Université de Montréal

Approximation de la distribution de la distance entre deux courbes empiriques

par

Nadine Ouellette

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Statistique

octobre 2004

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

Approximation de la distribution de la distance entre deux courbes empiriques

présenté par

Nadine Ouellette

a été évalué par un jury composé des personnes suivantes :

Pierre Duchesne
(président-rapporteur)

Jean-François Angers
(directeur de recherche)

Roch Roy
(membre du jury)

Mémoire accepté le:
13 octobre 2004

SOMMAIRE

Ce mémoire fait suite à un article (voir Angers, 2003) et il a pour but la comparaison de deux courbes empiriques. Ainsi, nous voulons déterminer si deux fonctions sont égales ou non. Tout comme l'article le suggère, la première étape consiste à estimer les fonctions en se basant sur deux échantillons indépendants. Dans un contexte de régression bayésienne non paramétrique, les fonctions sont décomposées à l'aide d'une base d'ondelettes et des densités *a priori* sont proposées pour les coefficients d'ondelettes.

Afin de tester l'égalité des fonctions, l'article suggère d'abord un test de Bayes, malgré le fait que le temps de calcul soit très long. Ensuite, une approximation en séries d'Edgeworth pour la distribution de la distance entre les fonctions, basée sur la notion de distance euclidienne, est proposée.

Étant donné qu'une telle approximation comporte également certains désavantages (voir section 2.3.1), nous proposons d'abord deux méthodes alternatives. Celles-ci approximent la distribution de la distance entre les deux fonctions à l'aide d'un mélange de distributions khi-deux non centrées. Dans un second temps, nous introduisons un test basé sur des intervalles de confiance simultanés pour la différence entre deux fonctions. Le comportement de ces méthodes, pour différentes paires de fonctions et différentes valeurs de tailles échantillonnales, est étudié au moyen de simulations. Finalement, un exemple utilisant des données réelles illustre la méthodologie.

SUMMARY

This Master thesis is a continuation of an article (see Angers, 2003) and it discusses the problem of detecting the difference between two sets of functional data. The first step is to estimate the functions based on the two independent samples. Using a Bayesian nonparametric regression paradigm, the functions are represented using wavelet decomposition and prior densities are put on the wavelet coefficients.

To test the equality of the functions, this article suggests a Bayes test, however it is computationally intense. Then, an approximation using Edgeworth expansion for the distribution of the distance between the functions, based on Euclidian distance, is proposed.

Since this approximation also has some drawbacks (see Section 2.3.1), we first introduce two alternative methods which are easier to evaluate. The distribution of the distance between the functions is approximated by a mixture of non-central chi-squared distributions. We also propose a test based on simultaneous confidence intervals for the difference between two functions. The behavior of these methods is studied through simulation for several pairs of functions and several sample sizes. Finally, the proposed methodology is further used to analyse a real data set.

MOTS CLÉS

Formes quadratiques, mélange de distributions khi-deux non centrées, estimation par point de selle, intervalles de confiance simultanés, ondelettes.

KEYWORDS

Quadratic forms, mixture of non-central chi-squared distributions, saddlepoint estimation, simultaneous confidence intervals, wavelets.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

REMERCIEMENTS

C'est avec beaucoup d'admiration que je tiens tout d'abord à remercier mon directeur de recherche Jean-François Angers. Sa patience remarquable, ainsi que son dévouement pour ses étudiants font de lui un directeur de recherche exemplaire.

J'aimerais également exprimer ma reconnaissance envers le Département de mathématiques et de statistique de l'Université de Montréal et envers la famille du Professeur Serge Tardif pour leur support financier.

Je désire aussi mentionner que mon parcours à la maîtrise n'aurait pu être aussi agréable sans la présence de mes amis au département. Les principaux intéressés sauront se reconnaître ; merci d'être ce que vous êtes.

Finalement, je dédie ce présent mémoire à mes très très chers parents, à ma petite soeur que j'adore et à mon merveilleux copain.

INTRODUCTION

La comparaison de deux ou plusieurs fonctions de régression est un problème couramment rencontré lors d'un travail expérimental. Citons l'exemple classique du chercheur qui désire comparer la variable réponse d'un groupe traitement à celle d'un groupe contrôle. Si en plus, une covariable est mesurée, alors l'expérimentateur peut décider de prédire la variable réponse de chaque groupe à l'aide de ces observations et vouloir comparer les courbes de régression obtenues.

Parmi les méthodes statistiques classiques, une modélisation paramétrique des fonctions de régression suivie d'une comparaison des paramètres des modèles estimés est parfois l'approche utilisée. Récemment, l'intérêt des chercheurs pour la généralisation des méthodes traditionnelles d'analyse de la variance (ANOVA) afin qu'elles puissent s'appliquer à l'analyse de données fonctionnelles (FANOVA) est grandissant (voir Abramovich et Angelini, 2003). Le désavantage de ces approches réside toutefois dans la spécification du modèle paramétrique. Celui-ci est parfois difficile à identifier.

Quelques méthodes de test non paramétrique ont déjà été proposées. La méthode de Kaplan-Meier est souvent celle utilisée lorsque les données se prêtent à l'analyse de survie (voir Klein et Moeschberger, 1997). Les travaux de Hall et Hart (1990), de Kulasekera (1995) et ceux de Koul et Schick (1997) introduisent d'autres méthodes de test. Toutefois, les valeurs critiques de ces tests doivent être obtenues au moyen de simulations, ou bien par calcul asymptotique. Une autre approche proposée par Fan et Lin (1998) consiste à utiliser les transformées de Fourier pour traduire le problème. Cependant, la méthode requiert des données recueillies à intervalles réguliers. Autrement, un regroupement en classes doit être effectué afin de préparer les données à cet effet.

Dans le présent mémoire, nous optons plutôt pour une modélisation bayésienne non paramétrique des fonctions de régression à l'aide de la base d'ondelettes de Daubechies (voir Daubechies, 1992). Ensuite, nous proposons différents tests permettant de confirmer ou de réfuter l'égalité de ces fonctions. L'estimation des fonctions est faite en se basant sur deux échantillons indépendants. Les tailles échantillonnales ne doivent pas nécessairement être égales et la seule hypothèse assumée est que les fonctions de régression soient lisses. Conséquemment, les tests proposés sont valides pour une grande variété de fonctions.

En résumé, ce mémoire est organisé de la façon suivante. Dans le premier chapitre, nous nous intéressons à l'estimation bayésienne non paramétrique d'une fonction. Nous présentons une introduction à quelques bases pouvant être utilisées pour représenter les fonctions. La base polynomiale ainsi que celle des splines sont les premières à être abordées. Par la suite, les bases d'ondelettes sont étudiées de manière plus approfondie étant donné qu'il s'agit du moyen d'estimation retenu.

Au chapitre 2, nous présentons différents tests servant à déterminer si les fonctions estimées sont égales ou non. Dans un premier temps, nous rappelons les méthodes employées dans Angers (2003), puis proposons quelques alternatives. Puisque la distribution de la distance entre les fonctions peut être exprimée comme un mélange de distributions khi-deux non centrées, nous décrivons deux approximations de la distribution d'un tel mélange : celle d'Imhof (1961) et une autre par point de selle (voir Kuonen, 1999). Basés sur ces approximations, nous suggérons d'abord trois tests. Par la suite, un quatrième test basé sur le concept d'intervalles de confiance simultanés pour la différence entre les fonctions est proposé.

Finalement, au dernier chapitre, nous étudions le comportement des méthodes développées à l'aide de simulations, puis nous comparons les résultats obtenus à ceux parus dans Angers (2003). De plus, un exemple utilisant des données réelles illustre la méthodologie.

Chapitre 1

ANALYSE FONCTIONNELLE

Dans un contexte d'estimation bayésienne non paramétrique d'une fonction, le modèle habituellement considéré est

$$y_i = g(t_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{1.0.1}$$

où g est la fonction inconnue à estimer, les t_i appartiennent à un ensemble borné $\mathcal{T} \subset \mathbb{R}$ et les ε_i représentent les termes d'erreur. L'étape suivante consiste à exprimer la fonction g à l'aide d'une base de fonctions quelconque, c'est-à-dire

$$g(t) = \sum_j \theta_j b_j(t),$$

où b_j appartient à la base $\forall j$. Par la suite, le but est d'estimer les coefficients θ_j et de faire de l'inférence sur ces derniers. Notons que les hypothèses de régularité concernant la fonction g varient selon la base de fonctions utilisée.

Le présent chapitre est consacré au survol de différentes bases de fonctions pouvant être utilisées pour exprimer la fonction g. Nous nous intéressons, dans l'ordre, aux bases polynomiales, à celles des splines et enfin à celles d'ondelettes.

1.1. LES BASES POLYNOMIALES

Malgré le fait qu'elle date de quelques siècles, la base polynomiale constitue souvent encore aujourd'hui l'approche la plus facile et la plus naturelle à utiliser. La méthode consiste à obtenir la base en développant la fonction g inconnue à l'aide des séries de Taylor autour d'un point choisi. L'idée de représenter des fonctions particulières par des séries de puissance remonte à l'époque de Newton. Pour

leur part, les séries de Taylor sont connues depuis les travaux des mathématiciens James Gregory et John Bernoulli datant du XVII^{e} siècle, mais elles doivent leur appellation au mathématicien anglais Brook Taylor. Ce dernier n'avait apparemment aucune connaissance des travaux de Gregory et Bernoulli lorsqu'il publia ses découvertes sur les séries en 1715 (voir Taylor, 1715).

Étant donné que la base polynomiale ne correspond pas à celle utilisée pour la modélisation dans ce mémoire, nous n'entrons pas dans la théorie derrière une telle approche. Nous optons plutôt pour un examen rapide des contributions jugées pertinentes sur le sujet.

Mentionnons d'abord les auteurs Weerahandi et Zidek (1988). Dans cet article, la fonction g est observée en n points n'étant pas nécessairement équidistants. Par hypothèse, g est une fonction lisse, c'est-à-dire $g \in \mathcal{C}^{\infty}(\mathcal{T})$ où $\mathcal{C}^{\infty}(\mathcal{T})$ représente l'espace des fonctions à valeurs dans $\mathcal{T}$ et qui sont infiniment continûment dérivables. De plus, g peut être développée en série de Taylor autour de t_0. Conséquemment, l'équation (1.0.1) devient

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

où $\mathbf{y} = (y_1, \ldots, y_n)^{\top}$; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\top}$, $\beta_j = \frac{\partial}{\partial t} g(t)\Big|_{t=t_0}$, $j = 1, \ldots, p$; $\mathbf{X}$ est une matrice $n \times (p+1)$ telle que $\mathbf{X} = (\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_p)$, $\mathbf{1}$ est un vecteur de taille n composé de 1, $\mathbf{x}_i = ((t_1 - t_0)^i/i!, \ldots, (t_n - t_0)^i/i!)^{\top}$, $i = 1, \ldots, n$; $\boldsymbol{\epsilon} = \boldsymbol{\eta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^{\top}$, $\varepsilon_i = \varepsilon(t_i)$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^{\top}$ et η_i est le reste du développement en série de Taylor de $g(t_i)$, c'est-à-dire $\eta_i = (t_i - t_0)^{p+1} D^{p+1} g(\theta_i)/(p+1)!$, où θ_i est un point situé entre t_i et t_0. Vu la structure du modèle, ϵ et β sont stochastiquement dépendants. Cependant, une transformation des données permet d'obtenir des erreurs ϵ^* non corrélées avec β. Une modélisation reposant sur la loi normale mène à l'obtention de l'estimateur de Bayes pour β. L'approche bayésienne empirique est celle retenue pour la spécification des paramètres. Celle-ci consiste à utiliser les observations pour estimer les paramètres de la densité *a priori* (voir Gelman *et al.*, 2004, chapitre 5). Finalement, une analyse de sensibilité est faite pour l'ordre p du développement en série de Taylor.

Citons maintenant la contribution des auteurs Angers et Delampady (1992) dans laquelle le modèle spécifié est semblable à celui employé par Weerahandi et Zidek (1988) qui est décrit ci-haut. Cependant, le modèle est étudié à l'aide d'une approche bayésienne hiérarchique et une densité *a priori diffuse* est considérée pour les paramètres. Une densité qui possède une très grande variance, voire parfois infinie, est dite diffuse (voir Robert, 2001, chapitre 1). Une méthode permettant de choisir l'ordre du développement de Taylor est présentée. De plus, une analyse de sensibilité est réalisée afin de déterminer l'influence de la densité *a priori* des hyperparamètres sur le lissage de la fonction g.

La prochaine sous-section est consacrée à l'exploration de la base des splines. Il s'agit d'une base un peu plus complexe que la base polynomiale, mais elle offre généralement une meilleure flexibilité.

1.2. Les bases splines

C'est surtout à partir du début des années 60 (1960) que de nombreux chercheurs ont commencé à exploiter le potentiel des fonctions splines. Ils ont découvert qu'elles constituaient un outil permettant de modéliser une fonction avec une approche entièrement mathématique. Pensons notamment aux travaux de Schoenberg (1964), qui ont servi d'introduction aux splines de lissage classiques. Mentionnons également quelques autres contributions de Wahba : Kimeldorf et Wahba (1971), Wahba et Wold (1975), puis Craven et Wahba (1979). Ces derniers ont grandement contribué au développement de telles fonctions splines.

Étroitement liée à celle des polynômes, la base des splines peut être considérée comme une alternative à cette dernière. Dans le cas où la représentativité de la vraie fonction g ne peut être assurée par un polynôme, les splines constituent une option envisageable. Mentionnons qu'il s'agit d'un outil ayant autant retenu l'attention des statisticiens bayésiens que celle des fréquentistes.

Les splines sont définis comme des polynômes par parties de degré m. Les abscisses où se produisent les jonctions de polynômes sont appelées les *noeuds*. Ces noeuds peuvent être simples ou multiples. La multiplicité des noeuds détermine les conditions de régularité de la fonction modélisée. Par exemple, si les noeuds

intérieurs sont simples à l'intérieur du domaine, alors les valeurs et les $(m-1)^{\text{e}}$ dérivées de la fonction sont les mêmes aux points de jonctions. Ainsi, les polynômes peuvent être vus comme un cas particulier de splines sans noeuds. Notons que le nombre total de polynômes par parties utilisés, leurs degrés respectifs et le positionnement des noeuds varient selon le type de modélisation voulue.

Aujourd'hui, toute une variété de fonctions splines existent et sont couramment utilisées à des fins de modélisation. Les fonctions B-splines font partie de celles étant fréquemment rencontrées.

Tout comme c'est le cas pour la base polynomiale, la base des splines n'est pas retenue comme méthode d'estimation dans ce mémoire. Nous faisons donc abstraction des détails théoriques sur le sujet. Voici plutôt un résumé de quelques contributions intéressantes traitant de l'estimation bayésienne non paramétrique d'une fonction au moyen de telles bases. Pour plus de détails théoriques se rapportant aux fonctions splines, nous suggérons les ouvrages de de Boor (1978) et de Schumaker (1981).

Dans un premier temps, citons la contribution de Wahba (1978). Basé sur les observations (t_i, y_i) satisfaisant l'équation (1.0.1) et telles que $t_i \in \mathcal{T} = [0,1]$, $i = 1, \ldots, n$, le spline polynomial de lissage de degré $(2m-1)$ est défini comme étant la solution au problème de minimisation suivant : trouver la fonction g qui minimise l'équation

$$\frac{1}{n}\sum_{i=1}^{n}(g(t_i)-y_i)^2 + \lambda \int_0^1 (g^{(m)}(u))^2 du, \tag{1.2.1}$$

où λ est un paramètre de lissage pour la fonction à spécifier, $g^{(j)}$ est absolument continue, $j = 0, \ldots, m-1$. De plus, $g^{(m)} \in \mathcal{L}_2[0,1]$. (La j^{e} dérivée de g est notée par $g^{(j)}$.) Notons que par définition, l'espace $\mathcal{L}_2(\mathcal{A})$ est celui constitué de fonctions f définies sur $\mathcal{A} \subseteq \mathbb{R}$ et de carré sommable (c'est-à-dire $\|f\|_2 = \left(\int_{\mathcal{A}} |f(x)|^2 dx\right)^{1/2} < \infty$). L'auteur montre que pour une valeur fixée de λ, la solution de l'équation (1.2.1) correspond à l'estimateur de Bayes de g sous une certaine distribution *a priori* pour une fonction de perte quadratique. Plus précisément, étant donné la faible quantité d'information sur les coefficients des polynômes de degré inférieur à m, la distribution *a priori* de ces derniers est

diffuse. Toutefois, elle demeure propre pour un ensemble de variables aléatoires excluant les coefficients mentionnés précédemment. L'article propose également une méthode généralisée de validation croisée (voir Craven et Wahba, 1979) pour le choix du paramètre de lissage λ.

