \delta_i$ and the δ_i's are i.i.d. Furthermore, there exist some positive reals σ_0 and σ_∞ such that

$$0 < \sigma_0 \le \min_{1 \le i \le n} \sigma_i \le \max_{1 \le i \le n} \sigma_i \le \sigma_\infty < \infty$$

2. f is continuous over $\mathcal{X} \times \Theta$.
3. Θ is a subset of $\mathbb{R}^p$ such that $\theta_0 \in \Theta$, and Θ satisfies either a) or b) or c)

   a) Θ is unbounded and as $||\theta|| \to \infty$, $|f(x, \theta)| \to \infty$ uniformly for x.

b) Θ is unbounded and there exists a constant a such that $|f(x,\theta)| \to a$ as $||\theta|| \to \infty$ uniformly for x and

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\{f_i(\theta) - a\}^2 > 0$$

c) Θ is compact.

4. $\mathcal{X}$ is a compact subset of $\mathbb{R}^d$ non reduced to a single point.
5. For any positive constants c and δ,

$$\liminf_{n\to\infty} \inf_{\theta\in\mathcal{A}(\delta,c)} \frac{1}{n}\sum_{i=1}^{n}(f_i(\theta) - f_i(\theta_0))^2 > 0$$

where $\mathcal{A}(\delta, c) = \{\theta \in \Theta : \delta \leq ||\theta - \theta_0|| \leq c\}$.

$\mathcal{S}_c$: There exists some open neighborhood $\mathcal{V}(\theta_0)$ of θ_0 over which f is twice differentiable with respect to θ and $\frac{\partial f}{\partial \theta}$ and $\frac{\partial^2 f}{\partial \theta^2}$ are continuous over $\mathcal{X} \times \mathcal{V}(\theta_0)$. Furthermore, there exists a non singular $p \times p$-matrix $\Sigma(\theta_0)$ depending on θ_0 such that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f_i'(\theta_0)\, {}^t f_i'(\theta_0) = \Sigma(\theta_0)$$

We turn now to the statement of our main result. In theorem 2.1, it is stated that under smoothness conditions, the jackknife studentized statistic is asymptotically gaussian with zero mean and unit variance, which is a sufficient condition for building confidence regions and tests as described in the introduction.

Theorem 2.1 *Suppose we are given model (2.1), where assumptions $\mathcal{S}_n$ and $\mathcal{S}_c$ hold true. Suppose that there exists some compact neihborhood $\mathcal{V}(\theta_0)$ of θ_0 such that f is three times differentiable with respect to θ over $\mathcal{V}(\theta_0)$ and $\frac{\partial^3 f}{\partial \theta^3}$ is bounded over $\mathcal{X} \times \mathcal{V}(\theta_0)$. If $\mathbb{E}|\delta_1|^6$ is finite then*

$$\sqrt{n}\tilde{s}^{-1}(\tilde{\theta} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_p(0, I_p) \text{ as } n \to \infty$$

where $\tilde{\theta}$ and $\tilde{s}^2$ are the jackknife estimators defined in (2.2) and (2.3).

Theorem 2.1 is a direct consequence of the following theorem 2.2. In this theorem, it is shown that, under suitable smoothness conditions on f and moments conditions on the errors ε_i, the jackknife estimator of the regression parameter is asymptotically gaussian, with asymptotic variance depending on the unknown parameter θ_0, and the variance-covariance matrix $\tilde{s}^2$ defined in (2.3) is a consistent estimator for the asymptotic variance of $\sqrt{n}(\tilde{\theta} - \theta_0)$

Note that under assumption $\mathcal{S}_c$, since there exist σ_0 and σ_∞ such that $0 < \sigma_0 \leq \sigma_i \leq \sigma_\infty < \infty$ for all i, the following limit (depending on θ_0) exists and is non singular. Let $S(\theta_0)$ denotes this limit.

$$S(\theta_0) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 f_i'(\theta_0)\, {}^t f_i'(\theta_0).$$

Theorem 2.2 *Suppose we are given model (2.1), where assumptions $\mathcal{S}_n$ and $\mathcal{S}_c$ hold true.*

1. If $\mathbb{E}|\delta_1|^6$ is finite then

$$\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_p(0, S(\theta_0)(\Sigma(\theta_0))^{-2}) \text{ as } n \to \infty$$

2. *If* $\mathbb{E}|\delta_1|^3$ *is finite and if there exists some open neihborhood* $\mathcal{V}(\theta_0)$ *of* θ_0 *such that* f *is three times differentiable with respect to* θ *over* $\mathcal{V}(\theta_0)$ *and* $\frac{\partial^3 f}{\partial \theta^3}$ *is bounded over* $\mathcal{X} \times \mathcal{V}(\theta_0)$ *then*

$$\tilde{s}^2 \xrightarrow{\mathbb{P}_{\theta_0}} S(\theta_0)\Sigma(\theta_0)^{-2} \qquad \text{as } n \to \infty$$

The main tool in proving theorem 2.2 is the following proposition, where it is shown that, on the one hand the jackknife estimator and the least-squares estimator are asymptotically equivalent at rate $n^{-1/2}$ and on the other hand $\tilde{s}^2$ is asymptotically equivalent to the following estimator $\hat{s}^2$ proposed by Shao.

$$\hat{s}^2 = (n-1)\sum_{i=1}^{n}(\hat{\theta}_{-i} - \hat{\theta})\, {}^t(\hat{\theta}_{-i} - \hat{\theta})$$

We can rewrite the jackknife estimator as

$$\tilde{\theta} = \hat{\theta} - \frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{-i} - \hat{\theta}).$$

It is clear from this identity that the asymptotic equivalence between $\tilde{\theta}$ and $\hat{\theta}$ will result from an appropriat control of $\sum_{i=1}^{n}(\hat{\theta}_{-i} - \hat{\theta})$.

Throughout the sequel, we shall use the same notation $||\,.\,||$ for denoting either the Euclidean norm in $\mathbb{R}^p$ defined by $||\theta||^2 = {}^t\theta\theta$, or the associated norms for linear or bilinear operators from $\mathbb{R}^p$ or $\mathbb{R}^p \times \mathbb{R}^p$ to $\mathbb{R}^p$.

Proposition 2.1 *Suppose we are given model (2.1), where assumptions* $\mathcal{S}_n$ *and* $\mathcal{S}_c$ *hold true.*

1. *If* $\mathbb{E}|\delta_1|^m$ *is finite for some* $m \geq 2$ *then*

$$\max_{1\leq i\leq n} ||\hat{\theta}_{-i} - \hat{\theta}|| = o_{\mathbb{P}}(n^{1/m-1}) \tag{2.6}$$

2. *If* $\mathbb{E}|\delta_1|^m$ *is finite for some* $m \geq 2$ *and if there exists some open neighborhood* $\mathcal{V}(\theta_0)$ *of* θ_0 *over which* f *is three times differentiable with respect to* θ *and* $\frac{\partial^3 f}{\partial \theta^3}$ *is bounded over* $\mathcal{X} \times \mathcal{V}(\theta_0)$ *then*

$$||\sum_{i=1}^{n}(\hat{\theta}_{-i} - \hat{\theta})|| = o_{\mathbb{P}}(n^{3/m-1}) \tag{2.7}$$

Moreover, under this assumptions,

(a) $||\hat{\theta} - \tilde{\theta}|| = o_{\mathbb{P}}(n^{\frac{3}{m}-1})$

(b) $||\hat{s}^2 - \tilde{s}^2|| = o_{\mathbb{P}}(n^{\frac{6}{m}-2})$

Thus, $\hat{\theta}$ and $\tilde{\theta}$ are asymptotically equivalent at rate $n^{-1/2}$ whenever the ε_i's admit moments of order 6, since $\hat{s}^2$ and $\tilde{s}^2$ are asymptotically equivalent whenever the ε_i's admit moments of order 3.

Suppose now that the parameter of interest is $\beta_0 = h(\theta_0)$ where h is a smooth function of the model parameter defined over a region R of Θ which contains θ_0. Let $\hat{\beta}$ be the naive least-squares estimator for θ_0 defined by $\hat{\beta} = h(\hat{\theta})$, and $\hat{\beta}_{-i} = h(\hat{\theta}_{-i})$. Then, the pseudo-values

$\tilde{\beta}_{-i}$ $(1 \leq i \leq n)$ are defined by $\tilde{\beta}_{-i} = n\hat{\beta} - (n-1)\hat{\beta}_{-i}$ and the jackknife estimator $\tilde{\beta}$ of β_0 constructed on the naive least-squares estimators is

$$\tilde{\beta} = n\hat{\beta} - \frac{n-1}{n}\sum_{i=1}^{n}\hat{\beta}_{-i}$$

The jackknife estimator for the asymptotic variance of $\tilde{\beta}$ is defined by

$$\tilde{s}_h^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\tilde{\beta}_i - \tilde{\beta})\,{}^t(\tilde{\beta}_i - \tilde{\beta})$$

We state in theorem 2.3 that the jackknife estimator $\tilde{\beta}$ is asymptotically equivalent to the least-squares estimator $\hat{\beta}$. So, $\tilde{\beta}$ has an asymptotic gaussian distribution provided h is smooth enough. Moreover, $\tilde{s}_h^2$ is a consistent estimator of the asymptotic variance of $\sqrt{n}(\tilde{\beta} - \beta_0)$, asymptotically equivalent to the estimator of variance proposed by Shao which is defined by

$$\hat{s}_h^2 = (n-1)\sum_{i=1}^{n}(\hat{\beta}_{-i} - \hat{\beta})\,{}^t(\hat{\beta}_{-i} - \hat{\beta})$$

Theorem 2.3 *Suppose we are given model (2.1) where assumptions S_n and S_c hold true. Let h be a twice differentiable function defined from R to $\mathbb{R}^q$ $(q \leq p)$, where R is an open subset of Θ such that $\theta_0 \in R$. Suppose that h'' is bounded over an open neighborhood of θ_0.*

1. *If $\mathbb{E}|\delta_1|^m$ is finite for some $m \geq 2$, then $\|\hat{\beta} - \tilde{\beta}\| = o_{\mathbb{P}}(n^{\frac{3}{m}-1})$*
2. *If $\mathbb{E}|\delta_1|^m$ is finite for some $m \geq 2$, then $\|\hat{s}_h^2 - \tilde{s}_h^2\| = o_{\mathbb{P}}(n^{\frac{6}{m}-2})$*
3. *If $\mathbb{E}|\delta_1|^6$ is finite, then*

$$\sqrt{n}\tilde{s}_h^{-1}(\tilde{\beta} - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_q(0, I_q) \text{ as } n \to \infty$$

Corollary 2.3 is the analoguous of theorems 1 and 2 in Miller [26]. In Miller's theorem 1, asymptotic normality of a function for the regression parameter in linear models is stated. The same smoothness assumptions for h are required in the two cases. Asumptions related to error terms are quite differents. In the linear case, Miller assumed that the ε_i's are i.i.d. with finite fourth moment, since in nonlinear models we assume that the ε_i's are i.i.d. except for an unknown scalar constant, with uniformly bounded 6th moment.

2.3 Proofs

Proposition 2.1 is proved in section 2.3.1. Theorems 2.2 and 2.3 are derived in 2.3.2. Througout the proofs, we shall use the following notation: for each ψ and ξ in $\mathbb{R}^p$,

$$(\psi, \xi) = \{\lambda\xi + (1-\lambda)\psi,\ \lambda \in [0,1]\}.$$

2.3.1 Order of $\max_{1\leq i\leq n}||\hat\theta_{-i}-\hat\theta||$ and $\sum_{1\leq i\leq n}(\hat\theta_{-i}-\hat\theta)$

This section is devoted to the proof of the proposition 2.1. Throughout the proof, we shall use the following notations

1. $\Sigma_n(\theta)=\sum_{i=1}^n(y_i-f_i(\theta))^2$
2. for each $i\in\{1,...,n\}$, $\Sigma_{-i}(\theta)=\sum_{j\neq i}(y_j-f_j(\theta))^2$.

Shao [37] proved in his theorem 1 and lemma 3 that under assumptions $\mathcal{S}_n$, $\hat\theta$ converges almost surely to θ_0 as $n\to\infty$ and $\max_{1\leq i\leq n}||\theta_0-\hat\theta_{-i}||$ converges almost surely to zero as $n\to\infty$. Let r be a positive real number such that the closed ball $B_r(\theta_0)$ with center θ_0 and radius r is included in $\mathcal{V}(\theta_0)$. Fix $\varepsilon>0$ and let n_0 be a positive integer such that $\max_{1\leq i\leq n}||\hat\theta_{-i}-\theta_0||\leq r$ and $||\hat\theta-\theta_0||\leq r$ with probability greater than $1-\varepsilon$ whenever $n\geq n_0$. Throughout the proof, we shall assume that $n\geq n_0$.

Let us prove (2.6). Since (with probability greater than $1-\varepsilon$) for all i, Σ_{-i} is twice differentiable over $(\hat\theta,\hat\theta_{-i})$, we can expand Σ_{-i} in Taylor formula at the second order. With probability greater than $1-\varepsilon$, for each $i\in\{1,...,n\}$ there exists some $\beta_i\in(\hat\theta,\hat\theta_{-i})$ such that

$$\Sigma_{-i}(\hat\theta)=\Sigma_{-i}(\hat\theta_{-i})+{}^t\Sigma'_{-i}(\hat\theta_{-i})(\hat\theta-\hat\theta_{-i})+\frac{1}{2}\,{}^t(\hat\theta-\hat\theta_{-i})\Sigma''_{-i}(\beta_i)(\hat\theta-\hat\theta_{-i})$$

where by definition $\hat\theta_{-i}$ minimizes Σ_{-i}. Thus $\Sigma'_{-i}(\hat\theta_{-i})=0$ and

$$\Sigma_{-i}(\hat\theta)-\Sigma_{-i}(\hat\theta_{-i})=\frac{1}{2}\,{}^t(\hat\theta-\hat\theta_{-i})\Sigma''_{-i}(\beta_i)(\hat\theta-\hat\theta_{-i})$$

For each $i\in\{1,...,n\}$, let ϕ_i denotes the function from Θ to $\mathbb{R}$ defined by $\phi_i(\theta)=(y_i-f_i(\theta))^2$. Then, $\Sigma_{-i}=\Sigma_n-\phi_i$ and we get, since $\hat\theta$ minimizes Σ_n

$$\begin{aligned}\Sigma_{-i}(\hat\theta)-\Sigma_{-i}(\hat\theta_{-i}) &= \Sigma_n(\hat\theta)-\phi_i(\hat\theta)-\Sigma_n(\hat\theta_{-i})+\phi_i(\hat\theta_{-i})\\ &\leq \phi_i(\hat\theta_{-i})-\phi_i(\hat\theta)\end{aligned}$$

So, there exists some $\gamma_i\in(\hat\theta_{-i},\hat\theta)$ such that

$$\frac{1}{2}\,{}^t(\hat\theta-\hat\theta_{-i})\Sigma''_{-i}(\beta_i)(\hat\theta-\hat\theta_{-i})\leq||\phi'_i(\gamma_i)||\;\;||\hat\theta_{-i}-\hat\theta|| \tag{2.8}$$

where by the following lemma (which proof is postponed to the end of this subsection),

$$\max_{1\leq i\leq n}||\Sigma''_{-i}(\beta_i)/n-2\Sigma(\theta_0)||\xrightarrow{\mathbb{P}}0 \qquad as\ n\to\infty$$

Lemma 2.1 *Suppose we are given model (2.1) where assumptions $\mathcal{S}_n$ and $\mathcal{S}_c$ hold true. Let r be a positive real number and such that the closed ball $B_r(\theta_0)$ with center θ_0 and radius r is included in $\mathcal{V}(\theta_0)$. Then*

1. *For each familly $\{\theta_{in}\}_{i,n\in\mathbb{N}}$ of random variables in $B_r(\theta_0)$ such that $\max_{1\leq i\leq n}||\theta_{in}-\theta_0||$ converges in probability to zero as $n\to\infty$,*

$$\max_{1\leq i\leq n}||\frac{1}{n}\Sigma''_{-i}(\theta_{in})-2\Sigma(\theta_0)||\xrightarrow{\mathbb{P}}0 \qquad as\ n\to\infty \tag{2.9}$$

2. *For each sequence $\{\theta_n\}_{n\in\mathbb{N}}$ of random variables in $B_r(\theta_0)$ which converges in probability to θ_0 as $n \to \infty$,*
$$\frac{1}{n}\Sigma''_n(\theta_n) \xrightarrow{\mathbb{P}} 2\Sigma(\theta_0) \qquad \text{as } n \to \infty \tag{2.10}$$

Let R be a positive real number such that $R < 2r$ and C_R denote the centered sphere in $\mathbb{R}^p$ with radius R: $C_R = \{\theta \in \mathbb{R}^p,\ ||\theta|| = R\}$. Then, $\inf_{\theta\in C_R}\ {}^t\theta\Sigma(\theta_0)\theta/R^2$ is achieved and positive. Let λ be this infimum. We get

$$\lambda = \inf_{\theta\in B_{2r}(0)\setminus\{0\}} \frac{{}^t\theta\Sigma(\theta_0)\theta}{||\theta||^2} = \frac{1}{R^2}\inf_{\theta\in C_R}\ {}^t\theta\Sigma(\theta_0)\theta$$

where $B_{2r}(0)$ denotes the closed ball in $\mathbb{R}^p$ with center 0 and radius $2r$. We suppose n large enough so that $\max_{1\le i\le n}||\Sigma''_{-i}(\beta_i)/n - 2\Sigma(\theta_0)|| \le \lambda$ with probability greater than $1-\varepsilon$. Then, with probability greater than $1-2\varepsilon$, for each non zero $\theta \in B_{2r}(0)$

$$\frac{n\lambda}{2} < n\left(\frac{{}^t\theta\Sigma(\theta_0)\theta}{||\theta||^2} - \frac{\lambda}{2}\right) < \frac{1}{2}\frac{{}^t\theta\Sigma''_{-i}(\beta_i)\theta}{||\theta||^2} \qquad \forall i \in \{1,...,n\}$$

and therefore for each non zero $\theta \in B_{2r}(\theta_0)$,

$$||\theta||^2 \le \frac{{}^t\theta\Sigma_{-i}(\beta_i)\theta}{n\lambda}$$

So, (2.8) yields with probability greater than $1-2\varepsilon$,

$$||\hat\theta - \hat\theta_{-i}||^2 \le \frac{2}{n\lambda}||\phi'_i(\gamma_i)||\ ||\hat\theta - \hat\theta_{-i}|| \qquad \forall i \in \{1,...,n\}$$

and therefore

$$\max_{1\le i\le n}||\hat\theta - \hat\theta_{-i}|| \le \frac{2}{n\lambda}\max_{1\le i\le n}||\phi'_i(\gamma_i)||$$

For each $\theta \in \mathcal{V}(\theta_0)$ and each i, $\phi'_i(\theta) = -2(\varepsilon_i + f_i(\theta_0) - f_i(\theta))f'_i(\theta)$. Since f and $\partial f/\partial\theta$ are continuous over $\mathcal{X} \times B_r(\theta_0)$, there exist some positive constants C_0 and C_1 such that

$$\sup_{\theta\in B_r(\theta_0)}\max_{1\le i\le n}||\phi'_i(\theta)|| \le C_0\max_{1\le i\le n}|\varepsilon_i| + C_1$$

Hence,

$$\max_{1\le i\le n}||\hat\theta - \hat\theta_{-i}|| \le \frac{2}{n\lambda}\left(C_0\max_{1\le i\le n}|\varepsilon_i| + C_1\right) \tag{2.11}$$

with probability greater than $1-2\varepsilon$. We have to control $\max_{1\le i\le n}|\varepsilon_i|$. This will be done by noticing that our assumptions ensure that the variables $|\varepsilon_i|^m$ are uniformly integrable, which allows to use the following elementary fact

Lemma 2.2 *Let $\{\varepsilon_i\}_{i\in\mathbb{N}}$ be a sequence of random variables such that $\{\varepsilon_i^m\}_{i\in\mathbb{N}}$ are uniformly integrable for some positive m. Then*

$$\max_{1\le i\le n}|\varepsilon_i| = o_{\mathbb{P}}(n^{1/m})$$

Proof of lemma 2.2:

Let δ be some positive real number. For all $t \in \mathbb{R}_+$,

$$\begin{aligned}
\mathbb{P}(\max_{1\le i\le n} |\varepsilon_i| > \delta n^{1/m}) &\le n \max_{1\le i\le n} \mathbb{P}(|\varepsilon_i| > \delta n^{1/m}) \\
&\le \frac{1}{\delta^m} \max_{1\le i\le n} \mathbb{E}\left(|\varepsilon_i|^m \mathbb{I}_{|\varepsilon_i|^m > n\delta^m}\right)
\end{aligned}$$

where the right hand term converges to zero as $n \to \infty$ since the variables $\{\varepsilon_i^m\}_{i\in\mathbb{N}}$ are uniformly integrable. ◊

We turn back to the proof of proposition 2.1. (2.11) yields (2.6) via lemma 2.2 since ε is arbitrarily small.

Let us prove now (2.7). Recall that we suppose that $n \ge n_0$, which implies $\max_{1\le i\le n} ||\hat\theta_{-i} - \theta_0|| \le r$ and $||\hat\theta - \theta_0|| \le r$ with probability greater than $1-\varepsilon$, r being some positive real number such that $B_r(\theta_0) \subset \mathcal{V}(\theta_0)$. With probability greater than $1-\varepsilon$, we can expand Σ_n' by Taylor's formula

$$\Sigma_n'(\hat\theta_{-i}) = \Sigma_n'(\hat\theta) + \Sigma_n''(\hat\theta)(\hat\theta_{-i} - \hat\theta) + R_i$$

where $\Sigma_n'(\hat\theta) = 0$ and R_i satisfies

$$||R_i|| \le \frac{1}{2}||\hat\theta_{-i} - \hat\theta||^2 \sup_{\theta\in B_r(\theta_0)} ||\Sigma_n'''(\theta)||$$

On the other hand, we can expand ϕ_i' by Taylor's formula and get

$$\phi_i'(\hat\theta_{-i}) = \phi_i'(\hat\theta) + \phi_i''(\hat\theta)(\hat\theta_{-i} - \hat\theta) + \tilde R_i$$

where $\tilde R_i$ satisfies

$$||\tilde R_i|| \le \frac{1}{2}||\hat\theta_{-i} - \hat\theta||^2 \sup_{\theta\in B_r(\theta_0)} ||\phi_i'''(\theta)||$$

But $\phi_i'(\hat\theta_{-i}) = \Sigma_n'(\hat\theta_{-i})$ which yields by identifying the two Taylor expansions above:

$$\Sigma_n''(\hat\theta)(\hat\theta_{-i} - \hat\theta) + R_i = \phi_i'(\hat\theta) + \phi_i''(\hat\theta)(\hat\theta_{-i} - \hat\theta) + \tilde R_i$$

Dividing by n and summing over i we get, since $\sum_{i=1}^n \phi_i'(\hat\theta) = \Sigma_n'(\hat\theta) = 0$,

$$\left(\frac{1}{n}\Sigma_n''(\hat\theta)\right)\sum_{i=1}^n(\hat\theta_{-i} - \hat\theta) = a_n + b_n$$

where

$$\begin{aligned}
a_n &= \frac{1}{n}\sum_{i=1}^n(\tilde R_i - R_i) \\
b_n &= \frac{1}{n}\sum_{i=1}^n \phi_i''(\hat\theta)(\hat\theta_{-i} - \hat\theta)
\end{aligned}$$

Taking the norm square in (2.12), one deduce that

$$ {}^{t}\left(\sum_{i=1}^{n}(\hat{\theta}_{-i}-\hat{\theta})\right)\left(\frac{1}{n}\Sigma_n''(\hat{\theta})\right)^2\left(\sum_{i=1}^{n}(\hat{\theta}_{-i}-\hat{\theta})\right) = ||a_n+b_n|| $$

But $\left(\frac{1}{n}\Sigma_n''(\hat{\theta})\right)^2$ converges in probability to $\Sigma(\theta_0)^2$ and arguing as in the proof of (2.6), it follows that with probability greater than $1-\varepsilon$, for each non zero $\theta \in B_r(\theta_0)$

$$ ||\theta||^2 \le \frac{1}{\lambda}\,{}^{t}\theta\left(\frac{1}{n}\Sigma_n''(\hat{\theta})\right)^2\theta $$

which yields with probability greater than $1-2\varepsilon$

$$ \begin{aligned} ||\sum_{i=1}^{n}(\hat{\theta}_{-i}-\hat{\theta})|| &\le \frac{1}{\lambda}||a_n+b_n|| \\ &\le \frac{1}{\lambda}(||a_n||+||b_n||) \end{aligned} \qquad (2.12) $$

By definition,

$$ \begin{aligned} ||a_n|| &\le \frac{n+1}{2}\max_{1\le i\le n}||\hat{\theta}_{-i}-\hat{\theta}||^2\max_{1\le i\le n}\sup_{\theta\in B_r(\theta_0)}||\phi_i'''(\theta)|| \\ ||b_n|| &\le \max_{1\le i\le n}||\hat{\theta}_{-i}-\hat{\theta}||\max_{1\le i\le n}\sup_{\theta\in B_r(\theta_0)}||\phi_i''(\theta)|| \end{aligned} $$

For all $\theta \in B_r(\theta_0)$

$$ \phi_i''(\theta) = 2f_i'(\theta)\,{}^{t}f_i'(\theta) - 2(\varepsilon_i + f_i(\theta_0) - f_i(\theta))f_i''(\theta) $$

$$ \phi_i'''(\theta) = 2f_i''(\theta)\,{}^{t}f_i'(\theta) + 4f_i'(\theta)f_i''(\theta) - 2(\varepsilon_i + f_i(\theta_0) - f_i(\theta))f_i'''(\theta) $$

and since $\partial f/\partial\theta$, $\partial^2 f/\partial\theta^2$ and $\partial^3 f/\partial\theta^3$ are bounded over $\mathcal{X}\times B_r(\theta_0)$, there exists some constants C_0 and C_1 both positive such that

$$ \max_{1\le i\le n}\sup_{\theta\in B_r(\theta_0)}||\phi_i'''(\theta)|| \le C_1 + C_0\max_{1\le i\le n}|\varepsilon_i| $$

$$ \max_{1\le i\le n}\sup_{\theta\in B_r(\theta_0)}||\phi_i''(\theta)|| \le C_1 + C_0\max_{1\le i\le n}|\varepsilon_i| $$

Now, the proof of (2.7) can be completed by combining (2.12), (2.6) and lemma 2.2.

We prove now the last part of proposition 2.1. By definition, $\tilde{\theta}-\hat{\theta} = \frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{-i}-\hat{\theta})$ and therefore (2.7) yields

$$ ||\hat{\theta}-\tilde{\theta}|| = o_{\mathbb{P}}(n^{3/m-1}). $$

Let $\hat{\theta}_{(.)}$ be the empirical mean of the least squares estimators in the reduced models : $\hat{\theta}_{(.)} = \sum_{i=1}^{n}\hat{\theta}_{-i}/n$. By definition,

$$ \tilde{s}^2 = (n-1)\sum_{i=1}^{n}(\hat{\theta}_{-i}-\hat{\theta}_{(.)})\,{}^{t}(\hat{\theta}_{-i}-\hat{\theta}_{(.)}) $$

So, $\tilde{s}^2 = \hat{s}^2 + r_n$ where r_n is defined by

$$\begin{aligned} r_n &= n(n-1)(\hat{\theta}_{(.)} - \hat{\theta})\ {}^t(\hat{\theta} - \hat{\theta}_{(.)}) \\ &= \frac{n-1}{n}\left(\sum_{i=1}^{n}(\hat{\theta}_{-i} - \hat{\theta})\right)\left(\sum_{i=1}^{n}{}^t(\hat{\theta}_{-i} - \hat{\theta})\right) \end{aligned}$$

We get $||\hat{s}^2 - \tilde{s}^2|| = \frac{n-1}{n}||\sum_{i=1}^{n}(\hat{\theta}_{-i} - \hat{\theta})||^2$ and therefore

$$||\hat{s}^2 - \tilde{s}^2|| = o_{\mathbb{P}}(n^{6/m-2}).$$

This completes the proof of proposition 2.1. ◇

<u>*Proof of lemma 2.1:*</u>

Let G_n be the function defined over $\mathcal{V}(\theta_0)$ which takes θ into $\sum_{i=1}^{n} f_i'(\theta)\ {}^t f_i'(\theta)$. We know from assumption $\mathcal{S}_c$ that $G_n(\theta_0)/n$ converges as $n \to \infty$ to $\Sigma(\theta_0)$. So, it is enough to prove that

$$\max_{1\leq i\leq n} \|\Sigma''_{-i}(\theta_{in}) - 2G_n(\theta_0)\| = o_{\mathbb{P}}(n) \tag{2.13}$$

to get (2.13). By definition,

$$\frac{1}{n}\Sigma''_{-i}(\theta_{in}) - \frac{2}{n}G_n(\theta_0) = \frac{2}{n}G_n(\theta_{in}) - \frac{2}{n}G_n(\theta_0) - \frac{2}{n}f_i'(\theta_{in})\ {}^t f_i'(\theta_{in}) + a_{in}$$

where

$$a_{in} = -\frac{2}{n}\sum_{j\neq i}\varepsilon_j f_j''(\theta_{in}) - \frac{2}{n}\sum_{j\neq i}(f_j(\theta_0) - f_j(\theta_{in}))f_j''(\theta_{in})$$

Since $\partial f/\partial\theta$, $\partial^2 f/\partial\theta^2$ and $\partial^3 f/\partial\theta^3$ are bounded over $\mathcal{X} \times B_r(\theta_0)$, there exists some positive constants C_1, C_2 and C_3 such that

$$\begin{aligned} \sup_{(x,\theta)\in\mathcal{X}\times B_r(\theta_0)} \left\|\frac{\partial f}{\partial\theta}\right\| &\leq C_1 \\ \sup_{(x,\theta)\in\mathcal{X}\times B_r(\theta_0)} \left\|\frac{\partial^2 f}{\partial\theta^2}\right\| &\leq C_2 \\ \sup_{(x,\theta)\in\mathcal{X}\times B_r(\theta_0)} \left\|\frac{\partial^3 f}{\partial\theta^3}\right\| &\leq C_3 \end{aligned}$$

So, we get

$$\begin{aligned} \max_{1\leq i\leq n} ||G_n(\theta_{in}) - G_n(\theta_0)|| &\leq \max_{1\leq i\leq n}\sum_{j=1}^{n}||(f_j'(\theta_{in}) - f_j'(\theta_0))\ {}^t f_j'(\theta_{in}) + f_j'(\theta_0)\ {}^t(f_j'(\theta_{in}) - f_j'(\theta_0))|| \\ &\leq 2nC_1C_2 \max_{1\leq i\leq n}||\theta_{in} - \theta_0|| \end{aligned}$$

and therefore

$$\max_{1\leq i\leq n}\left\|\frac{1}{n}G_n(\theta_{in}) - \frac{1}{n}G_n(\theta_0)\right\| \xrightarrow{\mathbb{P}} 0 \qquad as\ n \to \infty$$

Since $\partial f/\partial\theta$ is bounded over $\mathcal{X} \times B_r(\theta_0)$, it is clear that

$$\lim_{n\to\infty} \max_{1\le i\le n} \left\| \frac{2}{n} f_i'(\theta_{in})\, {}^t f_i'(\theta_{in}) \right\| = 0$$

Furthermore,

$$\max_{1\le i\le n} ||a_{in}|| \le \max_{1\le i\le n} \left\| \frac{2}{n} \sum_{j=1}^{n} \varepsilon_j f_j''(\theta_{in}) \right\| + \frac{2C_2}{n} \max_{1\le i\le n} |\varepsilon_i| + 2C_1C_2 \max_{1\le i\le n} ||\theta_0 - \theta_{in}||$$

By the law of large numbers,

$$\max_{1\le i\le n} \left\| \frac{2}{n} \sum_{j=1}^{n} \varepsilon_j f_j''(\theta_0) \right\| \xrightarrow{\mathbb{P}} 0 \qquad as\ n \to \infty$$

and since

$$\max_{1\le i\le n} \left\| \frac{2}{n} \sum_{j=1}^{n} \varepsilon_j (f_j''(\theta_0) - f_j''(\theta_{in})) \right\| \le 2C_3 \max_{1\le i\le n} ||\theta_0 - \theta_{in}|| \frac{1}{n} \sum_{i=1}^{n} |\varepsilon_i|$$

where $\frac{1}{n}\sum_{i=1}^{n} |\varepsilon_i|$ is almost surely finite, we get

$$\max_{1\le i\le n} \left\| \frac{2}{n} \sum_{j=1}^{n} \varepsilon_j f_j''(\theta_{in}) \right\| \xrightarrow{\mathbb{P}} 0 \qquad as\ n \to \infty \tag{2.14}$$

By combining (2.14), lemma 2.2 and the fact that $\max_{1\le i\le n} ||\theta_0 - \theta_{in}||$ converges to zero in probability as $n \to \infty$, we get

$$\max_{1\le i\le n} ||a_{in}|| \xrightarrow{\mathbb{P}} 0 \qquad as\ n \to \infty$$

(2.9) follows. It is clear from the proof that (2.10) holds true. ◇

2.3.2 Proof of main results

Proof of theorem 2.2:

By proposition 2.1, $\sqrt{n}(\tilde\theta - \theta_0)$ has the same asymptotic behaviour as $\sqrt{n}(\hat\theta - \theta_0)$. By theorem 2 in shao [37], $\sqrt{n}(\hat\theta - \theta_0)$ converges in distribution under assumptions of theorem 2.2 to a centered p-dimentionnal gaussian variable with variance $S(\theta_0)(\Sigma(\theta_0))^{-2}$, which proves the first part of theorem 2.2. The second point follows from proposition 2.1 and theorem 4 of Shao [37], where it is proved that under assumptions of theorem 2.2 (where the assumption $\mathbb{E}|\delta_1|^3 < \infty$ is useness) $\hat s^2$ converges almost surely to $S(\theta_0)(\Sigma(\theta_0))^{-2}$ as $n \to \infty$. ◇

Proof of theorem 2.3:

Let r be a positive real number such that h'' exists and is bounded over the closed ball $B_r(\theta_0)$ in $\mathbb{R}^p$ with center θ_0 and radius r. Note that since for each $\varepsilon > 0$, there exists some integer n_0 such that whenever $n \ge n_0$ $||\hat\theta - \theta_0|| \le r$ and $\max_{1\le i\le n} ||\hat\theta_{-i} - \theta_0|| \le r$ with probability greater than $1-\varepsilon$, it suffices to prove theorem 2.3 whenever $||\hat\theta - \theta_0|| \le r$ and $\max_{1\le i\le n} ||\hat\theta_{-i} - \theta_0|| \le r$

(see the proof of proposition 2.1).

We expand h in Taylor formula at the second order.

$$h(\hat{\theta}_{-i}) = h(\hat{\theta}) + {}^t h'(\hat{\theta})(\hat{\theta}_{-i} - \hat{\theta}) + R_i \tag{2.15}$$

where R_i satisfies

$$||R_i|| \leq \frac{1}{2} \sup_{\theta \in B_r(\theta_0)} ||h''(\theta)||\ ||\hat{\theta}_{-i} - \hat{\theta}||^2$$

By definition,

$$\tilde{\beta} = h(\hat{\theta}) - \frac{n-1}{n} \sum_{i=1}^{n} (h(\hat{\theta}_{-i}) - h(\hat{\theta}))$$

and therefore

$$\tilde{\beta} = \hat{\beta} - \frac{n-1}{n}\ {}^t h'(\hat{\theta}) \sum_{i=1}^{n} (\hat{\theta}_{-i} - \hat{\theta}) + \frac{n-1}{n} \sum_{i=1}^{n} R_i$$

By proposition 2.2

$$\left\| {}^t h'(\hat{\theta}) \sum_{i=1}^{n} (\hat{\theta}_{-i} - \hat{\theta}) \right\| = o_{\mathbb{P}}(n^{3/m-1})$$

Since h" is bounded over $B_r(\theta_0)$, there exists some $H > 0$ such that

$$\left\| \frac{n-1}{n} \sum_{i=1}^{n} R_i \right\| \leq H(n-1) \max_{1 \leq i \leq n} ||\hat{\theta}_{-i} - \hat{\theta}||^2$$

which is $o_{\mathbb{P}}(n^{2/m-1})$ by proposition 2.1. So, $||\tilde{\beta} - \hat{\beta}|| = o_{\mathbb{P}}(n^{3/m-1})$.

Let $\hat{\beta}_{(.)}$ be the empirical mean of the least squares estimators in reduced models : $\hat{\beta}_{(.)} = \sum_{i=1}^{n} \hat{\beta}_{-i}/n$. By definition

$$\tilde{s}_h^2 = (n-1) \sum_{i=1}^{n} (\hat{\beta}_{-i} - \hat{\beta}_{(.)})\ {}^t(\hat{\beta}_{-i} - \hat{\beta}_{(.)})$$

and therefore, $\tilde{s}_h^2 = \hat{s}_h^2 + r_{hn}$ where $r_{hn} = n(n-1)(\hat{\beta}_{(.)} - \hat{\beta})\ {}^t(\tilde{\beta} - \hat{\beta}_{(.)})$.
Shao [37] shown in his theorem 2 that $\hat{s}_h^2$ converges in probability to ${}^t h'(\theta_0)s(\theta_0)(\Sigma(\theta_0))^{-2}h'(\theta_0)$ as $n \to \infty$. By (2.15),

$$\sum_{i=1}^{n} (\hat{\beta}_{-i} - \hat{\beta}) = {}^t h'(\hat{\theta}) \sum_{i=1}^{n} (\hat{\theta}_{-i} - \hat{\theta}) + R_i$$

So by proposition 2.1, $||r_{hn}|| = o_{\mathbb{P}}(n^{6/m-2})$.

If $m \geq 6$, $\sqrt{n}(\tilde{\beta} - \hat{\beta})$ converges in probability to zero as $n \to \infty$. By theorem 2.2 and the following classical result (see theorem 2.4),

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_q(0, {}^t h'(\theta_0)S(\theta_0)(\Sigma(\theta_0))^{-2}\ h'(\theta_0) \qquad as\ n \to \infty$$

Shao proved in his theorem 4 that $\hat{s}^2$ converges almost surely to ${}^t h'(\theta_0)S(\theta_0)(\Sigma(\theta_0))^{-2}\ h'(\theta_0)$ under the assumptions of theorem 2.3, which completes the proof of theorem 2.3.

Theorem 2.4 *Let Γ be a $p \times p$ positive definite matrix. Let $\{\theta_n\}_{n\in\mathbb{N}}$ be a sequence of p-dimensional r.v. such that*

$$\sqrt{n}(\theta_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_p(0, \Gamma) \text{ as } n \to \infty$$

Let h be a function defined from $\mathbb{R}^p$ to $\mathbb{R}^q$. Suppose h is twice differentiable with second differential h'' bounded over a neighborhood for θ_0. Then

$$\sqrt{n}(h(\theta_n) - h(\theta_0)) \xrightarrow{\mathcal{D}} \mathcal{N}_q(0, {}^t h'(\theta_0)\Gamma\, h'(\theta_0)) \text{ as } n \to \infty$$

Chapitre 3

Mise en oeuvre de la méthode de jackknife. Comparaison de courbes de croissance

Nous illustrons ici par l'étude de données réelles la méthode de jackknife décrite dans le second chapitre. Après avoir décrit les données initiales, nous présentons le modèle de régression retenu. Nous effectuons alors les estimations jackknife, ainsi que quelques tests d'homogénéité.

Le programme, réalisé à l'aide du logiciel `S.A.S.`, qui a permis d'effectuer les calculs, ainsi que les données initiales sont donnés au paragraphe 3.4.

3.1 Modélisation

Les données présentées en figure 3.1 sont les résultats d'expériences portant sur l'évolution au cours du temps du taux d'une bactérie (nommée Lactobacillus Delbrueckii Lactis) en milieu lacté. Ces expériences ont pour but de déterminer l'influence de la température à laquelle est placé le milieu sur l'évolution de la bactérie, afin d'en améliorer les qualités nutritives.

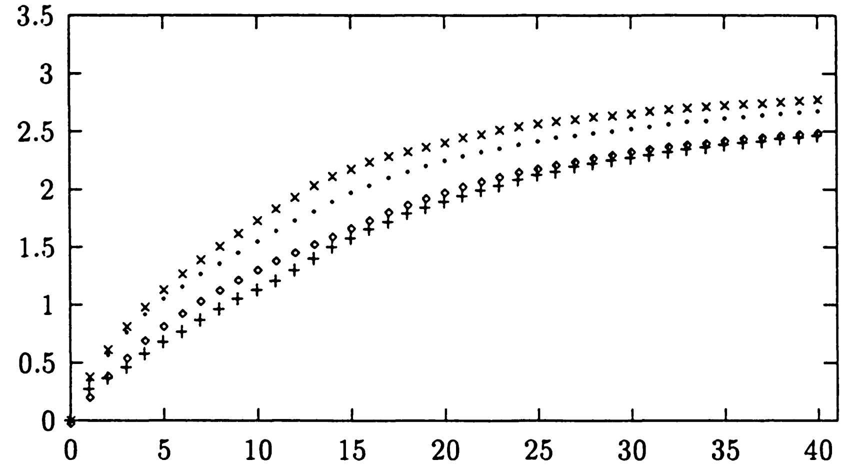
Le taux de bactéries dans le milieu de croissance est évalué au moyen de mesures de conductivité, la conductivité étant étroitement liée au taux de bactéries. Chaque expérience a été réalisée de la façon suivante: Le milieu étant placé à une température donnée, une mesure de conductivité est réalisée chaque heure, pendant 40 heures. La date de la première mesure est considérée comme date initiale, soit $x_1 = 0$. Ainsi, la i^{ème} mesure (pour $1 \leq i \leq 41$) est effectuée à la date $x_i = i - 1$.

Quatre expériences similaires ont été réalisées où, pour chacune, le milieu de croissance est placé à une température différente. Les températures envisagées sont

$$\begin{aligned} t_1 &= 33.85°C \\ t_2 &= 36.45°C \\ t_3 &= 39.45°C \\ t_4 &= 42.35°C \end{aligned}$$

Sur la figure 3.1, nous avons porté le temps en abscisse, et la mesure de conductivité correspondante en ordonnée, pour chacune de ces températures.

Figure 3.1: Mesures de conductivité: t_1 : + , t_2 : ⋄ , t_3 : . , t_4 : ×



Les questions que nous nous posons relativement à l'influence de la température sur l'évolution du taux de bactéries sont les suivantes:

1. La température joue-t-elle un rôle dans l'évolution globale du taux de bactéries? Plus précisément, peut-on considérer comme semblables les courbes de conductivité obtenues pour deux températures différentes?

2. Le taux maximal de bactéries est-il significativement sensible à un changement de température?

Nous devons tout d'abord modéliser les données, afin de formaliser ces questions. En l'absence d'une connaissance a priori de l'évolution du taux de bactéries, c'est l'allure même des données qui guide notre choix du modèle.

Un examen qualitatif des courbes de conductivité de la figure 3.1 conduit à la remarque suivante: les points expérimentaux se répartissent sur le graphique suivant une forme simple et régulière. Ceci justifie l'emploi de la régression non linéaire comme outil d'analyse. En outre, les courbes de conductivité obtenues pour chacune des températures t_1, t_2, t_3 et t_4 ont une forme comparable. Aussi, si $y_i^{(j)}$ désigne la i^{ème} mesure de conductivité observée lorsque le milieu de croissance est placé à la température t_j, nous considérons un modèle de la forme

$$y_i^{(j)} = f(x_i, \theta_0^{(j)}) + \varepsilon_i^{(j)} \qquad 1 \leq i \leq 41 \qquad 1 \leq j \leq 4$$

où $\theta_0^{(j)}$ est un paramètre inconnu à estimer, f est une fonction ne dépendant pas de j, et $\varepsilon_i^{(j)}$ est un terme résiduel. Ici, les résidus $(\varepsilon_i^{(j)})$ sont liés à l'imprécision de la mesure conduisant à l'observation $y_i^{(j)}$. Aussi pouvons nous considérer les hypothèses d'indépendance et d'équidistribution comme raisonnables. Nous supposons en outre les résidus gaussiens.

Nous observons que la forme générale de chaque courbe est comparable à celle des fonctions

$$x \mapsto a(1 - \exp(b + cx)) \tag{3.1}$$

où $a > 0$, $b < 0$ et $b + c < 0$. En effet, les fonctions de la forme (3.1)

- ont une pente comparable à celle des courbes de conductivité dès que c est un réel négatif,
- prennent des valeurs positives si $b + c$ est négatif,
- admettent une asymptote de hauteur a.

L'asymptote n'étant vraisemblablement atteinte pour aucune des courbes de conductivité, nous supposons que la hauteur d'asymptote est supérieure à la plus grande mesure observée. Une modélisation par la famille de fonctions définie par (3.1) conduit à de mauvaises estimations de la mesure de conductivité pour les premières dates d'observation. En effet, cette modélisation ne prend pas en compte le fait que chaque taux est nul en zéro. Nous introduisons donc un terme régularisant $\frac{x}{1+x}$. Ainsi, la fonction de régression retenue est:

$$f(x, \theta) = \frac{ax}{1+x}(1 - \exp(b + cx))$$

où $\theta = {}^t(a, b, c)$ appartient à $\mathbb{R}^3$. Compte tenu des remarques précédentes, nous supposons en outre que la vraie valeur $\theta_0^{(j)}$ du paramètre de régression correspondant à la température t_j appartient au sous ensemble $\Theta^{(j)}$ de $\mathbb{R}^3$ défini par

$$\Theta^{(j)} = \{(a, b, c) \in \mathbb{R}^3 : \; a \geq \max y_i^{(j)}, \; c < 0, \; b + c < 0\}$$

Nous pouvons à présent formaliser les questions ci-dessus.

1. Notre premier problème est le test d'homogénéité de deux courbes. Pour tout couple (j, k) d'éléments de $\{1, ..., 4\}$ tel que $j \neq k$, nous voulons tester l'hypothèse $\theta_0^{(j)} = \theta_0^{(k)}$ contre l'alternative $\theta_0^{(j)} \neq \theta_0^{(k)}$.

2. Notre second problème est le test d'homogénéité de deux asymptotes. Pour tout couple (j, k) d'éléments de $\{1, ..., 4\}$ tel que $j \neq k$, nous voulons tester l'hypothèse $a_0^{(j)} = a_0^{(k)}$ contre l'alternative $a_0^{(j)} \neq a_0^{(k)}$. Ceci revient à tester l'hypothèse $h(\theta_0^{(j)}) = h(\theta_0^{(k)})$ contre l'alternative $h(\theta_0^{(j)}) \neq h(\theta_0^{(k)})$, où la fonction h est la projection définie de $\mathbb{R}^3$ dans $\mathbb{R}$ par $h(\theta) = a$, lorsque $\theta = {}^t(a, b, c)$.

Nous avons en vue d'effectuer ces tests suivant la méthode d'analyse de variance décrite dans le second chapitre. Cette méthode se fonde sur l'heuristique de Tukey [39] qui, définissant les pseudo-valeurs $(\tilde{\theta}_i)$, postule qu'elles se comportent comme des variables indépendantes et de même loi gaussienne. Nous déterminons tout d'abord les estimations jackknife, afin de valider qualitativement notre équation de régression.

3.2 Estimations jackknife

Afin d'alléger les notations, nous omettons ici l'exposant (j), et notons n le nombre d'observations effectuées pour chaque courbe (soit $n = 41$). Nous considérons donc le modèle

$$y_i = f(x_i, \theta_0) + \varepsilon_i \qquad 1 \leq i \leq n \tag{3.2}$$

où les résidus $\{\varepsilon_i\}_{1 \leq i \leq n}$ sont indépendants et équi-distribués $\mathcal{N}(0, \sigma^2)$ pour un réel positif σ inconnu, $\theta_0 = {}^t(a_0, b_0, c_0)$ est un paramètre inconnu appartenant au sous ensemble Θ de $\mathbb{R}^3$ défini par

$$\Theta = \{(a, b, c) \in \mathbb{R}^3 : \; a \geq \max y_i, \; c < 0, \; b + c < 0\}$$

et f est la fonction définie sur $\mathbb{R} \times \mathbb{R}^3$ par $f(x,\theta) = \frac{ax}{1+x}(1 - \exp(b + cx))$.

Nous appliquons la méthode de jackknife à l'estimation par moindres carrés. Rappelons que l'estimateur des moindres carrés $\hat{\theta}$ de θ_0 est un vecteur de Θ qui minimise $\sum_{i=1}^{n}(y_i - f(x_i,\theta))^2$ sur Θ. On définit, pour tout $i \in \{1, ..., n\}$, $\hat{\theta}_{-i}$ comme l'estimateur des moindres carrés obtenu dans le modèle privé de la i^{ème} observation, c'est à dire que $\hat{\theta}_{-i}$ est un vecteur de Θ minimisant $\sum_{j \neq i}(y_j - f(x_j,\theta))^2$ sur Θ. L'estimateur jackknife de θ_0 est alors défini par

$$\tilde{\theta} = n\hat{\theta} - \frac{n-1}{n}\sum_{i=1}^{n}\hat{\theta}_{-i} \tag{3.3}$$

Nous calculons les estimations par moindres carrés au moyen de la procédure `nlin` du logiciel `S.A.S`. Cette procédure utilise la méthode itérative de minimisation de Gauss-Newton, et nécessite la définition d'une valeur initiale pour chaque paramètre. Nous devons nous appuyer sur les données pour fournir des initialisation aussi proches que possible de la vraie valeur des paramètres, afin de garantir la convergence de la procédure de Gauss-Newton.

Compte tenu de l'interprétation géométrique du paramètre a comme hauteur de l'asymptote, nous proposons

$$a_{init} = \max_{1 \leq i \leq n} y_i$$

comme valeur initiale pour l'estimation de a_0. Les valeurs obtenues sont consignées dans le tableau suivant

	a_{init}
t_1	2.46
t_2	2.51
t_3	2.67
t_4	2.77

Nous utilisons la formule donnant b_0 et c_0 pour déterminer une valeur initiale pour l'estimation de ces deux paramètres. Pour tout $i \in \{1, ..., n\}$ tel que $1 - \frac{1+x}{a_{init}x}y_i > 0$, soit

$$z_i = \log(1 - \frac{1+x}{a_{init}x}y_i)$$

Si l'on considère comme négligeables les résidus du modèle (3.2), alors, pour tout i, $z_i \approx b_0 + c_0x$. Ainsi, une régression linéaire sur les variables z_i fournit une première estimation des paramètres b_0 et c_0. Considérons en effet le modèle de régression linéaire

$$z_i = b_0 + c_0x_i + \zeta_i \tag{3.4}$$

où nous supposons les résidus (ζ_i) indépendants et de même loi gaussienne. La procédure `reg` de `S.A.S.` permet d'obtenir les estimations par moindres carrés b_{init0} et c_{init0} de b_0 et c_0 ajustés dans le modèle (3.4):

	b_{init0}	c_{init0}
t_1	0.428865	-0.120017
t_2	0.613294	-0.146276
t_3	0.338799	-0.140880
t_4	0.166340	-0.141558

Dans l'exemple que nous étudions ici, ces estimations ne peuvent pas être utilisées comme valeurs initiales pour l'estimation par moindres carrés de b_0 et c_0, dans la mesure où elles ne satisfont pas les contraintes de notre modèle (3.2). En effet, les estimations b_{init0} et c_{init0} obtenues ne vérifient pas la contrainte $b_{init0} + c_{init0} < 0$.

Nous devons proposer de meilleures initialisations. Nous définissons pour ce faire les variables v_i $(1 \leq i \leq n)$ par

$$v_i = 1 - \frac{1+x}{a_{init}x} y_i$$

Si l'on considère les résidus du modèle (3.2) comme négligeables, alors pour tout i,

$$v_i \approx \exp(b_0 + c_0 x).$$

Nous considérons donc le modèle de régression non linéaire

$$v_i = \exp(b_0 + c_0 x_i) + \delta_i \qquad 1 \leq i \leq n \tag{3.5}$$

où nous supposons les résidus (δ_i) indépendants et de même loi gaussienne. Nous déterminons les estimations par moindres carrés de b_0 et c_0 dans le modèle (3.5) grâce à la procédure `nlin` de `S.A.S`. Nous devons pour ce faire proposer des valeurs initiales pour les paramètres b_0 et c_0. Nous retenons comme valeurs initiales les estimations b_{init0} et c_{init0} obtenues dans le modèle de régression linéaire (3.4).

Les initialisations que nous obtenons finalement pour chacun des paramètres a_0, b_0 et c_0 sont fournies dans le tableau suivant

	a_{init}	b_{init}	c_{init}
t_1	2.46	-0.030200	-0.080094
t_2	2.51	-0.073023	-0.087370
t_3	2.67	-0.16274	-0.095521
t_4	2.77	-0.15497	-0.10924

Nous utilisons ces initialisations pour déterminer la valeur prise par l'estimateur des moindres carrés $\hat{\theta}$. Nous utilisons alors cette réalisation de la variable $\hat{\theta}$ comme valeur iniale pour le calcul des estimations par moindres carrés dans les modèles incomplets, c'est à dire des réalisations de $\hat{\theta}_{-i}$ $(1 \leq i \leq n)$. Nous calculons ensuite l'estimation jackknife grâce à la formule (3.3).

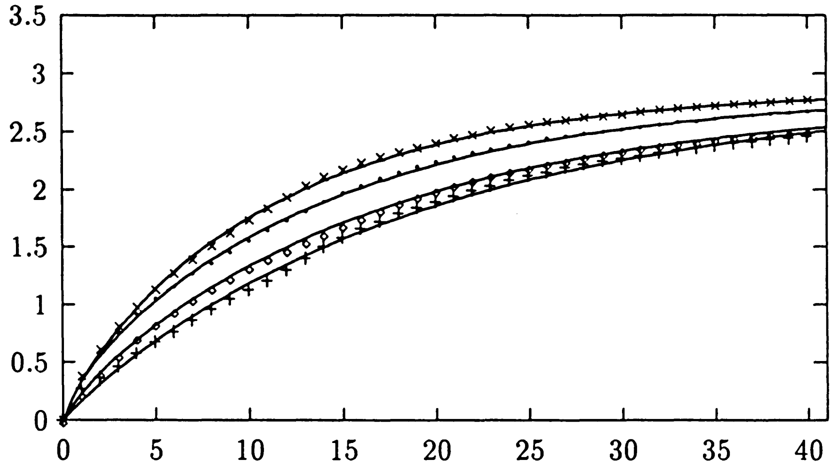
Les estimations jackknife obtenues pour chacun des paramètres sont consignées dans le tableau suivant

	$\tilde{a}$	$\tilde{b}$	$\tilde{c}$
t_1	2.8650529	-0.0719907	-0.0537202
t_2	2.7958263	-0.1261184	-0.0621865
t_3	2.8700630	-0.2006438	-0.0734319
t_4	2.9049411	-0.1826466	-0.0910303

La figure 3.2 représente simultanément les données et les courbes ajustées. Par exemple, pour la température t_1, la courbe ajustée est la courbe d'équation

$$y = \frac{2.865x}{1+x}(1 - \exp(-0.072 - 0.054x)).$$

Figure 3.2: Ajustements jackknife



Les estimations jackknife fournissent de bons ajustements de chacune des courbes de conductivité, Ceci valide qualitativement notre choix d'équation de régression.

Testons à présent l'homogénéité des courbes de conductivité, et l'homogénéité de leur hauteur d'asymptote.

3.3 Comparaison de courbes de croissance

Dans ce paragraphe, nous effectuons des tests d'homogénéité des courbes de croissance, grâce à un modèle d'analyse de variance sur les pseudo-valeurs. Rappelons que si $\hat{\theta}^{(j)}$ désigne l'estimateur des moindres carrés du paramètre $\theta_0^{(j)}$ et pour tout $i \in \{1, ..., n\}$, $\hat{\theta}_{-i}^{(j)}$ désigne l'estimateur des moindres carrés de $\theta_0^{(j)}$ obtenu lorsque la i^{ème} observation est omise, les pseudo-valeurs sont définies par

$$\forall i \in \{1, ..., n\} \qquad \tilde{\theta}_i^{(j)} = n\hat{\theta}^{(j)} - (n-1)\hat{\theta}_{-i}^{(j)}$$

L'estimateur jackknife $\tilde{\theta}^{(j)}$ de $\theta_0^{(j)}$ est alors la moyenne empirique des pseudo-valeurs, c'est à dire

$$\tilde{\theta}^{(j)} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\theta}_i^{(j)}$$

Nous nous appuyons ici sur l'heuristique de Tukey [39], selon laquelle les pseudo-valeurs se comportent comme des variables indépendantes et de même loi gaussienne (heuristique que nous avons validé asymptotiquement dans le second chapitre pour des modèles de régression non linéaires suffisament réguliers).

Nous souhaitons tout d'abord tester, pour un couple (j, k) d'éléments de $\{1, ..., 4\}$ tel que $j \neq k$, l'hypothèse $\theta_0^{(j)} = \theta_0^{(k)}$ contre l'alternative $\theta_0^{(j)} \neq \theta_0^{(k)}$. Nous considérons pour cela le modèle d'analyse de variance suivant

$$\begin{pmatrix} {}^t\tilde{\theta}_1^{(j)} \\ \vdots \\ {}^t\tilde{\theta}_n^{(j)} \\ {}^t\tilde{\theta}_1^{(k)} \\ \vdots \\ {}^t\tilde{\theta}_n^{(k)} \end{pmatrix} = \begin{pmatrix} 1_n & 0 \\ 0 & 1_n \end{pmatrix} \begin{pmatrix} {}^t\theta_0^{(j)} \\ {}^t\theta_0^{(k)} \end{pmatrix} + \begin{pmatrix} {}^t\gamma_1^{(j)} \\ \vdots \\ {}^t\gamma_n^{(j)} \\ {}^t\gamma_1^{(k)} \\ \vdots \\ {}^t\gamma_n^{(k)} \end{pmatrix} \tag{3.6}$$

où les vecteurs résiduels $(\gamma_i^{(j)})$ et $(\gamma_i^{(k)})$ sont supposés indépendants et de même loi gaussienne $\mathcal{N}_3(0, V)$ pour une matrice inconnue V de rang 3. L'estimateur des moindres carrés du paramètre

$$\begin{pmatrix} {}^t\theta_0^{(j)} \\ {}^t\theta_0^{(k)} \end{pmatrix}$$

obtenu dans ce modèle est

$$\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n {}^t\tilde{\theta}_i^{(j)} \\ \frac{1}{n}\sum_{i=1}^n {}^t\tilde{\theta}_i^{(k)} \end{pmatrix} = \begin{pmatrix} {}^t\tilde{\theta}^{(j)} \\ {}^t\tilde{\theta}^{(k)} \end{pmatrix}$$

Notre problème consiste à tester le modèle (3.6) contre le sous modèle

$$\begin{pmatrix} {}^t\tilde{\theta}_1^{(j)} \\ \vdots \\ {}^t\tilde{\theta}_n^{(j)} \\ {}^t\tilde{\theta}_1^{(k)} \\ \vdots \\ {}^t\tilde{\theta}_n^{(k)} \end{pmatrix} = \begin{pmatrix} 1_n & 0 \\ 0 & 1_n \end{pmatrix} \begin{pmatrix} {}^t\theta_0 \\ {}^t\theta_0 \end{pmatrix} + \begin{pmatrix} {}^t\gamma_1^{(j)} \\ \vdots \\ {}^t\gamma_n^{(j)} \\ {}^t\gamma_1^{(k)} \\ \vdots \\ {}^t\gamma_n^{(k)} \end{pmatrix} \tag{3.7}$$

L'estimateur des moindres carrés de θ_0 obtenu dans ce sous modèle est

$$\tilde{\theta}_c = \frac{1}{2}(\tilde{\theta}^{(j)} + \tilde{\theta}^{(k)})$$

Nous testons le modèle (3.6) contre le sous modèle (3.7) par la méthode de Hötelling Lawley (pour une présentation détaillée du test de Hötelling Lawley, on pourra par exemple consulter Coursol [9]). La statistique de Hötelling Lawley s'écrit ici

$$HL = \frac{n}{2(n-1)}\left[{}^t(\tilde{\theta}^{(j)} - \tilde{\theta}_c)\tilde{s}_c^{-2}(\tilde{\theta}^{(j)} - \tilde{\theta}_c) + {}^t(\tilde{\theta}^{(k)} - \tilde{\theta}_c)\tilde{s}_c^{-2}(\tilde{\theta}^{(k)} - \tilde{\theta}_c)\right]$$

où $\tilde{s}_c^2$ est la matrice définie par

$$\tilde{s}_c^2 = \frac{1}{2(n-1)}\left[\sum_{i=1}^n(\tilde{\theta}_i^{(j)} - \tilde{\theta}^{(j)})\,{}^t(\tilde{\theta}_i^{(j)} - \tilde{\theta}^{(j)}) + \sum_{i=1}^n(\tilde{\theta}_i^{(k)} - \tilde{\theta}^{(k)})\,{}^t(\tilde{\theta}_i^{(k)} - \tilde{\theta}^{(k)})\right]$$

La procédure `glm` de `S.A.S.` permet de calculer les estimations par moindres carrés des paramètres d'un modèle de régression linéaire multidimensionnelle. Il est également possible, grâce à cette procédure, de tester un modèle linéaire contre un sous modèle suivant la méthode de Hötelling Lawley. `glm` détermine en effet deux valeurs essentielles: la valeur de la statistique de Hötelling

Lawley pour le test proposé, et une probabilité P précisant le niveau auquel le sous modèle est accepté. Plus précisément, le sous modèle est accepté au niveau $1 - \alpha$ si la probabilité P est supérieure à α.

Ainsi, sous l'hypothèse d'équirépartition gaussienne des pseudo-valeurs, nous acceptons l'hypothèse $\theta_0^{(j)} = \theta_0^{(k)}$ au niveau $1-\alpha$ si $P \geq \alpha$. Rappelons cependant que, bien que les pseudo-valeurs se comportent asymptotiquement comme des variables indépendantes et de même loi gaussienne, elles ne vérifient certainement pas cette propriété dans le cas d'un échantillon fini. Le test défini par la région de rejet $\{P \geq \alpha\}$ est donc de niveau inconnu, et $1 - \alpha$ n'est qu'une estimation du niveau réel de ce test.