En relation avec l'article que nous venons de résumer, citons maintenant van der Linde (1993). Puisque les estimateurs des coefficients des fonctions splines peuvent être vus comme des estimateurs de Bayes pour une certaine densité *a priori*, l'auteur met l'accent sur la différence entre les modèles *a priori* introduits par Wahba (1978) et Silverman (1985). Malgré le fait que les deux modèles ont la même espérance *a posteriori*, leurs variances *a posteriori* diffèrent par l'erreur d'interpolation.

Ceci met fin à notre discussion sur les bases splines. Les bases d'ondelettes font l'objet de la sous-section suivante et elles retiendront notre attention jusqu'à la fin du présent chapitre. Mentionnons immédiatement que les bases d'ondelettes sont celles retenues pour l'estimation bayésienne non paramétrique des fonctions à estimer dans ce mémoire. Pour cette raison, notre intention est de les présenter avec un plus grand souci du détail.

1.3. Les bases d'ondelettes

L'origine des ondelettes remonte au début du XX^{e} siècle. Par contre, la compréhension de celles-ci pour ce qui est de leur utilité dans la construction de bases orthogonales est relativement récente. Cette compréhension est le fruit d'une union entre certaines théories déjà existantes dans divers domaines et de nouvelles découvertes. De nos jours, les ondelettes se retrouvent au sein de domaines d'application très variés. Notamment, elles sont utilisées dans le traitement des signaux, dans la compression de données et dans l'estimation non paramétrique de données. Depuis le début des années 90 (1990), les différents articles publiés par Donoho, Johnstone et coauteurs démontrent que les ondelettes constituent également un outil approprié pour les problèmes de débruitage, de régression et d'estimation de densité. Grâce à ces publications, les ondelettes ne cessent de gagner en popularité dans le domaine de la statistique.

Les bases d'ondelettes constituent une version améliorée des bases orthogonales classiques (Fourier, l'Hermite, Legendre, etc.). Malgré le fait qu'elles soient habituellement un peu plus difficiles à appliquer que les bases classiques, elles possèdent quelques caractéristiques fort intéressantes qui leurs sont exclusives. Soulignons, par exemple, qu'il est possible de construire des bases d'ondelettes dont tous les éléments ont un support compact. Cela n'est pas le cas de plusieurs bases classiques qui sont composées de fonctions définies sur $\mathbb{R}$. De plus, les bases d'ondelettes ont généralement la capacité de saisir le comportement local et global des fonctions plus efficacement et plus précisément que les bases classiques.

Pour l'ensemble de ces raisons, nous avons choisi d'utiliser les bases d'ondelettes pour l'estimation non paramétrique de la fonction g de l'équation (1.0.1). Les prochaines sous-sections visent donc à familiariser le lecteur avec cet outil. Pour plus de détails, nous suggérons Meyer (1990, 1992), Daubechies (1992), Mallat (1998) et Walnut (2002). Ces ouvrages présentent la théorie des ondelettes de façon plus approfondie. Pour un lecteur intéressé par une introduction centrée sur l'application des ondelettes en statistique, nous lui conseillons Ogden (1997), Härdle *et al.* (1998) et Vidakovic (1999).

1.3.1. L'analyse multirésolution

Commençons par explorer quelques notions relatives à la construction d'une base d'ondelettes. Une telle construction repose sur le concept d'analyse multirésolution. Étant donné que plusieurs auteurs s'y sont spécialement intéressés, nous avons choisi d'introduire ce concept à la manière de Mallat (1989) et de Meyer (1990).

Définition 1.3.1. *Une analyse multirésolution de* $\mathcal{L}_2(\mathbb{R})$ *est une suite* $\{V_j\}_{j\in\mathbb{Z}}$ *de sous-espaces fermés de* $\mathcal{L}_2(\mathbb{R})$ *telle que*

(i) $V_j \subset V_{j+1} \qquad \forall j \in \mathbb{Z}$,

(ii) $\bigcup_{j\in\mathbb{Z}} V_j$ *est dense dans* $\mathcal{L}_2(\mathbb{R})$,

(iii) $\bigcap_{j\in\mathbb{Z}} V_j = \{\mathbf{0}\}$,

(iv) $f(x) \in V_j \Longleftrightarrow f(2x) \in V_{j+1} \qquad \forall j \in \mathbb{Z}$,

(v) $f(x) \in V_j \Longrightarrow f(x - 2^{-j}k) \in V_j \qquad \forall j, k \in \mathbb{Z}$,

(vi) $\exists$ *une fonction* $h \in V_0$ *telle que* $\{h(x-k)\}_{k\in\mathbb{Z}}$ *est une base orthonormée de* V_0, *c'est-à-dire* V_0 *peut s'écrire comme*

$$V_0 = \left\{ f \in \mathcal{L}_2(\mathbb{R}) \;\middle|\; f(x) = \sum_{k\in\mathbb{Z}} c_k h(x-k), \sum_{k\in\mathbb{Z}} c_k^2 < +\infty \right\}.$$

Remarquons que le « **0** » de la propriété (*iii*) doit être vu comme la fonction identiquement nulle. De plus, la propriété (*iv*) ci-haut est présentée d'après les travaux de Mallat (1989), mais elle peut également s'écrire sous la forme

$$f(x) \in V_0 \Longleftrightarrow f(2^j x) \in V_j \qquad \forall j \in \mathbb{Z}.$$

Définissons maintenant une fonction bien particulière nommée la *fonction d'échelle* (ou *père des ondelettes*). Nous verrons que cette dernière joue un rôle central dans la construction d'une analyse multirésolution.

Définition 1.3.2. *Une fonction* $\phi \in \mathcal{L}_2(\mathbb{R})$ *est dite une fonction d'échelle si elle satisfait les deux propriétés suivantes*

(i) les sous-espaces V_j *de* $\mathcal{L}_2(\mathbb{R})$ *définis par*

$$V_j = \left\{ f \in \mathcal{L}_2(\mathbb{R}) \;\middle|\; f(x) = \sum_{k\in\mathbb{Z}} c_{j,k}\phi_{j,k}(x) \right\},$$

où

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k) \qquad \forall j,k \in \mathbb{Z},$$

forment une analyse multirésolution de $\mathcal{L}_2(\mathbb{R})$;

(ii) la suite $\{\phi(x-k)\}_{k\in\mathbb{Z}} = \{\phi_{0.k}\}_{k\in\mathbb{Z}}$ *forme une base orthonormée de* V_0, *c'est-à-dire* ϕ *satisfait la propriété* (*vi*) *de la définition 1.3.1.*

Dans une telle situation, nous disons que $\{V_j\}_{j\in\mathbb{Z}}$ *est l'analyse multirésolution engendrée par* ϕ.

1.3.2. L'équation de dilatation

Étudions maintenant le lien entre la fonctions d'échelle ϕ et une autre fonction ψ bien importante nommée la mère des ondelettes. Plus précisément, à partir d'une fonction d'échelle ϕ donnée, montrons comment construire la mère des ondelettes ψ correspondante. À titre d'exemple, prenons $\{V_j\}_{j\in\mathbb{Z}}$, une analyse

multirésolution engendrée par la fonction d'échelle ϕ. Ainsi, nous avons

$$\phi \in V_0 \subset V_1,$$

et la fonction ϕ peut être décomposée en série à partir des éléments de la base de V_1 dénotés $\{\phi_{1,k}\}_{k\in\mathbb{Z}}$. En effet, il existe une suite $\{A_k\}_{k\in\mathbb{Z}}$ de coefficients de carrés sommable (c'est-à-dire, $\sum_{k\in\mathbb{Z}} A_k^2 < +\infty$) telle que

$$\phi(x) = \sum_{k\in\mathbb{Z}} A_k \phi_{1,k}(x). \tag{1.3.1}$$

D'après la propriété (i) de la définition 1.3.2,

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k) \qquad \forall j,k \in \mathbb{Z},$$

alors nous obtenons

$$\phi_{1,k}(x) = \sqrt{2}\phi(2x-k) \qquad \forall k \in \mathbb{Z}. \tag{1.3.2}$$

L'équation (1.3.1) devient ainsi

$$\begin{aligned} \phi(x) &= \sum_{k\in\mathbb{Z}} \sqrt{2} A_k \phi(2x-k) \\ &= \sum_{k\in\mathbb{Z}} a_k \phi(2x-k). \end{aligned} \tag{1.3.3}$$

Cette dernière équation est connue sous le nom de l'*équation de dilatation.*

À l'aide d'un développement tout à fait similaire, nous pouvons obtenir une équation semblable à l'équation (1.3.3) pour la fonction ψ. Définissons W_0 comme étant le complément orthogonal de V_0 dans V_1. Si

$$\psi \in W_0 \subset V_1,$$

alors nous pouvons également développer ψ en série à l'aide des éléments de la base de V_1. Notons par $\{B_k\}_{k\in\mathbb{Z}}$, la suite de coefficients de carré sommable telle que

$$\psi(x) = \sum_{k\in\mathbb{Z}} B_k \phi_{1.k}(x)$$

$$= \sum_{k\in\mathbb{Z}} \sqrt{2} B_k \phi(2x-k)$$
$$= \sum_{k\in\mathbb{Z}} b_k \phi(2x-k). \tag{1.3.4}$$

Remarquons que ce sont les fonctions $\psi_{j,k}$ définies par

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \tag{1.3.5}$$

qui forment les *ondelettes*. Les auteurs Grossman et Morlet (1984) furent les premiers à les nommer ainsi. Puisque toutes les ondelettes sont issues de la même fonction ψ, cette dernière porte le nom de *mère des ondelettes* (voir Meyer, 1990). Dans la même optique, la fonction d'échelle ϕ est aussi appelée *père des ondelettes*.

Les coefficients a_k et b_k, $k \in \mathbb{Z}$ des équations (1.3.3) et (1.3.4) sont nommés *filtres* et ils permettent d'obtenir la mère des ondelettes ψ à partir du père des ondelettes ϕ. Effectivement, si ϕ est connu, alors la suite de filtres $\{a_k\}_{k\in\mathbb{Z}}$ l'est également. L'étape suivante consiste à choisir les b_k de façon à ce que la mère des ondelettes ψ ait les propriétés voulues. Par sa structure, l'analyse multirésolution nous permet automatiquement de fixer les valeurs des filtres b_k en fonction des a_k. En fait, nous pouvons montrer (voir Leblanc, 2001) que la solution générale à la construction de ψ consiste à choisir

$$b_k = (-1)^k a_{1-k} \qquad \forall k \in \mathbb{Z}.$$

Ainsi, l'équation (1.3.4) devient

$$\psi(x) = \sum_{k\in\mathbb{Z}} (-1)^k a_{1-k} \phi(2x-k). \tag{1.3.6}$$

1.3.3. Les ondelettes de Haar

Dans cette sous-section, nous illustrons la théorie des ondelettes introduite jusqu'à maintenant. Pour ce faire, utilisons la fonction d'échelle ϕ la plus simple qu'il soit possible de trouver, c'est-à-dire

$$\phi(x) = \begin{cases} 1 & \text{si } x \in [0,1), \\ 0 & \text{sinon.} \end{cases} \tag{1.3.7}$$

Cette fonction fut introduite par Haar (1910) plusieurs décennies avant l'émergence du concept de fonction d'échelle. À cette époque, les travaux de Haar portaient sur la construction d'une base orthonormée de $\mathcal{L}_2(\mathbb{R})$ à partir de la fonction ϕ de l'équation (1.3.7). Cette construction peut ainsi être vue comme un signe précurseur à l'analyse multirésolution et à la fonction d'échelle.

Débutons l'illustration en vérifiant si ϕ est bien une fonction d'échelle, c'est-à-dire assurons-nous que ϕ engendre une analyse multirésolution. Comme auparavant, notons

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k) \qquad \forall j, k \in \mathbb{Z}, \tag{1.3.8}$$

et remarquons que dans le cas particulier de la base de Haar où la fonctions d'échelle ϕ correspond à celle décrite par l'équation (1.3.7),

$$\phi_{j,k}(x) = \begin{cases} 2^{j/2} & \text{si } x \in \left[\frac{k}{2^j}, \frac{k+1}{2^j}\right), \\ 0 & \text{sinon.} \end{cases} \tag{1.3.9}$$

Considérons d'abord le sous-espace V_0 de $\mathcal{L}_2(\mathbb{R})$ défini par

$$V_0 = \left\{ f \in \mathcal{L}_2(\mathbb{R}) \;\middle|\; f \text{ est constante sur } [k, k+1), k \in \mathbb{Z} \right\}. \tag{1.3.10}$$

Ainsi, V_0 est un sous-espace composé de fonctions constantes sur des intervalles unitaires. La translation par un entier relatif étant une opération fermée sur V_0, nous avons

$$f \in V_0 \Longrightarrow f(x-k) \in V_0 \qquad \forall k \in \mathbb{Z}.$$

Pour toute fonction $f \in \mathcal{L}_2(\mathbb{R})$ et pour une certaine suite $\{c_k\}_{k\in\mathbb{Z}}$ de carré sommable,

$$f \in V_0 \Longleftrightarrow f(x) = c_k \qquad \text{pour } x \in [k, k+1), k \in \mathbb{Z}.$$

Cela revient à écrire

$$f \in V_0 \Longleftrightarrow f(x) = \sum_{k\in\mathbb{Z}} c_k \phi_{0,k}(x), \tag{1.3.11}$$

car des équations (1.3.8) et (1.3.9), nous avons

$$\phi_{0,k}(x) \;=\; \phi(x-k)$$

$$= \begin{cases} 1 & \text{si } x \in [k, k+1), \\ 0 & \text{sinon.} \end{cases}$$

D'après l'équation (1.3.11), l'expression (1.3.10) devient

$$V_0 = \left\{ f \;\middle|\; f(x) = \sum_{k\in\mathbb{Z}} c_k \phi(x-k), \sum_{k\in\mathbb{Z}} c_k^2 < +\infty \right\},$$

et puisque

$$\begin{aligned} < \phi_{0,k}, \phi_{0,l} > &= \int_{\mathbb{R}} \phi_{0,k}(x)\phi_{0,l}(x)dx \\ &= \int_{\mathbb{R}} \phi(x-k)\phi(x-l)dx \\ &= \begin{cases} 1 & \text{si } k = l, \\ 0 & \text{sinon,} \end{cases} \end{aligned}$$

alors $\{\phi_{0,k}\}_{k\in\mathbb{Z}}$, forme une base orthonormée de V_0.

Considérons maintenant un second sous-espace de $\mathcal{L}_2(\mathbb{R})$, noté V_1, tel que

$$V_1 = \left\{ f \in \mathcal{L}_2(\mathbb{R}) \;\middle|\; f \text{ est constante sur } \left[\frac{k}{2}, \frac{k+1}{2}\right), k \in \mathbb{Z} \right\}. \tag{1.3.12}$$

Ce nouveau sous-espace est composé des fonctions constantes sur les intervalles de largeur $1/2$. La translation par un multiple de $1/2$ étant une opération fermée sur V_1,

$$f \in V_1 \Longrightarrow f\left(x - \frac{k}{2}\right) \in V_1, \qquad \forall k \in \mathbb{Z}.$$

Comme précédemment, pour toute fonction $f \in \mathcal{L}_2(\mathbb{R})$ et pour une certaine suite $\{c_k\}_{k\in\mathbb{Z}}$ de carré sommable,

$$f \in V_1 \Longleftrightarrow f(x) = c_k \qquad \text{pour } x \in \left[\frac{k}{2}, \frac{k+1}{2}\right), k \in \mathbb{Z},$$

ou encore

$$f \in V_1 \Longleftrightarrow f(x) = \sum_{k\in\mathbb{Z}} c_k \phi_{1,k}(x).$$

L'expression (1.3.12) pour V_1 peut dorénavant s'écrire comme

$$V_1 = \left\{ f \;\middle|\; f(x) = \sum_{k\in\mathbb{Z}} c_k \phi_{1,k}(x), \sum_{k\in\mathbb{Z}} c_k^2 < +\infty \right\},$$

et puisque la suite de fonctions $\{\phi_{1,k}\}_{k\in\mathbb{Z}}$ est orthonormée, alors elle constitue une base orthonormée pour V_1.