Nous obtenons pour chacun des couples (j, k) d'éléments de $\{1, ..., 4\}$ tel que $j \neq k$ les valeurs suivantes de HL et de P:

	HL	P
$j = 1,\ k = 2$	107.0093	0,0001
$j = 1,\ k = 3$	435.4232	0.0001
$j - 1,\ k = 4$	688.7432	0.0001
$j - 2,\ k = 3$	658.2813	0.0001
$j = 2,\ k = 4$	1099.420	0.0001
$j = 3,\ k = 4$	478.5985	0.0001

Pour tout couple (j, k), l'hypothèse $\theta_0^{(j)} = \theta_0^{(k)}$ est donc rejetée au niveau (approximatif) $1 - \alpha = 0.95$. Cela signifie que la température du milieu de croissance influe sur l'évolution au cours du temps du taux de bactéries dans le milieu.

Nous souhaitons à présent tester l'homogénéité des taux maximaux de bactéries. Nous testons, pour un couple (j, k) d'éléments de $\{1, ..., 4\}$ tel que $j \neq k$, l'hypothèse $a_0^{(j)} = a_0^{(k)}$ contre l'alternative $a_0^{(j)} \neq a_0^{(k)}$. Nous considérons à nouveau le modèle d'analyse de variance (3.6), et testons ce modèle contre le sous modèle

$$\begin{pmatrix} {}^t\tilde{\theta}_1^{(j)} \\ \vdots \\ {}^t\tilde{\theta}_n^{(j)} \\ {}^t\tilde{\theta}_1^{(k)} \\ \vdots \\ {}^t\tilde{\theta}_n^{(k)} \end{pmatrix} = \begin{pmatrix} 1_n & 0 \\ 0 & 1_n \end{pmatrix} \begin{pmatrix} a_0 & b_0^{(j)} & c_0^{(j)} \\ a_0 & b_0^{(k)} & c_0^{(k)} \end{pmatrix} + \begin{pmatrix} {}^t\gamma_1^{(j)} \\ \vdots \\ {}^t\gamma_n^{(j)} \\ {}^t\gamma_1^{(k)} \\ \vdots \\ {}^t\gamma_n^{(k)} \end{pmatrix}$$

au moyen de la méthode de Hötelling Lawley et de la procédure **glm**. Nous obtenons les résultats suivants:

	HL	P
$j = 1,\ k = 2$	3.9996	0,0472
$j = 1,\ k = 3$	0.0209	0.8851
$j = 1,\ k = 4$	1.3279	0.2509
$j = 2,\ k = 3$	4.5994	0.0335
$j = 2,\ k = 4$	9.9365	0.0019
$j = 3,\ k = 4$	1.0153	0.3152

L'hypothèse d'homogénéité $a_0^{(2)} = a_0^{(k)}$ est rejetée au niveau (approximatif) $1 - \alpha = 0.95$ pour chaque $k \in \{1,3,4\}$. D'autre part, l'hypothèse $a_0^{(j)} = a_0^{(k)}$ est acceptée au niveau 0.95 pour tout couple d'éléments de $\{1,3,4\}$. Notons qu'en effet, l'estimation jackknife du paramètre d'asymptote $a_0^{(2)}$ est sensiblement inférieure à l'estimation jackknife de $a_0^{(j)}$, pour $j \neq 2$. La température t_2 semble donc jouer un rôle particulier dans l'évolution du taux de bactéries, conduisant à un taux maximal faible. Il semble par contre que les températures t_1, t_3 et t_4 soient équivalentes vis à vis du taux maximal. Nous testons cette dernière hypothèse $a_0^{(1)} = a_0^{(3)} = a_0^{(4)}$ contre l'alternative $\exists j,k \in \{1,3,4\} : a_0^{(j)} \neq a_0^{(k)}$. Nous considérons pour cela le modèle d'analyse de variance

$$\begin{pmatrix} {}^t\tilde{\theta}_1^{(1)} \\ \vdots \\ {}^t\tilde{\theta}_n^{(1)} \\ {}^t\tilde{\theta}_1^{(3)} \\ \vdots \\ {}^t\tilde{\theta}_n^{(3)} \\ {}^t\tilde{\theta}_1^{(4)} \\ \vdots \\ {}^t\tilde{\theta}_n^{(4)} \end{pmatrix} = \begin{pmatrix} 1_n & 0 & 0 \\ 0 & 1_n & 0 \\ 0 & 0 & 1_n \end{pmatrix} \begin{pmatrix} {}^t\theta_0^{(1)} \\ {}^t\theta_0^{(3)} \\ {}^t\theta_0^{(4)} \end{pmatrix} + \begin{pmatrix} {}^t\gamma_1^{(1)} \\ \vdots \\ {}^t\gamma_n^{(1)} \\ {}^t\gamma_1^{(3)} \\ \vdots \\ {}^t\gamma_n^{(3)} \\ {}^t\gamma_1^{(4)} \\ \vdots \\ {}^t\gamma_n^{(4)} \end{pmatrix}$$

où les vecteur résiduels sont supposé indépendants et de même loi centrée gaussienne $\mathcal{N}_3(0,V)$ pour une matrice inconnue V de rang 3, et testons par la méthode de Hötelling Lawley ce modèle contre le sous modèle

$$\begin{pmatrix} {}^t\tilde{\theta}_1^{(1)} \\ \vdots \\ {}^t\tilde{\theta}_n^{(1)} \\ {}^t\tilde{\theta}_1^{(3)} \\ \vdots \\ {}^t\tilde{\theta}_n^{(3)} \\ {}^t\tilde{\theta}_1^{(4)} \\ \vdots \\ {}^t\tilde{\theta}_n^{(4)} \end{pmatrix} = \begin{pmatrix} 1_n & 0 & 0 \\ 0 & 1_n & 0 \\ 0 & 0 & 1_n \end{pmatrix} \begin{pmatrix} a_0 & b_0^{(1)} & c_0^{(1)} \\ a_0 & b_0^{(3)} & c_0^{(3)} \\ a_0 & b_0^{(4)} & c_0^{(4)} \end{pmatrix} + \begin{pmatrix} {}^t\gamma_1^{(1)} \\ \vdots \\ {}^t\gamma_n^{(1)} \\ {}^t\gamma_1^{(3)} \\ \vdots \\ {}^t\gamma_n^{(3)} \\ {}^t\gamma_1^{(4)} \\ \vdots \\ {}^t\gamma_n^{(4)} \end{pmatrix}$$

Nous obtenons:

HL	P
0.9902	0.4565

ce qui confirme l'équivalence des température t_1, t_3 et t_4 du point de vue du taux maximal de batéries.

3.4 Programme S.A.S.

Dans ce paragraphe, nous présentons les données initiales ainsi que le programme S.A.S. grâce auquel ont été effectués les calculs décrits précédemment.

Données initiales

Le tableau `ldl.don` contient les données initiales, c'est à dire l'ensemble des mesures décrites par la figure (3.1). Le vecteur des observations recueillies à la température t_j est noté `yj`; le vecteur des dates d'observations est noté `x`.

```
data ldl.don;
    input x y1 y2 y3 y4;
    cards;
       0      0.00     0.00     0.00     0.00
       1      0.27     0.23     0.35     0.38
       2      0.37     0.41     0.57     0.61
       3      0.46     0.57     0.76     0.81
       4      0.58     0.72     0.92     0.98
       5      0.68     0.84     1.05     1.13
       6      0.77     0.95     1.16     1.27
       7      0.87     1.06     1.27     1.39
       8      0.96     1.15     1.36     1.51
       9      1.05     1.24     1.45     1.62
       10     1.13     1.33     1.55     1.73
       11     1.21     1.41     1.64     1.83
       12     1.30     1.48     1.73     1.93
       13     1.40     1.55     1.81     2.03
       14     1.50     1.62     1.89     2.11
       15     1.58     1.69     1.97     2.17
       16     1.66     1.76     2.03     2.23
       17     1.72     1.83     2.10     2.28
       18     1.79     1.89     2.15     2.32
       19     1.84     1.95     2.20     2.36
       20     1.89     2.00     2.24     2.40
       21     1.94     2.05     2.28     2.44
       22     1.99     2.09     2.32     2.47
       23     2.03     2.13     2.35     2.51
       24     2.08     2.17     2.38     2.54
       25     2.12     2.20     2.41     2.56
       26     2.15     2.23     2.44     2.58
       27     2.19     2.26     2.46     2.60
       28     2.22     2.29     2.48     2.62
       29     2.25     2.32     2.50     2.63
       30     2.27     2.35     2.52     2.65
       31     2.29     2.37     2.54     2.67
       32     2.32     2.39     2.56     2.69
       33     2.34     2.41     2.58     2.70
       34     2.36     2.42     2.59     2.71
       35     2.38     2.44     2.61     2.72
       36     2.40     2.46     2.62     2.73
       37     2.41     2.47     2.64     2.74
```

```
        38      2.43      2.49      2.65      2.75
        39      2.44      2.50      2.66      2.76
        40      2.46      2.51      2.67      2.77
   ;
run;
```

Initialisations

La procédure **means** détermine les valeurs $\max_{1\leq i\leq 41} y_i^{(j)}$, $1 \leq j \leq 4$ et fournit les valeurs initiales pour l'estimation des paramètres $a_0^{(j)}$, $1 \leq j \leq 4$. Les macroprocédures `init1` et `init2` calculent les estimations par moindres carrés de $b_0^{(j)}$ et $c_0^{(j)}$ ajustés dans le modèle (3.4) et (3.5) respectivement. Nous obtenons ainsi les valeurs initiales pour l'estimation des paramètres $b_0^{(j)}$ et $c_0^{(j)}$. Les valeurs initiales sont toutes sauvegardées dans le tableau `ldl.init`.

```
proc means data=ldl.don noprint;
     var y1-y4;
     output out=max
        max=a1 a2 a3 a4;

%macro init;
%do j=1 %to 4;
     data max2;
        set max;
        call symput('a',a&j);
     data init&j;
        set ldl.don;
        if _N_=1 then delete;
        v&j=1-((1+x)/(x*&a))*y&j;
        if v&j>0 then z&j=log(v&j);
     proc reg data=init&j;
        title''regression lineaire &j'';
        model z&j=x ;
%end;
%mend init;
%init
run;

%macro init2;
%do j=1 %to 4;
     data reg;
        b1=0.428865; c1=-0.120017;
        b2=0.613294; c2=-0.146276;
        b3=0.338799; c3=-0.140880;
        b4=0.166340; c4=-0.141558;
        call symput('b',b&j);
        call symput('c',c&j);
     proc nlin data=init&j noprint;
```

```
        parms b=&b c=&c;
        bounds c<0;
        model v&j= exp(b+c*x);
           der.b=exp(b+c*x);
           der.c=x*exp(b+c*x);
        output out=bc&j
           parms=b&j c&j;
%end;
data ldl.init;
     set max;
     %do j=1 %to 4;
        set bc&j;
     %end;
     if _N_>1 then delete;
     keep a1-a4 b1-b4 c1-c4;
proc print data=ldl.init;
     title'initialisations a';
     var a1-a4;
proc print data=ldl.init;
     title'initialisations b';
     var b1-b4;
proc print data=ldl.init;
     title'initialisations c';
     var c1-c4;
%mend init2;
%init2
run;
```

Estimation par moindres carrés dans le modèle complet

La macroprocédure `reg` ci-dessous détermine les estimations par moindres carrés du paramètre $\theta_0^{(j)}$ $(1 \leq j \leq 4)$ ajustés dans le modèle (3.2), en utilisant comme valeurs initiales les valeurs contenues dans `ldl.init`. Les estimations obtenues (notées `ea`, `eb` et `ec`) pour le $j^{ème}$ modèle sont sauvegardées dans `ldl.t&j`. Ce fichier contient également l'estimation `y` du vecteur `yj`, c'est à dire le vecteur défini par

$$y = \frac{ea\ x}{1+x}(1 - \exp(eb + ec\ x)$$

```
%macro reg;
%do j=1 %to 4;
     data init(replace=yes);
        set ldl.init;
        call symput ('a',a&j);
        call symput ('b',b&j);
        call symput ('c',c&j);
     proc nlin data= ldl.don noprint;
        title''regression courbe &j'';
```

```
        parms a=&a b=&b c=&c;
        d=b+c;
        bounds a>=&a, d<0, c<0;
        model y&j=(a*x/(1+x))*(1-exp(b+c*x));
            der.a=(x/(1+x))*(1-exp(b+c*x));
            der.b=-a*(x/(1+x))*exp(b+c*x);
            der.c=-(a*x/(1+x))*(x*exp(b+c*x));
        output out=mct&j parms=ea eb ec r=res ess=SC;
    data ldl.t&j;
        set mct&j;
        y= ea*x/(1+x)*(1-exp(eb+ec*x));
        keep x y y&j res SC ea eb ec;
    data mc&j;
        set mct&j;
        if _N_>1 then delete;
        keep ea eb ec;
    proc print data=mc&j;
        title ''estimateurs des moindres carres pour la courbe &j'';
    proc plot data=ldl.t&j;
        plot res*x='.'/VPOS=25 HPOS=60;
        title''graphe des residus courbe &j'';
    proc plot data=ldl.t&j;
        plot y*x='.'/VPOS=25 HPOS=60;
        title''estimation &j'';
%end;
%mend reg;
%reg
run;
```

Estimation par moindres carrés sur les modèles incomplets

La macroprocédure `regi` ci-dessous détermine les estimations par moindres carrés de $\theta_0^{(j)}$ $(1 \leq j \leq 4)$ dans les modèles privés de la $i^{ème}$ observation $(1 \leq i \leq 41)$. On utilise les estimations obtenues dans le modèle complet comme valeurs initiales (c'est à dire les valeurs contenues dans les tableaux `ldl.t&j`, $1 \leq j \leq 4$).

Les estimations de $\theta_0^{(j)}$ obtenues dans chacun des modèles incomplets sont sauvegardées dans les tableaux `theta&j`, $1 \leq j \leq 4$, qu'il convient d'initialiser.

```
data theta1; input rega regb regc; cards; ;
data theta2; input rega regb regc; cards; ;
data theta3; input rega regb regc; cards; ;
data theta4; input rega regb regc; cards; ;
data mc; input ea eb ec SC; cards; ;

%macro regi;
%do j=1 %to 4;
```

```
    data init;
        set ldl.init;
        call symput ('bounda',a&j);
        set ldl.t&j;
        call symput ('a',ea);
        call symput ('b',eb);
        call symput ('c',ec);
    %do i=1 %to 41;
        data regres (replace=yes);
            set ldl.t&j;
            if _N_ =&i then delete;
        proc nlin data= regres noprint;
            title''regression pour le &j eme modele, lorsque la &i eme observation
est omise'';
            parms a=&a
                b=&b
                c=&c;
                d=b+c;
            bounds a>=&bounda,
                d<0,
                c<0;
            model y&j=(a*x/(1+x))*(1-exp(b+c*x));
                der.a=(x/(1+x))*(1-exp(b+c*x));
                der.b=-a*(x/(1+x))*exp(b+c*x);
                der.c=-(a*x/(1+x))*(x*exp(b+c*x));
            output out=reg&j
                parms=rega regb regc;
        data estim&j;
            set reg&j;
            keep rega regb regc;
            if _N_>1 then delete;
        data theta&j (replace=yes);
            set theta&j estim&j;
    %end;
%end;
%mend regi;
%regi
run;
```

Estimations jackknife. Détermination des pseudo-valeurs

La macro `jackknif` détermine les vecteurs de pseudo-valeurs (notés `pva&j`, `pvb&j` et `pvc&j` dans le $j^{ème}$ modèle) puis calcule les estimations jackknife (notées `ja`, `jb` et `jc`) comme moyenne empirique des pseudo-valeurs. L'ensemble des pseudo-valeurs calculées dans le quatre modèles est sauvegardé dans le tableau `ldl.jk`

```
%macro jackknif;
```

```
%do j=1 %to 4;
    data pseudo&j;
        set theta&j;
        set ldl.t&j;
        pva&j=41*ea-40*rega;
        pvb&j=41*eb-40*regb;
        pvc&j=41*ec-40*regc;
        keep x y&j pva&j pvb&j pvc&j;
    proc means data=pseudo&j;
        title''estimateurs jackknife pour la courbe &j'';
        var pva&j pvb&j pvc&j;
        output out=jk&j
            mean=ja jb jc;
    data call&j;
        set jk&j;
        call symput('a',ja);
        call symput('b',jb);
        call symput('c',jc);
    data resjk&j;
        set ldl.don;
        r&j=y&j-(&a*x/(1+x))*(1-exp(&b+&c*x));
        r2&j=(r&j)**2;
    proc plot data=resjk&j;
        title''graphe des residus jackknife, courbe &j'';
        plot r&j*x='.'  /vpos=25 hpos=60;
    proc means data=resjk&j noprint;
        var r2&j;
        output out=sum&j
            sum=SCjk&j;
%end;
data ldl.jk;
    %do j=1 %to 4;
        set pseudo&j;
    %end;
data sumjk;
    %do j=1 %to 4;
        set sum&j;
    %end;
proc print data=sumjk;
    title''somme des carres residuels jackknife'';
%mend jackknif;
%jackknif
run;

%macro mc;
%do j=1 %to 4;
    data mc (replace=yes);
```

```
        set mc ldl.t&j;
        if _N_>&j then delete;
        keep ea eb ec SC;
%end;
proc print data=mc;
    title''estimations m.c.'';
    var ea eb ec;
proc print data=mc;
    title''somme des carres residuels m.c.'';
    var SC;
%mend mc;
%mc
run;
```

Comparaison de courbes de croissance

Les tests de comparaison décrits dans le paragraphe précédent sont effectués par la macro-procédure **tests**. Pour un couple fixé (j, k) d'éléments de $\{1, ..., 4\}$ tel que $j \neq k$, le tableau de données **glm** contient l'ensemble des pseudo valeurs calculées dans les j^{ème} et k^{ème} modèles. La procédure GLM effectue alors l'analyse de variance (3.6), puis teste l'hypothèse $\theta_0^{(j)} = \theta_0^{(k)}$ contre l'alternative $\theta_0^{(j)} \neq \theta_0^{(k)}$ suivant le méthode décrite dans le paragraphe précédent. Le tableau glm2 contient l'ensemble des pseudo-valeurs calculées dans les quatre modèles. La procédure GLM suivante effectue les tests d'homogénéité des hauteurs d'asymptotes.

```
%macro tests;
title''test des pseudo-valeurs'';
%do i=j %to 4;
    data test&j;
        set ldl.jk;
        ind=&j;
        pva=pva&j;
        pvb=pvb&j;
        pvc=pvc&j;
        keep ind pva pvb pvc;
%end;
%do k=1 %to 3;
    %do k=&j+1 %to 4;
        data glm (replace=yes);
            set test&j test&k;
        proc GLM data=glm;
            title2''comparaison des courbes &j et &k'';
            class ind;
            model pva pvb pvc=ind/ solution nouni;
            manova H=ind;
    %end;
%end;
data glm2 (replace=yes);
```

```
    set test1 test2 test3 test4;
proc GLM data=glm2;
    title2''comparaison des asymptotes'';
    class ind;
    model pva pvb pvc=ind/ nouni solution noint;
    contrast 'asympt 1 et 2'
       ind 1 -1 0 0 ;
    contrast 'asympt 1 et 3'
       ind 1 0 -1 0;
    contrast 'asympt 1 et 4'
       ind 1 0 0 -1 ;
    contrast 'asympt 2 et 3'
       ind 0 1 -1 0 ;
    contrast 'asympt 2 et 4'
       ind 0 1 0 -1;
    contrast 'asympt 3 et 4'
       ind 0 0 1 -1 ;
    contrast 'asympt 1, 3 et 4'
       ind 0 0 1 -1 ,
       ind 1 0 0 -1 ;
    manova H=ind M=(1 0 0 , 0 0 0 , 0 0 0 ) ;
%mend tests;
%tests run;
```

Partie II

Estimation isotonique en régression non paramétrique

Chapitre 4

Sharp asymptotics for isotonic regression

In this paper, we prove global convergence (in the $\mathbb{L}_1$-distance sense) of the isotonic estimator for a monotone regression function. We prove that the $\mathbb{L}_1$-distance between the isotonic estimator and the true function is $O_{\mathbb{P}}(n^{-1/3})$. Moreover, we prove that a centered version of this $\mathbb{L}_1$-distance converges with rate $n^{-1/2}$ to a Gaussian variable.

4.1 Introduction

The aim of this paper is to study the asymptotic behaviour of the $\mathbb{L}_1$-distance between the isotonic estimator and the unknown monotone regression function. More precisely, we consider the following regression model

$$Y_i = f(x_i) + \varepsilon_i \qquad 1 \leq i \leq n$$

where the x_i's are deterministic observation times, Y_i is the observation at time x_i and the ε_i's are independent and identically distributed errors centered at expectation. Moreover, the regression function f is monotone. We prove that the centered $\mathbb{L}_1$-distance between the isotonic estimator and the true function converges at the $n^{-1/2}$ rate to a Gaussian variable. Moreover we compute the asymptotic expectation of the $\mathbb{L}_1$-loss.

In the context of density estimation, Grenander [16] has suggested to use as an estimator of a monotone density (nonincreasing say) the maximum likelihood estimator under the order restriction that the density function is nonincreasing. He showed that this estimator is given by the left-continuous slope of the smallest concave majorant of the empirical distribution function, and can be obtained by a max-min operator. Then Prakasa Rao [30] has established the pointwise convergence of this estimator: with a norming factor of order $n^{1/3}$ (where n is the number of observations), the difference between the Grenander estimator and the density at a fixed point has a nondegenerate limiting distribution which is the distribution of the location of the maximum of the process $\{W(u) - u^2,\ u \in \mathbb{R}\}$, where W is the standard two-sided Brownian motion originating from zero.

Brunk [5] has proposed an analogous estimator for a monotone regression function, which is now usually called the isotonic estimator. This estimator may be presented as the solution of some weighted least square minimization under order restriction (see Barlow et al [1] for more details about that). As a matter of fact it can be also defined (in the nonincreasing case) as the left continuous slope of the smallest concave majorant of F_n, where $F_n(t) = \frac{1}{n}\sum_{i=1}^n Y_i \mathbb{I}_{x_i \leq t}$. Brunk [6] established the pointwise convergence of the isotonic estimator with rate of convergence $n^{-1/3}$, whenever the regression function is differentiable. Wright [41] generalized this result to Hölderian regression functions, making different rates of convergence appear.

The purpose of this paper is to study the global asymptotic behaviour of the isotonic estimator with respect to the $\mathbb{L}_1$-distance. Groeneboom [17] was first to realize that, in the context of density estimation, sharp asymptotics for the $\mathbb{L}_1$-loss of the Grenander estimator could be obtained. He stated that the expectation of the $\mathbb{L}_1$-loss converges at the rate $n^{-1/3}$ to some functional $\phi(f)$ of the density f. Moreover, more remarkably, substracting $n^{-1/3}\phi(f)$ to the $\mathbb{L}_1$-risk, one gets a convergence in distribution at the $n^{-1/2}$ rate to a centered Gaussian law with fixed variance (we mean independent of f). He proposed a sketch of proof of these results [1]. Our approach here will be to follow the line of proof proposed by Groeneboom. Our main result is theorem 4.2 below, which ensures the convergence in distribution of a properly centered and normalized version of the $\mathbb{L}_1$-risk of the isotonic estimator under adequate smoothness assumption on the regression function and moments conditions on the errors. The two main steps of the proof are the following:

- First, we establish a similar result for the isotonic estimator of a signal function in a white noise model.

- Next, using a strong approximation argument (which is a refinement due to Sakhanenko of the Komlós, Major and Tusnády construction) we build a white noise model which approximates the regression model. The quality of approximation depends on the integrability of the errors. We show that, if this approximation is good enough, then the isotonic estimator f_n of f in the regression model and the corresponding estimator $\tilde{f}_n$ in the white noise model are asymptotically equivalent for the normal convergence of the $\mathbb{L}_1$-loss. Namely we prove that

$$\mathbb{E}\int |\tilde{f}_n(t) - f_n(t)|dt \;=\; o(n^{-1/2}) \qquad (4.1)$$

The article is organized as follows:
The isotonic estimator of a monotone regression function is defined in section 4.2. The main theorem is stated in section 4.2. The rest of the article is devoted to the proof of theorem 4.2. The main tools used in the proof are presented in section 4.3. The exponential inequality obtained in theorem 4.4 is a fundamental tool that we shall use repeatedly. In section 4.4, we prove global convergence of the isotonic estimator of a signal function in a white noise model. The construction of a white noise model with isotonic estimator close enough to the isotonic estimator of the regression function and the proof of (4.1) are performed in section 4.5.

[1] At the time of the final writing of this paper we are aware that Groeneboom et al [19] have established a complete and new proof of these results.

4.2 Definitions and main results

We first define the isotonic estimator of a monotone regression function and its inverse process. Let f be a monotone function defined over $[0,1]$ and consider the regression model

$$Y_i = f(x_i) + \varepsilon_i \qquad 1 \leq i \leq n \tag{4.2}$$

where the x_i's are deterministic observation times, Y_i is the observation at time x_i and the ε_i's are independent and identically distributed errors centered at expectation. The isotonic estimator of f under the order restriction that f be nonincreasing is defined as:

$$\forall t \in [0,1] \qquad f_n(t) = \max_{x_i \geq t} \min_{t \geq x_j} \sum_{i \leq k \leq j} \frac{Y_k}{j-i}$$

Analogously, under the order restriction that f be nondecreasing the isotonic estimator of f is

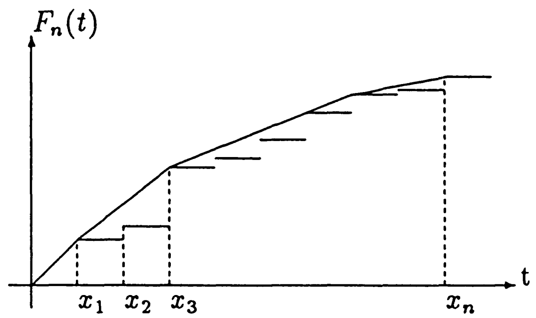
$$\forall t \in [0,1] \qquad f_n(t) = \max_{x_i \leq t} \min_{t \leq x_j} \sum_{j \leq k \leq i} \frac{Y_k}{i-j}$$

Throughout the sequel, we will focus on the nonincreasing case. Of course, the nondecreasing case may be straightforwardly deduced from that case by changing f into $-f$. Let F_n be the empirical process defined by

$$\forall t \in [0,1] \qquad F_n(t) = \frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{1}_{x_i \leq t} \tag{4.3}$$

Then, f_n is the left-continuous slope of the smallest concave majorant of F_n. In figure 4.2 below, F_n and its concave majorant are represented: F_n is the step function and its concave majorant is the continuous piecewise affine function.

Figure 4.1: The empirical process and its concave majorant



The isotonic estimator of a monotone regression function has been studied by Brunk [5] and Wright [41]. They proved pointwise convergence of the isotonic estimator and determined the asymptotic distribution of $f_n(t)$ for a fixed t. Theorem 2.2 of Brunk [6] in the particular case where the random errors are i.i.d. and Gaussian is the following

Theorem 4.1 *Assume we are given regression model above, where the ε_i's are i.i.d. centered Gaussian variables and f is decreasing over $[0,1]$. For fixed n, let one observation be made at* *each of the observation points* $x_i = i/n$. *Let* $t \in]0,1[$ *and let there be a neighbourhood of* t *in which* f' *is continuous. Then*

$$n^{1/3}\left|\frac{2}{\sigma^2 f'(t)}\right|^{1/3}(f_n(t) - f(t))$$

converges in distribution to $V(0)$.

(see definition 4.2 below for a definition of $V(0)$). Wright [41] generalized this theorem. In his theorem 1, he assumed f satisfies $|f(t) - f(x)| = A|x - t|^\alpha(1 + o(1))$ as $x \to t$ for some α and A both positive. He proved that the rate of convergence is in that case of order $n^{-\alpha/2\alpha+1}$.

The key idea of Groeneboom [17] for studying the global convergence of the Grenander estimator is to consider the generalized inverse functions of both the estimator and the density. The point is that the $\mathbb{L}_1$-distance is invariant under this transformation and that the so-defined inverse process is more directly describable from the empirical distribution function. Let us recall that the (generalized) inverse h^{-1} of a non increasing function h is defined in such a way that the following equivalence holds true for each $t \in [0,1]$ and $a \in \mathbb{R}$

$$h(t) \le a \Longleftrightarrow h^{-1}(a) \le t$$

We shall stick to this approach and consider the inverse function V_n of the isotonic estimator f_n and the inverse function g of f. It follows that (see appendix B)

$$\int_0^1 |f_n(t) - f(t)|dt = \int_{\mathbb{R}} |V_n(a) - g(a)|da$$

Moreover, V_n may be computed as the location of the maximum of some empirical process. Let us first define what we mean by location of the maximum of a process.

Definition 4.1 *Let* X *be a process indexed by a real subset* I. *The* argmax *of* X *over* I *is the random variable defined as the greatest location of the maximum of* X *over* I *(using the usual convention that the supremum of an empty set is* ∞*).*

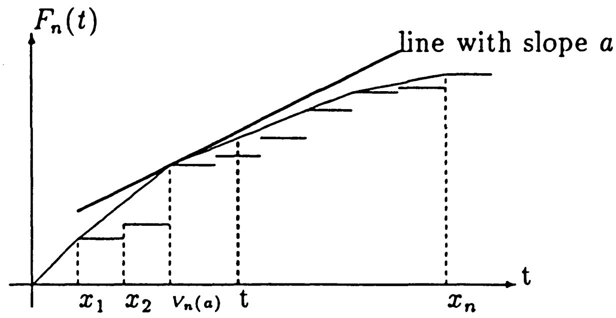
$$\operatorname*{argmax}_{u\in I}\{X(u)\} = \sup\{u \in I, X(u) = \sup_{x\in I} X(x)\}$$

It can easily be shown that

$$\forall a \in \mathbb{R} \qquad V_n(a) = \operatorname*{argmax}_{u\in[0,1]}\{F_n(u) - au\}$$

where F_n is defined by (4.3). This fact is illustrated by figure 4.2 below. One can see on this figure that V_n is the inverse process of f_n: fix $a \in \mathbb{R}$ and consider $t \ge V_n(a)$. Then, the slope at t of the concave majorant of F_n (which is $f_n(t)$) is no more than a.