Faisons maintenant le lien entre les fonctions des sous-espaces V_0 et V_1. Remarquons que

$$f \in V_1 \Longleftrightarrow \exists \text{ une fonction } h \in V_0 \text{ telle que } f(x) = h(2x),$$

c'est-à-dire que la fonction f est dans V_1 si et seulement si elle correspond à une fonction h de V_0 contractée d'un facteur de 2. De plus, étant donné qu'une fonction constante sur les intervalles de la forme $[k, k+1)\,, \forall k \in \mathbb{Z}$, l'est nécessairement sur ceux de la forme $\left[\frac{k}{2}, \frac{k+1}{2}\right)\,, \forall k \in \mathbb{Z}$, alors

$$V_0 \subset V_1.$$

La répétition de ce même procédé pour $j = 2, 3, \ldots$, nous permet de définir les sous-espaces de $\mathcal{L}_2(\mathbb{R})$ suivants

$$V_j = \left\{ f \;\middle|\; f(x) = h(2^j x), h \in V_0 \right\}. \tag{1.3.13}$$

Ces sous-espaces V_j sont tous composés de fonctions constantes sur les différents intervalles $\left[\frac{k}{2^j}, \frac{k+1}{2^j}\right)$, $k \in \mathbb{Z}$, et ont chacun $\{\phi_{j,k}\}_{k\in\mathbb{Z}}$ pour base orthonormée. De plus, tous respectent les propriétés

$$V_0 \subset V_1 \subset V_2 \subset \ldots,$$

et

$$f \in V_j \Longrightarrow f\left(x - \frac{k}{2^j}\right) \in V_j. \tag{1.3.14}$$

Nous pouvons même élargir ce procédé de façon à obtenir les sous-espaces V_j, pour $j < 0$, composés des fonctions dilatées de V_0. Dans ce cas, nous avons

$$\ldots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \ldots,$$

ou bien

$$V_j \subset V_{j+1} \qquad \forall j \in \mathbb{Z}. \tag{1.3.15}$$

Remarquons que l'intersection de tels sous-espaces mène à la fonction identiquement nulle, soit l'unique fonction de $\mathcal{L}_2(\mathbb{R})$ étant constante sur toute la droite réelle. Ceci est noté par

$$\bigcap_{j\in\mathbb{Z}} V_j = \{\mathbf{0}\}. \tag{1.3.16}$$

Notons finalement que d'après Rudin (1987, chapitre 3), les fonctions simples sont denses dans $\mathcal{L}_2(\mathbb{R})$. En considérant cette affirmation, le fait que $\{\phi_{0,k}\}$ constitue une base orthonormale de V_0, ainsi que les équations (1.3.13) à (1.3.16), nous constatons que ϕ et la suite de sous-espaces $\{V_j\}_{j\in\mathbb{Z}}$ satisfont aux définitions 1.3.1 et 1.3.2.

Par conséquent, nous pouvons affirmer que $\{V_j\}_{j\in\mathbb{Z}}$ forme une analyse multirésolution engendrée par la fonction d'échelle ϕ. L'équation de dilatation (voir équation (1.3.3)) est donc respectée par ϕ et de l'équation (1.3.2), nous déduisons

$$\phi(2x-k) = \frac{1}{\sqrt{2}}\phi_{1,k}(x). \tag{1.3.17}$$

Poursuivons maintenant notre illustration en montrant comment obtenir la mère des ondelettes ψ à l'aide de la fonction d'échelle ϕ. Puisque dans le cas particulier de la base de Haar, nous avons

$$\phi_{1,k} = \begin{cases} \sqrt{2} & \text{si } x \in \left[\frac{k}{2}, \frac{k+1}{2}\right), \\ 0 & \text{sinon,} \end{cases}$$

alors l'équation (1.3.17) devient

$$\phi(2x-k) = \begin{cases} 1 & \text{si } x \in \left[\frac{k}{2}, \frac{k+1}{2}\right), \\ 0 & \text{sinon,} \end{cases}$$

pour $k \in \mathbb{Z}$. Lorsque $k = 0$ ou $k = 1$, il en découle que

$$\phi(2x) = \begin{cases} 1 & \text{si } x \in \left[0, \frac{1}{2}\right), \\ 0 & \text{sinon,} \end{cases} \tag{1.3.18}$$

et

$$\phi(2x-1) = \begin{cases} 1 & \text{si } x \in \left[\frac{1}{2}, 1\right), \\ 0 & \text{sinon.} \end{cases} \tag{1.3.19}$$

Tout en respectant la fonction d'échelle de Haar définie par les équations (1.3.7) et (1.3.18) à (1.3.19), l'équation de dilatation se réduit à

$$\phi(x) = \phi(2x) + \phi(2x-1).$$

Par l'équation (1.3.3), nous en déduisons que la suite de filtres associée à la fonction d'échelle de Haar est

$$a_k = \begin{cases} 1 & \text{si } k = 0 \text{ ou } 1, \\ 0 & \text{sinon.} \end{cases}$$

À ce moment, nous pouvons également obtenir la mère des ondelettes associée à la base de Haar. En effet, d'après l'équation (1.3.6), nous avons

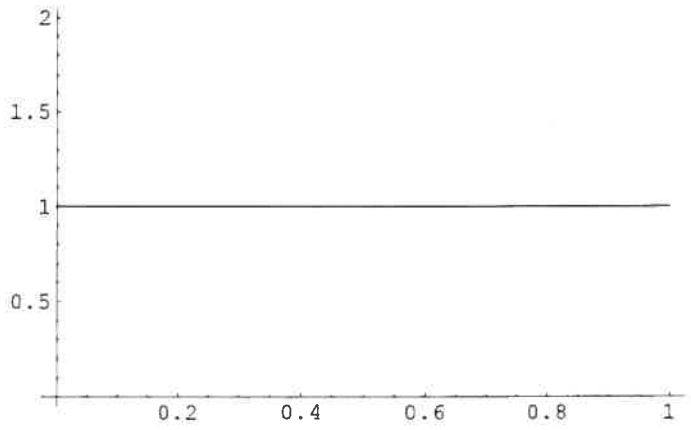
$$\begin{aligned} \psi(x) &= a_1\phi(2x) - a_0\phi(2x-1) \\ &= \phi(2x) - \phi(2x-1), \end{aligned}$$

et les filtres $\{b_k\}_{k\in\mathbb{Z}}$ associés à la mère des ondelettes sont donnés par

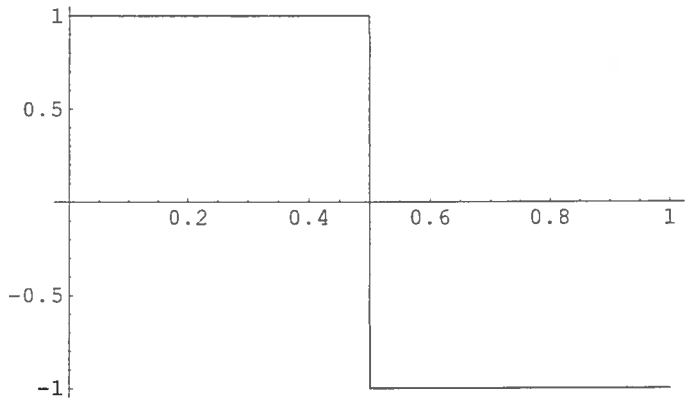
$$b_k = \begin{cases} 1 & \text{si } k = 0, \\ -1 & \text{si } k = 1, \\ 0 & \text{sinon.} \end{cases}$$

La forme explicite de ψ peut aussi être trouvée et possède la forme suivante

$$\begin{aligned} \psi(x) &= \phi(x) - \phi(2x-1) \\ &= \begin{cases} 1 & \text{si } x \in \left[0, \frac{1}{2}\right), \\ -1 & \text{si } x \in \left[\frac{1}{2}, 1\right), \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$



(a)



(b)

FIGURE 1.1. Ondelettes de Haar : (a) fonction d'échelle ϕ (b) mère des ondelettes ψ.

La figure 1.1 présente les graphes de la fonction d'échelle ϕ de Haar, ainsi que la mère des ondelettes ψ lui étant associée. Cette figure sera commentée un peu plus tard.

Pour leur part, les figures 1.2 et 1.3 présentent les fonctions $\phi_{j,k}$ et $\psi_{j,k}$ pour quelques valeurs de j et de k. Notons immédiatement que les différentes remarques que nous désirons apporter font référence à la figure 1.3, mais elles valent également pour la comparaison des fonctions $\phi_{j,k}$ entres elles. Tout d'abord, observons ce qui se produit lorsque l'indice j augmente. Pour ce faire, comparons l'allure

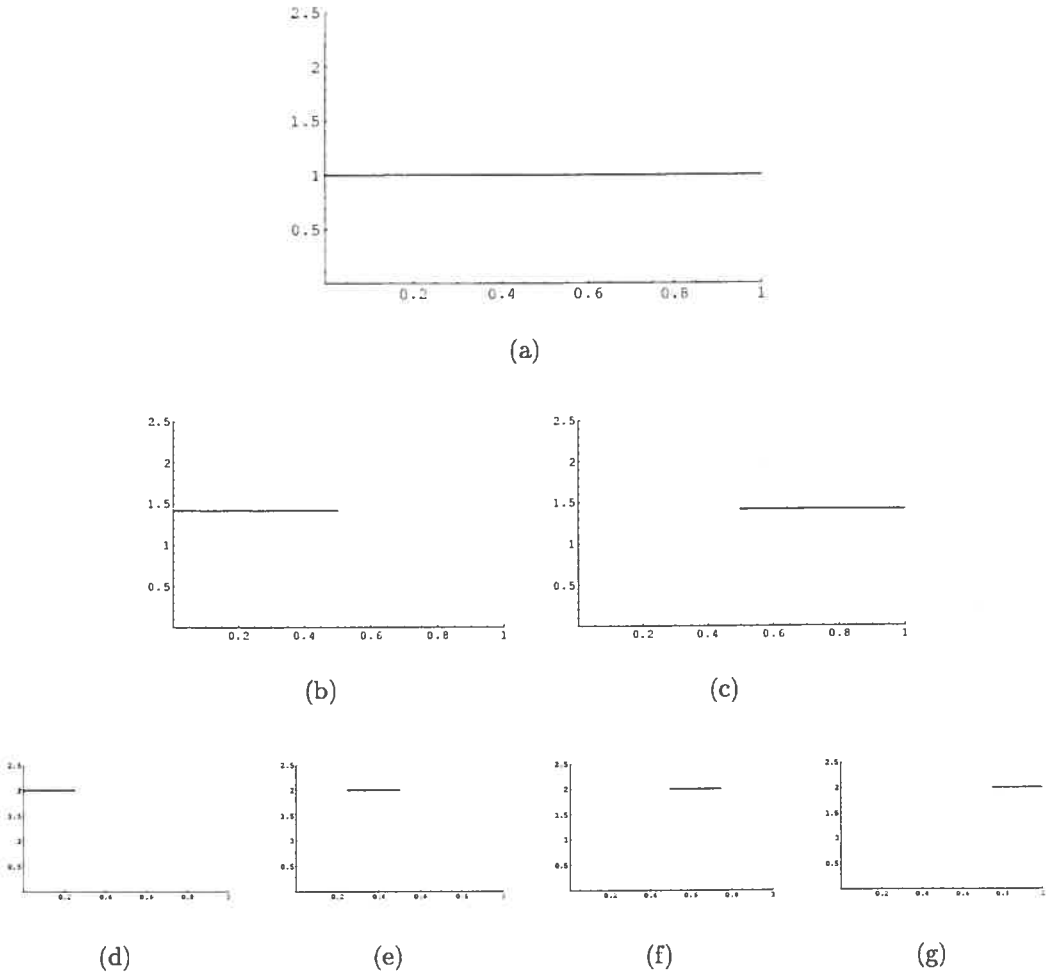


FIGURE 1.2. Ondelettes de Haar : (a) $\phi_{0,0}$ (b) $\phi_{1,0}$ (c) $\phi_{1,1}$ (d) $\phi_{2,0}$ (e) $\phi_{2,1}$ (f) $\phi_{2,2}$ (g) $\phi_{2,3}$.

des fonctions $\psi_{0,0}$ et $\psi_{1,0}$ apparaissant aux figures 1.3 (a) et 1.3 (b). D'une figure à l'autre, la forme générale des fonctions est conservée. Toutefois, comparativement au domaine de la fonction $\psi_{0,0}$, celui de $\psi_{1,0}$ est contracté. De plus, par rapport à $\psi_{0,0}$, la fonction $\psi_{1,0}$ est étirée le long de l'axe vertical. Observons maintenant ce qui se produit lorsque l'indice k augmente. Ainsi, comparons les figures 1.3 (b) et 1.3 (c) illustrant les fonctions $\psi_{1,0}$ et $\psi_{1,1}$. Nous constatons que la fonction $\psi_{1,1}$ est le résultat d'une simple translation horizontale de $\psi_{1,0}$. Ainsi, lorsque j et k augmentent, l'ensemble de ces observations devraient être combinées. Effectivement, c'est ce que nous remarquons en comparant, par exemple, les fonctions $\psi_{0,0}$

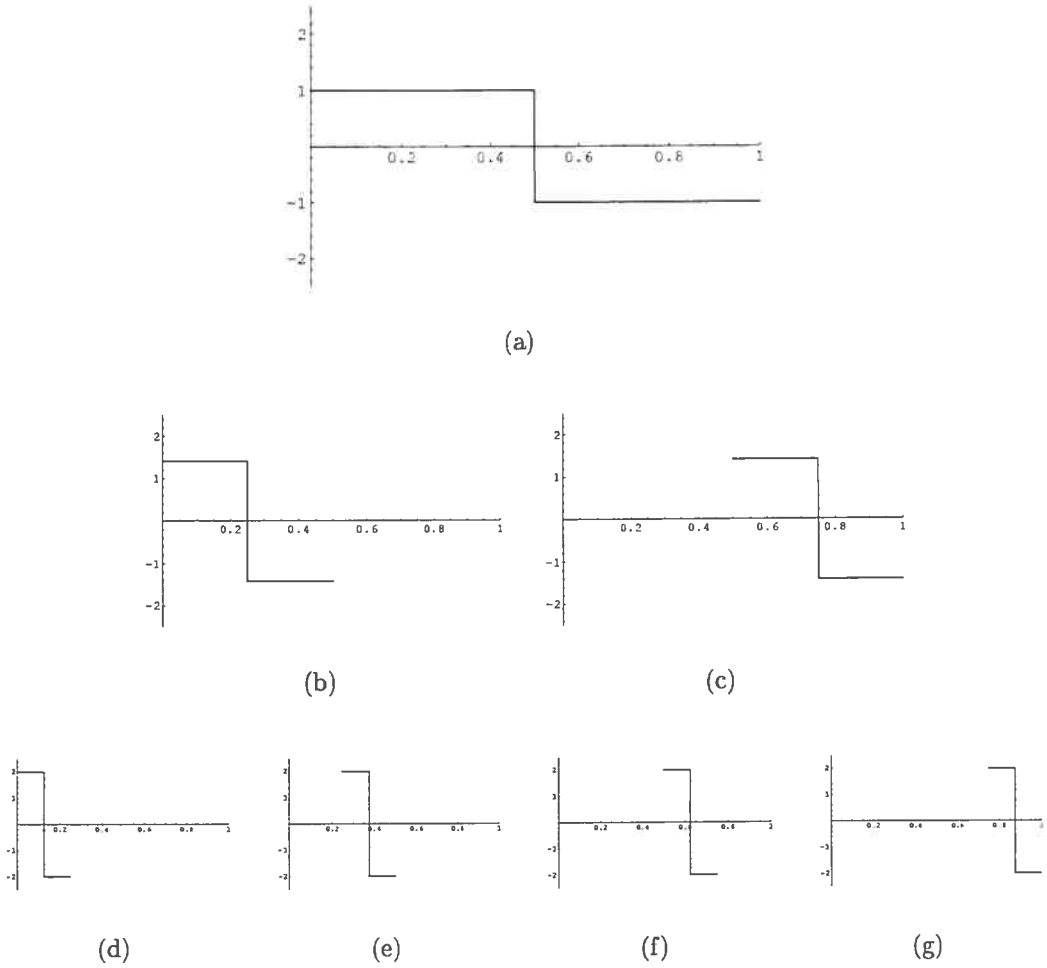


FIGURE 1.3. Ondelettes de Haar : (a) $\psi_{0,0}$ (b) $\psi_{1,0}$ (c) $\psi_{1,1}$ (d) $\psi_{2,0}$ (e) $\psi_{2,1}$ (f) $\psi_{2,2}$ (g) $\psi_{2,3}$.

et $\psi_{1,1}$ des figures 1.3 (a) et 1.3 (c). Le passage de la fonction $\psi_{0,0}$ à $\psi_{1,1}$ nécessite une contraction du domaine, un étirement vertical et une translation horizontale.

Afin d'avoir une compréhension plus approfondie du fonctionnement des fonctions $\phi_{j,k}$ et $\psi_{j,k}$, exprimons maintenant les équations (1.3.5) et (1.3.8) sous la forme suivante

$$\begin{aligned} \phi_{j,k}(x) &= 2^{j/2} \times \phi\left(2^j\left(x - \frac{k}{2^j}\right)\right), \\ \psi_{j,k}(x) &= 2^{j/2} \times \psi\left(2^j\left(x - \frac{k}{2^j}\right)\right). \end{aligned}$$

Le premier terme rencontré dans ces équations, soit $2^{j/2}$, est un *facteur de normalisation.* Celui-ci est responsable de l'étirement vertical des fonctions dont il était question ci-haut. Ensuite, le second facteur est 2^j et nous le nommons *facteur d'échelle.* Ce dernier a pour rôle la contraction du domaine sur lequel les fonctions sont définies. Finalement, le facteur $k/2^j$ est appelé *facteur de translation* (ou position) et comme son nom l'indique, il a pour rôle la translation des fonctions le long de l'axe horizontal.

Ceci termine l'illustration de la théorie vue aux sections précédentes à l'aide de la fonction d'échelle de Haar. Comme nous l'avons déjà mentionné, la forme de cette fonction est la plus simple qu'il soit possible de trouver. À des fins illustratrices, elle offre l'avantage d'alléger les calculs à effectuer. D'autre part, le plus grand désavantage des constructions obtenues à partir de la base de Haar demeure leur discontinuité. À l'heure actuelle, une multitude de fonctions d'échelles palliant à cet inconvénient existent. Parmi les plus populaires figurent celles de Daubechies, de Shannon, de Meyer et finalement celles de Franklin. Dans les pages à venir, nous présentons brièvement les bases d'ondelettes de Daubechies et nous nous limitons à celles-ci.

1.3.4. Les ondelettes de Daubechies

Daubechies fut la première à construire une base d'ondelettes dont tous les éléments orthogonaux ont un support compact et un certain ordre de régularité préétabli. La compacité des éléments découle directement d'une contrainte élaborée par Daubechies elle-même : la suite de filtres doit contenir un nombre fini d'éléments non nuls. Basée sur ce principe, elle créa toute une classe de fonctions d'échelle de différents ordres. Plus le nombre de coefficients composant la suite de filtres est grand, plus nous dirons que l'ordre est élevé.

La fonction d'échelle de Daubechies d'ordre N ($N \geq 1$) est notée ${}_N\phi(x)$. Remarquons que le choix du mot "ordre" vient du fait que les polynômes de degré inférieur à N peuvent s'exprimer directement à l'aide des translatés de ${}_N\phi(x)$. Cette fonction d'échelle possède $2N$ coefficients non nuls dans sa suite de filtres. Ces coefficients sont notés $a_0, a_1, a_2, \ldots, a_{2N-1}$ et l'équation de dilatation associée

à ${}_N\phi$ est donnée par

$$ {}_N\phi(x) = \sum_{k=0}^{2N-1} a_k \times {}_N\phi(2x-k). $$

Notons que son support est $[0, 2N-1)$ (voir Alpert, 1992). La mère des ondelettes correspondante est donnée par l'expression

$$ {}_N\psi(x) = \sum_{k=2-2N}^{1} (-1)^k a_{1-k} \times {}_N\phi(2x-k), $$

et son support est $[1-N, N)$.

Lorsque $N = 1$, il est facile de constater (voir Leblanc, 2001) que l'expression pour ${}_1\phi(x)$ nous ramène directement à la fonction d'échelle de Haar. Cela nous indique que la base d'ondelettes de Daubechies d'ordre 1 correspond tout simplement à la base de Haar. Conséquemment, les bases d'ondelettes de Daubechies d'ordre supérieur à 1 peuvent être vues comme une généralisation de la base de Haar.

La figure 1.4 montre la fonction d'échelle et la mère des ondelettes de Daubechies d'ordre 2. Ces graphiques sont obtenus à l'aide de l'*algorithme pyramidal* de Daubechies-Lagarias (ou « Daubechies-Lagarias *local pyramidal algorithm* »). Pour obtenir plus de détails au sujet de cet algorithme, nous recommandons Daubechies et Lagarias (1991, 1992) ou encore Vidakovic (1999, section 3.5.4).

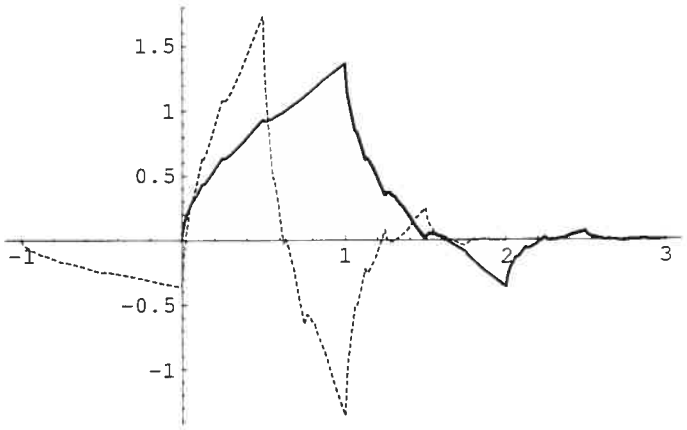


FIGURE 1.4. Ondelettes de Daubechies d'ordre 2 : fonction d'échelle ${}_2\phi$ (trait plein) et mère des ondelettes ${}_2\psi$ (pointillés).

En comparant l'allure des fonctions d'échelle de Daubechies d'ordres 1 et 2 (voir les figures 1.1 et 1.4), voici principalement ce que nous remarquons. La fonction ${}_2\phi$ apparaît plus régulière, mais moins symétrique que ${}_1\phi$. En fait, Daubechies (1992, chapitre 7) s'est intéressée à la régularité des fonctions d'échelle ${}_N\phi$, $N \geq 1$. Elle a démontré que si N est grand et si $\mathcal{C}^M(\mathbb{R})$ représente l'espace de fonctions à valeurs réelles et M fois continûment dérivables, alors ${}_N\phi \in \mathcal{C}^{\mu N}(\mathbb{R})$ où $\mu \simeq 0.2$. Par conséquent, la régularité des fonctions ${}_N\phi$ ($N \geq 1$) augmente avec l'ordre N. En ce qui a trait à la symétrie des fonctions d'échelle ${}_N\phi$, $N \geq 1$, elle montre que la fonction d'échelle de Haar, ${}_1\phi$, est la seule à posséder un support compact et à être à la fois symétrique.

1.3.5. Décomposition d'une fonction à l'aide d'ondelettes

La présente section répond à la question suivante : quel usage des bases d'ondelettes peut-on faire pour représenter une fonction $g \in \mathcal{L}_2(\mathbb{R})$? De façon générale, nous savons que si $\{w_k(x)\}_{k\in\mathbb{Z}}$ est une suite orthonormée de fonctions constituant une base de $\mathcal{L}_2(\mathbb{R})$, alors toute fonction $g \in \mathcal{L}_2(\mathbb{R})$ peut être exprimée comme

$$g(x) = \sum_{k\in\mathbb{Z}} c_k w_k(x),$$

où les coefficients c_k sont donnés par

$$c_k = < g, w_k > = \int_{\mathbb{R}} g(x) w_k(x) dx \qquad \forall k \in \mathbb{Z},$$

et où $x \in \mathcal{T}$, un sous-ensemble borné de $\mathbb{R}$. Remarquons que ce résultat est vérifié quelle que soit la base orthonormale utilisée. Appliquons-le aux bases d'ondelettes. Il est possible de démontrer (voir Leblanc, 2001) que pour n'importe quel $J \in \mathbb{Z}$, $\{\phi_{J,k}, \psi_{j,k}\}_{j\geq J, k\in\mathbb{Z}}$ constitue une base orthonormée de $\mathcal{L}_2(\mathbb{R})$.

Par conséquent, toute fonction g définie sur $\mathbb{R}$ et de carré sommable peut être décomposée en une combinaison linéaire de base d'ondelettes comme suit

$$g(x) = \sum_{k\in\mathbb{Z}} \alpha_{J,k}\phi_{J,k}(x) + \sum_{j=J}^{\infty}\sum_{k\in\mathbb{Z}} \beta_{j,k}\psi_{j,k}(x), \tag{1.3.20}$$

où les coefficients $\alpha_{J,k}$ et $\beta_{j,k}$ sont donnés par

$$\alpha_{J,k} = < g, \phi_{J,k} > = \int_{\mathbb{R}} g(x)\phi_{J,k}(x)dx,$$

et

$$\beta_{j,k} = < g, \psi_{j,k} > = \int_{\mathbb{R}} g(x)\psi_{j,k}(x)dx.$$

Parmi les caractéristiques intéressantes de certaines bases d'ondelettes, nous avons déjà mentionné la compacité de leurs éléments. En supposant l'utilisation d'une telle fonction d'ondelettes à support compact $\psi \in \mathcal{C}^{\varpi}(\mathbb{R})$, c'est-à-dire ψ est à valeurs réelles et ϖ fois différentiable, l'équation (1.3.20) devient

$$g(x) = \sum_{|k|\leq K_J} \alpha_{J,k}\phi_{J,k}(x) + \sum_{j=J}^{\infty}\sum_{|k|\leq K_j} \beta_{j,k}\psi_{j,k}(x). \tag{1.3.21}$$

Cela provient du fait qu'en supposant un support compact pour les fonctions ϕ et ψ, il existe un certain K_j, $j \geq J$, tel que $\phi_{J,k}(x)$ et $\psi_{j,k}(x)$ s'estompent lorsque $|k| > K_j$, et ce, $\forall x \in \mathcal{T}$. Ainsi, en fixant $J = 0$, l'équation (1.3.21) s'écrit

$$\begin{aligned} g(x) &= \sum_{|k|\leq K_0} \alpha_{0,k}\phi_{0,k}(x) + \sum_{j=0}^{\infty}\sum_{|k|\leq K_j} \beta_{j,k}\psi_{j,k}(x) \\ &= \sum_{|k|\leq K_0} \alpha_k\phi_k(x) + \sum_{j=0}^{\infty}\sum_{|k|\leq K_j} \beta_{j,k}\psi_{j,k}(x), \end{aligned} \tag{1.3.22}$$

où

$$\phi_k(x) = \phi(x-k),$$

et

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k).$$

Pour un niveau de résolution L spécifié, l'équation (1.3.22) devient

$$\begin{aligned} g(x) &= \sum_{|k|\leq K_0} \alpha_k\phi_k(x) + \sum_{j=0}^{L}\sum_{|k|\leq K_j} \beta_{j,k}\psi_{j,k}(x) + \sum_{j=L+1}^{\infty}\sum_{|k|\leq K_j} \beta_{j,k}\psi_{j,k}(x) \\ &= g_L(x) + R_L(x), \end{aligned}$$

où

$$g_L(x) = \sum_{|k|\leq K_0} \alpha_k\phi_k(x) + \sum_{j=0}^{L}\sum_{|k|\leq K_j} \beta_{j,k}\psi_{j,k}(x),$$

et

$$R_L(x) = \sum_{L+1}^{\infty} \sum_{|k| \leq K_j} \beta_{j,k} \psi_{j,k}(x).$$

Nous aurons l'occasion d'approfondir ces relations au prochain chapitre.

Pour l'instant, clôturons ce premier chapitre en citant quelques travaux publiés portant sur les ondelettes en estimation bayésienne non paramétrique. L'ensemble de ces publications considèrent un modèle identique ou encore très similaire à celui décrit par l'équation (1.0.1) où la fonction inconnue à estimer $g \in \mathcal{L}_2(\mathbb{R})$. Sauf si spécifié autrement, les hypothèses sont les suivantes. Tout d'abord, les t_i sont équidistants, c'est-à-dire $t_i = i/n$. Ensuite, n est un multiple de 2. Enfin, les termes d'erreur ε_i sont indépendants et normalement distribués.

Ces présupposés sont tels qu'ils permettent de passer aisément de l'espace des observations à celui des ondelettes. Il s'agit d'une particularité bien intéressante lorsque des méthodes de seuillage sont utilisées pour estimer la fonction inconnue g. Cela puisque ces méthodes requièrent une transformation des observations y_i, $i = 1, \ldots, n$, nommée la *transformation discrète d'ondelettes*, qui permet de les ramener au domaine des ondelettes. Il en résulte une série de coefficients d'ondelettes notés d_i, $i = 1, \ldots, n$. De nos jours, les différentes méthodes de seuillage sont nombreuses, mais remarquons qu'elles ont la caractéristique commune de réduire à 0 tous les coefficients d_i inférieurs, en valeur absolue, à une valeur non négative fixée à l'avance, nommée le *seuil* et notée λ. Parmi tous les articles que nous allons résumer, seul le dernier ne considère pas une méthode de seuillage pour l'estimation de g.

Mentionnons que les présupposés énumérés ci-haut peuvent apparaître contraignants, mais ils n'empêchent pas la fonction g d'être assez complexe et inhomogène dans l'espace. Cependant, il reste que dans un contexte de données réelles, ces hypothèses sont rarement toutes satisfaites.

Commençons par citer le travail de Chipman *et al.* (1997) où le modèle *a priori* correspond à un mélange de deux densités normales centrées en 0. Ces densités diffèrent par leurs variances. L'une d'entre elles est très concentrée et contribue à répartir la masse autour de 0. L'autre, plus étendue, met l'accent sur les valeurs plus éloignées. Les auteurs proposent des estimateurs de Bayes pour

les paramètres. Une étude de simulation compare les estimateurs des coefficients d'ondelettes à ceux dits « VisuShrink » et « SureShrink » introduits par Donoho et Johnstone (1994, 1995) pour quelques *fonctions tests*. Les fonctions tests sont des fonctions couramment utilisées lors des simulations pour fins de comparaison.

Une seconde contribution intéressante est celle de Donoho et Johnstone (1998). Les auteurs présentent d'abord une introduction détaillée à la décomposition d'une fonction par ondelettes. Ils montrent ensuite que les méthodes de seuillage pour les coefficients d'ondelettes peuvent atteindre le taux minimax de convergence. Finalement, une discussion portant sur la capacité qu'ont les ondelettes à saisir simultanément le comportement local et global des fonctions a lieu.

Les quatre prochains articles proviennent d'un compte-rendu de Müller et Vidakovic (1999a) portant exclusivement sur la régression bayésienne non paramétrique par bases d'ondelettes.

La publication d'Abramovich et Sapatinas (1999) utilise un modèle *a priori* composé d'une densité normale et d'un point de masse situé en 0. Les hyperparamètres sont estimés en prenant le logarithme naturel de la densité marginale des *coefficients importants*. Ces coefficients correspondent à ceux obtenus suite à l'application du « seuil universel », $\lambda = \sqrt{2\log(n)}\sigma$, décrit par Dohono et Johnstone (1994).

Les auteurs Yau et Kohn (1999) se servent d'un modèle *a priori* semblable à celui décrit par Chipman *et al.* (1997). Différentes bases sont considérées pour le lissage des fonctions : la base d'ondelettes de Haar, celle de Daubechies d'ordre 4, les symlets d'ordre 8 et finalement, les séries de Fourier. En utilisant une approche bayésienne empirique, la sélection de modèles ainsi que le moyennage de modèles sont comparées pour plusieurs fonctions tests.

L'étude menée par Müller et Vidakovic (1999b) repose sur un modèle *a priori* semblable à ceux des contributions précédentes. Cependant, une approche bayésienne hiérarchique plutôt qu'empirique est exploitée afin de déterminer les hyperparamètres. De plus, par opposition à tous les articles cités jusqu'à présent, la contrainte d'équidistance des t_i ne doit pas nécessairement être satisfaite. Une

étude de simulation basée sur la méthode MCMC est proposée et permet de faire de l'inférence *a posteriori* sur les paramètres.

Le dernier article du compte-rendu de Müller et Vidakovic (1999a) à mentionner est celui de Clyde et Georges (1999). Le résultat prédominant de cette parution est l'utilisation de l'algorithme E-M pour l'estimation des paramètres. Par l'entremise d'une étude de simulation, les estimateurs bayésiens empiriques obtenus sont comparés aux estimateurs « Hard thresholding », « SureShrink » et « Risk Inflation Criterion » respectivement introduits par Donoho et Johnstone (1994,1995) et Foster et Georges (1994) pour différentes fonctions tests.

En terminant, citons la contribution des auteurs Angers et Delampady (2001). Par opposition aux articles cités précédemment, ainsi qu'à la majorité des travaux sur le sujet, l'estimation est directement faite sur les observations plutôt qu'à partir du domaine des éléments de la base d'ondelettes. Par conséquent, l'hypothèse d'équidistance des t_i, $i = 1, \ldots, n$, ainsi que la contrainte sur n n'ont plus à êtres remplies. L'utilisation d'une approche hiérarchique mène à l'obtention de l'estimateur de Bayes. Ce dernier est ensuite comparé, au moyen de simulations, aux estimateurs « VisuShrink », « SureShrink » et à celui introduit par Chipman *et al.* (1997). Les auteurs Angers et Delampady (2001) arrivent a la conclusion suivante : leur estimateur a une performance supérieure à ceux déjà existants.

Chapitre 2

TESTS D'ÉGALITÉ POUR DEUX FONCTIONS EMPIRIQUES

Au tout début du premier chapitre, nous avons décrit le contexte d'estimation bayésienne non paramétrique dans lequel, le modèle habituellement considéré est

$$y_i = g(t_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, n$$

où g est la fonction à estimer, les t_i appartiennent à un ensemble borné $\mathcal{T} \subset \mathbb{R}$ et les ε_i représentent les termes d'erreur. Dans ce second chapitre, nous maintenons un tel contexte d'estimation, mais l'élargissons quelque peu en considérant deux jeux de données provenant d'échantillons indépendants. Les données sont observées sous la forme $\{(t_{1i}, y_{1i}), i = 1, 2, \ldots, n_1\}$ et $\{(t_{2i}, y_{2i}), i = 1, 2, \ldots, n_2\}$ où

$$y_{1i} = g_1(t_{1i}) + \varepsilon_{1i}, \qquad i = 1, 2, \ldots, n_1, \qquad t_{1i} \in \mathcal{T}, \tag{2.0.1}$$

et

$$y_{2i} = g_2(t_{2i}) + \varepsilon_{2i}, \qquad i = 1, 2, \ldots, n_2, \qquad t_{2i} \in \mathcal{T}. \tag{2.0.2}$$

Les erreurs des deux groupes sont indépendantes et normalement distribuées, c'est-à-dire $\varepsilon_{\ell i} \sim N(0, \sigma_\ell^2)$, $i = 1, \ldots, n_\ell$, $\ell = 1, 2$. Remarquons que les tailles

échantillonnales n_ℓ ne doivent pas nécessairement être égales et qu'aucune restriction d'équidistance n'est faite sur les $t_{\ell i}$.