Figure 4.2: The inverse process



The asymptotic distribution of the $\mathbb{L}_1$-loss of the isotonic estimator will crucially depend, via some delicate approximation arguments that will be detailed later on, on a location process for the maximum of drifted Brownian motion. More precisely the process involved to decribe the asymptotic distribution has been introduced by Groeneboom (see [17] and [18]). We shall call it *Groeneboom's process* in the sequel and it can be defined as follows:

Definition 4.2 *Let W be the standard two-sided Brownian motion originating from zero. Then, Groeneboom's process V is defined by: $\forall a \in \mathbb{R}$, $V(a) = \underset{u \in \mathbb{R}}{\operatorname{argmax}}\{W(u) - (u-a)^2\}$.*

It has been shown by Groeneboom that

Proposition 4.1 *1. $\{V(a) - a\}_{a \in \mathbb{R}}$ is a stationnary process*

2. $k = \int_0^\infty \operatorname{cov}(|V(0)|, |V(b) - b|)db$ is finite

Some more properties are given in Groeneboom [17] and [18]. In particular, Groeneboom discussed the analytical properties of the process $(V(a),\ a \in \mathbb{R})$ and showed that the density of $V(0)$ can be found from solution of an integral equation. The density of $V(0)$ can be characterized in terms of Airy functions.

We consider the following regularity conditions for f for some m and $M \in \mathbb{R}$

$\mathcal{R}$: f is decreasing from $[0,1]$ onto $[m, M]$
f is twice differentiable with non vanishing first derivative and bounded second derivative over $[0,1]$.

We turn now to the statement of our main result.

Theorem 4.2 *Assume we are given regression model (4.2), where f satisfies $\mathcal{R}$. Suppose that*

1. *the ε_i's are i.i.d. variables with zero mean and variance σ^2, and $\mathbb{E}|\varepsilon_1|^p$ is finite for some $p > 12$*

2. *for each $i \in \{1, ..., n\}$, $x_i = i/n$.*

Let f_n be the isotonic estimator of f. Then

$$n^{1/6}\left\{n^{1/3}\int_0^1 |f_n(t)-f(t)|dt - C_f\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 8k\sigma^2) \text{ as } n\to\infty$$

where C_f and k are defined from the Groeneboom process V by the following equations $C_f = 2\mathbb{E}|V(0)|\int_0^1 \left|\frac{\sigma^2}{2}f'(t)\right|^{1/3} dt$ and $k = \int_0^\infty \operatorname{cov}(|V(0)|, |V(b)-b|)db$.

Remark 4.1

- By numerical integration, Groeneboom [17] gave approximations for the expectation $2\mathbb{E}|V(0)|$ and the constant $8k$. These constants are approximately equal to 0.82 and 0.17 respectively.
- For sake of simplicity, we have assumed the design points to be uniformly spread, but it is clear from our proof that this assumption may be replaced by the weaker condition that

$$\exists c_0 \in]0,1/2[, \qquad \left|x_i - \frac{i}{n}\right| \le \frac{c_0}{n} \qquad \forall i \in \{1,\dots,n\}$$

- We have assumed in theorem 4.2 a rather stringent integrability condition on the errors. We do not think that this condition is optimal. However, as a by-product of our proof it is easy to derive some weaker result than the above limit theorem but which holds under the square integrability of the errors.

Theorem 4.3 *Assume we are given regression model (4.2), where f satisfies $\mathcal{R}$. Suppose that*

1. *the ε_i's are i.i.d. variables with zero mean and finite variance*
2. *for each $i \in \{1,\dots,n\}$, $x_i = i/n$.*

Let f_n be the isotonic estimator of f. Then

$$\mathbb{E}\int_0^1 |f_n(t)-f(t)|dt = O(n^{-1/3})$$

4.3 Location of the maximum of a drifted Brownian motion

The location of the maximum of a process X is well defined, even if X either achieves its supremum at two distinct points or does not achieve its supremum (see definition 4.1). However, it will be useful in some cases to know if the supremum of a process can be achieved at two distinct points. Kim and Pollard (see [22], lemma 2.6) stated that Gaussian processes cannot achieve their supremum at two distinct points. Let us recall their result:

Lemma 4.1 *Let $\{X(t), t\in T\}$ be a Gaussian process with continuous sample paths, indexed by a σ-compact metric space T. If $\operatorname{var}(X(s)-X(t)) \ne 0$ for $s\ne t$, then, for almost all ω, no sample path can achieve its supremum at two distinct points of T.*

The two processes to be useful in the proof of the main theorem 4.2 are Groeneboom's process (see definition 4.2) and the inverse process of the isotonic estimator in a white noise model. By definition, Groeneboom's process is the location of the maximum of a Brownian motion with parabolic drift, and one can prove that a well normalized version of the inverse process in a white noise model is the location of the maximum of a Brownian motion with drift bounded from below by some parabola. This is the reason why we are interested by the properties of such processes.

Let W be the standard two-sided Brownian motion originating from zero, and D some deterministic function. D will be assumed to satisfy the following

$\mathcal{D}$: $D \in C^1(I)$ where $I \subset \mathbb{R}$ contains 0,
$D(0) = 0$,
There exists some positive C_0 such that for each $t \in I$, $D(t) \geq C_0 t^2$.

Let Z denote the Brownian motion with drift D:

$$\forall t \in I \qquad Z(t) = W(t) - D(t) \tag{4.4}$$

and z the location of the maximum of Z: $z = \text{argmax}_{u \in I}\{Z(u)\}$.

We first provide a probability bound for the location of the maximum of the drifted Brownian motion Z, whenever the drift D satisfies $\mathcal{D}$. This inequality turns out to be a fundamental tool in the proof of the main theorems, since it yields exponential inequalities for both the Groeneboom process at zero and the inverse process in a white noise model at a fixed point (see corollary 4.1 and lemma 4.4 below).

Theorem 4.4 *Let Z be the drifted Brownian motion defined in (4.4) with D satisfying $\mathcal{D}$. Then*

$$\forall t \geq 0 \qquad \mathbb{P}\left(|z| \geq t\right) \leq 2\exp\left(-\frac{C_0^2 t^3}{2}\right)$$

Remark 4.2 By a Borel-Cantelli argument we can derive from theorem 4.4 that z is almost surely finite, which proves that Z achieves its supremum. On the other hand by lemma 4.1, with probability one Z cannot achieve its supremum at two distinct points. Thus, with probability one, a Brownian motion with drift satisfying $\mathcal{D}$ achieves its supremum at a unique point.

Proof of theorem 4.4

Since $W(0) = 0$ almost surely, $0 \in I$ and $D(0) = 0$, the supremum of $W - D$ over I is almost surely nonnegative. Let $t > 0$. There exists some positive C_0 such that $D(u) \geq C_0 u^2$ for each $u \in \mathbb{R}$, thus

$$\begin{aligned}
\mathbb{P}(|z| \geq t) &\leq \mathbb{P}\left(\sup_{|u|\geq t}\{W(u) - D(u)\} \geq 0\right) \\
&\leq \mathbb{P}\left(\sup_{|u|\geq t}\left\{\frac{W(u)}{u^2}\right\} \geq C_0\right) \\
&\leq \mathbb{P}\left(\sup_{|u|\leq 1/t}\{u^2 W(1/u)\} \geq C_0\right)
\end{aligned}$$

But W has the same distribution as $uW(1/u)$ and the distribution of W is symmetric thus

$$\begin{aligned} \mathbb{P}(|z| \geq t) &\leq \mathbb{P}\left(\sup_{|u|\leq 1/t}\{uW(u)\} \geq C_0\right) \\ &\leq 2\mathbb{P}\left(\sup_{0\leq u\leq 1/t}\{W(u)\} \geq C_0 t\right) \end{aligned}$$

and since $W(\; . \; /t)$ has the same distribution as $t^{-1/2}W(\; . \;)$ we get

$$\begin{aligned} \mathbb{P}(|z| \geq t) &\leq 2\mathbb{P}\left(\sup_{0\leq u\leq 1}\{W(u)\} \geq C_0 t^{3/2}\right) \\ &\leq 2\exp\left(-\frac{C_0^2 t^3}{2}\right) \end{aligned}$$

and we get theorem 4.4. ◊

In the particular case where $I = \mathbb{R}$ and $\forall u \in \mathbb{R}$, $D(u) = u^2$ we get the following

Corollary 4.1 *Let V be Groeneboom's process. Then for each $t > 0$,*

$$\mathbb{P}\left(|V(0)| \geq t\right) \leq 2\exp\left(-\frac{t^3}{2}\right)$$

The proof of theorem 4.4 (and thus of its corollary 4.1) is very simple: the only tools we use here are standard properties of the two-sided Brownian motion originating from zero. Thus the proof may seem to be crude. However, it provides the right exponential decay up to a constant. In fact, Groeneboom proved that the density function ϕ of $V(0)$ satisfies

$$\phi(t) \sim c_0|t|\exp\left(-\frac{2}{3}|t|^3 + c_1|t|\right)$$

as $|t| \to \infty$ for some constants $c_0 > 0$ and $c_1 < 0$ (see [18], corollary 3.4). Thus, for large t, $\log\mathbb{P}(|V(0)| > t)$ is of order $-\frac{2}{3}t^3$. In theorem 4.4 we obtain that $\log\mathbb{P}(|V(0)| > t)$ is of order $-t^3/2$.

The last result of this section is the evaluation of the smallest distance between the maximum of the drifted Brownian motion Z and the maximum of this process outside some neighbourhood of the location of its maximum. We have in fact in view to guarantee that with high probability, if δ is small enough, for each t with $|t - z| > \delta$, $Z(z) - Z(t)$ is of order at least $\delta^{3/2}$. Lemma 4.2 will allow us to prove that if Y_n is a sequence of processes which converges uniformly in probability to Z and if Y_n is smooth enough, then the sequence of the location of the maximum of Y_n converges in probability to z. Moreover, it relates the rate of convergence of the sequence of locations of maximum to the rate of convergence of the sequence of processes Y_n (see lemma 4.9 below).

Lemma 4.2 *Let Z be the drifted Brownian motion defined in (4.4) where D satisfies $\mathcal{D}$ and z be the almost surely unique and finite location of the maximum of Z. If the derivative s of D satisfies*

$$\sup_{|t|\leq T} s^2(t) \leq \frac{1}{2\delta\log\frac{1}{2\pi\delta}}$$

for some $\delta > 0$, $x > 0$ and $T \geq 0$ then

$$\mathbb{P}\left(\left|\sup_{t\in I} Z(t) - \sup_{|t-z|>\delta} Z(t)\right| \leq x\delta^{3/2}\right) \leq 8eTx + 12ex\delta + \mathbb{P}(|z| > T)$$

We prove lemma 4.2 by using the following lemma, which relates the probability that the supremum of a drifted Brownian motion lies in some real interval to the probability that a Brownian motion lies in this interval.

Lemma 4.3 *Let d be some real valued function such that $\int_0^1 d^2(x)dx$ is finite, and Z be the drifted Brownian motion defined by*

$$\forall t \in [0,1] \qquad Z(t) = W(t) - \int_0^t d(x)dx$$

Then, for each $a, b \in \mathbb{R}$

$$\mathbb{P}\left(\sup_{t\in[0,1]} Z(t) \in [a,b]\right) \leq e|b-a|$$

whenever $0 < |b-a| < 1$ and

$$\int_0^1 d^2(x)dx \leq \frac{1}{2\log\frac{1}{|b-a|}}$$

Proof of lemma 4.3

Let a and b be real numbers such that $a \neq b$. By Cameron-Martin-Girsanov's theorem,

$$\mathbb{E}\left(\mathbb{1}_{\sup_{t\in[0,1]} Z(t)\in[a,b]}\right) = \mathbb{E}\left(\mathbb{1}_{\sup_{t\in[0,1]} W(t)\in[a,b]} \exp(\sqrt{v}G - v/2)\right)$$

where $v = \int_0^1 d^2(x)dx$ and $G = -\frac{1}{\sqrt{v}}\int_0^1 d(x)dW(x)$. By Hölder's inequality, for each $q > 1$,

$$\mathbb{P}\left(\sup_{t\in[0,1]} Z(t) \in [a,b]\right) \leq \exp(-v/2)\left[\mathbb{E}\left(\exp(q\sqrt{v}G)\right)\right]^{1/q}\left[\mathbb{P}\left(\sup_{t\in[0,1]} W(t) \in [a,b]\right)\right]^{1-1/q}$$

where

$$\mathbb{E}\left(\exp(q\sqrt{v}G)\right) = \exp(q^2v/2)$$

since G is a standard Gaussian variable. Let ϕ be the density function of $|G|$. Then, $\sup_{t\in\mathbb{R}}|\phi(t)| \leq 1$, and since $\sup_{t\in[0,1]} W(t)$ has the same distribution as $|G|$ we get

$$\mathbb{P}\left(\sup_{t\in[0,1]} W(t) \in [a,b]\right) \leq |b-a|$$

Therefore,

$$\mathbb{P}\left(\sup_{t\in[0,1]} Z(t) \in [a,b]\right) \leq |b-a|\exp\left(\frac{qv}{2} + \frac{1}{q}\log\frac{1}{|b-a|}\right)$$

Suppose $0 < |b-a| < 1$ and $v < 2\log(\frac{1}{|b-a|})$. Taking $q = \sqrt{\frac{2}{v}\log\frac{1}{|b-a|}}$ yields

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in[0,1]} Z(t) \in [a,b]\right) &\leq |b-a|\exp\left(\sqrt{2v\log\frac{1}{|b-a|}}\right)\\ &\leq e|b-a|\end{aligned}$$

whenever $v \leq \frac{1}{2\log\frac{1}{|b-a|}}$. Suppose now $v \geq 2\log\frac{1}{|b-a|}$. Then, $e|b-a| \geq 1$ whenever $v \leq \frac{1}{2\log\frac{1}{|b-a|}}$, which completes the proof of lemma 4.3. ◇

Proof of lemma 4.2

Let N_0 be some real interval such that $N_0 \subset [-T,T]$ and $N_0 = [t_0, t_0+\delta]$ for some $t_0 \in \mathbb{R}$. The increments of Z are independent since the increments of a Brownian motion are independent thus for each positive α

$$\begin{aligned}
&\mathbb{P}\left(\left|\sup_{t\in N_0} Z(t) - \sup_{t<t_0} Z(t)\right| \leq \alpha \Big/ \sup_{t<t_0} Z(t) - Z(t_0) = c\right) = \\
&\quad = \mathbb{P}\left(\left|\sup_{t\in N_0} Z(t) - Z(t_0) - c\right| \leq \alpha \Big/ \sup_{t<t_0} Z(t) - Z(t_0) = c\right) \\
&\quad = \mathbb{P}\left(\left|\sup_{t\in N_0} Z(t) - Z(t_0) - c\right| \leq \alpha\right)
\end{aligned}$$

Averaging over the possible realizations of $\sup_{t<t_0}\{Z(t) - Z(t_0)\}$ we get

$$\mathbb{P}\left(\left|\sup_{t\in N_0} Z(t) - \sup_{t<t_0} Z(t)\right| \leq \alpha\right) \leq \sup_{c\in\mathbb{R}} \mathbb{P}\left(\left|\sup_{t\in N_0} Z(t) - Z(t_0) - c\right| \leq \alpha\right)$$

Changes of origine and changes of scale in te Brownian motion W yields for each real number c

$$\begin{aligned}
\mathbb{P}\left(\left|\sup_{t\in N_0} Z(t) - \sup_{t<t_0} Z(t)\right| \leq \alpha\right) &\leq \mathbb{P}\left(\left|\sup_{t\in[0,1]} W(t) - \delta^{-1/2}(D(\delta t + t_0) - D(t_0) - c)\right| \leq \alpha\delta^{-1/2}\right) \\
&\leq \mathbb{P}\left(\sup_{t\in[0,1]} W(t) - D_\delta(t) \in [(c-\alpha)\delta^{-1/2}, (c+\alpha)\delta^{-1/2}]\right)
\end{aligned}$$

where D_δ is the function of $C^1[0,1]$ defined by $\forall t \in [0,1]$, $D_\delta(t) = \delta^{-1/2}(D(\delta t + t_0) - D(t_0))$. Let s_δ be the derivative of D_δ. Then, $\int_0^1 s_\delta^2(x)dx = \int_{t_0}^{t_0+\delta} s^2(x)dx$ and lemma 4.3 yields

$$\mathbb{P}\left(\left|\sup_{t\in N_0} Z(t) - \sup_{t<t_0} Z(t)\right| \leq \alpha\right) \leq 2e\alpha\delta^{-1/2}$$

whenever s satisfies

$$\sup_{|t|\leq T} s^2(t) \leq \frac{1}{2\delta\log\frac{1}{2\alpha\delta^{-1/2}}} \tag{4.5}$$

Using the same arguments and since the distribution of a two-sided Brownian motion originating from zero is symmetric we get

$$\begin{aligned}
\mathbb{P}\left(\left|\sup_{t\in N_0} Z(t) - \sup_{t>t_0+\delta} Z(t)\right| \leq \alpha\right) &\leq \sup_{c\in\mathbb{R}} \mathbb{P}\left(\left|\sup_{t\in N_0} Z(t) - Z(t_0+\delta) - c\right| \leq \alpha\right) \\
&\leq 2e\alpha\delta^{-1/2}
\end{aligned}$$

whenever s satisfies (4.5), and therefore

$$\mathbb{P}\left(\left|\sup_{t\in N_0} Z(t) - \sup_{t\notin N_0} Z(t)\right| \leq \alpha\right) \leq 4e\alpha\delta^{-1/2} \tag{4.6}$$

Let $z = \text{argmax}_{u\in I}\{Z(u)\}$ and suppose first that $|z| \le T$. We will finish the proof of lemma 4.2 by separating $[-T,T]$ into intervals of length δ. For each integer k such that $-T\delta^{-1}-1 \le k \le T\delta^{-1}+1$, let N_k be the interval $[k\delta,(k+1)\delta]\cap[-T,T]$. We get for each $\alpha > 0$

$$\begin{aligned}\mathbb{P}\left(\left|\sup_{t\in I} Z(t) - \sup_{|t-z|>\delta} Z(t)\right| \le \alpha \;;\; |z|\le T\right) &\le \\ &\le \sum_k \mathbb{P}\left(\left|\sup_{t\in I} Z(t) - \sup_{|t-z|>\delta} Z(t)\right| \le \alpha \;;\; z\in N_k\right) \\ &\le \sum_k \mathbb{P}\left(\left|\sup_{t\in N_k} Z(t) - \sup_{t\notin N_k} Z(t)\right| \le \alpha\right)\end{aligned}$$

where the sum is taken over all the integers k such that $-T\delta^{-1}-1 \le k \le T\delta^{-1}+1$. (4.6) then yields

$$\mathbb{P}\left(\left|\sup_{t\in I} Z(t) - \sup_{|t-z|>\delta} Z(t)\right| \le \alpha\right) \;\le\; 4e\alpha\delta^{-1/2}(2T\delta^{-1}+3) + \mathbb{P}(|z|>T)$$

whenever s satisfies (4.5). Choosing $\alpha = x\delta^{3/2}$ yields lemma 4.2. $\diamond$

4.4 $\mathbb{L}_1$-convergence of the isotonic estimator in a white noise model

As mentioned before, the key idea in the proof of $\mathbb{L}_1$-convergence of the isotonic estimator of a monotone regression function is to build a white noise model in such a way that the isotonic estimator obtained in this white noise model

- satisfies to a limit theorem which is an analogue of theorem 4.2.
- is asymptotically equivalent to the isotonic estimator in the regression model in the sense that (4.1) holds.

We begin by defining the isotonic estimator of a signal function in a white noise model and then prove a limit theorem for its $\mathbb{L}_1$-loss. The construction of a white noise model which approximates the regression model and the proof of (4.1) will be performed in the next section.

Let f be a decreasing function from [0,1] to $[m, M]$ (f is the signal function), W the two-sided Brownian motion originating from zero and let $Y_n(t)$ ($t \in [0,1]$) be the observations in the white noise model

$$dY_n(t) = f(t)dt + \frac{1}{\sqrt{n}}dW(t) \tag{4.7}$$

Let F be defined by $\forall t \in [0,1]$, $F(t) = \int_0^t f(x)dx$. The isotonic estimator f_n of f is the left continuous slope of the least concave majorant of $F(.)+W(.)/\sqrt{n}$. The following results provide a central limit theorem for the $\mathbb{L}_1$-loss of f_n and a computation of the asymptotic bias.

Theorem 4.5 *Assume we are given the white noise model (4.7) where the signal function f satisfies $\mathcal{R}$ and let f_n be the isotonic estimator of f. Then*

$$n^{1/2}\left\{\int_0^1 |f_n(t)-f(t)|dt - \mathbb{E}\int_0^1 |f_n(t)-f(t)|dt\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 8k) \text{ as } n\to\infty$$

Theorem 4.6 *Assume we are given the white noise model (4.7) where the signal function f satisfies $\mathcal{R}$ and let f_n be the isotonic estimator of f. Then*

$$\lim_{n\to\infty} n^{1/6}\left(n^{1/3}\mathbb{E}\int_0^1 |f_n(t)-f(t)|dt - C_f\right) = 0$$

where $C_f = 2\mathbb{E}|V(0)|\int_0^1 |f'(t)/2|^{1/3}\,dt$ *and* $k = \int_0^\infty cov(|V(0)|, |V(b)-b|)db.$

Theorem 4.6 means that the expectation of the $\mathbb{L}_1$-distance between the signal and its isotonic estimate normalized by $n^{1/2}$ is asymptotically equivalent to $n^{1/6}C_f$, and as a consequence of theorems 4.5 and 4.6 we get

$$n^{1/6}\left\{n^{1/3}\int_0^1 |f_n(t)-f(t)|dt - C_f\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,8k) \text{ as } n\to\infty$$

So, we obtain in the context of white noise the kind of result we want to prove in the context of regression.

In order to study the $\mathbb{L}_1$-distance between f_n and f we use the identity

$$\int_0^1 |f_n(t)-f(t)|dt = \int_{\mathbb{R}} |U_n(a)-g(a)|da$$

where U_n is the inverse function of f_n and g is the inverse function of f. We shall also use the expression of U_n as

$$\forall a\in\mathbb{R} \qquad U_n(a) = \operatorname*{argmax}_{u\in[0,1]}\{W(u)+\sqrt{n}(F(u)-au)\} \tag{4.8}$$

and we shall prove in lemma 4.12 that the $\mathbb{L}_1$-distance between f_n and f is asymptotically equivalent to $\int_m^M |U_n(a)-g(a)|da$, namely

$$\mathbb{E}\int_{\mathbb{R}} |U_n(a)-g(a)|da = \mathbb{E}\int_m^M |U_n(a)-g(a)|da + o(n^{-1/2}) \tag{4.9}$$

which means that we only have to deal with $\int_m^M |U_n(a)-g(a)|da$. Section 4.4.1 is devoted to the proof of theorem 4.5 while section 4.4.2 is devoted to the proof of theorem 4.6.

4.4.1 Asymptotic normality

We derive properties of the $\mathbb{L}_1$-distance $\int_m^M |U_n(b)-g(b)|db$ by embedding it in a process B_n defined as follows

Definition 4.3 *Let U_n be the process defined by (4.8). Let define the process B_n by*

$$\forall t\in[m,M] \qquad B_n(t) = \int_m^t |U_n(a)-g(a)|da$$

Let B_n^* be its centered associated process: $\forall t\in[m,M]$, $B_n^*(t) = B_n(t) - \mathbb{E}B_n(t)$.

In this section, we prove a more general result than asymptotic normality of the centered $\mathbb{L}_1$-distance between U_n and g: actually we show that $\sqrt{n}B_n^*$ converges in $\mathcal{C}[m,M]$ to a Gaussian process. More precisely, section 4.4.1 is devoted to the proof of the following

Proposition 4.2 *Let f be a decreasing function from $[0,1]$ onto $[m,M]$, g its inverse and F the primitive of f such that $F(0) = 0$. Let B_n^* be the process defined in definition 4.3. If f satisfies $\mathcal{R}$ then*

$$\sqrt{n}B_n^* \xrightarrow{\mathcal{D}} W \circ v$$

where $\forall t \in [m,M]$, $v(t) = 8k(1-g(t))$ and $k = \int_0^\infty \operatorname{cov}(|V(0)|, |V(b)-b|)db$.

As a particular case of proposition 4.2, the random variable $B_n^*(M)$ converges in distribution with rate of convergence $n^{-1/2}$ to a centered Gaussian variable with variance $8k$, which proves, since (4.9) holds true, that theorem 4.5 is a direct consequence of proposition 4.2.

To prove proposition 4.2, we shall use a refinement of Billingsley's theorem (see [3], theorem 19.1). The statement that we give here differs from the original one in two points:

- We use a slightly different weak dependency assumption on the increments (see definition 4.4 below).
- We study the convergence to $W \circ v$ instead of W, where W is a Brownian motion and v is a variance function.

The proof can be straightforwardly performed by following the lines of the proof of Billingsley's theorem (see appendix A).

Definition 4.4 *Let $\{X_n,\ n \in \mathbb{N}\}$ and $\{Y_n,\ n \in \mathbb{N}\}$ be sequences of random variables. These sequences are said weakly asymptotically independent if for each $x, y \in \mathbb{R}$*

$$\lim_{n\to\infty} \left| \mathbb{E}\left(e^{i(xX_n + yY_n)}\right) - \mathbb{E}\left(e^{ixX_n}\right)\mathbb{E}\left(e^{iyY_n}\right) \right| = 0$$

Proposition 4.3 *Let m, M be real numbers such that $m \le M$. Let $\{X_n,\ n \in \mathbb{N}\}$ be a tight sequence in $C[m,M]$ and for each s and $t \in [m,M]$, let $\Delta_n(t,s) = X_n(t) - X_n(s)$. Suppose*

h_1 - *For each $n \in \mathbb{N}$, $X_n(0) = 0$ a.s.*

h_2 - *For each s_1, t_1, s_2, t_2 in $[m,M]$ with $s_1 \le t_1 < s_2 \le t_2$, $\Delta_n(t_1,s_1)$ and $\Delta_n(s_2,t_2)$ are weakly asymptotically independent.*

h_3 - *For each $t,s \in [m,M]$, $\lim_{n\to\infty} \mathbb{E}\Delta_n(t,s) = 0$*

h_4 - *There exists some increasing positive function v differentiable over $[m,M]$ satisfying $v(0) = 0$ such that for each $s,t \in [m,M]$, $\lim_{n\to\infty} \mathbb{E}\Delta_n^2(t,s) = |v(t) - v(s)|$*

h_5 - *For each $t,s \in [m,M]$, $\{\Delta_n^2(s,t),\ n \in \mathbb{N}\}$ is uniformly integrable.*

Then, X_n converges in distribution in $C[m,M]$ to $W \circ v$ as $n \to \infty$.

In 4.4.1.1 we state an exponential inequality which proves that $n^{1/3}|U_n(a) - g(a)|$ is almost surely finite. This inequality is the main tool of the proof of proposition 4.2. The exponential inequality and a first approximation lemma allow us to compute the asymptotic variance of the increments of $\sqrt{n}B_n^*$. This is done in 4.4.1.2. In this subsection, we also establish a non-asymptotic inequality for the variance of the increments of $\sqrt{n}B_n^*$, which will be useful to prove inequalities for moments of greater order. Inequalities for moments of B_n and B_n^* are proved in 4.4.1.3: we prove that the $\mathbb{L}_p$-norm (which will be useful for $p \ge 2$) of an increment of length τ of $\sqrt{n}B_n^*$ is of order $\tau^{1/2}$. Tightness of B_n^* is derived in 4.4.1.4 and we complete the proof of proposition 4.2 in 4.4.1.6.

4.4.1.1 Exponential inequality for U_n

Theorem 4.4 provides an exponential inequality for the location of the maximum of a Brownian motion with drift bounded from below by some parabola. By the standard properties of the Brownian motion (change of variables and change of scales), we can prove that a version of $U_n(a)$ (for each $a \in [m, M]$) is such a variable. Thus, we derive an exponential inequality for $U_n(a)$. This inequality will be useful to compute the asymptotic variance of $\sqrt{n}B_n^*$ and to establish inequalities for moments of B_n^*, and thus to prove that $\sqrt{n}B_n^*$ satisfies assumptions h_4 and h_5 of proposition 4.3.

Lemma 4.4 *Let f be some decreasing continuous function from $[0,1]$ onto $[m, M]$ and g its inverse function. Let U_n be the process defined by (4.8). If f is differentiable over $[0,1]$ and there exists some positive m_f such that $\inf_{t\in[0,1]} |f'(t)| \geq m_f$ then for each $t > 0$*

$$\forall a \in [m, M] \qquad \mathbb{P}(n^{1/3}|U_n(a) - g(a)| > t) \leq 2\exp\left(-\frac{m_f^2 t^3}{8}\right)$$

and if p is any positive integer, there exists some positive constant C_p which is independent depend of a and n such that $\mathbb{E}n^{p/3}|U_n(a) - g(a)|^p \leq C_p$ for each $a \in [m, M]$.

<u>*Proof of lemma 4.4*</u>:

Fix $a \in [m, M]$ and define $I_n(a) = [-n^{1/3}g(a), n^{1/3}(1 - g(a))]$. The change of variables $v = n^{1/3}(u - g(a))$ yields

$$n^{1/3}(U_n(a) - g(a)) = \underset{v\in I_n(a)}{\operatorname{argmax}}\left\{W(g(a) + n^{-1/3}v) + \sqrt{n}(F(g(a) + n^{-1/3}v) - a(g(a) + n^{-1/3}v))\right\}$$

Let $W_a^{(n)}$ be the Brownian motion defined by $W_a^{(n)}(.) = n^{1/6}\left(W(g(a) + \ . n^{-1/3}) - W(g(a))\right)$. We get

$$n^{1/3}(U_n(a) - g(a)) =$$

$$= \underset{v\in I_n(a)}{\operatorname{argmax}}\left\{n^{-1/6}W_a^{(n)}(v) + W(g(a)) + \sqrt{n}(F(g(a) + n^{-1/3}v) - a(g(a) + n^{-1/3}v))\right\}$$

But the location of the maximum of a process is invariant by deleting terms which do not depend on v or by renormalization, thus

$$n^{1/3}(U_n(a) - g(a)) = \underset{v\in I_n(a)}{\operatorname{argmax}}\left\{W_a^{(n)}(v) + n^{2/3}(F(g(a) + n^{-1/3}v) - an^{-1/3}v)\right\}$$

Substracting the term $Fg(a)$ yields

$$n^{1/3}(U_n(a) - g(a)) = \underset{v\in I_n(a)}{\operatorname{argmax}}\left\{W_a^{(n)}(v) - D_a^{(n)}(v)\right\}$$

where

$$D_a^{(n)}(v) = -n^{2/3}(F(g(a) + n^{-1/3}v) - Fg(a) - an^{-1/3}v)$$

For each $v \in I_n(a)$, there exists some $\xi_{a,v} \in]g(a), g(a)+n^{-1/3}v[$ such that $D_a^{(n)}(v) = |f'(\xi_{a,v})|v^2/2$. and therefore for each $a \in [m, M]$ and each $v \in I_n(a)$

$$D_a^{(n)}(v) \geq \frac{m_f}{2}v^2$$

Since $\forall a \in \mathbb{R}$, $0 \in I_n(a)$ and $D_a^{(n)}(0) = 0$, theorem 4.4 yields the first part of lemma 4.4. Let $C = \frac{m_f^2}{8}$; then for each $a \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}n^{p/3}|U_n(a) - g(a)|^p &= \int_0^\infty \mathbb{P}\left(n^{1/3}|U_n(a) - g(a)| > x^{1/p}\right) dx \\ &\leq 2\int_0^\infty \exp\left(-Cx^{3/p}\right) dx \end{aligned}$$

which is finite and does not depend on a. This completes the proof of lemma 4.4. ◇

4.4.1.2 Asymptotic and non asymptotic evaluations for the variance of the increments of $n^{1/2}B_n^*$

In lemma 4.6, we study the variance of the increments of $\sqrt{n}B_n^*$. We compute the asymptotic variance of the increments of $\sqrt{n}B_n^*$ and prove that $\sqrt{n}B_n^*$ satisfies assumption h_4 of proposition 4.3. The function v of assumption h_4 is found to be $8k(1-g)$, where g is the inverse function of f and k depends on the moments of Groeneboom's process. We also establish a non-asymptotic inequality for the variance of the increments of $\sqrt{n}B_n^*$. This inequality will be used in 4.4.1.3 to prove inequalities for moments of greater order.