Notre objectif est la comparaison des fonctions g_ℓ, $\ell = 1, 2$. Autrement dit, nous voulons confronter les hypothèses

$$H_0 : g_1 = g_2 \quad \text{contre} \quad H_a : g_1 \neq g_2 \tag{2.0.3}$$

sur le domaine $\mathcal{T}$.

La littérature portant sur la comparaison fonctionnelle de données au moyen d'un test d'hypothèses tel celui décrit en (2.0.3) est assez abondante. Notamment, Brillinger (1973, 1981) a développé une variété de techniques pour l'analyse fonctionnelle de données. Dans un contexte où les données se prêtent à une modélisation au moyen de séries chronologiques, mentionnons Shumway (1988). En estimation non paramétrique, les travaux de Hall et Hart (1990), Kulasekera (1995) et Koul et Schick (1997) sont à citer et de nombreuses références supplémentaires sont fournies par Fan et Lin (1998).

Pour le moment, concentrons nous sur l'estimation des fonctions g_ℓ, $\ell = 1, 2$ à l'aide des observations $\{(t_{\ell i}, y_{\ell i}), i = 1, 2, \ldots, n_\ell\}$, en utilisant la base d'ondelettes de Daubechies.

2.1. Modélisation par base d'ondelettes

Dans la présente sous-section, nous résumons la méthode d'estimation non paramétrique par base d'ondelettes proposée dans la publication de Angers et Delampady (2001). Quelques-uns des résultats déjà obtenus au chapitre précédent seront également utilisés. Afin d'alléger la notation, l'indice $\ell = 1, 2$ faisant référence à l'échantillon considéré est omis.

Précisons à nouveau que la base d'ondelettes retenue est celle de Daubechies, puisqu'elle offre l'avantage d'un support compact pour les fonctions ϕ et ψ. En faisant référence à l'équation (1.3.22), nous décomposons les fonctions à estimer des équations (2.0.1) et (2.0.2) de la façon suivante

$$g(t) = \sum_{|k| \leq K_0} \alpha_k \phi_k(t) + \sum_{j=0}^{L} \sum_{|k| \leq K_j} \beta_{j,k} \psi_{j,k}(t) + \sum_{j=L+1}^{\infty} \sum_{|k| \leq K_j} \beta_{j,k} \psi_{j,k}(t)$$

$$= g_L(t) + R_L(t), \tag{2.1.1}$$

où

$$g_L(t) = \sum_{|k|\leq K_0} \alpha_k \phi_k(t) + \sum_{j=0}^{L} \sum_{|k|\leq K_j} \beta_{j,k}\psi_{j,k}(t),$$

$$R_L(t) = \sum_{j=L+1}^{\infty} \sum_{|k|\leq K_j} \beta_{j,k}\psi_{j,k}(t),$$

et

$$\phi_k(t) = \phi(t-k),$$

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k).$$

Par l'algorithme pyramidal de Daubechies-Lagarias discuté préalablement à la sous-section 1.3.4, nous savons que les fonctions ϕ_k et $\psi_{j,k}$ sont connues. Toutefois, les coefficients α_k et $\beta_{j,k}$ demeurent inconnus et doivent être estimés.

2.1.1. Modèle *a priori*

Dans un premier temps, notons qu'au niveau de résolution L, l'égalité (2.1.1) permet d'écrire les équations (2.0.1) et (2.0.2) comme suit

$$y_i = g_L(t_i) + \eta_i + \varepsilon_i,$$

où $\eta_i = R_L(t_i)$, $i = 1, \ldots, n$. Mentionnons que le niveau de résolution correspond à la plus grande valeur de L satisfaisant la contrainte

$$l_t 2^{L+1} + L(l_\psi + 1) + (l_\phi + l_\psi + 2) \leq \min(n_1, n_2), \tag{2.1.2}$$

où l_ϕ, l_ψ et l_t dénotent respectivement la longueur du support pour ϕ, celle pour ψ et la longueur de $\mathcal{T}$. Cela provient du fait que nous pouvons montrer que le nombre total de coefficients α_k et $\beta_{j,k}$ à estimer est borné par la quantité figurant à gauche de l'inégalité (2.1.2). Lorsque cette inégalité n'est pas respectée, la matrice de design risque d'être singulière et les estimateurs de α_k et $\beta_{j,k}$ peuvent ainsi devenir fortement influencés par le choix de la distribution *a priori* (voir Angers et Delampady, 2001).

Dans le but de se définir une distribution *a priori* conjointe pour les coefficients α_k et $\beta_{j,k}$, nous faisons l'hypothèse que ces derniers sont des variables

aléatoires indépendantes normalement distribuées. Notons que l'indépendance *a priori* des coefficients α_k et $\beta_{j,k}$ ne limite pas la dépendance *a posteriori* entre ces derniers. Comme nous n'avons pas d'information relative à la moyenne de ces coefficients, nous avons choisi de fixer leur moyenne *a priori* à 0. Afin de déterminer leur variance *a priori* respectives, remarquons d'abord que dans la décomposition (2.1.1), il y a les fonctions ϕ_k discernant les caractéristiques globales de g, puis les fonctions $\psi_{j,k}$ qui identifient les détails de g. De plus, les auteurs Abramovich et Sapatinas (1999) montrent que les coefficients $\beta_{j,k}$ sont $O(2^{-2j\varpi})$. Conséquemment, nous supposons que la variance *a priori* des α_k est τ^2, alors que celle des coefficients $\beta_{j,k}$ est naturellement décroissante en j et elle vaut $\tau^2/2^{2j\varpi}$.

Définissons maintenant les vecteurs $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ suivants

$$
\begin{aligned}
\boldsymbol{\alpha} &= (\alpha_k)_{|k|\leq K_0} \\
&= (\alpha_{-K_0},\ \alpha_{-K_0+1},\ \cdots,\ \alpha_0,\ \cdots,\ \alpha_{K_0-1},\ \alpha_{K_0})^{\top}, \\
\boldsymbol{\beta} &= (\beta_{j,k})_{0\leq j\leq L, |k|\leq K_j} \\
&= \left((\beta_{0,k})_{|k|\leq K_0},\ (\beta_{1,k})_{|k|\leq K_1},\ \cdots,\ (\beta_{L,k})_{|k|\leq K_L}\right)^{\top} \\
&= (\beta_{0,-K_0},\ \cdots,\ \beta_{0,0},\ \cdots,\ \beta_{0,K_0}, \\
&\qquad \beta_{1,-K_1},\ \cdots,\ \beta_{1,0},\ \cdots,\ \beta_{1,K_1},\ \cdots,\ \beta_{L,-K_L},\ \cdots,\ \beta_{L,0},\ \cdots,\ \beta_{L,K_L})^{\top}.
\end{aligned}
$$

Ainsi, nous avons

$$
\begin{aligned}
\boldsymbol{\alpha}|\tau^2 &\sim \mathcal{N}(0, \tau^2 \boldsymbol{I}_{2K_0+1}), \\
\boldsymbol{\beta}|\tau^2 &\sim \mathcal{N}(0, \tau^2 \boldsymbol{\Delta}_{M_\beta}),
\end{aligned}
$$

où $\boldsymbol{I}_{2K_0+1}$ représente la matrice identité d'ordre $2K_0+1$, $M_\beta = \sum_{j=0}^{L}(2K_j+1)$ et $\boldsymbol{\Delta}$ est la matrice de variance-covariance associée à $\boldsymbol{\beta}$. Plus précisément, la matrice $\boldsymbol{\Delta}$ est diagonale d'ordre M_β et ses éléments diagonaux correspondent à $1/2^{2j\varpi}$, $|k| \leq K_j$, $j = 0, \ldots, L$.

Étant donné que les quantités σ^2 et τ^2 sont inconnues, nous adoptons une approche bayésienne empirique pour les estimer. Ainsi, σ^2 et τ^2 seront estimés à partir de la densité marginale des $\{(t_{1i}, y_{1i}), i = 1, 2, \ldots, n_1\}$ et $\{(t_{2i}, y_{2i}), i = 1, 2, \ldots, n_2\}$.

2.1.2. Densités *a posteriori*

À présent, définissons le vecteur de coefficients $\boldsymbol{\gamma} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$. En se basant sur les densités *a priori* des vecteurs de coefficients $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ mentionnées précédemment, nous avons

$$\boldsymbol{\gamma}|\tau^2 \sim \mathcal{N}(0, \tau^2\boldsymbol{\Gamma}),$$

où

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{I}_{2K_0+1} & 0 \\ 0 & \boldsymbol{\Delta}_{M_\beta} \end{bmatrix}.$$

De plus, la densité *a priori* du vecteur $\boldsymbol{\eta} = (\eta_1 \ \dots \ \eta_n)^\top = (R_L(t_1) \ \dots \ R_L(t_n))^\top$ est donnée par

$$\boldsymbol{\eta}|\tau^2 \sim \mathcal{N}(0, \tau^2\boldsymbol{Q}_n),$$

où la matrice $\boldsymbol{Q}_n$ dépend de la structure de variance des $(\beta_{j,k})_{j\geq L+1, |k|\leq K_j}$. Étant donné que les $\beta_{j,k}$ sont supposés indépendants, l'élément (i,l) de $\boldsymbol{Q}_n$ est donné par

$$\begin{aligned}
(\boldsymbol{Q}_n)_{i,l} &= \tau^{-2}\mathbb{C}\mathrm{ov}(\eta_i, \eta_l) \\
&= \tau^{-2}\mathbb{C}\mathrm{ov}\left(\sum_{j\geq L+1}\sum_{|k|\leq K_j}\beta_{j,k}\psi_{j,k}(t_i), \sum_{p\geq L+1}\sum_{|q|\leq K_p}\beta_{p,q}\psi_{p,q}(t_l)\right) \\
&= \tau^{-2}\sum_{j\geq L+1}\sum_{|k|\leq K_j}\sum_{p\geq L+1}\sum_{|q|\leq K_p}\mathbb{C}\mathrm{ov}(\beta_{j,k}, \beta_{p,q})\psi_{j,k}(t_i)\psi_{p,q}(t_l) \\
&= \tau^{-2}\sum_{j\geq L+1}\sum_{|k|\leq K_j}\mathbb{V}\mathrm{ar}(\beta_{j,k})\psi_{j,k}(t_i)\psi_{j,k}(t_l) \\
&= \sum_{j\geq L+1}\sum_{|k|\leq K_j} 2^{-2j\varpi}\psi_{j,k}(t_i)\psi_{j,k}(t_l).
\end{aligned}$$

Puisque la fonction ψ est bornée, nous pouvons aussi montrer que tout élément (i,l) de $\boldsymbol{Q}_n$ est lui-même borné par

$$|(\boldsymbol{Q}_n)_{i,l}| \leq \frac{(l_\psi + 1)\left(\max_z \psi^2(z)\right)}{2^{2L\varpi}(2^{2\varpi} - 1)}.$$

Par conséquent, nous sommes assurés que la matrice $\boldsymbol{Q}_n$ est bien définie. Les auteurs Angers et Delampady (2001) présentent une analyse de sensibilité pour le choix de $\boldsymbol{Q}_n$ et ils démontrent qu'un tel choix n'a pas d'influence majeure sur le lissage par ondelettes.

Définissons maintenant la matrice des ondelettes $\boldsymbol{T} = (\boldsymbol{\Phi}^\top, \boldsymbol{S}^\top)$ où

$$\boldsymbol{\Phi}^\top = \begin{bmatrix} \phi_{-K_0}(t_1) & \dots & \phi_0(t_1) & \dots & \phi_{K_0}(t_1) \\ \phi_{-K_0}(t_2) & \dots & \phi_0(t_2) & \dots & \phi_{K_0}(t_2) \\ \vdots & & & & \\ \phi_{-K_0}(t_n) & \dots & \phi_0(t_n) & \dots & \phi_{K_0}(t_n) \end{bmatrix},$$

et

$$\boldsymbol{S}^\top = \begin{bmatrix} (\psi_{0,k}(t_1))_{|k|\leq K_0}^\top & (\psi_{1,k}(t_1))_{|k|\leq K_1}^\top & \dots & (\psi_{L,k}(t_1))_{|k|\leq K_L}^\top \\ (\psi_{0,k}(t_2))_{|k|\leq K_0}^\top & (\psi_{1,k}(t_2))_{|k|\leq K_1}^\top & \dots & (\psi_{L,k}(t_2))_{|k|\leq K_L}^\top \\ \vdots & & & \\ (\psi_{0,k}(t_n))_{|k|\leq K_0}^\top & (\psi_{1,k}(t_n))_{|k|\leq K_1}^\top & \dots & (\psi_{L,k}(t_n))_{|k|\leq K_L}^\top \end{bmatrix}.$$

Ainsi, nous pouvons décrire les observations $\boldsymbol{y} = (y_1, y_2, \dots, y_n)^\top$ à l'aide du modèle linéaire suivant

$$\boldsymbol{y} = \boldsymbol{T}\boldsymbol{\gamma} + \boldsymbol{u},$$

où $\boldsymbol{u} = \boldsymbol{\eta} + \boldsymbol{\varepsilon}$. Étant donné que

$$\begin{aligned} \boldsymbol{y}|\boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma^2, \tau^2 &\sim \mathcal{N}(\boldsymbol{T}\boldsymbol{\gamma} + \boldsymbol{\eta}, \sigma^2 \boldsymbol{I}_n), \\ \boldsymbol{\eta}|\tau^2 &\sim \mathcal{N}(0, \tau^2 \boldsymbol{Q}_n), \end{aligned} \tag{2.1.3}$$

alors $\boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{\Xi})$, où $\boldsymbol{\Xi} = \tau^2 \boldsymbol{Q}_n + \sigma^2 \boldsymbol{I}_n$. Remarquons que le vecteur des observations $\boldsymbol{y}$ peut aussi être exprimé de façon plus détaillée par

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{bmatrix} \sum_{|k|\leq K_0} \alpha_k \phi_k(t_1) + \sum_{j=0}^{L} \sum_{|k|\leq K_j} \beta_{j,k} \psi_{j,k}(t_1) + u_1 \\ \sum_{|k|\leq K_0} \alpha_k \phi_k(t_2) + \sum_{j=0}^{L} \sum_{|k|\leq K_j} \beta_{j,k} \psi_{j,k}(t_2) + u_2 \\ \vdots \\ \sum_{|k|\leq K_0} \alpha_k \phi_k(t_n) + \sum_{j=0}^{L} \sum_{|k|\leq K_j} \beta_{j,k} \psi_{j,k}(t_n) + u_n \end{bmatrix}.$$

Combinée à l'utilisation de modèles hiérarchiques bayésiens (voir Lindley et Smith, 1972) et d'identités matricielles (voir Searle, 1982, chapitre 5), l'équation (2.1.3) nous permet d'écrire

$$\begin{aligned} \boldsymbol{y}|\sigma^2,\tau^2 &\sim \mathcal{N}(0,\sigma^2\boldsymbol{I}_n+\tau^2[\boldsymbol{T}\Gamma\boldsymbol{T}^\top+\boldsymbol{Q}_n]), \\ \boldsymbol{\gamma}|\boldsymbol{y},\sigma^2,\tau^2 &\sim \mathcal{N}(\boldsymbol{AY},\boldsymbol{B}), \end{aligned} \tag{2.1.4}$$

où

$$\begin{aligned} \boldsymbol{A} &= \tau^2\Gamma\boldsymbol{T}^\top(\sigma^2\boldsymbol{I}_n+\tau^2[\boldsymbol{T}\Gamma\boldsymbol{T}^\top+\boldsymbol{Q}_n])^{-1}, \\ \boldsymbol{B} &= \tau^2\Gamma-\tau^4\Gamma\boldsymbol{T}^\top(\sigma^2\boldsymbol{I}_n+\tau^2[\boldsymbol{T}\Gamma\boldsymbol{T}^\top+\boldsymbol{Q}_n])^{-1}\boldsymbol{T}\Gamma. \end{aligned}$$

Afin d'arriver à estimer σ^2 et τ^2, encore quelques simplifications algébriques sont nécessaires. Comme $\boldsymbol{T}\Gamma\boldsymbol{T}^\top+\boldsymbol{Q}_n$ est une matrice symétrique définie positive, nous avons par décomposition spectrale

$$\boldsymbol{T}\Gamma\boldsymbol{T}^\top+\boldsymbol{Q}_n=\boldsymbol{HDH}^\top, \tag{2.1.5}$$

où $\boldsymbol{D}$ est une matrice diagonale dont les éléments diagonaux $(d_1,\ldots,d_n)$, $d_i>0,\ \forall\ i=1,\ldots,n$, correspondent aux valeurs propres de $\boldsymbol{T}\Gamma\boldsymbol{T}^\top+\boldsymbol{Q}_n$ et $\boldsymbol{H}$ est la matrice des vecteurs propres normalisés correspondants. Ainsi, nous avons les égalités suivantes

$$\boldsymbol{HH}^\top = \boldsymbol{I}_n, \tag{2.1.6}$$

$$\boldsymbol{H}^{-1} = \boldsymbol{H}^\top, \tag{2.1.7}$$

$$(\boldsymbol{H}^\top)^{-1} = \boldsymbol{H}. \tag{2.1.8}$$

En utilisant les équations (2.1.5) et (2.1.6), puis en posant $\nu=\sigma^2/\tau^2$,

$$\begin{aligned} \sigma^2\boldsymbol{I}_n+\tau^2(\boldsymbol{T}\Gamma\boldsymbol{T}^\top+\boldsymbol{Q}_n) &= \tau^2(\nu\boldsymbol{I}_n+\boldsymbol{HDH}^\top) \\ &= \tau^2(\nu\boldsymbol{HI}_n\boldsymbol{H}^\top+\boldsymbol{HDH}^\top) \\ &= \tau^2\boldsymbol{H}(\nu\boldsymbol{I}_n+\boldsymbol{D})\boldsymbol{H}^\top. \end{aligned}$$

D'après cette dernière décomposition, la densité marginale des observations conditionnellement à σ^2 et τ^2 décrite par (2.1.4) devient

$$\boldsymbol{y}|\tau^2, \nu \sim \mathcal{N}(0, \tau^2 \boldsymbol{H}(\nu \boldsymbol{I}_n + \boldsymbol{D})\boldsymbol{H}^\top).$$

Plus précisément, en posant $\boldsymbol{s} = (s_1, \ldots, s_n)^\top = \boldsymbol{H}^\top \boldsymbol{y}$ et en utilisant les égalités (2.1.7) et (2.1.8), nous pouvons écrire

$$\begin{aligned} m(\boldsymbol{y}|\tau^2, \nu) &= \frac{1}{(2\pi\tau^2)^{n/2}} \frac{1}{\det(\nu \boldsymbol{I}_n + \boldsymbol{D})^{1/2}} \exp\left\{-\frac{1}{2\tau^2} \boldsymbol{y}^\top \boldsymbol{H}(\nu \boldsymbol{I}_n + \boldsymbol{D})^{-1} \boldsymbol{H}^\top \boldsymbol{y}\right\} \\ &= \frac{1}{(2\pi\tau^2)^{n/2}} \frac{1}{\prod_{i=1}^n (\nu + d_i)^{1/2}} \exp\left\{-\frac{1}{2\tau^2} \sum_{i=1}^n \frac{s_i^2}{\nu + d_i}\right\}. \end{aligned} \tag{2.1.9}$$

À présent, nous sommes en mesure d'estimer σ^2 et τ^2 par l'entremise d'une approche bayésienne empirique en se servant de l'équation (2.1.9). En effet, il suffit de suivre les étapes décrites ci-dessous.

(1) Résoudre

$$\sum_{i=1}^n \frac{s_i^2}{(\nu + d_i)^2} \left[\sum_{j=1}^n \left(\frac{d_j - d_i}{\nu + d_j}\right)\right] = 0$$

par rapport à $\nu = \sigma^2/\tau^2$ et dénoter la solution obtenue par $\widehat{\nu}$.

(2) Poser

$$\begin{aligned} \widehat{\tau}^2 &= n^{-1} \sum_{i=1}^n s_i^2/(\widehat{\nu} + d_i), \\ \widehat{\sigma}^2 &= \widehat{\nu}\widehat{\tau}^2. \end{aligned}$$

En remplaçant σ^2 et τ^2 par leurs estimateurs $\widehat{\sigma}^2$ et $\widehat{\tau}^2$ dans le modèle *a priori*, les densités marginale et *a posteriori* deviennent respectivement

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \widehat{\tau}^2 \boldsymbol{H}[\widehat{\nu} \boldsymbol{I}_n + \boldsymbol{D}]\boldsymbol{H}^\top), \tag{2.1.10}$$

$$\begin{aligned} \boldsymbol{\gamma}|\boldsymbol{y} \sim \mathcal{N}(&\boldsymbol{\Gamma}\boldsymbol{T}^\top \boldsymbol{H}[\widehat{\nu} \boldsymbol{I}_n + \boldsymbol{D}]^{-1} \boldsymbol{H}^\top \boldsymbol{y}, \\ &\widehat{\tau}^2(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{T}^\top \boldsymbol{H}[\widehat{\nu} \boldsymbol{I}_n + \boldsymbol{D}]^{-1} \boldsymbol{H}^\top \boldsymbol{T}\boldsymbol{\Gamma})). \end{aligned} \tag{2.1.11}$$

Sous la fonction de perte quadratique, l'estimateur de Bayes non paramétrique de la fonction g est donné par

$$\begin{aligned}\widehat{g}(t) &= \sum_{|k|\leq K_0} \widehat{\alpha}_k \phi_k(t) + \sum_{j=0}^{L} \sum_{|k|\leq K_j} \widehat{\beta}_{j,k} \psi_{j,k}(t) \\ &= \boldsymbol{D}(t)\widehat{\boldsymbol{\gamma}},\end{aligned}$$

où

$$\begin{aligned}\boldsymbol{D}(t) &= \left((\phi_k(t))_{|k|\leq K_0}^{\top}, (\psi_{j,k}(t))_{0\leq j\leq L, |k|\leq K_j}^{\top}\right), \\ \widehat{\boldsymbol{\gamma}} &= \boldsymbol{\Gamma}\boldsymbol{T}^{\top}\boldsymbol{H}(\widehat{\nu}\boldsymbol{I}_n + \boldsymbol{D})^{-1}\boldsymbol{H}^{\top}\boldsymbol{y}.\end{aligned}$$

2.1.3. Distance observée entre deux fonctions

Supposons maintenant que les fonctions g_ℓ, $\ell = 1, 2$ figurant aux expressions (2.0.1) et (2.0.2) sont estimées par

$$\begin{aligned}\widehat{g}_\ell(t) &= \sum_{|k|\leq K_0} \widehat{\alpha}_{(\ell)k} \phi_k(t) + \sum_{j=0}^{L} \sum_{|k|\leq K_j} \widehat{\beta}_{(\ell)j,k} \psi_{j,k}(t) \\ &= \boldsymbol{D}(t)\widehat{\boldsymbol{\gamma}}_\ell,\end{aligned}$$

où

$$\widehat{\boldsymbol{\gamma}}_\ell = \boldsymbol{\Gamma}\boldsymbol{T}_\ell^{\top}\boldsymbol{H}_\ell(\widehat{\nu}_\ell\boldsymbol{I}_{n_\ell} + \boldsymbol{D}_\ell)^{-1}\boldsymbol{H}_\ell^{\top}\boldsymbol{y}_\ell.$$

Précisons que les quantités $\boldsymbol{T}_\ell, \boldsymbol{H}_\ell, \widehat{\nu}_\ell, \boldsymbol{D}_\ell$ et $\boldsymbol{y}_\ell$ font référence à celles définies à la sous-section précédente et qu'elles sont calculées à partir des observations du ℓ^{e} jeu de données.

Par conséquent, la distance observée, au sens de $\mathcal{L}_2$, entre les fonctions $\widehat{g}_1$ et $\widehat{g}_2$ est donnée par

$$\begin{aligned}d(\widehat{\boldsymbol{\gamma}}_1, \widehat{\boldsymbol{\gamma}}_2) &= \int_{\mathbb{R}} \|\widehat{g}_1(t) - \widehat{g}_2(t)\|^2 dt \\ &= \int_{\mathbb{R}} \Bigg\| \sum_{|k|\leq K_0} \widehat{\alpha}_{(1)k} \phi_k(t) + \sum_{j=0}^{L} \sum_{|k|\leq K_j} \widehat{\beta}_{(1)j,k} \psi_{j,k}(t) \\ &\qquad - \left(\sum_{|k|\leq K_0} \widehat{\alpha}_{(2)k} \phi_k(t) + \sum_{j=0}^{L} \sum_{|k|\leq K_j} \widehat{\beta}_{(2)j,k} \psi_{j,k}(t) \right) \Bigg\|^2 dt\end{aligned}$$

$$= \int_{\mathbb{R}} \Big\| \sum_{|k|\leq K_0} (\widehat{\alpha}_{(1)k} - \widehat{\alpha}_{(2)k})\phi_k(t) + \sum_{j=0}^{L} \sum_{|k|\leq K_j} (\widehat{\beta}_{(1)j,k} - \widehat{\beta}_{(2)j.k})\psi_{j,k}(t) \Big\|^2 dt. \quad (2.1.12)$$

Puisque $\{\phi_k, \psi_{j,k}\}_{j\geq 0, k\in\mathbb{Z}}$ est une base orthonormée de $\mathcal{L}_2(\mathbb{R})$, alors l'équation (2.1.12) devient

$$\begin{aligned} d(\widehat{\gamma}_1, \widehat{\gamma}_2) &= \sum_{|k|\leq K_0} (\widehat{\alpha}_{(1)k} - \widehat{\alpha}_{(2)k})^2 \int_{\mathbb{R}} \|\phi_k(t)\|^2 dt \\ &\quad + \sum_{j=0}^{L} \sum_{|k|\leq K_j} (\widehat{\beta}_{(1)j,k} - \widehat{\beta}_{(2)j.k})^2 \int_{\mathbb{R}} \|\psi_{j,k}(t)\|^2 dt \\ &= \sum_{|k|\leq K_0} (\widehat{\alpha}_{(1)k} - \widehat{\alpha}_{(2)k})^2 + \sum_{j=0}^{L} \sum_{|k|\leq K_j} (\widehat{\beta}_{(1)j,k} - \widehat{\beta}_{(2)j,k})^2 \quad (2.1.13) \\ &= (\widehat{\gamma}_1 - \widehat{\gamma}_2)^\top (\widehat{\gamma}_1 - \widehat{\gamma}_2). \quad (2.1.14) \end{aligned}$$

Ceci met fin à la première partie de ce second chapitre. À présent, décrivons les différents tests pouvant être utilisés afin de confronter les hypothèses H_0 et H_a décrites par (2.0.3).

2.2. Test basé sur le facteur de Bayes

Le premier test que nous présentons est basé sur le facteur de Bayes. Commençons par définir ce qu'est un facteur de Bayes.

Définition 2.2.1. *Soit les hypothèses suivantes :*

$$H_i : \theta \in \Theta_i \quad \text{contre} \quad H_j : \theta \in \Theta_j,$$

où $\Theta_i \cap \Theta_j = \emptyset$. *Ainsi, nous appelons facteur de Bayes le rapport*

$$B_{ij}(x) = \frac{\Pr\{\theta \in \Theta_i | x\}}{\Pr\{\theta \in \Theta_j | x\}} \Big/ \frac{\Pr\{\theta \in \Theta_i\}}{\Pr\{\theta \in \Theta_j\}}.$$

Ce rapport évalue donc la modification de la vraisemblance relative de l'hypothèse i par rapport à l'hypothèse j qui est due aux observations et il se compare

à la valeur 1. Par exemple, si nous obtenons

$$B_{ij}(x) > 1 \Longleftrightarrow \frac{\Pr\{\theta \in \Theta_i | x\}}{\Pr\{\theta \in \Theta_i\}} > \frac{\Pr\{\theta \in \Theta_j | x\}}{\Pr\{\theta \in \Theta_j\}},$$

alors cela signifie que les observations augmentent la cote de $H_i : \theta \in \Theta_i$ comparativement à celle de $H_j : \theta \in \Theta_j$. Par conséquent, nous devrions accepter l'hypothèse H_i.

Dans la situation où les hypothèses H_i et H_j correspondent aux hypothèses H_0 et H_a décrites en (2.0.3), nous pouvons montrer que le facteur de Bayes se réduit à

$$B(\boldsymbol{y}_1, \boldsymbol{y}_2) = \frac{m_0(\boldsymbol{y}_1, \boldsymbol{y}_2)}{m_1(\boldsymbol{y}_1, \boldsymbol{y}_2)},$$

où $m_0(\boldsymbol{y}_1, \boldsymbol{y}_2)$ et $m_1(\boldsymbol{y}_1, \boldsymbol{y}_2)$ représentent respectivement les marginales sous H_0 et H_a. Ces marginales s'obtiennent à partir de (2.1.10) (voir Angers, 2003).

Dans un contexte bayésien, nous préférons habituellement utiliser un test basé sur le facteur de Bayes pour effectuer un test d'hypothèse. Cependant, dans notre situation, Angers (2003) montre que ce test peut être difficile à évaluer étant donné qu'il requiert la décomposition spectrale d'une matrice carrée de dimension $(n_1 + n_2)$. Ainsi, nous ne passons pas plus de temps sur le test de Bayes et proposons plutôt diverses alternatives plus rapides à exécuter et qui ne nécessitent pas autant de puissance informatique.

2.3. Tests basés sur l'approximation de la distribution de la distance entre deux fonctions empiriques

Les trois prochaines méthodes tentent de pallier aux inconvénients rencontrés avec l'approche du test de Bayes. Comme le titre de la sous-section l'indique, ces méthodes reposent toutes sur l'approximation de la distribution de la distance entre les deux fonctions empiriques.

À la sous-section 2.1.3, nous avons vu que la distance observée entre $\widehat{g}_1$ et $\widehat{g}_2$ est donnée par l'une ou l'autre des équations (2.1.13) et (2.1.14). En utilisant l'équation (2.1.1), nous pouvons aussi définir la distance, au sens de $\mathcal{L}_2$, entre les

fonctions g_1 et g_2 (voir Zhao, 2000) par

$$\begin{aligned} d(g_1, g_2) &= \int_{\mathbb{R}} \|g_1(t) - g_2(t)\|^2 dt \\ &= \int_{\mathbb{R}} \Big\| \sum_{|k|\leq K_0} \alpha_{(1)k}\phi_k(t) + \sum_{j\geq 0}\sum_{|k|\leq K_j} \beta_{(1)j,k}\psi_{j,k}(t) \\ &\qquad - \sum_{|k|\leq K_0} \alpha_{(2)k}\phi_k(t) - \sum_{j\geq 0}\sum_{|k|\leq K_j} \beta_{(2)j,k}\psi_{j,k}(t) \Big\|^2 dt \\ &= \int_{\mathbb{R}} \Big\| \sum_{|k|\leq K_0} (\alpha_{(1)k} - \alpha_{(2)k})\phi_k(t) \\ &\qquad + \sum_{j\geq 0}\sum_{|k|\leq K_j} (\beta_{(1)j,k} - \beta_{(2)j,k})\psi_{j,k}(t) \Big\|^2 dt. \end{aligned} \tag{2.3.1}$$

Une fois de plus, étant donné que $\{\phi_k, \psi_{j,k}\}_{j\geq 0, k\in\mathbb{Z}}$ constitue une base orthonormée de $\mathcal{L}_2(\mathbb{R})$, l'équation (2.3.1) devient

$$\begin{aligned} d(g_1, g_2) &= \sum_{|k|\leq K_0} (\alpha_{(1)k} - \alpha_{(2)k})^2 + \sum_{j\geq 0}\sum_{|k|\leq K_j} (\beta_{(1)j,k} - \beta_{(2)j,k})^2 \\ &= d(\gamma_1, \gamma_2) + \sum_{j\geq L+1}\sum_{|k|\leq K_j} (\beta_{(1)j,k} - \beta_{(2)j,k})^2 \\ &= d(\gamma_1, \gamma_2) + O(2^{-2(L+1)s}). \end{aligned} \tag{2.3.2}$$

Ainsi, la quantité $d(\gamma_1, \gamma_2)$ semble représenter une bonne approximation de la distance entre les fonctions g_1 et g_2. Notons $d(\gamma_1, \gamma_2) = \xi$ et mentionnons que notre objectif est donc d'approximer la distribution d'une telle quantité. La première méthode d'approximation présentée repose sur un développement en séries d'Edgeworth. Les méthodes suivantes consistent toutes les deux à exprimer ξ sous une forme quadratique puis à approximer la distribution de celle-ci par la suite.

2.3.1. Développement en séries d'Edgeworth

Le développement en séries d'Edgeworth est une approximation asymptotique de la fonction de densité ou de la fonction de distribution reposant sur la loi normale. Amorçons cette sous-section en énonçant quelques concepts et définitions nécessaires à la compréhension d'un tel développement.