We first state a weak version of a strong approximation lemma which will be proved in section 4.4.2. It provides a first example of convergence of a sequence of locations of the maximum of processes. Lemma 4.5 proves that moments of $n^{1/3}|U_n(a) - g(a)|$ (for each $a \in [m, M]$) can be approximated by moments of some Groeneboom process. This justifies dependence of the asymptotic variance of B_n^* on moments of Groeneboom's process.

Lemma 4.5 *Let f be some function satisfying $\mathcal{R}$ and c the function defined by*

$$\forall a \in \mathbb{R} \qquad c(a) = (|f'g(a)|/2)^{2/3}$$

Let U_n be the process defined by (4.8) and for each $l > 0$ let $T_n(l) = \{(a,b) \in [m, M],\ a \leq b \leq a + ln^{-1/3}\}$. Then for each $(a,b) \in T_n(l)$ there exists some Groeneboom's process $V_{a,b}^{(n)}$ satisfying:

$$\lim_{n\to\infty} \sup_{(a,b)\in T_n(l)} \mathbb{E}\left||V_{a,b}^{(n)}(0)| - c(a)n^{1/3}|U_n(b) - g(b)|\right|^2 = 0$$

Moreover, for each $(a,b) \in T_n(l)$, $V_{a,b}^{(n)}(0) = V_{a,a}^{(n)}(c(a)n^{1/3}(g(a) - g(b)))$.

The proof is postponed to section 4.4.2 (see remark 4.5).

Lemma 4.6 *Let f be some function satisfying $\mathcal{R}$, g its inverse function and F the primitive of f such that $F(0) = 0$. Let B_n be the process defined in definition 4.3 and B_n^* be its centered associated process. Then*

1. $\forall s,t \in [m,M]$, $\lim_{n\to\infty} n\mathbb{E}|B_n^*(t) - B_n^*(s)|^2 = 8k|g(s) - g(t)|$.

2. *there exists some positive* A *such that* $\forall s,t \in [m,M]$, $n\mathbb{E}|B_n^*(t) - B_n^*(s)|^2 \leq A|t-s|$.

Proof of lemma 4.6:

Let s and t be real numbers in $[m,M]$ such that $t > s$. For each $n \in \mathbb{N}$ let U_n^* be the centered process associated to $|U_n - g|$:

$$\forall a \in \mathbb{R} \qquad U_n^*(a) = |U_n(a) - g(a)| - \mathbb{E}|U_n(a) - g(a)|$$

By Fubini's theorem,

$$\begin{aligned}
\mathbb{E}\left(|B_n^*(t) - B_n^*(s)|^2\right) &= \mathbb{E}\left(\int_s^t U_n^*(a)da \int_s^t U_n^*(b)db\right) \\
&= \int_s^t \int_s^t \mathbb{E}\left(U_n^*(a)U_n^*(b)\right) dbda \\
&= \int_s^t \int_s^t \operatorname{cov}\left(|U_n(a) - g(a)|, |U_n(b) - g(b)|\right) dbda \\
&= 2\int\int_{s<a<b<t} \operatorname{cov}\left(|U_n(a) - g(a)|, |U_n(b) - g(b)|\right) dbda
\end{aligned}$$

Let l be some positive real number. We get

$$\begin{aligned}
n\mathbb{E}\left(|B_n^*(t) - B_n^*(s)|^2\right) = 2n\int_s^t \int_a^{a+ln^{-1/3}} &\operatorname{cov}\left(|U_n(a) - g(a)|, |U_n(b) - g(b)|\right) dbda \\
+2n\int_s^t \int_{a+ln^{-1/3}}^t &\operatorname{cov}\left(|U_n(a) - g(a)|, |U_n(b) - g(b)|\right) dbda \qquad (4.10)
\end{aligned}$$

We will prove that the first integral in (4.10) converges to $8k|g(s) - g(t)|$ and the second integral is negligible as n and l tend to infinity by building approximating variables of $|U_n(a) - g(a)|$ and $|U_n(b) - g(b)|$ and integrating the covariance of the approximating variables. We will use the following remark:

Remark 4.3 Let X, X', Y and Y' be centered random variables. Then

$$|\operatorname{cov}(X,Y) - \operatorname{cov}(X',Y')| = |\mathbb{E}X(Y - Y') + \mathbb{E}Y'(X - X')|$$

We can apply Hölder's inequality and get

$$|\operatorname{cov}(X,Y) - \operatorname{cov}(X',Y')| \leq ||X||_2||Y - Y'||_2 + ||Y'||_2||X - X'||_2$$

We need some more notations:

- Let $T_n = \{(a,b) \in [s,t],\ a \leq b \leq a + ln^{-1/3}\}$
- Let c be the function defined by $\forall a \in \mathbb{R}$, $c(a) = |f'g(a)/2|^{2/3}$

By lemma 4.5, for each $(a,b) \in T_n$, there exists some Groeneboom's process $V_{a,b}^{(n)}$ such that

$$\lim_{n\to\infty} \sup_{(a,b)\in T_n} \mathbb{E}\left||V_{a,b}^{(n)}(0)| - c(a)n^{1/3}|U_n(b) - g(b)|\right|^2 = 0$$

and since c is bounded

$$\lim_{n\to\infty} \sup_{(a,b)\in T_n} \mathbb{E}\left|c(a)^{-1}|V_{a,b}^{(n)}(0)| - n^{1/3}|U_n(b) - g(b)|\right|^2 = 0$$

In the particular case where $l = 0$ we get

$$\lim_{n\to\infty} \sup_{s\le a\le t} \mathbb{E}\left|c(a)^{-1}|V_{a,a}^{(n)}(0)| - n^{1/3}|U_n(a) - g(a)|\right|^2 = 0$$

Let $I_{l,n}^{(1)}$ be the first term in the right hand side of equality (4.10). We get

$$\left|I_{l,n}^{(1)} - 2n^{1/3}\int_s^t \int_a^{a+ln^{-1/3}} c(a)^{-2}\operatorname{cov}(|V_{a,a}^{(n)}(0)|, |V_{a,b}^{(n)}(0)|)dbda\right| \le$$

$$2l|t-s| \sup_{(a,b)\in T_n}\left|n^{2/3}\operatorname{cov}(|U_n(a) - g(a)|, |U_n(b) - g(b)|) - c(a)^{-2}\operatorname{cov}(|V_{a,a}^{(n)}(0)|, |V_{a,b}^{(n)}(0)|)\right|$$

By lemma 4.4 and corollary 4.1, both $n^{1/3}|U_n(a) - g(a)|$ (for each $a \in \mathbb{R}$) and $|V(0)|$ satisfy exponential inequalities, and thus admit uniformly bounded moments of order p ($p \in \mathbb{N}$). Therefore remark 4.3 yields

$$\lim_{n\to\infty} I_{l,n}^{(1)} = 2\lim_{n\to\infty} n^{1/3}\int_s^t\int_a^{a+ln^{-1/3}} c(a)^{-2}\operatorname{cov}(|V_{a,a}^{(n)}(0)|, |V_{a,b}^{(n)}(0)|)dbda$$

By construction, $V_{a,b}^{(n)}(0) = V_{a,a}^{(n)}(c(a)n^{1/3}(g(a) - g(b))$ and therefore

$$\operatorname{cov}(|V_{a,a}^{(n)}(0)|, |V_{a,b}^{(n)}(0)|) = \operatorname{cov}(|V(0)|, |V(c(a)n^{1/3}(g(a) - g(b)))|)$$

where V is some Groeneboom's process. By Hölder's inequality and proposition 4.1

$$\begin{aligned}
\sup_{(a,b)\in T_n} \operatorname{cov}(|V(0)|, |V(c(a)n^{1/3}(g(a) - g(b)))|) \le \\
&\le 2\mathbb{E}^{1/2}|V(0)|^2 \sup_{(a,b)\in T_n} \mathbb{E}^{1/2}|V(c(a)n^{1/3}(g(a) - g(b)))|^2 \\
&\le 2\mathbb{E}^{1/2}|V(0)|^2 \sup_{(a,b)\in T_n} \mathbb{E}^{1/2}|V(0) + c(a)n^{1/3}(g(a) - g(b)))|^2
\end{aligned}$$

which is finite by corollary 4.1. On the other hand, $\sup_{(a,b)\in T_n}|1 - \frac{f'g(a)}{f'g(b)}| = O(n^{-1/3})$. Thus,

$$\begin{aligned}
\lim_{n\to\infty} I_{l,n}^{(1)} &= 2\lim_{n\to\infty} n^{1/3}\int_s^t\int_a^{a+ln^{-1/3}} c(a)^{-2}\frac{f'g(a)}{f'g(b)}\operatorname{cov}(|V(0)|, |V(c(a)n^{1/3}(g(a) - g(b)))|)dbda \\
&= 2\lim_{n\to\infty}\int_s^t\int_0^{c(a)n^{1/3}(g(a)-g(a+ln^{-1/3}))} c(a)^{-3}|f'g(a)|\operatorname{cov}(|V(0)|, |V(b)|)dbda \\
&= 8\int_s^t\int_0^{lc(a)|g'(a)|} |g'(a)|\operatorname{cov}(|V(0)|, |V(b)|)\, dbda
\end{aligned}$$

We can compute the limit as n and l tend to infinity:

$$\begin{aligned}
\lim_{l\to\infty}\lim_{n\to\infty} I_{l,n}^{(1)} &= 8\int_s^t\int_0^\infty |g'(a)|\operatorname{cov}(|V(0)|, |V(b)|)\, dbda \\
&= 8k(g(s) - g(t)) \qquad (4.11)
\end{aligned}$$

To prove that the second term in (4.10) is negligible as l and n tend to infinity, we will consider independent approximating random variables of $n^{1/3}(U_n(b) - g(b))$ and $n^{1/3}(U_n(a) - g(a))$. Let a and b be real numbers such that $a, b \in [s,t]$ with $a < b$, and let

$$D_n(a,b) = \underset{u \in [\frac{3g(b)-g(a)}{2}, \frac{g(b)+g(a)}{2}] \cap [0,1]}{\operatorname{argmax}} \{W(u) + \sqrt{n}(F(u) - bu)\}$$

$$E_n(a,b) = \underset{u \in [\frac{g(b)+g(a)}{2}, \frac{3g(a)-g(b)}{2}] \cap [0,1]}{\operatorname{argmax}} \{W(u) + \sqrt{n}(F(u) - au)\}$$

For each $u \in [\frac{3g(b)-g(a)}{2}, \frac{g(b)+g(a)}{2}]$ and each $v \in [\frac{g(b)+g(a)}{2}, \frac{3g(a)-g(b)}{2}]$, since $g(b) \leq g(a)$, $u \leq \frac{g(a)+g(b)}{2} \leq v$. Thus $D_n(a,b)$ only depends upon W's increments before time $\frac{g(a)+g(b)}{2}$ whereas $E_n(a,b)$ depends upon W's increments after this time, and $D_n(a,b)$ and $E_n(a,b)$ are independent (recall that Brownian motion increments are independent). In particular

$$\int_s^t \int_{a+ln^{-1/3}}^t \operatorname{cov}(|D_n(a,b) - g(b)|, |E_n(a,b) - g(a)|)\, db da = 0 \qquad (4.12)$$

$D_n(a,b) - g(b)$ and $U_n(b) - g(b)$ are the location of the maximum of $\{W(u+g(b)) + \sqrt{n}(F(u+g(b)) - bu)\}$ over $[\frac{g(b)-g(a)}{2}, \frac{g(a)-g(b)}{2}] \cap [-g(b), 1-g(b)]$ and $[-g(b), 1-g(b)]$ respectively. Thus for each $b \in [s,t]$, $|D_n(a,b) - g(b)| \leq |U_n(b) - g(b)|$ and

$$\| \, |D_n(a,b) - g(b)| - |U_n(b) - g(b)| \, \|_2 =$$

$$\begin{aligned} &= \left\| (|D_n(a,b) - g(b)| - |U_n(b) - g(b)|) \mathbb{1}_{|D_n(a,b)-g(b)| \neq |U_n(b)-g(b)|} \right\|_2 \\ &\leq \left\| \, |U_n(b) - g(b)| \mathbb{1}_{|D_n(a,b)-g(b)| \neq |U_n(b)-g(b)|} \right\|_2 \end{aligned}$$

If $|D_n(a,b) - g(b)| \neq |U_n(b) - g(b)|$ then the supremum of $\{W(u+g(b)) + \sqrt{n}(F(u+g(b)) - bu), u \in [-g(b), 1-g(b)]\}$ is not achieved over $[\frac{g(b)-g(a)}{2}, \frac{g(a)-g(b)}{2}]$ and $|U_n(b) - g(b)| > \frac{g(a)-g(b)}{2}$. Hölder's inequality then yields

$$\begin{aligned} \| \, |D_n(a,b) - g(b)| - |U_n(b) - g(b)| \, \|_2 &\leq \left\| \, |U_n(b) - g(b)| \mathbb{1}_{|U_n(b)-g(b)| > \frac{g(a)-g(b)}{2}} \right\|_2 \\ &\leq \|U_n(b) - g(b)\|_4 \, \mathbb{P}^{1/4}\left(|U_n(b) - g(b)| > \frac{g(a)-g(b)}{2} \right) \end{aligned}$$

Let $T'_n = \{(a,b) \in [s,t]^2 : \; b > a + ln^{-1/3}\}$. Then,

$$\forall (a,b) \in T'_n \qquad g(a) - g(b) \;\geq\; (b-a) \left(\sup_{u \in [0,1]} |f'(u)| \right)^{-1}$$

and therefore by lemma 4.4 there exist some positive constants C and C' such that for each $(a,b) \in T'_n$

$$\| \, |D_n(a,b) - g(b)| - |U_n(b) - g(b)| \, \|_2 \leq C' n^{-1/3} \exp(-Cn(b-a)^3)$$

Using the same arguments, we can prove

$$\| \, |E_n(a,b) - g(a)| - |U_n(a) - g(a)| \, \|_2 \leq C' n^{-1/3} \exp(-Cn(b-a)^3)$$

On the other hand, $|D_n(a,b) - g(b)| \leq |U_n(b) - g(b)|$ and $|E_n(a,b) - g(a)| \leq |U_n(a) - g(a)|$ thus by lemma 4.4 there exists some constant C_2 (which is independent of a, b and n) such that for each $(a,b) \in T_n'$

$$\begin{aligned} n^{1/3} \|U_n(a) - g(a)\|_2 &\leq C_2 \\ n^{1/3} \|D_n(a,b) - g(b)\|_2 &\leq C_2 \\ n^{1/3} \|E_n(a,b) - g(a)\|_2 &\leq C_2 \end{aligned}$$

Thus by remark 4.3 and (4.12) we get: there exists some positive K such that

$$\begin{aligned} &\left| \int_s^t \int_{a+ln^{-1/3}}^t \operatorname{cov}\left(|U_n(a) - g(a)|, |U_n(b) - g(b)|\right) dbda \right| \\ &\leq Kn^{-2/3} \int_s^t \int_{a+ln^{-1/3}}^t \exp\left(-Cn(b-a)^3\right) dbda \\ &\leq Kn^{-1} \int_s^t \int_l^{n^{1/3}(t-a)} \exp\left(-Cx^3\right) dxda \\ &\leq Kn^{-1}|t-s| \int_l^\infty \exp\left(-Cx^3\right) dx \end{aligned}$$

Let $I_{l,n}^{(2)}$ be the second term in the right hand side of equality (4.10). We get

$$|I_{l,n}^{(2)}| \leq 2K|t-s| \int_l^\infty \exp\left(-Cx^3\right) dx \tag{4.13}$$

Thus the second term in (4.10) is negligible as l and n tend to infinity:

$$\lim_{l\to\infty} \lim_{n\to\infty} 2n \int_s^t \int_{a+ln^{-1/3}}^t \operatorname{cov}\left(|U_n(a) - g(a)|, |U_n(b) - g(b)|\right) dbda = 0 \tag{4.14}$$

(4.11) and (4.14) yield the first part of lemma 4.6. Let us compute a non asymptotic upper bound for the first term $I_{l,n}^{(1)}$ in the right hand side of equality (4.10): denoting $\{(a,b) \in [s,t]^2 : a \leq b \leq a + ln^{-1/3}\}$ by T_n we get by Cauchy-Schwartz inequality

$$\begin{aligned} |I_{l,n}^{(1)}| &\leq 2l|t-s|n^{2/3} \sup_{(a,b)\in T_n} |\operatorname{cov}(|U_n(a) - g(a)|, |U_n(b) - g(b)|)| \\ &\leq 4l|t-s|n^{2/3} \sup_{(a,b)\in T_n} \|U_n(a) - g(a)\|_2 \|U_n(b) - g(b)\|_2 \end{aligned}$$

Thus by lemma 4.4, there exists some positive C_l such that $|I_{l,n}^{(1)}| \leq C_l|t-s|$. On the other hand, (4.13) gives a non asymptotic upper bound for the second term of (4.10), which completes the proof of lemma 4.6. ◊

4.4.1.3 Moment inequalities for the increments of B_n and B_n^*

We have in view to apply proposition 4.3 to the process $\sqrt{n}B_n^*$. The following upper bounds for the moments of B_n and B_n^* will be crucial in order , on the one hand to warrant the uniform integrability of the square of the increments of $\sqrt{n}B_n^*$, and on the other hand to perform a chaining argument which will lead to the tightness of $\sqrt{n}B_n^*$.

Lemma 4.7 *Let f satisfy $\mathcal{R}$, g its inverse function and F the primitive of f such that $F(0) = 0$. Let p be a positive integer, B_n the processes defined in definition 4.3 and B_n^* its centered associated process. We get the two following upper bounds for the $\mathbb{L}_p$-norm of B_n's and B_n^*'s increments:*

1. *there exists some positive K_p' such that $\forall s, t \in [m, M]$*

$$n^{1/3} \|B_n(t) - B_n(s)\|_p \leq K_p'|t - s|$$

2. *there exist some real numbers C_p and K_p both positive depending only on f and p and some positive integer n_p such that for each $s, t \in [m, M]$ and $n \geq n_p$ satisfying $n^{1/3}|s - t| \geq C_p \log n$*

$$n^{1/2} \|B_n^*(t) - B_n^*(s)\|_p \leq K_p|t - s|^{1/2}$$

Proof of lemma 4.7

Fix s and t in $[m, M]$. By Jensen's inequality,

$$n^{1/3} \|B_n(t) - B_n(s)\|_p \leq n^{1/3} \left(|t - s|^{p-1} \int_s^t \mathbb{E}|U_n(a) - g(a)|^p da\right)^{1/p}$$

and by lemma 4.4 there exists some positive K_p' such that $n^{1/3} \|B_n(t) - B_n(s)\|_p \leq K_p'|t - s|$, which proves the first part of lemma 4.7.

We prove the second inequality of lemma 4.7 for any $p \geq 4$. It will be then obvious for any $p < 4$ since the $\mathbb{L}_q$ norm is no more than the $\mathbb{L}_p$ norm whenever $q \leq p$.

Fix $p \in \mathbb{N}$, $p \geq 4$ and s, t in $[m, M]$. Let U_n^* be the zero mean process associated to $|U_n - g|$:

$$\forall a \in \mathbb{R} \qquad U_n^*(a) = |U_n(a) - g(a)| - \mathbb{E}|U_n(a) - g(a)|$$

Let us explain $B_n^*(t) - B_n^*(s)$ as a sum of small increments with step $\delta = n^{-\frac{p}{6(p-1)}}|t - s|^{\frac{p-2}{2(p-1)}}$:

$$B_n^*(t) - B_n^*(s) = \sum_{i=0}^{I} \int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} U_n^*(a)da \tag{4.15}$$

where $I = [|t - s|\delta^{-1}]$. We will consider some independent approximating random variables of $\int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} U_n^*(a)da$. This way, $B_n^*(t) - B_n^*(s)$ will appear as a sum of *almost independent variables*. For each $i \in \{0, ..., I\}$, let $X_{n,i}$ be the process indexed by $[s + i\delta, s + (i + 1)\delta]$ defined by: $\forall a \in [s + i\delta, s + (i + 1)\delta]$,

$$X_{n,i}(a) = \underset{T_{i+1} \leq u \leq T_{i-1}}{\text{argmax}} \{W(u) + \sqrt{n}(F(u) - au)\}$$

where for each $i \in \{0, ..., I\}$,

$$T_i = \frac{g(s + i\delta) + g(s + (i + 1)\delta)}{2}$$

For each $i \in \{0, ..., I\}$, let $X_{n,i}^*$ be the zero mean process associated to $|X_{n,i} - g|$:

$$\forall a \in [s + i\delta, s + (i + 1)\delta] \qquad X_{n,i}^*(a) = |X_{n,i}(a) - g(a)| - \mathbb{E}|X_{n,i}(a) - g(a)|$$

To prove lemma 4.7, it suffices to upper bound the $\mathbb{L}_p$-distance between approximating and true variables:

$$\left\| n^{1/2} \sum_{i=0}^{I} \int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} (X_{n,i}^*(a) - U_n^*(a))da \right\|_p \tag{4.16}$$

and to control the $\mathbb{L}_p$-norm of the approximating variables:

$$\left\| n^{1/2} \sum_{i=0}^{I} \int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} X_{n,i}^*(a)da \right\|_p \tag{4.17}$$

We first evaluate the $\mathbb{L}_p$-distance between $X_{n,i}^*(a)$ and $U_n^*(a)$ for a fixed $a \in [s+i\delta, s+(i+1)\delta]$ to get an upper bound for (4.16). By definition $|X_{n,i}(a) - g(a)| \leq 1$ and $|U_n(a) - g(a)| \leq 1$ since g takes values in $[0,1]$. Thus $||X_{n,i}(a) - g(a)| - |U_n(a) - g(a)|| \leq 1$ and for each $i \in \{0, ..., I\}$ and $a \in [s+i\delta, s+(i+1)\delta]$

$$\begin{aligned} \left\| X_{n,i}^*(a) - U_n^*(a) \right\|_p &\leq 2 \| \, |X_{n,i}(a) - g(a)| - |U_n(a) - g(a)| \, \|_p \\ &\leq 2 \left\| (|X_{n,i}(a) - g(a)| - |U_n(a) - g(a)|) \mathbb{1}_{X_{n,i}(a) \neq U_n(a)} \right\|_p \\ &\leq 2\mathbb{P}^{1/p} (X_{n,i}(a) \neq U_n(a)) \end{aligned}$$

If $X_{n,i}(a) \neq U_n(a)$ then

- either $|U_n(a) - g(a))| > (g(s+(i-1)\delta) - g(s+i\delta))/2$
- or $|U_n(a) - g(a))| > (g(s+(i+1)\delta) - g(s+(i+2)\delta))/2$

Let M_f be some positive real number such that $M_f \geq \sup_{u\in[0,1]} |f'(u)|$. If $X_{n,i}(a) \neq U_n(a)$ then $|U_n(a) - g(a))| > \frac{\delta}{2M_f}$. Thus by lemma 4.4 there exists some positive C depending only on f such that

$$\begin{aligned} \left\| X_{n,i}^*(a) - U_n^*(a) \right\|_p &\leq 2\mathbb{P}^{1/p} \left(|U_n(a) - g(a)| > \frac{\delta}{2M_f} \right) \\ &\leq 4 \exp(-Cn\delta^3/p) \end{aligned}$$

Let $C_p = p/2C$ and suppose $n^{1/3}|t-s| \geq C_p \log n$ and n large enough so that $C_p \log n \geq 1$. By definition, $\delta = n^{-\frac{p}{6(p-1)}} |t-s|^{\frac{p-2}{2(p-1)}}$. Thus

$$\begin{aligned} \left\| X_{n,i}^*(a) - U_n^*(a) \right\|_p^p &\leq 4^p \exp\left(-C \left(n^{1/3}|t-s| \right)^{\frac{3(p-2)}{2(p-1)}} \right) \\ &\leq 4^p \exp\left(-C \left(C_p \log n \right)^{\frac{3(p-2)}{2(p-1)}} \right) \\ &\leq 4^p \exp(-CC_p \log n) \end{aligned}$$

since $\frac{3(p-2)}{2(p-1)} \geq 1$ whenever $p \geq 4$, and so

$$\sup_{s+i\delta \leq a \leq s+(i+1)\delta} \left\| X_{n,i}^*(a) - U_n^*(a) \right\|_p \leq 4n^{-1/2} \tag{4.18}$$

Let us now evaluate (4.16). By Jensen's inequality and (4.18) we get

$$\begin{aligned}\mathbb{E}\left|\sum_{i=0}^{I}\int_{s+i\delta}^{(s+(i+1)\delta)\wedge t}(X^*_{n,i}(a)-U^*_n(a))da\right|^p &\le |t-s|^{p-1}\mathbb{E}\sum_{i=0}^{I}\int_{s+i\delta}^{(s+(i+1)\delta)\wedge t}|X^*_{n,i}(a)-U^*_n(a)|^p da\\ &\le |t-s|^{p-1}\sum_{i=0}^{I}\int_{s+i\delta}^{(s+(i+1)\delta)\wedge t}\left\|X^*_{n,i}(a)-U^*_n(a)\right\|_p^p da\\ &\le 4^p n^{-p/2}|t-s|^p\end{aligned}$$

and since $|t-s| \le M$ we get for any positive integer p not less than 4

$$\left\|n^{1/2}\sum_{i=0}^{I}\int_{s+i\delta}^{(s+(i+1)\delta)\wedge t}(X^*_{n,i}(a)-U^*_n(a))da\right\|_p \le 4M^{1/2}|t-s|^{1/2} \tag{4.19}$$

We compute now an upper bound for (4.17). Fix $i \in \{0,...,I\}$ and a in $[s+i\delta, s+(i+1)\delta]$. The process $X_{n,i}$ only depends on W's increments between times T_{i+1} and T_{i-1} since by definition

$$X_{n,i}(a) = \underset{T_{i+1}\le u\le T_{i-1}}{\operatorname{argmax}}\ \{W(u)-W(T_{i+1})+\sqrt{n}(F(u)-au)\}$$

But Brownian motion's increments are independent, thus on the one hand variables

$$\int_{s+i\delta}^{(s+(i+1)\delta)\wedge t}|X_{n,i}(a)-g(a)|da$$

with odd i are independent and on the other hand these variables with even i are independent. Let S_0 and S_e be the sums of variables $\int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} X^*_{n,i}(a)da$ with respectively odd i and even i:

$$S_o = n^{1/2}\sum_{\substack{0\le i\le I\\ i\ odd}}\int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} X^*_{n,i}(a)da$$

$$S_e = n^{1/2}\sum_{\substack{0\le i\le I\\ i\ even}}\int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} X^*_{n,i}(a)da$$

Then, S_o and S_e are sums of independent variables with zero mean, and we get an upper bound for (4.17) as soon as we get upper bounds for the $\mathbb{L}_p$-norm of S_o and S_e. Rosenthal's inequality provides an upper bound for the p-order moment of sums of independent variables whenever $p \ge 2$. Recall this inequality (see [33]):

Theorem 4.7 *Let $p \ge 2$, $I \in \mathbb{N}$ and $S_I = \sum_{i=0}^{I}\xi_i$, where the ξ_i's are independent random variables with zero mean. If $|\xi_i|^p$ is integrable for each $i \in \{0,...,I\}$, then there exists some positive A_p depending only on p such that*

$$\mathbb{E}(|S_I|^p) \le A_p\left[\left(\sum_{i=0}^{I}\mathbb{E}|\xi_i|^2\right)^{p/2}+\sum_{i=0}^{I}\mathbb{E}|\xi_i|^p\right]$$

We apply Rosenthal's inequality to both S_o and S_e. Then

$$\xi_i = \begin{cases} n^{1/2} \int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} X^*_{n,i}(a)da & \text{if } i \text{ is odd (resp. even)} \\ 0 & \text{if } i \text{ is even (resp. odd)} \end{cases}$$

Fix $i \in \{0, ...I\}$ and a in $[s + i\delta, s + (i + 1)\delta]$. By lemma 4.4 and inequality (4.18), there exists some positive E such that

$$\begin{aligned} \left\|X^*_{n,i}(a)\right\|_p &\leq \left\|X^*_{n,i}(a) - U^*_n(a)\right\|_p + \|U^*_n(a)\|_p \\ &\leq 4n^{-1/2} + En^{-1/3} \\ &\leq 2En^{-1/3} \end{aligned}$$

for n large enough and therefore

$$\left\|n^{1/2} \int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} X^*_{n,i}(a)da\right\|_p \leq 2E\delta n^{1/6}$$

Suppose $n^{1/3}|t-s| \geq C_p \log n$ and n large enough so that $C_p \log n \geq 1$. By definition, $|t-s|\delta^{-1} = (n^{1/3}|t-s|)^{\frac{p}{2(p-1)}}$, thus $1 \leq |t-s|\delta^{-1}$ and $I + 1 \leq 3|t-s|\delta^{-1}$. So

$$\begin{aligned} \sum_{i=0}^{I} \mathbb{E}\left|n^{1/2} \int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} X^*_{n,i}(a)da\right|^p &\leq (I+1)(2E)^p\delta^p n^{p/6} \\ &\leq 3(2E)^p n^{p/6}\delta^{p-1}|t-s| \\ &\leq 3(2E)^p|t-s|^{p/2} \end{aligned} \quad (4.20)$$

On the other hand by definition for each $s, t \in [m, M]$, $B^*_n(t) - B^*_n(s) = \int_s^t U^*_n(a)da$. Thus the second part of lemma 4.6 and inequality (4.18) yield an upper bound for the $\mathbb{L}_2$-norm of ξ_i: there exists some positive A such that

$$\begin{aligned} &\left\|n^{1/2} \textstyle\int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} X^*_{n,i}(a)da\right\|_2 \leq \\ &\leq \left\|n^{1/2} \int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} \left(X^*_{n,i}(a) - U^*_n(a)\right) da\right\|_2 + \left\|n^{1/2} \int_{s+i\delta}^{s+(i+1)\delta} U^*_n(a)da\right\|_2 \\ &\leq 4\delta + A^{1/2}\delta^{1/2} \\ &\leq (4 + A^{1/2})\delta^{1/2} \end{aligned}$$

for n large enough, and so, by relabelling the constant we get

$$\begin{aligned} \left(\sum_{i=0}^{I} \mathbb{E}\left|n^{1/2} \int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} X^*_{n,i}(a)da\right|^2\right)^{p/2} &\leq (A(I+1)\delta)^{p/2} \\ &\leq (3A)^{p/2}|t-s|^{p/2} \end{aligned} \quad (4.21)$$

Theorem 4.7, (4.20) and (4.21) yield: there exists some positive K (depending only upon p) such that $\mathbb{E}|S_o|^p \leq K|t-s|^{p/2}$ and $\mathbb{E}|S_e|^p \leq K|t-s|^{p/2}$ which furnishes an upper bound for

(4.17):

$$\begin{aligned}\left\| n^{1/2}\sum_{i=0}^{I}\int_{s+i\delta}^{(s+(i+1)\delta)\wedge t} X_{n,i}^{*}(a)da\right\|_p &= \|S_e + S_o\|_p \\ &\leq 2K^{1/p}|t-s|^{1/2} \end{aligned} \tag{4.22}$$

Thus when n is large enough and $n^{1/3}|t-s| \geq C_p \log n$, (4.22) and (4.19) yield the second part of lemma 4.7. ◊

4.4.1.4 Asymptotic equicontinuity of B_n^*

Lemma 4.8 *Let f be some function satisfying $\mathcal{R}$, g its inverse function and F the primitive of f such that $F(0) = 0$. Let B_n be the process defined in definition 4.3 and B_n^* its centered associated process. Then*

1. *for each $s \in [m, M]$ and each positive integer p there exits some $n_p \in \mathbb{N}$ and $A_p > 0$ such that*
$$\left\| \sup_{s\leq t\leq s+\tau} \sqrt{n}\,|B_n^*(t) - B_n^*(s)| \right\|_p \leq A_p \tau^{1/2}$$
whenever $n \geq n_p$ and $\tau > n^{-1/3}\log^2 n$.