Définition 2.3.1. *Pour tout réel t, la fonction génératrice des moments de la variable aléatoire X continue de densité f, notée $M(t, X)$, est définie par*

$$\begin{aligned} M(t, X) &= \mathbb{E}_f(e^{tX}) \\ &= \int_{-\infty}^{\infty} e^{tx} f(x) dx. \end{aligned}$$

Définition 2.3.2. *Pour tout réel t, la fonction génératrice des cumulants de la variable aléatoire X continue de densité f, notée $C(t, X)$, est définie par*

$$C(t, X) = \log(M(t, X)).$$

Définition 2.3.3. *Pour tout réel t, la fonction caractéristique (ou la transformée de Fourier) de la variable aléatoire X continue de densité f, notée $\varrho(t, X)$, est définie par*

$$\varrho(t, X) = M(it, X) \qquad \text{où} \quad i = \sqrt{-1}.$$

Définition 2.3.4. *Les polynômes d'Hermite $H_n(x)$, $n \geq 1$, peuvent être obtenus à l'aide de la formule*

$$H_n(x) = (-1)^n e^{\frac{1}{2}x^2} \frac{d^n}{dx^n} e^{-\frac{1}{2}x^2}.$$

Les polynômes d'Hermite $H_n(x)$, $n = 1, \ldots, 6$ sont donnés par

$$\begin{aligned} H_1(x) &= x, \\ H_2(x) &= x^2 - 1, \\ H_3(x) &= x^3 - 3x, \\ H_4(x) &= x^4 - 6x^2 + 3, \\ H_5(x) &= x^5 - 10x^3 + 15x, \\ H_6(x) &= x^6 - 15x^4 + 45x^2 - 15. \end{aligned}$$

La combinaison de ces différentes notions nous permet d'obtenir un développement en séries d'Edgeworth (voir Lange, 1998, section 17.5) pouvant être utilisé pour approximer la fonction de densité f de toute variable aléatoire X continue.

Cette approximation s'écrit

$$f(x) \cong \varphi(x)\left[1 + \frac{\rho_3}{6}H_3(x) + \frac{\rho_4}{24}H_4(x) + \frac{\rho_3^2}{72}H_6(x)\right], \tag{2.3.3}$$

où $\varphi(\cdot)$ représente la fonction de densité d'une normale centrée et réduite et les ρ_j, $j = 3, 4$ correspondent aux cumulants standardisés (une définition plus élaborée est fournie à la page suivante). En intégrant l'équation (2.3.3), nous obtenons l'expression suivante pour la fonction de répartition F de X

$$\begin{aligned} F(x) &\cong \int_{-\infty}^{x}\left[\varphi(u) + \frac{\rho_3}{6}\varphi(u)H_3(u) + \frac{\rho_4}{24}\varphi(u)H_4(u) + \frac{\rho_3^2}{72}\varphi(u)H_6(u)\right]du \\ &= \Phi(x) + \frac{\rho_3}{6}\int_{-\infty}^{x}\varphi(u)H_3(u)du \\ &\quad + \frac{\rho_4}{24}\int_{-\infty}^{x}\varphi(u)H_4(u)du + \frac{\rho_3^2}{72}\int_{-\infty}^{x}\varphi(u)H_6(u)du, \end{aligned} \tag{2.3.4}$$

où $\Phi(\cdot)$ représente la fonction de répartition d'une normale centrée et réduite. Le théorème suivant nous permet de simplifier l'équation (2.3.4) davantage.

Théorème 2.3.1. *Si φ représente la fonction de densité d'une variable aléatoire $X \sim \mathcal{N}(0,1)$ et H_n le polynôme d'Hermite d'ordre n, alors*

$$\int_{-\infty}^{x}\varphi(u)H_n(u)du = -\varphi(x)H_{n-1}(x).$$

En utilisant le résultat de ce théorème, l'équation (2.3.4) devient

$$\begin{aligned} F(x) &\cong \Phi(x) - \frac{\rho_3}{6}\varphi(x)H_2(x) - \frac{\rho_4}{24}\varphi(x)H_3(x) - \frac{\rho_3^2}{72}\varphi(x)H_5(x) \\ &= \Phi(x) - \varphi(x)\left[\frac{\rho_3}{6}H_2(x) + \frac{\rho_4}{24}H_3(x) + \frac{\rho_3^2}{72}H_5(x)\right]. \end{aligned} \tag{2.3.5}$$

Les équations (2.3.3) et (2.3.5) peuvent ainsi être utilisées pour approximer la densité et la fonction de répartition de la variable aléatoire $d(\gamma_1, \gamma_2) = \xi$ de l'équation (2.3.2). Nous pouvons écrire

$$\begin{aligned} \xi &= \sum_{|k|\leq k_o}(\alpha_{(1)k} - \alpha_{(2)k})^2 + \sum_{j=0}^{L}\sum_{|k|\leq k_j}(\beta_{(1)j,k} - \beta_{(2)j,k})^2 \\ &= (\gamma_1 - \gamma_2)^\top(\gamma_1 - \gamma_2). \end{aligned} \tag{2.3.6}$$

De plus, d'après l'équation (2.1.11), nous avons

$$(\gamma_1 - \gamma_2) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

où

$$\begin{aligned} \boldsymbol{\mu} &= \boldsymbol{A}_1\boldsymbol{y}_1 - \boldsymbol{A}_2\boldsymbol{y}_2, \\ \boldsymbol{\Sigma} &= \boldsymbol{B}_1 + \boldsymbol{B}_2, \end{aligned}$$

et pour $\ell = 1, 2$

$$\begin{aligned} \boldsymbol{A}_\ell &= \boldsymbol{\Gamma}\boldsymbol{T}_\ell^\top \boldsymbol{H}_\ell[\widehat{\nu}_\ell \boldsymbol{I}_{n_\ell} + \boldsymbol{D}_\ell]^{-1}\boldsymbol{H}_\ell^\top \boldsymbol{y}_\ell, \\ \boldsymbol{B}_\ell &= \widehat{\tau}_\ell^2(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{T}_\ell^\top \boldsymbol{H}_\ell[\widehat{\nu}_\ell \boldsymbol{I}_{n_\ell} + \boldsymbol{D}_\ell]^{-1}\boldsymbol{H}_\ell^\top \boldsymbol{T}_\ell\boldsymbol{\Gamma}). \end{aligned}$$

Ainsi, le r^{e} cumulant *a posteriori* (voir Searle, 1971, p.55) de ξ peut facilement être calculé comme suit

$$\kappa_r(\boldsymbol{y}_1, \boldsymbol{y}_2) = 2^{r-1}(r-1)![\text{Tr}(\boldsymbol{\Sigma}^r) + r\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{r-1}\boldsymbol{\mu}],$$

où $\text{Tr}(\boldsymbol{\Sigma}^r)$ représente la trace de $\boldsymbol{\Sigma}^r$, c'est-à-dire la somme des éléments diagonaux de $\boldsymbol{\Sigma}^r$. Les équations (2.3.3) et (2.3.5) peuvent alors s'écrire

$$\begin{aligned} f(\xi|\boldsymbol{y}_1, \boldsymbol{y}_2) &\cong \varphi\left(\frac{\xi - a}{b}\right)\left[1 + \frac{\rho_3}{6}H_3\left(\frac{\xi - a}{b}\right)\right. \\ &\qquad \left. + \frac{\rho_4}{24}H_4\left(\frac{\xi - a}{b}\right) + \frac{\rho_3^2}{72}H_6\left(\frac{\xi - a}{b}\right)\right], \\ F(\xi|\boldsymbol{y}_1, \boldsymbol{y}_2) &\cong \Phi\left(\frac{\xi - a}{b}\right) - \varphi\left(\frac{\xi - a}{b}\right)\left[\frac{\rho_3}{6}H_2\left(\frac{\xi - a}{b}\right)\right. \\ &\qquad \left. + \frac{\rho_4}{24}H_3\left(\frac{\xi - a}{b}\right) + \frac{\rho_3^2}{72}H_5\left(\frac{\xi - a}{b}\right)\right]. \end{aligned}$$

où

$$\begin{aligned} a &= \kappa_1(\boldsymbol{y}_1, \boldsymbol{y}_2), \\ b &= \sqrt{\kappa_2(\boldsymbol{y}_1, \boldsymbol{y}_2)}, \\ \rho_r &= \frac{\kappa_r(\boldsymbol{y}_1, \boldsymbol{y}_2)}{b^r}. \end{aligned}$$

La méthode d'approximation d'Edgeworth pour la distribution de ξ correspond à celle utilisée dans l'article de Angers (2003). Cependant, celle-ci s'avère plutôt inexacte aux extrémités de la distribution. Entre autre, un phénomène d'ondulation est souvent observé et des valeurs négatives sont obtenues. Ces désagréments nous ont donc motivé à étudier d'autres méthodes d'approximation pour la distribution de ξ. En se basant sur l'équation (2.3.6) qui suggère qu'il est possible d'exprimer ξ sous une forme quadratique de variables normales, nous présentons deux manières d'approximer la distribution de celle-ci.

2.3.2. Formes quadratiques

Commençons par introduire la notion de forme quadratique.

Définition 2.3.5. *Soit* $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)^\top$, *un vecteur aléatoire suivant une distribution normale multivariée ayant pour vecteur moyen* $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^\top$ *et pour matrice de covariance* $\boldsymbol{\Sigma}$. *Ainsi, la forme quadratique* $Q(v_1, \ldots, v_n)$ *associée à la matrice symétrique* $\boldsymbol{C}$ *est définie comme*

$$Q(\boldsymbol{v}) = \boldsymbol{v}^\top \boldsymbol{C} \boldsymbol{v} = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} v_i v_j.$$

En se basant sur cette définition, la quantité $d(\gamma_1, \gamma_2) = \xi$ de l'équation (2.3.2) est une forme quadratique de variables normales, puisque de l'équation (2.3.6), nous avons

$$\begin{aligned} \xi &= (\gamma_1 - \gamma_2)^\top (\gamma_1 - \gamma_2) \\ &= \boldsymbol{v}^\top \boldsymbol{C} \boldsymbol{v} \end{aligned} \tag{2.3.7}$$

où $\boldsymbol{C} = \boldsymbol{I}_n$, $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ avec $\boldsymbol{\mu} = \boldsymbol{A}_1 \boldsymbol{y}_1 - \boldsymbol{A}_2 \boldsymbol{y}_2$ et $\boldsymbol{\Sigma} = \boldsymbol{B}_1 + \boldsymbol{B}_2$. Les expressions suivantes pour l'espérance et la variance d'une forme quadratique sont bien connues

$$\mathbb{E}(Q) = \mathrm{Tr}(\boldsymbol{C}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top \boldsymbol{C} \boldsymbol{\mu}, \tag{2.3.8}$$

$$\mathbb{V}\mathrm{ar}(Q) = 2\mathrm{Tr}(\boldsymbol{C}\boldsymbol{\Sigma})^2 + 4\boldsymbol{\mu}^\top \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}\boldsymbol{\mu}. \tag{2.3.9}$$

De plus, ce qui est d'autant plus intéressant pour nous, c'est que de nombreux auteurs se sont intéressés au calcul de $\Pr\{Q \leq q\}$ où q est un scalaire donné. Notamment, dans un chapitre entièrement consacré aux formes quadratiques, Johnson et Kotz (1970, chapitre 29) présentent le résultat suivant.

Théorème 2.3.2. *Si la matrice $\boldsymbol{\Sigma}$ est non singulière et définie positive, alors il existe une transformation linéaire orthogonale non singulière de $\boldsymbol{v}$, telle que $Q(\boldsymbol{v})$ possède la même distribution que celle de*

$$\begin{aligned} Q(\boldsymbol{x}) &= \sum_{r=1}^{m} \lambda_r x_r^2 \\ &= \sum_{r=1}^{m} \lambda_r \chi^2_{h_r;\delta_r^2} \qquad m \leq n. \end{aligned} \tag{2.3.10}$$

où

λ_r *représente les valeurs propres distinctes et ordonnées associées à* $\boldsymbol{C\Sigma}$,

m *est le nombre de valeurs propres distinctes associées à* $\boldsymbol{C\Sigma}$,

h_r *représente l'ordre de multiplicité respectif des valeurs propres,*

δ_r^2 *est obtenu en trouvant d'abord*

$$\delta'_j = \boldsymbol{P}\boldsymbol{L}^{-1}\boldsymbol{\mu},$$

tel que

$\boldsymbol{L}$ *est triangulaire inférieure, non singulière et satisfait* $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$,

$\boldsymbol{P}$ *contient les vecteurs propres normalisés de* $\boldsymbol{L}^\top\boldsymbol{C}\boldsymbol{L}$ *associée à* λ_r,

et en calculant

$$\delta_r^2 = \sum_{k=1}^{h_r} (\delta'_{k+h_{r-1}})^2, \qquad h_0 = 0.$$

Ainsi, les $\chi^2_{h_r;\delta_r^2}$ sont des variables aléatoires de densité khi-carrée non centrée ayant h_r degrés de liberté et δ_r^2 comme paramètre de non-centralité. De telles

variables sont définies par la relation

$$\chi^2_{h_r;\delta_r^2} = (z_1 + \delta_r)^2 + \sum_{i=2}^{h_r} z_i^2,$$

où les z_i sont des variables aléatoires indépendantes normales centrées réduites. Notons également que d'après l'équation (2.3.10), les expressions pour l'espérance et la variance de Q s'écrivent

$$\mathbb{E}(Q) = \sum_{r=1}^{m} \lambda_r (h_r + \delta_r^2),$$

$$\mathbb{V}\mathrm{ar}(Q) = 2 \sum_{r=1}^{m} \lambda_r^2 (h_r + 2\delta_r^2).$$

À présent, notre objectif consiste à trouver une approximation de la distribution d'un mélange de variables aléatoires khi-deux non centrées. Dans un premier temps, nous présentons une approximation due à Imhof (1961) et par la suite, nous abordons celle étudiée par Kuonen (1999). Remarquons que les travaux réalisés par Imhof (1961) fournissent également un algorithme exact pour évaluer la distribution de $Q(x)$ définie par (2.3.10). Cependant, comme Imhof montre que l'approximation est plus rapide à exécuter et qu'elle est très précise (même aux extrémités de la distribution de $Q(x)$), nous avons choisi d'utiliser cette dernière.

2.3.2.1. *Approximation d'Imhof*

Dans sa publication de 1961, Imhof généralise l'approximation de Pearson (1959) connue sous l'appellation « three-moment central chi-squared approximation » afin qu'elle puisse être applicable lorsque nous nous intéressons à la distribution de variables khi-deux non centrées. La forme quadratique décrite par l'équation (2.3.10) constitue un cas particulier de cette généralisation.

Suivant les travaux de Johnson (1959) et de Pearson (1959), nous avons le théorème suivant.

Théorème 2.3.3. *Si Q est définie positive, alors nous pouvons écrire*

$$Q \cong \frac{(\chi^2_{h'} - h')}{\sqrt{2h'}} \sqrt{\mathbb{V}\mathrm{ar}(Q)} + \mathbb{E}(Q)$$

en définissant h' de façon à ce que les deux membres de l'équation aient leurs troisièmes moments égaux. Remarquons que $\chi^2_{h'}$ fait référence à la distribution khi-deux centrée ayant h' degrés de liberté.

Faisant suite à ce théorème, Imhof déclare qu'une approximation de la distribution et de la fonction de densité de la forme quadratique Q décrite par l'équation (2.3.10) sont données par les théorèmes suivants.

Théorème 2.3.4. *La distribution de Q peut être approximée par*

$$F(q) = \Pr\{Q \leq q\} \cong \Pr\left\{\chi^2_{h'} \leq t\right\},$$

où

$$h' = e_2^3/e_3^2, \quad t = \sqrt{h'/e_2}(q - e_1) + h', \quad e_j = \sum_{r=1}^{m} \lambda_r^j (h_r + j\delta_r^2) \quad (j = 1, 2, 3).$$

Théorème 2.3.5. *La fonction de densité de Q peut être approximée par*

$$f(q) \cong \sqrt{h'/e_2}\; d(t),$$

où d représente la fonction de densité d'une variable aléatoire $\chi^2_{h'}$ et les quantités h' et e_2 correspondent à celles définies au théorème précédent.

Étant donné qu'une variable aléatoire khi-deux ne prend que des valeurs positives, alors la contrainte suivante doit être respectée

$$\begin{aligned} t \geq 0 &\iff \sqrt{\frac{h'}{e_2}}(q - e_1) + h' \geq 0 \\ &\iff q \geq e_1 - \sqrt{e_2 h'} \\ &\iff q \geq e_1 - \frac{e_2^2}{e_3}. \end{aligned}$$

Par conséquent, l'approximation d'Imhof vaut pour des valeurs de q supérieures ou égales à $e_1 - e_2^2/e_3$. Mentionnons que cela n'a pas d'implications dans le présent travail puisque nous ne nous intéressons pas à l'extrémité gauche de la distribution de Q (voir la sous-section 2.3.2.3 et la section 2.4).

2.3.2.2. *Approximation par point de selle*

Les méthodes d'approximation par point de selle ont été introduites par Daniels (1954) et étudiées par de nombreux auteurs ensuite. Ces méthodes fournissent généralement une approximation très précise des fonctions de densités et de distribution. Par opposition à l'approximation de Pearson qui fait uniquement intervenir les trois premiers cumulants, ces méthodes considèrent la fonction génératrice de tous les cumulants

$$K(\zeta) = -\frac{1}{2}\sum_{r=1}^{m} h_r \log(1-2\zeta\lambda_r) + \sum_{r=1}^{m} \frac{\zeta\delta_r^2\lambda_r}{1-2\zeta\lambda_r},$$

où $\zeta < 2^{-1}\min_r \lambda_r^{-1}$.