2. *The sequence $\{\sqrt{n}B_n^*,\ n \in \mathbb{N}\}$ is tight in $C[m, M]$.*

Proof of lemma 4.8

Fix $p \in \mathbb{N}$. Let C_p and K_p be positive real numbers and n_p be some positive integer such that
$$n^{1/2}\|B_n^*(t) - B_n^*(s)\|_p \leq K_p|t-s|^{1/2}$$
whenever $n \geq n_p$ and $n^{1/3}|t-s| \geq C_p \log n$ (the existence of such numbers is proved in the second part of lemma 4.7). Assume $C_p > 1$ and let τ be some positive real number such that $\tau \geq n^{-1/3}\log^2 n$ and k_p be the non negative integer defined as follows:

$$k_p = \inf\{k \in \mathbb{N}, 2^{-k}n^{1/3}\tau < C_p \log n\}$$

Since $C_p > 1$ the integer k_p is positive. Fix $s \in [m, M-\tau]$ and suppose $n \geq n_p$. For each $k \in \{0, ..., k_p\}$ we build a grid over $[s, s+\tau]$ with step $2^{-k}\tau$: we note $t_{k,0} = s$, and for each j, $0 \leq j \leq 2^k - 1$, $t_{k,j+1} = t_{k,j} + 2^{-k}\tau$. Let π_k be the projection over the kth grid:

$$\forall t \in [s, s+\tau] \qquad \pi_k(t) = \inf\{t_{k,j_0} : \inf_{0\leq j\leq 2^k} |t - t_{k,j}| = |t - t_{k,j_0}|\}$$

Then, for each $t \in [s, s+\tau]$

$$B_n^*(t) - B_n^*(s) = B_n^*(\pi_0(t)) - B_n^*(s) + \sum_{k=1}^{k_p-1} B_n^*(\pi_k(t)) - B_n^*(\pi_{k-1}(t)) + B_n^*(t) - B_n^*(\pi_{k_p-1}(t))$$

and therefore

$$\left\|\sup_{s\le t\le s+\tau}|B_n^*(t)-B_n^*(s)|\right\|_p \le \left\|\sup_{s\le t\le s+\tau}|B_n^*(\pi_0(t))-B_n^*(s)|\right\|_p$$
$$+\left\|\sum_{k=1}^{k_p-1}\sup_{s\le t\le s+\tau}|B_n^*(\pi_k(t))-B_n^*(\pi_{k-1}(t))|\right\|_p+\left\|\sup_{s\le t\le s+\tau}|B_n^*(t)-B_n^*(\pi_{k_p-1}(t))|\right\|_p \quad (4.23)$$

where the second term is equal to zero whenever $k_p = 1$. Thus it suffices to upper bound each term of the right hand side of inequality (4.23).

- *Upper bound for the first term:* For each $t \in [s, s+\tau]$, $\pi_0(t) \in \{s, s+\tau\}$. Thus

$$\sup_{s\le t\le s+\tau}|B_n^*(\pi_0(t))-B_n^*(s)| = |B_n^*(s+\tau)-B_n^*(s)|$$

By the second part of lemma 4.7 (recall $k_p > 0$, that is $n^{1/3}\tau \ge C_p\log n$)

$$\left\|n^{1/2}\sup_{s\le t\le s+\tau}|B_n^*(\pi_0(t))-B_n^*(s)|\right\|_p \le K_p\tau^{1/2}$$

- *Upper bound for the second term whenever $k_p > 1$:* $\pi_{k-1}(t)$ and $\pi_k(t)$ are elements of $\{t_{k,0},\dots,t_{k,2^k}\}$ either equal or consecutive thus

$$\begin{aligned}\sup_{s\le t\le s+\tau}|B_n^*(\pi_k(t))-B_n^*(\pi_{k-1}(t))|^p &= \sup_{1\le j\le 2^k}|B_n^*(t_{k,j})-B_n^*(t_{k,j-1})|^p\\ &\le \sum_{1\le j\le 2^k}|B_n^*(t_{k,j})-B_n^*(t_{k,j-1})|^p\end{aligned}$$

By construction of the kth grid (where $k < k_p$) and definition of k_p, for each $j \in \{1,\dots,2^k\}$, $|t_{k,j}-t_{k,j-1}| = 2^{-k}\tau$, and $n^{1/3}|t_{k,j}-t_{k,j-1}| \ge C_p\log n$. Thus by the second part of lemma 4.7

$$n^{p/2}\mathbb{E}\sup_{s\le t\le s+\tau}|B_n^*(\pi_k(t))-B_n^*(\pi_{k-1}(t))|^p \le 2^{k-kp/2}K_p^p\tau^{p/2}$$

and we get an upper bound for the second term in (4.23) whenever $k_p > 1$:

$$\begin{aligned}\left\|\sum_{k=1}^{k_p-1}n^{1/2}\sup_{s\le t\le s+\tau}|B_n^*(\pi_k(t))-B_n^*(\pi_{k-1}(t))|\right\|_p &\le \sum_{k=1}^{k_p-1}\left\|n^{1/2}\sup_{s\le t\le s+\tau}|B_n^*(\pi_k(t))-B_n^*(\pi_{k-1}(t))|\right\|_p\\ &\le K_p\tau^{1/2}\sum_{k\ge1}2^{k/p-k/2}\end{aligned}$$

Suppose first that $p > 2$. Then, $\sum_{k\ge1}2^{k/p-k/2}$ is finite and there exists some positive K_p' such that

$$\left\|\sum_{k=1}^{k_p-1}n^{1/2}\sup_{s\le t\le s+\tau}|B_n^*(\pi_k(t))-B_n^*(\pi_{k-1}(t))|\right\|_p \le K_p'\tau^{1/2}$$

But the $\mathbb{L}_q$ norm is no more than the $\mathbb{L}_p$ norm whenever $q \le p$ thus the previous inequality remains true for any $p \in \mathbb{N}$.

- *Upper bound for the third term:* By definition, $B_n^*(.) = B_n(.) - \mathbb{E}B_n(.)$ thus

$$|B_n^*(\pi_{k_p-1}(t)) - B_n^*(t)| \leq |B_n(\pi_{k_p-1}(t)) - B_n(t)| + \mathbb{E}|B_n(\pi_{k_p-1}(t)) - B_n(t)|$$

By construction of the grids, $\pi_{k_p-1}(t) - 2^{-k_p}\tau \leq t \leq \pi_{k_p-1}(t) + 2^{-k_p}\tau$ and by definition B_n is non decreasing, thus

$$|B_n(\pi_{k_p-1}(t)) - B_n(t)| \leq |B_n(\pi_{k_p-1}(t) + 2^{-k_p}\tau) - B_n(\pi_{k_p-1}(t) - 2^{-k_p}\tau)|$$

Thus we get

$$\begin{aligned} |B_n^*(\pi_{k_p-1}(t)) - B_n^*(t)| &\leq |B_n(\pi_{k_p-1}(t) + 2^{-k_p}\tau) - B_n(\pi_{k_p-1}(t) - 2^{-k_p}\tau)| \\ &\quad + \mathbb{E}|B_n(\pi_{k_p-1}(t) + 2^{-k_p}\tau) - B_n(\pi_{k_p-1}(t) - 2^{-k_p}\tau)| \\ &\leq |B_n^*(\pi_{k_p-1}(t) + 2^{-k_p}\tau) - B_n^*(\pi_{k_p-1}(t) - 2^{-k_p}\tau)| \\ &\quad + 2\mathbb{E}|B_n(\pi_{k_p-1}(t) + 2^{-k_p}\tau) - B_n(\pi_{k_p-1}(t) - 2^{-k_p}\tau)| \end{aligned}$$

The first part of lemma 4.7 with $p = 1$ yields: there exists some $K > 0$ such that

$$\begin{aligned} \mathbb{E}n^{1/2}|B_n(\pi_{k_p-1}(t) + 2^{-k_p}\tau) - B_n(\pi_{k_p-1}(t) - 2^{-k_p}\tau)| &\leq Kn^{1/6}2^{-k_p+1}\tau \\ &\leq 2KC_p n^{-1/6}\log n \\ &\leq 2KC_p\tau^{1/2} \end{aligned}$$

whenever $\tau \geq n^{-1/3}\log^2 n$. On the other hand, for any positive integer p we get

$$\begin{aligned} &\sup_{s\leq t\leq s+\tau} |B_n^*(\pi_{k_p-1}(t) + 2^{-k_p}\tau) - B_n^*(\pi_{k_p-1}(t) - 2^{-k_p}\tau)|^p \leq \\ &\quad\leq \sup_{0\leq j\leq 2^{k_p-1}} |B_n^*(t_{k_p-1,j} + 2^{-k_p}\tau) - B_n^*(t_{k_p-1,j} - 2^{-k_p}\tau)|^p \\ &\quad\leq \sum_{j=0}^{2^{k_p-1}} |B_n^*(t_{k_p-1,j} + 2^{-k_p}\tau) - B_n^*(t_{k_p-1,j} - 2^{-k_p}\tau)|^p \end{aligned}$$

Thus the second part of lemma 4.7 yields: there exists some positive K_p such that

$$\begin{aligned} &\left\| n^{1/2} \sup_{s\leq t\leq s+\tau} |B_n^*(\pi_{k_p-1}(t) + 2^{-k_p}\tau) - B_n^*(\pi_{k_p-1}(t) - 2^{-k_p}\tau)| \right\|_p \leq \\ &\quad\leq 2^{k_p/p} n^{1/2} \sup_{0\leq j\leq 2^{k_p-1}} \left\| B_n^*(t_{k_p-1,j} + 2^{-k_p}\tau) - B_n^*(t_{k_p-1,j} - 2^{-k_p}\tau) \right\|_p \\ &\quad\leq 2^{k_p/p} K_p 2^{(-k_p+1)/2}\tau^{1/2} \\ &\quad\leq 2^{1/2} K_p \tau^{1/2} \end{aligned}$$

whenever $p > 2$. But the $\mathbb{L}_q$ norm is no more than the $\mathbb{L}_p$ norm whenever $q \leq p$ thus the previous inequality remains true for any $p \in \mathbb{N}$ and we get for any $p \in \mathbb{N}$

$$\left\| n^{1/2} \sup_{s\leq t\leq s+\tau} |B_n^*(\pi_{k_p-1}(t)) - B_n^*(t)| \right\|_p \leq A\tau^{1/2}$$

for some positive A which completes the proof of the first part of lemma 4.8.

We turn now to the proof of tightness of $\{n^{1/2}B_n^*,\ n \in \mathbb{N}\}$. By Billingsley [3] (see theorem 8.3), it suffices to prove that for each positive fixed ε and η, there exist some τ with $0 < \tau < 1$ and some integer n_0 such that for each fixed $s \in [m, M]$,

$$\frac{1}{\tau}\mathbb{P}\left(\sup_{s \le t \le s+\tau} \sqrt{n}\,|B_n^*(t) - B_n^*(s)| > \varepsilon\right) \le \eta \quad \forall n \ge n_0 \tag{4.24}$$

Fix $s \in [m, M]$, $\varepsilon > 0$ and $\eta > 0$. For each positive τ, there exists some positive integer n_τ such that $\tau > n^{-1/3}\log^2 n$ whenever $n \ge n_\tau$. Thus by Markov's inequality and the first part of lemma 4.8 with $p = 4$, there exists some positive real number A_4 such that

$$\begin{aligned}\frac{1}{\tau}\mathbb{P}\left(\sup_{s \le t \le s+\tau} \sqrt{n}\,|B_n^*(t) - B_n^*(s)| > \varepsilon\right) &\le \frac{1}{\varepsilon^4\tau}\left\|\sup_{s \le t \le s+\tau} \sqrt{n}\,|B_n^*(t) - B_n^*(s)|\right\|_4^4 \\ &\le A_4^4\tau/\varepsilon^4\end{aligned}$$

whenever $n \ge n_\tau$. It suffices to choose τ small enough so that $A\tau/\varepsilon^4 \le \eta$ to get (4.24). ◇

4.4.1.5 Weak asymptotic independence of the increments of B_n^*

Let s_1, s_2, t_1, t_2 be real numbers in $[m, M]$ such that $s_1 \le t_1 < s_2 \le t_2$, and for each $s, t \in [m, M]$, let $\Delta_n(t, s) = \sqrt{n}(B_n^*(t) - B_n^*(s))$. We have to prove that for each $x, y \in \mathbb{R}$ and each positive ε, there exists some positive integer n_0 such that for each $n \ge n_0$,

$$\left|\mathbb{E}e^{i[x\Delta_n(t_1,s_1)+y\Delta_n(t_2,s_2)]} - \mathbb{E}e^{ix\Delta_n(t_1,s_1)}\mathbb{E}e^{iy\Delta_n(t_2,s_2)}\right| \le \varepsilon \tag{4.25}$$

We will use the following remark to prove convergence of characteristic functions.

Remark 4.4 Let p be some positive integer, $<\ ,\ >$ be the scalar product in $\mathbb{R}^p$ and $||\ ||$ be the associated norm. Let $\{X_n,\ \in \mathbb{N}\}$ and $\{Y_n,\ n \in \mathbb{N}\}$ be sequences of random variables which take values in $\mathbb{R}^p$ and φ_{X_n}, φ_{Y_n} be their characteristic functions: $\forall t \in \mathbb{R}^p, \varphi_{X_n}(t) = \mathbb{E}e^{i<t,X_n>}$ and $\varphi_{Y_n}(t) = \mathbb{E}e^{i<t,Y_n>}$. Since the exponential function is uniformly continuous and bounded by one we get

$$\begin{aligned}|\varphi_{X_n}(t) - \varphi_{Y_n}(t)| &\le \mathbb{E}\left|e^{i<t,X_n>} - e^{i<t,Y_n>}\right| \\ &\le ||t||\ \mathbb{E}||X_n - Y_n||\end{aligned}$$

Thus $\forall t \in \mathbb{R}^p$,

$$\lim_{n\to\infty} |\varphi_{X_n}(t) - \varphi_{Y_n}(t)| = 0$$

whenever $\lim_{n\to\infty} \mathbb{E}||X_n - Y_n|| = 0$.

Once more, we use approximating variables of the variables increments Δ_n. Let $X_{n,1}$ and $X_{n,2}$ be the processes defined by

$$\forall a \in [s_1, t_1] \qquad X_{n,1}(a) = \operatorname*{argmax}_{u \in [\frac{g(t_1)+g(s_2)}{2}, 2g(a) - \frac{g(t_1)+g(s_2)}{2}] \cap [0,1]} \{W(u) + \sqrt{n}(F(u) - au)\}$$

$$\forall a \in [s_2, t_2] \qquad X_{n,2}(a) = \operatorname*{argmax}_{u \in [2g(a) - \frac{g(t_1)+g(s_2)}{2}, \frac{g(t_1)+g(s_2)}{2}] \cap [0,1]} \{W(u) + \sqrt{n}(F(u) - au)\}$$

The processes $X_{n,1}$ and $X_{n,2}$ are independent since Brownian motion's increments are independent. For $i = 1,2$, let

$$\Gamma_n(t_i, s_i) = \sqrt{n}\int_{s_i}^{t_i} |X_{n,i}(a) - g(a)|da - \sqrt{n}\mathbb{E}\int_{s_i}^{t_i} |X_{n,i}(a) - g(a)|da$$

$\Gamma_n(t_1, s_1)$ and $\Gamma_n(t_2, s_2)$ are independent. Thus for each $x, y \in \mathbb{R}$

$$\mathbb{E}e^{i(x\Gamma_n(t_1,s_1)+y\Gamma_n(t_2,s_2))} = \mathbb{E}e^{ix\Gamma_n(t_1,s_1)}\mathbb{E}e^{iy\Gamma_n(t_2,s_2)} \tag{4.26}$$

Let i be either 1 or 2. By definition

$$\mathbb{E}\,|\Delta_n(t_i, s_i) - \Gamma_n(t_i, s_i)| \leq 2\sqrt{n}\mathbb{E}\int_{s_i}^{t_i} ||U_n(a) - g(a)| - |X_{n,i}(a) - g(a)||\,da$$

Fix a in $[s_i, t_i]$. By definition, either $X_{n,i}(a) = U_n(a)$ or

$$|X_{n,i}(a) - g(a)| \leq \left|g(a) - \frac{g(t_1) + g(s_2)}{2}\right| \leq |U_n(a) - g(a)|$$

Thus $|X_{n,i}(a) - g(a)| \leq |U_n(a) - g(a)|$ and we get

$$\begin{aligned}\mathbb{E}\,||U_n(a) - g(a)| - |X_{n,i}(a) - g(a)|| &\leq \mathbb{E}\left(||U_n(a) - g(a)| - |X_{n,i}(a) - g(a)||\,\mathbb{1}_{U_n(a)\neq X_{n,i}(a)}\right)\\ &\leq \mathbb{E}\left(|U_n(a) - g(a)|\mathbb{1}_{U_n(a)\neq X_{n,i}(a)}\right)\end{aligned}$$

Hölder's inequality then yields

$$\mathbb{E}\,||U_n(a) - g(a)| - |X_{n,i}(a) - g(a)|| \leq ||U_n(a) - g(a)||_2\mathbb{P}^{1/2}\left(U_n(a) \neq X_{n,i}(a)\right)$$

By definition, if $U_n(a) \neq X_{n,i}(a)$ then

$$|U_n(a) - g(a)| \geq \left|g(a) - \frac{g(t_1) + g(s_2)}{2}\right|$$

with $g(a) \neq \frac{g(t_1)+g(s_2)}{2}$ since g is decreasing. Thus by lemma 4.4, there exists some positive C_i such that

$$\mathbb{P}\left(U_n(a) \neq X_{n,i}(a)\right) \leq 2e^{-C_i n}$$

On the other hand, by lemma 4.4, $n^{1/3}||U_n(b) - g(b)||_2$ is uniformly bounded. Thus there exists some positive C which does not depend on n nor b such that

$$\mathbb{E}\,||U_n(a) - g(a)| - |X_{n,i}(a) - g(a)|| \leq Cn^{-1/3}e^{-C_i n/2}$$

Thus $\mathbb{E}\,|\Delta_n(t_i, s_i) - \Gamma_n(t_i, s_i)| \leq 2Cn^{1/6}|t_i - s_i|e^{-C_i n/2}$ and

$$\lim_{n\to\infty}\mathbb{E}\,|\Delta_n(t_i, s_i) - \Gamma_n(t_i, s_i)| = 0$$

Fix $\varepsilon > 0$, x and y in $\mathbb{R}$. By remark 4.4, there exists some positive integer n_0 such that for each $n \geq n_0$ and i equal to either 1 or 2

$$\left|\mathbb{E}e^{i[x\Delta_n(t_1,s_1)+y\Delta_n(t_2,s_2)]} - \mathbb{E}e^{i[x\Gamma_n(t_1,s_1)+y\Gamma_n(t_2,s_2)]}\right| \leq \frac{\varepsilon}{3} \tag{4.27}$$

$$\left|\mathbb{E}e^{ix\Delta_n(t_1,s_1)} - \mathbb{E}e^{ix\Gamma_n(t_1,s_1)}\right| \le \frac{\varepsilon}{3}$$

Thus if $n \ge n_0$ we get

$$\left|\mathbb{E}e^{ix\Gamma_n(t_1,s_1)}\mathbb{E}e^{iy\Gamma_n(t_2,s_2)} - \mathbb{E}e^{ix\Delta_n(t_1,s_1)}\mathbb{E}e^{iy\Delta_n(t_2,s_2)}\right| =$$

$$= \left|\mathbb{E}e^{ix\Gamma_n(t_1,s_1)}\left(\mathbb{E}e^{iy\Gamma_n(t_2,s_2)} - \mathbb{E}e^{iy\Delta_n(t_2,s_2)}\right) - \mathbb{E}e^{iy\Delta_n(t_2,s_2)}\left(\mathbb{E}e^{ix\Delta_n(t_1,s_1)} - \mathbb{E}e^{ix\Gamma_n(t_1,s_1)}\right)\right|$$

$$\le \left|\mathbb{E}e^{iy\Gamma_n(t_2,s_2)} - \mathbb{E}e^{iy\Delta_n(t_2,s_2)}\right| + \left|\mathbb{E}e^{ix\Delta_n(t_1,s_1)} - \mathbb{E}e^{ix\Gamma_n(t_1,s_1)}\right|$$

Thus $\left|\mathbb{E}e^{ix\Gamma_n(t_1,s_1)}\mathbb{E}e^{iy\Gamma_n(t_2,s_2)} - \mathbb{E}e^{ix\Delta_n(t_1,s_1)}\mathbb{E}e^{iy\Delta_n(t_2,s_2)}\right| \le \frac{2\varepsilon}{3}$ whenever $n \ge n_0$, and by (4.26) and (4.27) we get (4.25).

4.4.1.6 Proof of proposition 4.2

We have to prove that the sequence of processes $\{\sqrt{n}B_n^*,\ n \in \mathbb{N}\}$ satisfies the assumptions of proposition 4.3. h_1 and h_3 are obvious consequences of the definition of B_n^*. h_4 has been proved in the first part of lemma 4.6 with $v = 8k(1-g)$; tightness has been proved in the second part of lemma 4.8 and weak asymptotic independence of increments of $\sqrt{n}B_n^*$ is proved in 4.4.1.5. Thus it suffices to prove uniform integrability of the increments of $\sqrt{n}B_n^*$ raised to the square.

By the second part of lemma 4.7, there exist some positive real numbers C_4 and K_4 and some positive integer n_4 such that for each $n \ge n_4$ and each $s, t \in [m, M]$ satisfying $n^{1/3}|t-s| \ge C_4 \log n$

$$n^2\mathbb{E}|B_n^*(t) - B_n^*(s)|^4 \le K_4|t-s|^2$$

Fix s and t in $[m, M]$ and define $N = n_4 \vee \inf\{n \in \mathbb{N} :\ n^{1/3}|t-s| \ge C_4 \log n\}$. We get

$$\sup_{n \ge N} n^2\mathbb{E}|B_n^*(s) - B_n^*(t)|^4 \le K_4|t-s|^2$$

On the other hand, by the first part of lemma 4.7 there exists some positive real number K_4' such that $n^{4/3}\mathbb{E}|B_n^*(s) - B_n^*(t)|^4 \le K_4'|t-s|^4$. Thus

$$\sup_{n \le N} n^2\mathbb{E}|B_n^*(s) - B_n^*(t)|^4 \le K_4'N^{2/3}|t-s|^4$$

and therefore $\sup_{n\in\mathbb{N}} n^2\mathbb{E}|B_n^*(s) - B_n^*(t)|^4$ is finite and $\{n|B_n^*(t) - B_n^*(s)|^2,\ n \in \mathbb{N}\}$ is uniformly integrable which completes the proof of proposition 4.2. ◇

4.4.2 Asymptotic bias

In this section, we compute the asymptotic bias of the $\mathbb{L}_1$-distance between the isotonic estimator and the true function. This means that we intend to prove theorem 4.6, namely

$$\lim_{n\to\infty} n^{1/6}\left(C_f - n^{1/3}\mathbb{E}\int_{\mathbb{R}} |U_n(a) - g(a)|da\right) = 0$$

This convergence is proved in two steps. In 4.4.2.1 we state a strong approximation lemma. Lemma 4.9 provides sufficient conditions for a sequence of locations of the maximum of processes Y_n to converge to the location of the maximum of some process Z, whenever Z is a Brownian motion with parabolic drift. Moreover, this lemma yields rate of convergence of the sequence of location of the maximum when the rate of convergence of the sequence of processes Y_n to Z is known. We then derive lemma 4.10 from lemma 4.9. This lemma states convergence in probability of $n^{1/3}|U_n(a) - g(a)|$ to a Groeneboom process with rate of convergence $n^{-1/6}$. Moreover, it provides a probability inequality which it suffices to integrate to get convergence of $n^{1/3}|U_n(a) - g(a)|$ to a Groeneboom process in $\mathbb{L}_1$-sense with rate of convergence $n^{-1/6}$. This integration is performed in 4.4.2.2.

4.4.2.1 Approximation of argmax-processes

We consider a drifted Brownian motion $Z = W - D$ where D is defined over a real interval I and satisfies $\mathcal{D}$ stated in section 4.3. Given some process Y defined on a subinterval J of I, we want to understood how close is the location of the maximum of Y to that of Z whenever Y is uniformly close to Z. The following lemma provides an answer to that question and proposes two different probability bounds. The first one is well adapted to the situations where Y is also a drifted Brownian motion while the second one is meant to cover situations where the location if the maximum of Y cannot be exponentially controled.

To compute the asymptotic bias and prove theorem 4.6 we shall only deal with the first situation. The alternative bound will be useful to control the error when one substitute to the argmax process V_n defined from the isotonic estimator in the regression model, the argmax process U_n defined from an approximating white noise model. This will be performed in section 4.5.

Lemma 4.9 *Let Z be a Brownian motion with drift D satisfying $\mathcal{D}$, and z the almost surely unique and finite (see remark 4.2) location of the maximum of Z. Let Y be some process defined over a real interval $J \subset I$ and $y = \operatorname{argmax}_{u\in J}\{Y(t)\}$. Then the two following inequalities are available for each $\delta > 0$, $x > 0$ and $T \geq 0$ such that $[-T,T] \subset J$ and the derivative s of D satisfies*

$$\sup_{|t|\leq T} s^2(t) \leq \frac{1}{\delta \log \frac{1}{2x\delta}}$$

1. $\mathbb{P}(|y - z| > \delta) \leq \mathbb{P}\left(2 \sup_{|t|\leq T} |Z(t) - Y(t)| > x\delta^{3/2}\right) + 8eTx + 12ex\delta + 2\mathbb{P}(|z| > T) + \mathbb{P}(|y| > T)$

2. $\mathbb{P}(|y - z| > \delta) \leq \mathbb{P}\left(2 \sup_{t\in J} |Z(t) - Y(t)| > x\delta^{3/2}\right) + 8eTx + 12ex\delta + 2\mathbb{P}(|z| > T)$

Proof of lemma 4.9

Let x, δ and T be positive real numbers such that $[-T,T] \subset J$ and

$$\sup_{|t|\leq T} s^2(t) \leq \frac{1}{\delta \log \frac{1}{2x\delta}}$$

By lemma 4.2,

$$\begin{aligned}\mathbb{P}\left(|y - z| > \delta \; ; \left|\sup_{t\in I} Z(t) - \sup_{|t-z|>\delta} Z(t)\right| \leq x\delta^{3/2}\right) &\leq \mathbb{P}\left(\left|\sup_{t\in I} Z(t) - \sup_{|t-z|>\delta} Z(t)\right| \leq x\delta^{3/2}\right)\\ &\leq 8eTx + 12ex\delta + \mathbb{P}(|z| > T)\end{aligned}$$

On the other hand, since $J \subset I$

$$\mathbb{P}\left(|y-z|>\delta\ ;\ \left|\sup_{t\in I} Z(t) - \sup_{|t-z|>\delta} Z(t)\right| > x\delta^{3/2}\right) \leq \mathbb{P}(|Z(z)-Z(y)|>x\delta^{3/2})$$

But $|Z(z)-Z(y)| \leq |Z(z)-Y(y)| + |Y(y)-Z(y)|$. Since y and z are the location of the maximum of Y and Z respectively we get

1. $|Z(z)-Z(y)| \leq 2\sup_{t\in[-T,T]} |Z(t)-Y(t)|$ whenever $|z|\leq T$ and $|y|\leq T$. Therefore

$$\mathbb{P}(|Z(z)-Z(y)|>x\delta^{3/2}) \leq \mathbb{P}\left(2\sup_{t\in[-T,T]}|Z(t)-Y(t)|>x\delta^{3/2}\right)+\mathbb{P}(|z|>T)+\mathbb{P}(|y|>T)$$

which completes the proof of the first point of lemma 4.9.

2. $|Z(z)-Z(y)| \leq 2\sup_{t\in J}|Z(t)-Y(t)|$ whenever $z\in J$. Therefore

$$\mathbb{P}(|Z(y)-Z(z)|>x\delta^{3/2}) \leq \mathbb{P}\left(2\sup_J|Z(t)-Y(t)|>x\delta^{3/2}\right)+\mathbb{P}(|z|>T)$$

since $[-T,T]\subset J$, which completes the proof of the second point of lemma 4.9. ◊

In the particular case where $I=\mathbb{R}$ and $\forall t\in\mathbb{R}$, $D(t)=t^2$, lemma 4.9 provides sufficient conditions to converge in probability to $V(0)$.

Lemma 4.10 *Let f satisfy regularity conditions $\mathcal{R}$ and c the function defined by*

$$\forall a\in\mathbb{R} \qquad c(a)=\left|\frac{f'g(a)}{2}\right|^{2/3}$$

Let U_n be the process defined by (4.8). For each $l>0$ let $T_n(l)=\{(a,b)\in[m,M],\ a\leq b\leq a+ln^{-1/3}\}$. Then, there exist some positive constants A_0, A_1 and C such that for each $(a,b)\in T_n(l)$, there exists Groeneboom's process $V_{a,b}$ satisfying

$$\mathbb{P}\left(\left|c(a)n^{1/3}(U_n(b)-g(b))-V_{a,b}(0)\right|>\delta\right)\leq 8eTx+12ex\delta+6\exp(-\frac{C}{2}T^3)$$

whenever $x\delta^{3/2}\geq n^{-1/3}(A_0T^3+A_1l)$ and $\delta\log\frac{1}{2x\delta}\leq T^{-1}/2$.

<u>*Proof of lemma 4.10*</u>

For each $b\in[m,M]$, let r_b be the real valued function defined by:

$$\forall u\in[0,1] \qquad r_b(u)=F(u)-Fg(b)-b(u-g(b))-f'g(b)(u-g(b))^2/2$$

Fix a, b in $\mathbb{R}$. Then,

$$U_n(b)=\operatorname*{argmax}_{u\in[0,1]}\{W(u)+\sqrt{n}(f'g(b)(u-g(b))^2/2+r_b(u))\}$$

and we get

$$c(a)n^{1/3}(U_n(b)-g(b)) =$$

$$= \operatorname*{argmax}_{u\in I^{(n)}_{a,b}}\{W(g(b)+c(a)^{-1}n^{-1/3}v) - n^{-1/6}\left|\frac{f'g(b)}{2}\right|c(a)^{-2}v^2 + \sqrt{n}r_b(g(b)+c(a)^{-1}n^{-1/3}v)\}$$

where for each a, $b\in\mathbb{R}$, $I^{(n)}_{a,b} = [-c(a)n^{1/3}g(b), c(a)n^{1/3}(1-g(b)]$. Let $W^{(n)}_a$ and $W^{(n)}_{a,b}$ be the Brownian motions defined by

$$\begin{aligned}\forall t\in\mathbb{R}\qquad W^{(n)}_a(t) &= c(a)^{1/2}n^{1/6}W(c(a)^{-1}n^{-1/3}t)\\ W^{(n)}_{a,b}(t) &= W^{(n)}_a(t+g(b)c(a)n^{1/3}) - W^{(n)}_a(g(b)c(a)n^{1/3})\end{aligned}$$

and $V^{(n)}_{a,b}$ the Groeneboom process defined by

$$\forall t\in\mathbb{R}\qquad V^{(n)}_{a,b}(t) = \operatorname*{argmax}_{u\in\mathbb{R}}\{W^{(n)}_{a,b}(u)-(u-t)^2\}$$

Then,

$$V^{(n)}_{a,b}(0) = V^{(n)}_{a,a}(c(a)n^{1/3}(g(a)-g(b))) \tag{4.28}$$

and

$$\begin{aligned}&c(a)n^{1/3}(U_n(b)-g(b)) =\\ &= \operatorname*{argmax}_{u\in I^{(n)}_{a,b}}\{n^{-1/6}c(a)^{-1/2}W^{(n)}_{a,b}(v) - n^{-1/6}\left|\frac{f'g(b)}{2}\right|c(a)^{-2}v^2 + \sqrt{n}r_b(g(b)+c(a)^{-1}n^{-1/3}v)\}\\ &= \operatorname*{argmax}_{u\in I^{(n)}_{a,b}}\{W^{(n)}_{a,b}(v) - \frac{f'g(b)}{f'g(a)}v^2 + n^{2/3}c(a)^{1/2}r_b(g(b)+c(a)^{-1}n^{-1/3}v)\}\\ &= \operatorname*{argmax}_{u\in I^{(n)}_{a,b}}\{W^{(n)}_{a,b}(v) - v^2 + R^{(n)}_{a,b}(v)\}\end{aligned}$$

where

$$R^{(n)}_{a,b}(v) = (1-\frac{f'g(b)}{f'g(a)})v^2 + n^{2/3}c(a)^{1/2}r_b(g(b)+c(b)^{-1}n^{-1/3}v)$$

Let $D\in C^1(\mathbb{R})$ be defined by $\forall v\in\mathbb{R}$, $D(v)=v^2$. Then, for each positive T,

$$\sup_{|t|\le T}(D'(t))^2 = 2T$$

Let x, δ and T be some positive real numbers such that $\delta\log\frac{1}{2x\delta}\le T^{-1}/2$ and suppose n large enough so that $[-T,T]\subset I^n_{a,b}$. By the first part of lemma 4.9,

$$\mathbb{P}\left(|c(a)n^{1/3}(U_n(b)-g(b)) - V_{a,b}(0)|>\delta\right) \le \mathbb{P}\left(2\sup_{|t|\le T}|R^{(n)}_{a,b}(t)| > x\delta^{3/2}\right) +$$

$$+8eTx + 12ex\delta + 2\mathbb{P}(|V_{a,b}(0)|>T) + \mathbb{P}\left(c(a)n^{1/3}|U_n(b)-g(b)|>T\right)$$

By definition r_b is the rest of a Taylor's expansion of F of order 2. Since f' is bounded from below and f'' is bounded, there exists some positive constant A_0 (which does not depend on a nor b) such that for each $a, b\in\mathbb{R}$

$$n^{2/3}c(a)^{1/2}|r_b(g(b)+c(a)^{-1}n^{-1/3}u)| \;\le\; \frac{A_0}{2}n^{-1/3}u^3$$

On the other hand, since f'' is bounded and f' is bounded from below, there exists some positive constant A_1 such that for each $(a,b) \in T_n(l)$

$$\left|1 - \frac{f'g(b)}{f'g(a)}\right| = \frac{1}{|f'g(a)|}|f'g(a) - f'g(b)| \leq \frac{A_1}{2} l n^{-1/3}$$

and therefore

$$\mathbb{P}\left(|c(a)n^{1/3}(U_n(b) - g(b)) - V_{a,b}(0)| > \delta\right) \leq$$

$$\leq 8\epsilon Tx + 12ex\delta + 2\mathbb{P}(|V_{a,b}(0)| > T) + \mathbb{P}\left(c(a)n^{1/3}|U_n(b) - g(b)| > T\right)$$

whenever $x\delta^{3/2} \geq n^{-1/3}(A_0T^3 + A_1 l)$.
Since f' is bounded, c is bounded and by lemma 4.4 there exists some positive constant C such that for each $n \in \mathbb{N}$ and each $b \in \mathbb{R}$

$$\mathbb{P}\left(c(a)n^{1/3}|U_n(b) - g(b)| > T\right) \leq 2\exp(-CT^3/2)$$

Thus by corollary 4.1 and since we can assume $C \leq 1$

$$\mathbb{P}\left(c(a)n^{1/3}|U_n(b) - g(b)| > T\right) + 2\mathbb{P}(|V_{a,b}(0)| > T) \leq 6\exp\left(-\frac{C}{2}T^3\right)$$

which proves lemma 4.10. ◊

Remark 4.5 : To compute the asymptotic variance of B_n^*, we used a weak version of lemma 4.10, namely lemma 4.5, that we are now in position to prove.

Proof of lemma 4.5

Let A_0, A_1 and C be the positive constants defined in lemma 4.10. Let $T = \log n$, $\delta = n^{-\varepsilon}$ for some positive ε such that $\varepsilon < 2/9$ and $x = n^{-1/3}\delta^{-3/2}(A_0T^3 + A_1 l)$, and suppose n large enough so that $\delta \log \frac{1}{2x\delta} \leq T^{-1}/2$. Let

$$\mathbb{P}^{(n)} = \mathbb{P}\left(\left|V_{a,b}^{(n)}(0) - c(a)n^{1/3}(U_n(b) - g(b))\right| > n^{-\varepsilon}\right)$$

By lemma 4.10, we get for each $(a,b) \in T_n(l)$

$$\mathbb{P}^{(n)} \leq 8eTn^{-\frac{1}{3}+\frac{3\varepsilon}{2}}(A_0T^3 + A_1 l) + 12en^{-\frac{1}{3}+\frac{\varepsilon}{2}}(A_0T^3 + A_1 l) + 6\exp\left(-\frac{C}{2}T^3\right)$$

and therefore,

$$\lim_{n\to\infty} \mathbb{P}^{(n)} = 0. \tag{4.29}$$

Hölder's inequality yields

$$\mathbb{E}\left|V_{a,b}^{(n)}(0) - c(a)n^{1/3}(U_n(b) - g(b))\right|^2 \leq$$

$$\leq \mathbb{E}\left|\left(V_{a,b}^{(n)}(0) - c(a)n^{1/3}(U_n(b) - g(b))\right)\mathbb{1}_{|V_{a,b}^{(n)}(0) - c(a)n^{1/3}(U_n(b)-g(b))|>n^{-\varepsilon}}\right|^2 + n^{-2\varepsilon}$$

$$\leq \mathbb{E}^{1/2}\left|V_{a,b}^{(n)}(0) - c(a)n^{1/3}(U_n(b) - g(b))\right|^4 \left(\mathbb{P}^{(n)}\right)^{1/2} + n^{-2\varepsilon}$$

By definition, $V_{a,b}^{(n)}(0)$ has the same distribution as $V(0)$, thus it admits uniformly bounded moments of order 4 (see corollary 4.1). On the other hand, by lemma 4.4 and since c is bounded, $c(a)n^{1/3}(U_n(b) - g(b))$ admits uniformly bounded moments of order 4. Thus (4.29) yields

$$\lim_{n\to\infty} \sup_{(a,b)\in T_n(l)} \mathbb{E}\left||V_{a,b}^{(n)}(0)| - c(a)n^{1/3}|U_n(b) - g(b)|\right|^2 = 0$$

Moreover, by (4.28), $V_{a,b}^{(n)}(0) = V_{a,a}^{(n)}(c(a)n^{1/3}(g(a) - g(b))$. This completes the proof of lemma 4.5. ◊

4.4.2.2 Proof theorem 4.6

Let c be the function defined by $\forall a \in \mathbb{R}$, $c(a) = |f'g(a)/2|^{2/3}$. By lemma 4.10 with $l = 0$, there exists some $A > 0$ and $C > 0$ such that for each $a \in [m, M]$, there exists some Groeneboom's process V_a satisfying

$$\mathbb{P}\left(\left|c(a)n^{1/3}(U_n(a) - g(a)) - V_a(0)\right| > \delta\right) \leq 8eTx + 12ex\delta + 6\exp(-\frac{C}{2}T^3)$$

whenever $x\delta^{3/2} \geq n^{-1/3}AT^3$ and $\delta \log \frac{1}{2x\delta} \leq T^{-1}/2$.

Let $T = \log n$, ε be some positive real number such that $\varepsilon < 1/9$ and for each $\delta \leq n^{-\varepsilon}$, let $x_\delta = n^{-1/3}AT^3\delta^{-3/2}$. Suppose n large enough so that $(\log n)^{-1}n^{-1/6} \leq n^{-\varepsilon}$ and $\delta \log \frac{1}{2x_\delta \delta} \leq \frac{1}{2T}$ for each $\delta \in [(\log n)^{-1}n^{-1/6}, n^{-\varepsilon}]$. Then for each $\delta \in [(\log n)^{-1}n^{-1/6}, n^{-\varepsilon}]$

$$\mathbb{P}\left(\left|c(a)n^{1/3}(U_n(a) - g(a)) - V_a(0)\right| > \delta\right) \leq 8en^{-1/3}AT^4\delta^{-3/2} + 12en^{-1/3}AT^3\delta^{-1/2} + 6\exp(-\frac{C}{2}T^3)$$

and therefore

$$\begin{aligned}
&\int_{(\log n)^{-1}}^{n^{1/6-\varepsilon}} \mathbb{P}(n^{1/6}\left|c(a)n^{1/3}(U_n(a) - g(a)) - V_a(0)\right| > \rho)d\rho = \\
&= n^{1/6}\int_{(\log n)^{-1}n^{-1/6}}^{n^{-\varepsilon}} \mathbb{P}(\left|c(a)n^{1/3}(U_n(a) - g(a)) - V_a(0)\right| > \delta)d\delta \\
&\leq 8eAT^4n^{-1/6}\int_{(\log n)^{-1}n^{-1/6}}^{n^{-\varepsilon}} \delta^{-3/2}d\delta + 12eAT^3n^{-1/6}\int_{(\log n)^{-1}n^{-1/6}}^{n^{-\varepsilon}} \delta^{-1/2}d\delta + 6n^{1/6-\varepsilon}\exp(-\frac{C}{2}T^3) \\
&\leq 16eAT^4n^{-1/12}(\log n)^{1/2} + 24eAT^3n^{-1/3-\varepsilon/2} + 6n^{1/6-\varepsilon}\exp(-\frac{C}{2}T^3)
\end{aligned}$$

which is independent of a and converges to zero as $n \to \infty$. For each $\rho \geq n^{1/6-\varepsilon}$,

$$\mathbb{P}(n^{1/6}\left|c(a)n^{1/3}(U_n(a) - g(a)) - V_a(0)\right| > \rho) \leq \mathbb{P}(\left|c(a)n^{1/3}(U_n(a) - g(a)) - V_a(0)\right| > n^{-\varepsilon})$$

and therefore

$$\int_{n^{1/6-\varepsilon}}^{n^{1/6}\log n} \mathbb{P}(n^{1/6}\left|c(a)n^{1/3}(U_n(a) - g(a)) - V_a(0)\right| > \rho)d\rho \leq$$

$$n^{1/6}\log n\left[8eAT^4n^{-\frac{1}{3}+\frac{3\varepsilon}{2}} + 12eAT^3n^{-\frac{1}{3}+\frac{\varepsilon}{2}} + 6\exp(-\frac{C}{2}T^3)\right]$$

which is independent of a and converges to zero as $n \to \infty$ since $\varepsilon < 1/9$ and $T = \log n$. On the other hand, by corollary 4.1 and lemma 4.4 there exists some constant $C_0 > 0$ such that for each $\rho > 0$

$$\begin{aligned}
\mathbb{P}\left(n^{1/6}\left|c(a)n^{1/3}(U_n(a)-g(a))-V_a(0)\right|>\rho\right) \leq \\
&\leq \mathbb{P}\left(\left|c(a)n^{1/3}(U_n(a)-g(a))\right|>n^{-1/6}\frac{\rho}{2}\right)+\mathbb{P}\left(|V_a(0)|>n^{-1/6}\frac{\rho}{2}\right)\\
&\leq 4\exp(-C_0 n^{-1/2}\rho^3)
\end{aligned}$$

thus we get

$$\int_{n^{1/6}\log n}^{\infty}\mathbb{P}(n^{1/6}\left|c(a)n^{1/3}(U_n(a)-g(a))-V_b(0)\right|>\rho)d\rho \;\leq\; n^{1/6}\int_{\log n}^{\infty}4\exp(-C_0\delta^3)d\delta$$

which is independent of a and converges to zero as $n \to \infty$. But

$$\int_0^{(\log n)^{-1}}\mathbb{P}(n^{1/6}\left|c(a)n^{1/3}(U_n(a)-g(a))-V_a(0)\right|>\rho)d\rho \leq (\log n)^{-1}$$

and therefore

$$\lim_{n\to\infty}\sup_{a\in[m,M]}\int_0^{\infty}\mathbb{P}(n^{1/6}\left|c(a)n^{1/3}(U_n(a)-g(a))-V_a(0)\right|>\rho)d\rho = 0$$

which proves that

$$\lim_{n\to\infty}\sup_{a\in[m,M]}\mathbb{E}n^{1/6}|c(a)n^{1/3}(U_n(a)-g(a))-(V_a(0))| = 0$$

But c is bounded, thus

$$\lim_{n\to\infty} n^{1/6}\left|\mathbb{E}n^{1/3}\int_m^M |U_n(a)-g(a)|da - C_f\right| = 0 \tag{4.30}$$

where $C_f = \mathbb{E}\int_m^M c(a)^{-1}|V_a(0)|da$, and by lemma 4.12 below,

$$\lim_{n\to\infty} n^{1/6}\left|\mathbb{E}n^{1/3}\int_{\mathbb{R}} |U_n(a)-g(a)|da - C_f\right| = 0$$

By definition, $c(a) = |f'g(a)/2|^{2/3}$, thus

$$\begin{aligned}
C_f &= \mathbb{E}|V(0)|\int_m^M |2g'(a)|^{2/3}da\\
&= \mathbb{E}|V(0)|\int_0^1 |2g'f(t)|^{2/3}|f'(t)|dt\\
&= 2\mathbb{E}|V(0)|\int_0^1 |f'(t)/2|^{1/3}dt
\end{aligned}$$

which completes the proof of theorem 4.6.

4.5 Proof of $\mathbb{L}_1$-convergence in the context of regression

We prove theorem 4.2 whenever $\sigma = 1$ and f is decreasing from $[0,1]$ onto $[m, M]$, since the other cases follows from this case and the definition of the isotonic estimate. Let g be the inverse function of f and V_n the inverse process of the isotonic estimator of f. Recall that V_n satisfies

$$\forall a \in \mathbb{R} \qquad V_n(a) = \operatorname*{argmax}_{u\in[0,1]}\{F_n(u) - au\}$$

where F_n is the empirical process associated to the regression model (4.2):

$$\forall t \in [0,1] \qquad F_n(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i \mathbb{1}_{x_i \leq t}$$

We prove theorem 4.2 in two steps:

- In a first step, we establish a probability inequality for V_n. This inequality proves that for each $a \in \mathbb{R}$ the distance $|V_n(a) - g(a)|$ is of order no more than $n^{-1/3}$ (see lemma 4.11).
- In a second step, we build a white noise model such that the $\mathbb{L}_1$-distance beetwen V_n and the inverse process of the isotonic estimator of the signal function converges to zero with rate of convergence $n^{-1/2}$.

4.5.1 A probability inequality for V_n

Lemma 4.11 *Suppose we are given model (4.2), where f satisfies $\mathcal{R}$. Suppose that the ε_i's are i.i.d. variables with zero mean and finite variance and for each $i \in \{1, ..., n\}$, $x_i = i/n$. Let f_n be the isotonic estimator of f and V_n the inverse process of f_n. Then,*

1. *there exists some positive constant B such that for each $a \notin [m, M]$ and each $t > 0$*

$$\mathbb{P}(|V_n(a) - g(a)| > t) \leq \frac{B}{nt((m-a)_+ + (a-M)_+)^2} \tag{4.31}$$

2. *if $\mathbb{E}|\varepsilon_1|^p$ is finite for some $p \geq 2$, there exists some positive constant A_p such that for each $a \in \mathbb{R}$ and each $t > 0$*

$$\mathbb{P}(|V_n(a) - g(a)| > t) \leq \frac{A_p}{n^{p/2}t^{3p/2}} \tag{4.32}$$

Remark 4.6 *One can notice that in particular, for each $t > 0$ and each $a \in \mathbb{R}$*

$$\mathbb{P}(n^{1/3}|V_n(a) - g(a)| > t) \leq \frac{A}{t^3}$$

whenever $\mathbb{E}\varepsilon_1^2$ is finite.

Proof of lemma 4.11

Fix $a \in \mathbb{R}$. By definition,

$$V_n(a) - g(a) = \operatorname*{argmax}_{u\in[-g(a),1-g(a)]}\{n^{-1}\sum_{i=1}^{n}\varepsilon_i \mathbb{1}_{x_i\in]g(a),u+g(a)]} - D_a^{(n)}(u)\}$$

where $\forall u \in [-g(a), 1-g(a)]$,

$$D_a^{(n)}(u) \quad = \quad -\frac{1}{n}\sum_{i=1}^{n} f(x_i)\mathbb{I}_{x_i \in]g(a), u+g(a)]} + au$$

Let $Z_a^{(n)}$ be the process defined by

$$\forall u \in [-g(a), 1-g(a)] \qquad Z_a^{(n)}(u) = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \mathbb{I}_{x_i \in]g(a), u+g(a)]}$$

Then $D_a^{(n)}(0) = Z_a^{(n)}(0) = 0$ and therefore the supremum of $Z_a^{(n)} - D_a^{(n)}$ is almost surely nonnegative and we get for each $t > 0$

$$\mathbb{P}\left(|V_n(a) - g(a)| > t\right) \leq \mathbb{P}\left(\sup_{|u|>t}\{Z_a^{(n)}(u) - D_a^{(n)}(u)\} \geq 0\right) \tag{4.33}$$

In the course of the proof of lemma 4.11, we shall provide an upper bound for the number of elements of the sets

$$\mathcal{A}_a^n(u) = \{i : \ x_i \in]g(a), g(a)+u]\} \tag{4.34}$$

where $a \in \mathbb{R}$ and $u \in [-g(a), 1-g(a)]$. One can notice that for each a and u,

$$\#\mathcal{A}_a^n(u) \leq n|u| + 1 \tag{4.35}$$

since the design is unifomly spread over $[0,1]$. We can provide a better upper bound for $\#\mathcal{A}_a^n(u)$ whenever either $a > M$ or $a < m$. Suppose for instance that $a > M$. Then, $g(a) = 0$ and we get for each $u \in [0,1]$, $\mathcal{A}_a^n(u) = \{i : \ x_i \in]0, u]\}$, and therefore

$$\#\mathcal{A}_a^n(u) \leq nu \tag{4.36}$$

Let us prove the probability inequality (4.31). Let a be some real number such that either $a > M$ or $a < m$. Suppose for instance that $a > M$ (one can use the same arguments to prove (4.31) whenever $a < m$). Then, $g(a) = 0$ and we get for each $u \in [0,1]$

$$D_a^{(n)}(u) = -\frac{1}{n}\sum_{i=1}^{n} f(x_i)\mathbb{I}_{0<x_i\leq u} + au$$

Since f is decreasing, we get for each $x_i \in]0, u]$

$$f(x_i) \leq f(0)$$

where $f(0) = M$. So, by (4.36), for each $a > M$ and each $u \in [0,1]$,

$$D_a^{(n)}(u) \geq u(a - M).$$

By (4.33) we get for each $a > M$ and each $t > 0$

$$\begin{aligned}
\mathbb{P}\left(|V_n(a) - g(a)| > t\right) \quad &\leq \quad \mathbb{P}\left(\sup_{u>t}\{Z_a^{(n)}(u) - \frac{1}{2}u(a-M)\} \geq 0\right) \\
&\leq \quad \sum_{k\geq 0}\mathbb{P}\left(\sup_{t2^k<u\leq t2^{k+1}} Z_a^{(n)}(u) \geq \frac{1}{2}t2^k(a-M)\right)
\end{aligned}$$

For each $j \leq n$ and each $x > 0$ one has by Doob's inequality,

$$\mathbb{P}\left(\sup_{k\leq j}\sum_{i=1}^{k}\varepsilon_i > x\right) \leq \frac{\mathbb{E}\left|\sum_{i=1}^{j}\varepsilon_i\right|^p}{x^p}.$$

Moreover, by Rosenthal's inequality (see theorem 4.7) and Jensen's inequality, if $\mathbb{E}|\varepsilon_1|^p < \infty$ then there exists some positive constant C_p depending on $\mathbb{E}|\varepsilon_1|^p$ such that for each $j \leq n$ and $x > 0$, $\mathbb{E}|\sum_{i=1}^{j}\varepsilon_i|^p \leq C_p j^{p/2}$ and therefore

$$\forall j \leq n \qquad \forall x > 0 \qquad \mathbb{P}\left(\sup_{k\leq j}\sum_{i=1}^{k}\varepsilon_i > x\right) \leq C_p j^{p/2} x^{-p} \tag{4.37}$$

So, if $\mathbb{E}\varepsilon_1^2 < \infty$, (4.36) yields

$$\begin{aligned}\mathbb{P}\left(|V_n(a) - g(a)| > t\right) &\leq \sum_{k\geq 0}\frac{C_2}{n^2(a-M)^2t^2 2^{2k}}\#\{i:\ x_i \in]0, t2^{k+1}]\} \\ &\leq \frac{B}{nt(a-M)^2}\end{aligned}$$

for some positive constant B, which proves (4.31).

Let us prove now the probability inequality (4.32). Since f is differentiable with non vanishing derivative, there exists some positive constant c such that for each $x_i \in]g(a), g(a)+u]$,

$$f(x_i) \leq fg(a) - c(x_i - g(a))$$

By using the notation (4.34) and the upper bound (4.35), we thus get for each $a \in \mathbb{R}$ and $u \in [0, 1-g(a)]$

$$\begin{aligned}D_a^{(n)}(u) &\geq au - \frac{\#A_a^n(u)}{n}fg(a) + c\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - g(a))\mathbb{1}_{x_i\in]g(a),u+g(a)]} - \int_{g(a)}^{u+g(a)}(x-g(a))dx\right) \\ &\quad + c\int_{g(a)}^{u+g(a)}(x-g(a))dx \\ &\geq (a - fg(a))u - \frac{fg(a)}{n} - \frac{c}{n} + \frac{cu^2}{2}\end{aligned}$$

On the other hand, one can prove by using the same arguments that the same inequality holds true for each $a \in \mathbb{R}$ and each $u \in [-g(a), 0]$. Furthermore, $(a - fg(a))u$ is greater or equal to zero for each $a \in \mathbb{R}$ and each $u \in [-g(a), 1-g(a)]$: $fg(a) = a$ whenever $a \in [m, M]$ and in this case, $(a - fg(a))u = 0$, while $fg(a) = M$ (resp. $fg(a) = m$) whenever $a > M$ (resp. $a < m$) and in this case, $D_a^{(n)}$ is defined only for nonnegative (resp. nonpositive) u. So, we get for each $a \in \mathbb{R}$ and $u \in [-g(a), 1-g(a)]$ which satifies $cu^2 \geq 4(fg(a)+c)/n$

$$D_a^{(n)}(u) \geq \frac{c}{4}u^2$$

By (4.33) and (4.37), there exists some positive C_p such that for each positive t satisfying $ct^2 \geq 4(fg(a)+c)/n$

$$\mathbb{P}\left(|V_n(a) - g(a)| > t\right) \leq \mathbb{P}\left(\sup_{|u|>t}\{Z_a^{(n)}(u) - \frac{1}{4}cu^2\} \geq 0\right)$$

$$
\begin{aligned}
&\leq \sum_{k\geq 0} \mathbb{P}\left(\sup_{t2^k<|u|\leq t2^{k+1}} Z_a^{(n)}(u) \geq \frac{1}{4}ct^2 2^{2k} \right) \\
&\leq C_p 4^p \sum_{k\geq 0} \frac{1}{n^p c^p t^{2p} 2^{2pk}} (\#\mathcal{A}_a(t2^{k+1}) + \#\mathcal{A}_a(-t2^{2k+1}))^{p/2}
\end{aligned}
$$

By (4.35), $\#\mathcal{A}_a(t2^k) + \#\mathcal{A}_a(-t2^{2k}) \leq 2(nt2^{k+1}+1)$ and therefore, for each $t \in]0,1]$ such that $ct^2 \geq 4(fg(a)+c)/n$, we get

$$
\mathbb{P}\left(|V_n(a) - g(a)| > t\right) \leq \frac{A}{n^p t^{3p/2}}
$$

for some positive constant A. The above inequality remains true for each $t > 1$ since $\mathbb{P}\left(|V_n(a) - g(a)| > t\right)$ 0 whenever $t > 1$ (recall that V_n and g take values in $[0,1]$). Moreover, possibly enlarging the constant A and taking into account the fact that a probability is not greater than 1 finally yields (4.32) for each $t > 0$. ◊

Lemma 4.12 *1. Suppose we are given model (4.2), where f satisfies $\mathcal{R}$. Suppose that the ε_i's are i.i.d. variables with zero mean and finite variance and for each $i \in \{1,...,n\}$, $x_i = i/n$. Let f_n be the isotonic estimator of f and V_n the inverse process of f_n. Then,*

$$
\mathbb{E}\int_{\mathbb{R}} |V_n(a) - g(a)|da = \mathbb{E}\int_m^M |V_n(a) - g(a)|da + O(n^{-2/3})
$$

2. Let f be some function satisfying $\mathcal{R}$. Let U_n be the inverse process of the isotonic estimator of f in a white noise model. Then

$$
\mathbb{E}\int_{\mathbb{R}} |U_n(a) - g(a)|da = \mathbb{E}\int_m^M |U_n(a) - g(a)|da + O(n^{-2/3})
$$

Proof of lemma 4.12:

For each $a \in \mathbb{R}$, let $\mu(a)$ denote $(m-a)_+ + (a-M)_+$. By lemma 4.11 there exists some positive constants A and B such that for each a with $\mu(a) > 0$

$$
\begin{aligned}
\mathbb{E}|V_n(a) - g(a)| &\leq n^{-1}\mu(a)^{-2} + \int_{n^{-1}\mu(a)^{-2}}^{\mu(a)} \frac{B}{nt\mu(a)^2}dt + \int_{\mu(a)}^{\infty} \frac{A}{nt^3} \\
&\leq n^{-1}\mu(a)^{-2}\left(1 + \frac{A}{2} + B\log(n\mu(a)^3)\right)
\end{aligned}
$$

On the other hand by lemma 4.11, there exists some $A_0 > 0$ such that

$$
\mathbb{E}|V_n(a) - g(a)| \leq A_0 n^{-1/3}
$$

and therefore

$$
\int_M^{\infty} \mathbb{E}|V_n(a) - g(a)|da \leq A_0 n^{-2/3} + \int_{M+n^{-1/3}}^{\infty} \frac{1 + A/2 + B\log(n(a-M)^3)}{n(a-M)^2}da
$$

Setting $(a - M) = n^{-1/3}x$ in the above integral yields

$$
\int_M^{\infty} \mathbb{E}|V_n(a) - g(a)|da = O(n^{-2/3})
$$

Similarly,

$$
\int_{-\infty}^{m} \mathbb{E}|V_n(a) - g(a)|da = O(n^{-2/3})
$$

which proves the first part of lemma 4.12. The second part of lemma 4.12 follows from the fact that lemma 4.11 holds true for U_n as V_n.

4.5.2 A Komlós-Major-Tusnády type approximation

As mentionned before, we prove theorem 4.2 by building some white noise model with isotonic estimator close to the isotonic estimator of the regression function. We state a version of Sakhanenko's [34] theorem that can be easily derived from his theorem 5.

Theorem 4.8 *For each $i \in \{1,..,n\}$ let $x_i = i/n$. Let $\varepsilon_i, ..., \varepsilon_n$ be i.i.d. random variables such that $\mathbb{E}|\varepsilon_1|^p$ is finite for some $p > 2$. Assuming $\mathbb{E}\varepsilon_1 = 0$ and $\mathbb{E}\varepsilon_1^2 = 1$, provided that the ε_i's are defined on some rich enough probability space, there exists some Brownian motion W such that*

$$\mathbb{P}\left(\sup_{u\in[0,1]}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\varepsilon_i \mathbb{1}_{x_i\leq u} - W(u)\right| > x\right) \leq A_0 n^{1-p/2}x^{-p}$$

for some $A_0 > 0$ which depends only on $\mathbb{E}|\varepsilon_1|^p$.

Proof of theorem 4.8

By Sakhanenko's theorem (see [34], theorem 5), there exists some Brownian motion W_0 such that

$$\mathbb{E}\left(\sup_{k\leq n}\left|\sum_{i=1}^{k}\varepsilon_i - W_0(k)\right|^p\right) \leq n\mathbb{E}|\varepsilon_1|^p$$

Let W be the Brownian motion defined by $\forall u \in \mathbb{R}$, $W(u) = W_0(nu)/\sqrt{n}$. Then

$$\mathbb{E}\left(\sup_{k\leq n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\varepsilon_i - W(n^{-1}k)\right|^p\right) \leq n^{1-p/2}\mathbb{E}|\varepsilon_1|^p$$

For each $u \in [0,1]$, let x_u be the greatest x_i, $i \in \{1,...,n\}$, such that $x_i \leq u$. Then for each $x > 0$

$$\begin{aligned}\mathbb{P}\left(\sup_{u\in[0,1]}\left|W(x_u) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i \mathbb{1}_{x_i\leq u}\right| > x\right) &= \mathbb{P}\left(\sup_{k\leq n}\left|W(n^{-1}k) - \frac{1}{\sqrt{n}}\sum_{i=1}^{k}\varepsilon_i\right| > x\right)\\ &\leq \frac{\mathbb{E}|\varepsilon_1|^p}{x^p}n^{1-p/2}\end{aligned}$$

On the other hand $u - x_u \leq 1/n$ and therefore there exists some positive constant A such that for each positive x

$$\begin{aligned}\mathbb{P}\left(\sup_{u\in[0,1]}|W(x_u) - W(u)| > x\right) &\leq \mathbb{P}\left(\sup_{|u-v|\leq 1/n,\ 0\leq v\leq 1}|W(u) - W(v)| > x\right)\\ &\leq An\exp\left(-nx^2/2\right)\end{aligned}$$

For each $\eta > 0$, there exists some positive integer n_η such that

$$An\exp\left(-nx^2/2\right) \leq \frac{\mathbb{E}|\varepsilon_1|^p}{x^p}n^{1-p/2}$$

whenever $x \geq n^{-\eta}$ and $n \geq n_\eta$.. Possibly enlarging the constant A_0 and taking into account the fact that a probability is not greater than one finally yields theorem 4.8. ◇

4.5.3 Proof of theorem 4.2

By theorem 4.8, there exists some Brownian motion W such that

$$\mathbb{P}\left(\sup_{u\in[0,1]}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\varepsilon_i \mathbb{1}_{x_i\le u} - W(u)\right| > x\right) \le An^{1-p/2}x^{-p}$$

for some $A > 0$ which only depends on $\mathbb{E}|\varepsilon_1|^p$. Throughout the proof, W will denote such a Brownian motion. Let F be the function defined by $\forall t \in [0,1]$, $F(t) = \int_0^t f(s)ds$. For each $a \in \mathbb{R}$, let $Z_a^{(n)}$ and $R^{(n)}$ be the processes defined by $\forall u \in [0,1]$,

$$\begin{aligned} Z_a^{(n)}(u) &= W(u) + \sqrt{n}(F(u) - au) \\ R^{(n)}(u) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\varepsilon_i + f(x_i))\mathbb{1}_{x_i\le u} - W(u) - \sqrt{n}F(u) \end{aligned}$$

Let U_n and V_n be the processes defined by $\forall a \in \mathbb{R}$,

$$\begin{aligned} U_n(a) &= \operatorname*{argmax}_{u\in[0,1]}\{Z_a^{(n)}(u)\} \\ V_n(a) &= \operatorname*{argmax}_{u\in[0,1]}\{Z_a^{(n)}(u) + R^{(n)}(u)\} \end{aligned}$$

Then, U_n is the inverse process of the isotonic estimator of f in a white noise model and V_n is the inverse process of the isotonic estimator of f in model (4.2). By theorem 4.5 and the second part of lemma 4.12

$$n^{1/6}\left(n^{1/3}\int_m^M |U_n(a) - g(a)|da - C_f\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 8k) \quad as\ n \to \infty$$

Thus, by lemma 4.12 it suffices to prove that

$$\lim_{n\to\infty} \mathbb{E}n^{1/2}\int_m^M |V_n(a) - U_n(a)|da = 0 \tag{4.38}$$

Fix $a \in \mathbb{R}$. Let W_1 and W_2 be the Brownian motion defined by $\forall u \in \mathbb{R}$, $W_1(u) = W(u + g(a)) - W(g(a))$ and $W_2(u) = n^{1/6}W_1(un^{-1/3})$. Let $I_a^{(n)} = [-n^{1/3}g(a), n^{1/3}(1-g(a))]$. Changes of variables yield:

$$\begin{aligned} n^{1/3}(U_n(a) - g(a)) &= \operatorname*{argmax}_{u\in I_a^{(n)}}\{W(g(a) + n^{-1/3}u) + \sqrt{n}F(g(a) + n^{-1/3}u) - n^{1/6}au\} \\ &= \operatorname*{argmax}_{u\in I_a^{(n)}}\{W_1(n^{-1/3}u) + \sqrt{n}F(g(a) + n^{-1/3}u) - n^{1/6}au\} \\ &= \operatorname*{argmax}_{u\in I_a^{(n)}}\{n^{-1/6}W_2(u) + \sqrt{n}F(g(a) + n^{-1/3}u) - n^{1/6}au\} \\ &= \operatorname*{argmax}_{u\in I_a^{(n)}}\{W_2(u) + n^{2/3}F(g(a) + n^{-1/3}u) - n^{1/3}au\} \end{aligned}$$

One can prove using the same arguments

$$n^{1/3}(V_n(a) - g(a)) = \operatorname*{argmax}_{u\in I_a^{(n)}}\{W_2(u) + n^{2/3}F(g(a) + n^{-1/3}u) - n^{1/3}au + n^{1/6}R^{(n)}(g(a) + n^{-1/3}u)\}$$

Since f is bounded, there exists some $c > 0$ such that

$$\sup_{u\in[0,1]} \left| \sum_{i=1}^{n} \frac{f(x_i)}{n} \mathbb{I}_{x_i \le u} - F(u) \right| \le \frac{c}{n}$$

Thus for each positive real number x we get

$$\begin{aligned} \mathbb{P}\left(2 \sup_{u\in I_a^{(n)}} n^{1/6} |R^{(n)}(g(a)+n^{-1/3}u)| > x\right) = \\ &= \mathbb{P}\left(2 \sup_{u\in[0,1]} n^{1/6} |R^{(n)}(u)| > x\right) \\ &\le \mathbb{P}\left(2 \sup_{u\in[0,1]} n^{1/6} \left| W(u) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \mathbb{I}_{x_i \le u} \right| > x - 2cn^{-1/3}\right) \\ &\le \mathbb{P}\left(\sup_{u\in[0,1]} n^{1/6} \left| W(u) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \mathbb{I}_{x_i \le u} \right| > x/4\right) \end{aligned}$$

whenever $x \ge 4cn^{-1/3}$. Since a probability is not greater than 1 and possibly enlarging the constant A, we get for each $x > 0$

$$\mathbb{P}\left(2 \sup_{u\in I_n(a)} n^{1/6} |R^{(n)}(g(a)+n^{-1/3}u)| > x\right) \le An^{1-p/3}x^{-p}$$

Let $D_a^{(n)}$ be the function of $C^1(I_a^{(n)})$ defined by

$$\forall u \in I_a^{(n)} \qquad D_a^{(n)}(u) = n^{2/3}\left[-F(g(a)+n^{-1/3}u) + Fg(a) + n^{-1/3}au\right]$$

and $s_a^{(n)}$ be the derivative of $D_a^{(n)}$. Then, $n^{1/3}(U_n(a) - g(a)) = \operatorname{argmax}_{u\in I_a^{(n)}}\{W_2(u) - D_a^{(n)}(u)\}$. Since f' is bounded, there exists some positive S such that for each $T > 0$ and $a \in [m, M]$

$$\sup_{|t|\le T} (s_a^{(n)}(t))^2 \le ST.$$

Let T denotes $\log n$ and δ be some positive real number such that $\delta \le T$. By the second part of lemma 4.9 and lemma 4.4, there exists some positive C such that

$$\mathbb{P}\left(n^{1/3}|U_n(a) - V_n(a)| > \delta\right) \le An^{1-p/3}x^{-p}\delta^{-3p/2} + 20eTx + 4\exp(-\frac{C}{2}T^3) \qquad (4.39)$$

whenever $\delta \log \frac{1}{2x\delta} \le S^{-1}T^{-1}$.

Let ε be some positive real number and for each $\delta \in [(\log n)^{-1}n^{-1/6}, n^{-\varepsilon}]$, let x_δ be

$$x_\delta = \left(\frac{A}{20eT}\right)^{\frac{1}{p+1}} n^{\frac{3-p}{3(p+1)}} \delta^{-\frac{3p}{2(p+1)}}$$

Suppose n large enough (say $n \ge n_0$) so that $\delta \log \frac{1}{2x_\delta\delta} \le S^{-1}T^{-1}$ for each $\delta \in [(\log n)^{-1}n^{-1/6}, n^{-\varepsilon}]$. We get for each $\delta \in [(\log n)^{-1}n^{-1/6}, n^{-\varepsilon}]$

$$\mathbb{P}\left(n^{1/3}|U_n(a) - V_n(a)| > \delta\right) \le 40eT\left(\frac{A}{20eT}\right)^{\frac{1}{p+1}} n^{\frac{3-p}{3(p+1)}} \delta^{-\frac{3p}{2(p+1)}} + 4\exp(-\frac{C}{2}T^3)$$

whenever $n \geq n_0$ and therefore

$$\int_{(\log n)^{-1}}^{n^{1/6-\epsilon}} \mathbb{P}(n^{1/2}|U_n(a) - V_n(a)| > \rho)d\rho =$$

$$\begin{aligned}
&= n^{1/6} \int_{(\log n)^{-1}n^{-1/6}}^{n^{-\epsilon}} \mathbb{P}(n^{1/3}|U_n(a) - V_n(a)| > \delta)d\delta \\
&\leq 40eT \left(\frac{A}{20eT}\right)^{\frac{1}{p+1}} n^{\frac{1}{6}+\frac{3-p}{3(p+1)}} \int_{(\log n)^{-1}n^{-1/6}}^{n^{-\epsilon}} \delta^{-\frac{3p}{2(p+1)}} d\delta + 4n^{1/6-\epsilon}\exp(-\frac{C}{2}T^3) \\
&\leq 40eT\frac{2(p+1)}{p-2}\left(\frac{A}{20eT}\right)^{\frac{1}{p+1}} n^{\frac{12-p}{12(p+1)}}(\log n)^{\frac{p-2}{2(p+1)}} + 4n^{1/6-\epsilon}\exp(-\frac{C}{2}T^3)
\end{aligned}$$

which does not depend on a and converges to zero as $n \to \infty$ since $p > 12$. On the other hand,

$$\int_0^{(\log n)^{-1}} \mathbb{P}(n^{1/2}|U_n(a) - V_n(a)| > \rho)d\rho \leq (\log n)^{-1}$$

which does not depend on a and converges to zero as $n \to \infty$. Moreover, if θ denotes $\frac{p}{6p-4}$,

$$\int_{n^{1/6-\epsilon}}^{n^\theta \log n} \mathbb{P}(n^{1/2}|U_n(a) - V_n(a)| > \rho)d\rho \leq$$

$$\begin{aligned}
&\leq n^\theta \log n \mathbb{P}(n^{1/3}|U_n(a) - V_n(a)| > n^{-\epsilon}) \\
&\leq 40eT\left(\frac{A}{20eT}\right)^{\frac{1}{p+1}} n^{\frac{3-p}{3(p+1)}+\frac{3p\epsilon}{2(p+1)}+\theta} \log n + 4n^\theta \log n \exp(-\frac{C}{2}T^3)
\end{aligned}$$

Since $\frac{3-p}{3(p+1)} + \frac{p}{6p-4} < 0$ whenever $p \geq 8$, it suffices to choose ε small enough to get

$$\lim_{n\to\infty} \sup_{a\in[m,M]} \int_{n^{1/6-\epsilon}}^{n^\theta \log n} \mathbb{P}\left(n^{1/2}|U_n(a) - V_n(a)| > \rho\right) = 0$$

On the other hand, by lemmas 4.4 and 4.11 there exist some constants $C_0 > 0$ and $C > 0$ such that for each $\delta > 0$

$$\mathbb{P}\left(n^{1/3}|U_n(a) - V_n(a)| > \delta\right) \leq$$

$$\begin{aligned}
&\leq \mathbb{P}\left(n^{1/3}|U_n(a) - g(a)| > \frac{\delta}{2}\right) + \mathbb{P}\left(n^{1/3}|V_n(a) - g(a)| > \frac{\delta}{2}\right) \\
&\leq 2\exp(-C\delta^3) + \frac{C_0}{\delta^{3p/2}}
\end{aligned}$$

Thus we get

$$\begin{aligned}
\int_{n^\theta \log n}^{\infty} \mathbb{P}(n^{1/2}|U_n(a) - V_n(a)| > \rho)d\rho &= n^{1/6}\int_{n^{\theta-1/6}\log n}^{\infty} \mathbb{P}(n^{1/3}|U_n(a) - V_n(a)| > \delta)d\delta \\
&\leq n^{1/6}\int_{n^{\theta-1/6}\log n}^{\infty} 2\exp(-C\delta^3) + \frac{C_0}{\delta^{3p/2}}d\delta
\end{aligned}$$

where $\theta - 1/6 > 0$ and

$$n^{1/6}\int_{n^{\theta-1/6}\log n}^{\infty} \delta^{-3p/2}d\delta = \frac{2}{3p-2}(\log n)^{\frac{2-3p}{2}}$$

and therefore

$$\lim_{n\to\infty} \sup_{a\in[m,M]} \int_0^\infty \mathbb{P}(n^{1/2}|U_n(a) - g(a)| > \rho)d\rho = 0$$

which proves that

$$\lim_{n\to\infty} \sup_{a\in[m,M]} \mathbb{E}n^{1/2}|U_n(a) - V_n(a)| = 0$$

and completes the proof of theorem 4.2. ◇

4.5.4 Proof of theorem 4.3

We simply use $\int_0^1 |f_n(t) - f(t)|dt = \int_{\mathbb{R}} |V_n(a) - g(a)|da$, where V_n is the inverse process of f_n and g is the inverse function of f. Lemma 4.12 ensures that $\mathbb{E}\int_0^1 |f_n(t) - f(t)|dt = \mathbb{E}\int_m^M |V_n(a) - g(a)|da + O(n^{-2/3})$ and the result follows by integrating inequality (4.32). ◇

Appendice A

Convergence to a gaussian process with independent increments

In order to prove the asymptotic normality of the $\mathbb{L}_1$-distance between the isotonic estimator of a monotone signal function in a white noise model and the true function, we used the proposition 4.3, which is a refinement of Billingsley's theorem (see [3], theorem 19.1). This appendix is devoted to the proof of proposition 4.3. Let us recall this result and the definition we use for weak asymptotically independence.

Definition A.1 *Let $\{X_n,\ n \in \mathbb{N}\}$ and $\{Y_n,\ n \in \mathbb{N}\}$ be sequences of random variables. These sequences are said weakly asymptotically independent if for each $x, y \in \mathbb{R}$*

$$\lim_{n\to\infty} \left| \mathbb{E}\left(e^{i(xX_n+yY_n)}\right) - \mathbb{E}\left(e^{ixX_n}\right)\mathbb{E}\left(e^{iyY_n}\right)\right| = 0$$

Proposition A.1 *Let m and M be real numbers such that $m < M$, and let $\{X_n,\ n \in \mathbb{N}\}$ be a tight sequence in $C[m, M]$. For each s and $t \in [m, M]$, let $\Delta_n(t,s) = X_n(t) - X_n(s)$. Suppose*

h_1 - *For each $n \in \mathbb{N}$, $X_n(0) = 0$ a.s.*

h_2 - *For each s_1, t_1, s_2, t_2 in $[m, M]$ with $s_1 \le t_1 < s_2 \le t_2$, $\Delta_n(t_1, s_1)$ and $\Delta_n(s_2, t_2)$ are weakly asymptotically independent.*

h_3 - *For each $t, s \in [m, M]$, $\lim_{n\to\infty} \mathbb{E}\Delta_n(t,s) = 0$*

h_4 - *There exists some increasing positive function v differentiable over $[m, M]$ satisfying $v(0) = 0$ such that for each $s, t \in [m, M]$, $\lim_{n\to\infty} \mathbb{E}\Delta_n^2(t,s) = |v(t) - v(s)|$*

h_5 - *For each $t, s \in [m, M]$, $\{\Delta_n^2(s,t),\ n \in \mathbb{N}\}$ is uniformly integrable.*

Then, X_n converges in distribution in $C[m, M]$ to $W \circ v$ as $n \to \infty$.

Proof: Let X be the limit of a convergente subsequence of $\{X_n,\ n \in \mathbb{N}\}$. For each $s, t \in [m, M]$, let $\Delta(s,t)$ denote the increment $X(t) - X(s)$ and let $\varphi(t,\ .\ )$ be the caracteristic function of $X(t)$:

$$\forall u \in \mathbb{R}, \quad \varphi(t, u) = \mathbb{E}\left(e^{iuX(t)}\right)$$

Fix $t \in [m, M]$ and $h \in [m, M-t]$. Then,

$$\begin{aligned}\varphi(t+h,u)-\varphi(t,u) &= \mathbb{E}\left(e^{iuX(t+h)}-e^{iuX(t)}\right)\\ &= \mathbb{E}\left[e^{iuX(t)}\left(e^{iu\Delta(t,t+h)}-1\right)\right]\end{aligned}$$

But X's sample paths are continuous, so X's increments are independent and we get

$$\varphi(t+h,u)-\varphi(t,u) = \mathbb{E}\left[e^{iuX(t)}\right]\mathbb{E}\left[e^{iu\Delta(t,t+h)}-1\right]$$

For each $x \in \mathbb{R}$, $e^{ix} = 1 + ix - \frac{1}{2}x^2 + c(x)$ where $|c(x)| \le x^3$, thus

$$\varphi(t+h,u)-\varphi(t,u) = \varphi(t,u)\mathbb{E}\left[iu\Delta(t,t+h)-\frac{1}{2}u^2\Delta(t,t+h)^2+c(u\Delta(t,t+h))\right]$$

For each s and t in $[m, M]$, $\mathbb{E}\Delta(t,s) = 0$ and $\mathbb{E}\Delta(t,s)^2 = |v(t)-v(s)|$. Since v is increasing and differentiable over $[m, M]$, we get

$$\begin{aligned}\lim_{h\searrow 0}\frac{1}{h}\mathbb{E}\left[iu\Delta(t,t+h)-\frac{1}{2}u^2\Delta(t,t+h)^2\right] &= -\frac{1}{2}u^2\lim_{h\searrow 0}\frac{v(t+h)-v(t)}{h}\\ &= -\frac{1}{2}u^2v'(t)\end{aligned}$$

On the other haand, we get since $|c(x)| \le x^3$,

$$\mathbb{E}\left(c(u\Delta(t,t+h))\right) \le \mathbb{E}\left(|c(u\Delta(t,t+h))|\right) \le |u|^3\mathbb{E}\left(|\Delta(t,t+h)|^3\right)$$

To compute this third order moment, we make use of some results of Billingsley about fluctuation of partial sums of dependant or independent variables (see [3], theorem 12.1)

Theorem A.1 *Let $\xi_1, \dots, \xi_p$ be random variables, $S_0 = 0$ and for each $i \in \{1, \dots, p\}$ $S_i = \sum_{k=1}^{i}\xi_k$. Suppose there exist some nonnegative numbers $u_1, \dots, u_p$ such that $\forall 0 \le i \le j \le k \le p$,*

$$\mathbb{E}\left(|S_j - S_i|^2|S_k - S_j|^2\right) \le \left(\sum_{i<l\le k} u_l\right)^2 \tag{A.1}$$

Then there exists some $K \in \mathbb{R}^+$ such that $\forall\lambda > 0$,

$$\mathbb{P}\left(|S_p| \ge \lambda\right) \le \frac{K}{\lambda^4}\left(\sum_{1\le l\le p} u_l\right)^2 + \mathbb{P}\left(\max_{1\le i\le p}|\xi_i| \ge \frac{\lambda}{4}\right)$$

Let p be a positive integer and for each $i \in \{1, \dots, j\}$, define $\xi_i = \Delta(t + \frac{i-1}{p}h, t + \frac{i}{p}h)$. Using the notations of theorem A.1, we get (for $i \in \{0, \dots, p\}$) $S_i = \Delta(t, t + \frac{i}{p}h)$. For each s and

t in $[m, M]$, $\mathbb{E}\Delta(t,s)^2 = |v(t) - v(s)|$ and X's increments are independent. Thus for each $0 \leq i \leq j \leq k \leq p$

$$\begin{aligned}
\mathbb{E}\left(|S_j - S_i|^2 |S_k - S_j|^2\right) &= \mathbb{E}\left(|S_j - S_i|^2\right) \mathbb{E}\left(|S_k - S_j|^2\right) \\
&= \left(v(t + \frac{j}{j}h) - v(t + \frac{i}{j}h)\right)\left(v(t + \frac{k}{j}h) - v(t + \frac{j}{j}h)\right) \\
&\leq \left(v(t + \frac{k}{j}h) - v(t + \frac{i}{j}h)\right)^2 \\
&\leq \left(\sum_{i<l\leq k} u_l\right)^2
\end{aligned}$$

where $\forall\ l \in \{1, ..., p\}$, $u_l = v(t + \frac{l}{p}h) - v(t + \frac{l-1}{p}h)$. By theorem A.1, for each $\lambda > 0$

$$\mathbb{P}(|\Delta(t, t+h)| \geq \lambda) \leq \frac{K}{\lambda^4}(v(t+h) - v(t))^2 + \mathbb{P}\left(\max_{1\leq i\leq p} |\xi_i| \geq \frac{\lambda}{4}\right) \tag{A.2}$$

Leaving p goes to infinity we get

$$\mathbb{P}(|\Delta(t, t+h)| \geq \lambda) \leq \frac{K}{\lambda^4}\left(v(t+h) - v(t)\right)^2$$

We can now overestimate the third order moment of an increment with step h using the following equality

$$\begin{aligned}
\mathbb{E}(|\Delta(t, t+h)|^3) &= \int_0^{+\infty} \mathbb{P}(|\Delta(t, t+h)| \geq \lambda^{1/3}) \wedge 1\ d\lambda \\
&\leq \int_0^{+\infty} \left(\frac{K}{\lambda^{4/3}}(v(t+h) - v(t))^2\right) \wedge 1\ d\lambda \\
&\leq 4K^{3/4}(v(t+h) - v(t))^{3/2}
\end{aligned}$$

and since v is differentiable over $[m, M]$, $\lim_{h \searrow 0} \frac{1}{h}\mathbb{E}(|\Delta(t, t+h)|^3) = 0$. We complete the proof by using the same arguments as those of Billingsley.

Appendice B

$\mathbb{L}_1$-distance between inverses of monotone functions

A crucial equality in the proof of the central limit theorem 4.2 for the $\mathbb{L}_1$-distance between a monotone function f and its isotonic estimate f_n (either in a white noise model or in a regression model) is the following

Lemma B.1 *Let l and h be decreasing (resp. increasing) functions defined over [0,1] and l^{-1}, h^{-1} their inverse functions. Then,*

$$\int_0^1 |h(t) - l(t)|dt = \int_{\mathbb{R}} |h^{-1}(a) - l^{-1}(a)|da$$

This appendix is devoted to the proof of this equality. We shall prove it whenever l and h are decreasing, since it can be proved using the same arguments if l and h are increasing. Recall that if h is a decreasing function defined over $[0, 1]$, the inverse function h^{-1} of h is defined over $\mathbb{R}$ by

$$\forall t \in [0,1],\ \forall a \in \mathbb{R}, \qquad h(t) \leq a \Longleftrightarrow h^{-1}(a) \leq t$$

We can notice that $\int_{\mathbb{R}_+} \mathbb{1}_{|h(t)-l(t)|>a}\, da = |h(t) - l(t)|$ and therefore

$$\int_0^1 |h(t) - l(t)|\, dt = \int_0^1 \int_{\mathbb{R}_+} \mathbb{1}_{h(t)>l(t)+a}\, da\, dt + \int_0^1 \int_{\mathbb{R}_+} \mathbb{1}_{l(t)>h(t)+a}\, da\, dt$$

We compute the first integral, since the second one can be computed in the same way. Using changes of variables and Fubini's theorem we get

$$\begin{aligned} \int_0^1 \int_{\mathbb{R}_+} \mathbb{1}_{h(t)>l(t)+a}\, da\, dt &= \int_0^1 \int_{l(t)}^{\infty} \mathbb{1}_{h(t)>a}\, da\, dt \\ &= \int_0^1 \int_{l(t)}^{\infty} \mathbb{1}_{h^{-1}(a)>t}\, da\, dt \\ &= \int_{l(1)}^{\infty} \int_{l^{-1}(a)}^{1} \mathbb{1}_{h^{-1}(a)>t}\, dt\, da \end{aligned}$$

But h^{-1} takes values in $[0, 1]$, thus $\mathbb{1}_{h^{-1}(a)>t}$ is equal to zero whenever $t \geq 1$ and therefore

$$\int_0^1 \int_{\mathbb{R}_+} \mathbb{1}_{h(t)>l(t)+a}\, da\, dt = \int_{l(1)}^{\infty} \int_0^{\infty} \mathbb{1}_{h^{-1}(a)-l^{-1}(a)>t}\, dt\, da$$

But $l^{-1}(a)$ is equal to 1 whenever $a \leq l(1)$, thus $\mathbb{I}_{h^{-1}(a)-l^{-1}(a)>t}$ is zero for each $t > 0$ and $a \leq l(1)$ and therefore

$$\int_0^1 \int_{\mathbb{R}_+} \mathbb{I}_{h(t)>l(t)+a} \, da \, dt = \int_{\mathbb{R}} \int_0^\infty \mathbb{I}_{h^{-1}(a)-l^{-1}(a)>t} \, dt \, da$$

We get

$$\begin{aligned} \int_0^1 |h(t) - l(t)| \, dt &= \int_{\mathbb{R}} \int_0^\infty \mathbb{I}_{h^{-1}(a)-l^{-1}(a)>t} \, dt \, da + \int_{\mathbb{R}} \int_0^\infty \mathbb{I}_{h^{-1}(a)-l^{-1}(a)>t} \, dt \, da \\ &= \int_{\mathbb{R}} \int_0^\infty \mathbb{I}_{|h^{-1}(a)-l^{-1}(a)|>t} \, dt \, da \\ &= \int_{\mathbb{R}} |h^{-1}(a) - l^{-1}(a)| da \end{aligned}$$

which proves lemma B.1. $\diamond$

Références

[1] R. E. Barlow, D. J. Bartholomew, J. M. Bremmer, and H. D. Brunk. *Statistical Inference under Order Restrictions.* Wiley, 1972.

[2] J. K. Benedetti. On the nonparametric estimation of regression functions. *Journ. Roy. Statist. Soc.*, 39:248–253, 1977. Series B.

[3] P. Billingsley. *Convergence of probability measures.* Wiley, 1968.

[4] L. Birgé. The grenander estimator: a nonasymptotic approach. *Ann. Statist.*, 17(4):1532–1549, 1989.

[5] H. D. Brunk. On the estimation of parameters restricted by inequalities. *Ann. Math. statist.*, pages 437–454, 1958.

[6] H. D. Brunk. Estimation of isotonic regression. *Non parametric techniques in Statistical Inference*, pages 177–195, 1970.

[7] N. N. Cencov. Evaluation of an unknown distribution density from observations. *Soviet Math. Dokl.*, 3:1559–1562, 1962.

[8] K. F. Cheng and P. E. Lin. Nonparametric estimation of a regression functions. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 57:223–233, 1981.

[9] J. Coursol. *Techniques statistiques des modèles linéaires*, volume 1. Aspects théoriques. Les cours du C.I.M.P.A., 1980.

[10] M. Csorgo and L. Horváth. Central limit theorems for l_p-norms of density estimators. *Probab. Th. Rel. Fields*, 80:269–291, 1988.

[11] D. Dacunha-Castelle and M. Duflo. *Probabilités et Statistiques*, volume 2. Problèmes à temps mobile. Masson, Paris, 1983.

[12] G.T. Duncan. An empirical study of jackknife-constructed confidence regions in nonlinear models. *Technometrics*, 20(2):123–129, 1978.

[13] B. Efron. *The jackknife, the bootstrap, and other resampling plans.* Philadelphia: Society of industrial and applied mathematics, 1992.

[14] T. Fox, D. Hinkley, and K. Larntz. Jackknifing in nonlinear regression. *Technometrics*, 22(1):29–33, 1980.

[15] H. L. Gray and W. R. Schucany. *The generalized jackknife statistics.* M. Dekker, INC. New York, 1972.

[16] U. Grenander. On the theory of mortality measurement ii. *Skand. Aktuarietidskr*, 39:125–153, 1956.

[17] P. Groeneboom. Estimating a monotone density. In Richard A. Olsen Lucien Le Cam, editor, *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume 2, pages 539–554, Wadsworth, 1985.

[18] P. Groeneboom. Brownian motion with parabolic drift and airy functions. *Probab. Th. Rel. Fields*, pages 79–109, 1989.

[19] P. Groeneboom, G. Hooghiemstra, and H. P. Lopuhaa. Asymptotic normality of the l_1-error of the grenander estimator. Technical report, Delft University of Technology, 1996.

[20] D.V. Hinkley. Jackknifing in unbalanced situations. *Technometrics*, 19(3):285–292, 1977.

[21] R.I. Jenrich. Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.*, 40(2):633–643, 1969.

[22] J. Kim and D. Pollard. Cube root asymptotics. *Annals of Statistics*, 18:191–219, 1990.

[23] A. Kogure. Asymptotically optimal cells for a histogram. *Ann. Statis*, 15:1023–1030, 1987.

[24] E. Malinvaud. The consistency of nonlinear regressions. *Ann. Math. Statist.*, 41:956–969, 1969.

[25] R.G. Miller. The jackknife - a review. *Biometrika*, 61(1):1–15, 1974.

[26] R.G. Miller. An unbalanced jackknife. *Ann. Statis.*, 2(5):880–891, 1974.

[27] Pfanzagl. On the mesurability and consistancy of minimum contrats estimates. *Metrika*, 14:249–272, 1969.

[28] Pfanzagl. The berry-essen bound for minimum of contraste estimates. *Metrika*, 17:82–91, 1971.

[29] M. B. Priestley and M. T. Chao. Nonparametric function fitting. *Journ. Roy. Statist. Soc.*, 34:385–392, 1972. series B.

[30] B. L. S. Prakasa Rao. Estimation of a unimodal density. *Sankhya Ser*, 31:23–36, 1969.

[31] B. L. S. Prakaso Rao. Asymptotic theory of the least squares estimator in a nonregular nonlinear regression model. *Statist. Probab. Lett.*, 3:15–18, 1985.

[32] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 37:642–669, 1956.

[33] H. P. Rosenthal. On the subspace of l_p, $p > 2$, spanned by sequences of independent random variables. *Israel J. Math.*, 8:273–303, 1970.

[34] A. I. Sakhanenko. Estimates in the variance principle. *Trudy. Inst. Mat. Sibirsk. Otdel*, pages 27–44, 1972.

[35] G. A. F. Seber and C. J. Wild. *Nonlinear regression.* Wiley, 1989.

[36] J. Shao. On resamplig methods for variance and bias estimation in linear models. *Ann. Statist.*, 16(3):986–1008, 1988.

[37] J. Shao. Consistency of the least-squares estimator and its jackknife variance estimator in nonlinear models. *Canadian J. Statist.*, 20(4):415–428, 1992.

[38] J. Shao. Jackknifing in generalized linear models. *Ann. Inst. Statist. Math.*, 44(4):673–686, 1992.

[39] J. Tukey. Bias and confidence in not quite large samples. *Ann. Math. Statist*, 29:614, 1958.

[40] N.C. Weber and H. Welsh. Jackknifing the general linear model. *Austral. J. Statist.*, 25(3):425–436, 1983.

[41] F. T. Wright. Monotone regression estimates. *Annals of statistics*, 9(2):449–452, 1981.

[42] C. F. J. Wu. Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.*, 9:501–513, 1981.