Dans sa publication, Kuonen (1999) montre que l'approximation par point de selle de Barndorff-Nielsen (1990) pour la distribution de Q décrite par l'équation (2.3.10) s'avère particulièrement efficace. Celle-ci est décrite au théorème suivant.

Théorème 2.3.6. *La distribution de Q peut être approximée par*

$$F(q) = \Pr\{Q \leq q\} \cong \Phi\left\{w + \frac{1}{w}\log\left(\frac{v}{w}\right)\right\}, \tag{2.3.11}$$

où

$$v = \widehat{\zeta}\sqrt{K''(\widehat{\zeta})}, \quad w = \mathrm{Sign}(\widehat{\zeta})\sqrt{2\left(\widehat{\zeta}q - K(\widehat{\zeta})\right)},$$

$\mathrm{Sign}(\widehat{\zeta})$ *dénote le signe de* $\widehat{\zeta}$ *et* $\widehat{\zeta} = \widehat{\zeta}(q)$*, le point de selle, est la valeur de* ζ *telle que* $K'(\zeta) = q$*. De plus, comme nous le verrons plus bas, les quantités* $K'(\zeta)$ *et* $K''(\zeta)$ *sont les première et seconde dérivées de* $K(\zeta)$ *par rapport à* ζ*.*

Présentons maintenant les expressions pour $K'(\zeta)$ et $K''(\zeta)$ vues précédemment, puis celle pour $K'''(\zeta)$ dont nous aurons besoin un peu plus loin :

$$\begin{aligned} K^{(j)}(\zeta) &= \frac{\partial^j K(\zeta)}{\partial\zeta^j} \\ &= 2^{j-1}(j-1)!\sum_{r=1}^{m}\frac{\lambda_r^j}{(1-2\zeta\lambda_r)^j}\left[h_r + \frac{j\delta_r^2}{1-2\zeta\lambda_r}\right], \end{aligned}$$

où $j = 1, 2, 3$ et $K^{(1)}(\zeta)$, $K^{(2)}(\zeta)$, $K^{(3)}(\zeta)$ correspondent respectivement aux expressions pour $K'(\zeta)$, $K''(\zeta)$ et $K'''(\zeta)$.

Notons que le calcul de l'équation (2.3.11) requiert la résolution de $K'(\zeta) = q$, menant à $\widehat{\zeta} = \widehat{\zeta}(q)$, pour chaque valeur q d'intérêt. L'existence et l'unicité de $\widehat{\zeta}$ sont dues à Daniels (1954). À titre d'exemple, la figure 2.1 montre l'allure de la relation entre ζ et q.

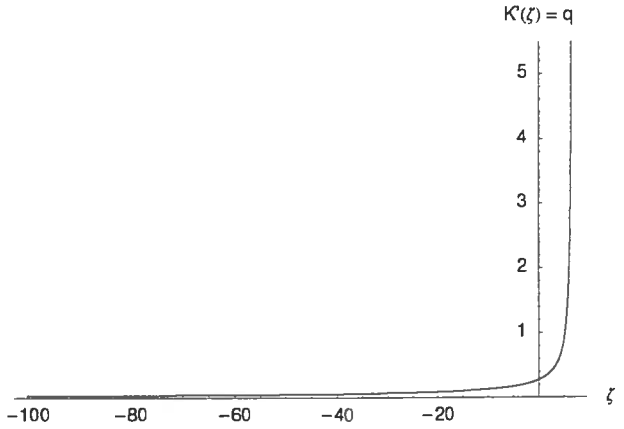


FIGURE 2.1. Approximation par point de selle : relation entre ζ et q.

Énonçons maintenant le théorème suivant portant sur la densité de la forme quadratique Q décrite en (2.3.10).

Théorème 2.3.7. *La densité de Q peut être approximée par*

$$f(q) \cong \left(\left[\frac{\partial w}{\partial q} \right] \left[1 - \frac{1}{w^2} \left(\log\left(\frac{v}{w}\right) + 1 \right) \right] + \frac{1}{vw} \frac{\partial v}{\partial q} \right) \phi \left\{ w + \frac{1}{w} \log\left(\frac{v}{w}\right) \right\}.$$

Les expressions pour $\partial v/\partial q$ et $\partial w/\partial q$ sont

$$\frac{\partial v}{\partial q} = \frac{\partial \widehat{\zeta}}{\partial q} \left[\sqrt{K''(\widehat{\zeta})} + \frac{\widehat{\zeta} K'''(\widehat{\zeta})}{2\sqrt{K''(\widehat{\zeta})}} \right], \tag{2.3.12}$$

et

$$\frac{\partial w}{\partial q} = \frac{\text{Sign}(\widehat{\zeta})}{\sqrt{2\{\widehat{\zeta} - K(\widehat{\zeta})\}}} \left[\frac{\partial \widehat{\zeta}}{\partial q} q + \widehat{\zeta} - K'(\widehat{\zeta}) \frac{\partial \widehat{\zeta}}{\partial q} \right]. \tag{2.3.13}$$

Toutefois, nous savons que

$$\begin{aligned} K'(\widehat{\zeta}) = q &\iff \frac{\partial K'(\widehat{\zeta})}{\partial q} = \frac{\partial q}{\partial q} \\ &\iff K''(\widehat{\zeta}) \frac{\partial \widehat{\zeta}}{\partial q} = 1 \end{aligned}$$

$$\iff \frac{\partial \widehat{\zeta}}{\partial q} = \frac{1}{K''(\widehat{\zeta})}. \tag{2.3.14}$$

Ainsi, en utilisant l'équation (2.3.14), les équations (2.3.12) et (2.3.13) deviennent

$$\frac{\partial v}{\partial q} = \frac{1}{\sqrt{K''(\widehat{\zeta})}} + \frac{\widehat{\zeta} K'''(\widehat{\zeta})}{2\left(K''(\widehat{\zeta})\right)^{3/2}},$$

et

$$\frac{\partial w}{\partial q} = \frac{\text{Sign}(\widehat{\zeta})\,\widehat{\zeta}}{\sqrt{2\{\widehat{\zeta} q - K(\widehat{\zeta})\}}}.$$

Tout comme c'était le cas pour l'équation (2.3.11), soulignons que le calcul de la densité $f(q)$ de la forme quadratique Q au point d'intérêt q requiert la valeur de $\widehat{\zeta}$ correspondante.

2.3.2.3. *Tests basés sur l'approximation de la distribution de Q*

Nous disposons maintenant de deux méthodes d'approximation pour la distribution et la densité de la forme quadratique Q de l'équation (2.3.10). Basé sur ces approximations, nous désirons développer un test qui nous permettra de confronter les hypothèses

$$H_0 : g_1 = g_2 \quad \text{contre} \quad H_a : g_1 \neq g_2 \tag{2.3.15}$$

telles que décrites auparavant.

Pour ce faire, trouvons d'abord un moyen d'exprimer la distribution de Q sous H_0. Étant donné que sous l'hypothèse nulle, les fonctions sous-jacentes g_1 et g_2 sont égales, notre objectif consiste à diminuer l'écart vertical observé entre les données provenant des deux échantillons. Tout juste avant de décrire comment nous procédons pour y arriver, mentionnons à nouveau que les données originales des deux échantillons sont observées sous la forme $\{(t_{1i}, y_{1i}), i = 1, 2, \ldots, n_1\}$ et $\{(t_{2i}, y_{2i}), i = 1, 2, \ldots, n_2\}$.

Pour $i = 1, \ldots, n_1$,

1) déterminer la plus petite distance (non nulle), notée Δ_{1i}, entre t_{1i} et t_{2j}, $j = 1, \ldots, n_2$, c'est-à-dire

$$\Delta_{1i} = \min_{j=1,\ldots,n_2} \{t_{1i} - t_{2j}\};$$

2) calculer la moyenne des y_{2j}, $j = 1, \ldots, n_2$ situés à une distance de t_{1i} inférieure à $2 \times \Delta_{1i}$ et la noter moy_{1i} ;

3) calculer

$$y'_{1i} = \frac{y_{1i} + \text{moy}_{1i}}{2}.$$

Ensuite, pour $j = 1, \ldots, n_2$,

1) déterminer la plus petite distance (non nulle), notée Δ_{2j}, entre t_{2j} et t_{1i}, $i = 1, \ldots, n_1$, c'est-à-dire

$$\Delta_{2j} = \min_{i=1,\ldots,n_1} \{t_{2j} - t_{1i}\};$$

2) calculer la moyenne des y_{1i}, $i = 1, \ldots, n_1$ situés à une distance de t_{2j} inférieure à $2 \times \Delta_{2j}$ et dénoter celle-ci par moy_{2j} ;

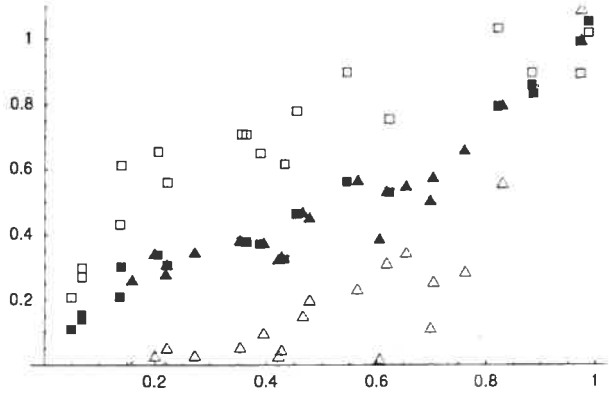
3) calculer

$$y'_{2j} = \frac{y_{2j} + \text{moy}_{2j}}{2}.$$

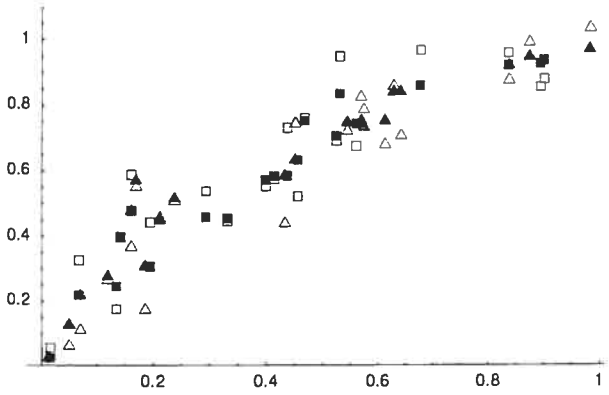
Ces opérations nous permettent d'obtenir deux jeux de données transformées, soit $\{(t_{1i}, y'_{1i}), i = 1, 2, \ldots, n_1\}$ et $\{(t_{2i}, y'_{2i}), i = 1, 2, \ldots, n_2\}$. Nous ferons référence à ces derniers au moyen de l'appellation « données sous l'hypothèse que les fonctions sont égales » ou encore tout simplement « données sous H_0 ».

La figure 2.2 illustre le résultat de ces manipulations pour deux situations différentes. Tout d'abord, la figure 2.2 (a) suggère que si les données observées des échantillons semblent issues de fonctions différentes, alors les donnés sous H_0 sont éloignées de celles observées. À l'opposé, la figure 2.2 (b) montre que lorsque les fonctions sous-jacentes des échantillons semblent égales, il y a proximité entre les données sous H_0 et celles observées.

Suivant cet ordre d'idées, nous avons choisi de baser les deux premiers tests proposés sur la comparaison de l'approximation de la distribution de Q observée



(a)



(b)

FIGURE 2.2. (a) Données observées : premier échantillon (□) et deuxième échantillon (△). (b) Données sous H_0 : premier échantillon (■) et deuxième échantillon (▲).

avec celle de Q sous H_0. Plus les approximations seront rapprochées l'une de l'autre, plus nous pencherons en faveur de H_0 et vice-versa.

Afin d'éviter toute confusion, précisons immédiatement que la distribution de Q observée correspond à celle obtenue à partir des données $\{(t_{1i}, y_{1i}), i = 1, 2, \ldots, n_1\}$ et $\{(t_{2i}, y_{2i}), i = 1, 2, \ldots, n_2\}$. Pour sa part, la distribution de Q sous H_0 est celle calculée à partir des données $\{(t_{1i}, y'_{1i}), i = 1, 2, \ldots, n_1\}$ et $\{(t_{2i}, y'_{2i}), i = 1, 2, \ldots, n_2\}$.

Le premier test qui nous permet de confronter les hypothèses décrites en (2.3.15) tient compte du mode de l'approximation de la fonction de densité de Q observée.

Plus précisément, nous calculons la probabilité suivante

$$\Pr\{Q \geq \text{mode}_{\text{observé}} \mid H_0\}. \tag{2.3.16}$$

La figure 2.3 fournit une illustration graphique de ce test. Nous savons que pour des fonctions $\widehat{g}_1$ et $\widehat{g}_2$ distantes, la densité de Q sous H_0 est éloignée de celle observée et nous obtenons une petite région ombragée. Par conséquent, nous rejetons H_0 lorsque la probabilité calculée en (2.3.16) est faible. Il est à noter que le choix du mode dans l'équation (2.3.16) est arbitraire. Par exemple, nous aurions pu utiliser la médiane ou encore la moyenne de la distribution plutôt que le mode de cette dernière.

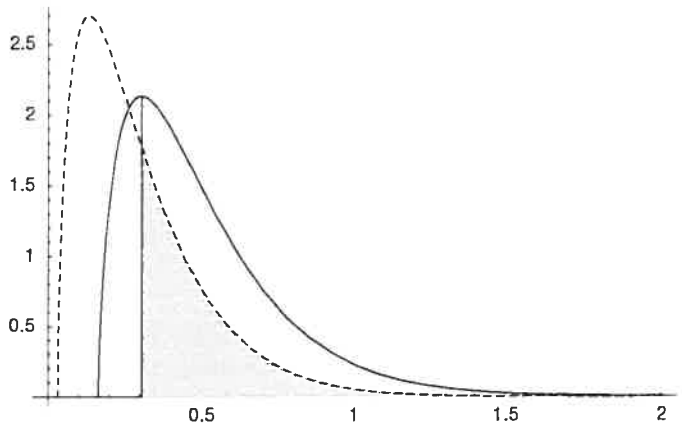


FIGURE 2.3. Illustration du premier test : approximation de la densité de Q observée (trait plein), approximation de la densité de Q sous H_0 (pointillés) et résultat du premier test (région ombragée).

Le second test proposé tient compte du mode de la densité de Q observée ainsi que de celui de la densité de Q sous H_0. Intuitivement, ce critère constitue une sorte de mesure de la distance entre ces modes en terme de probabilité. Plus précisément, le test se calcule comme suit

$$\Pr\{Q < \text{mode}_{\text{observé}} \mid H_0\} - \Pr\{Q < \text{mode}_{\text{sous } H_0} \mid H_0\}. \tag{2.3.17}$$

La figure 2.4 présente ce test sous forme graphique. Puisque la région ombragée sur ce graphe est proportionnelle à l'éloignement des densités, nous rejetons H_0 lorsque le résultat de l'expression (2.3.17) est élevé.

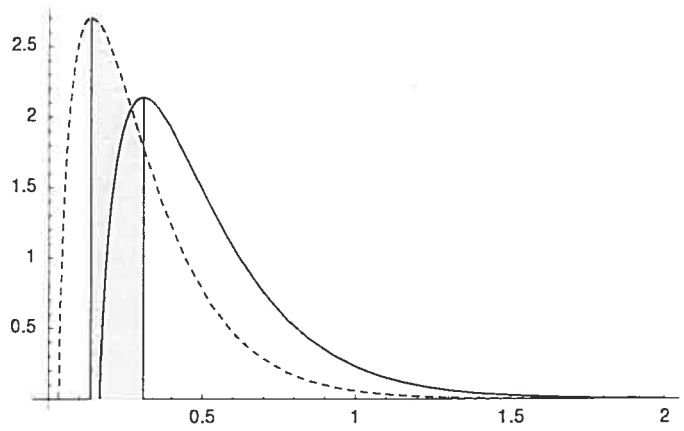


FIGURE 2.4. Illustration du deuxième test : approximation de la densité de Q observée (trait plein), approximation de la densité de Q sous H_0 (pointillés) et résultat du deuxième test (région ombragée).

Pour sa part, le troisième test se distingue des précédents puisqu'il ne repose pas sur la comparaison de l'approximation des distributions de Q observée et sous H_0. Celui-ci fait plutôt intervenir la notion de distance observée entre les fonctions estimées $d(\widehat{\gamma}_1, \widehat{\gamma}_2)$ de l'équation (2.1.14). Plus précisément, nous calculons la probabilité

$$\Pr\{Q \geq d(\widehat{\gamma}_1, \widehat{\gamma}_2) \mid H_0\}. \tag{2.3.18}$$

La figure 2.5 illustre l'expression décrite en (2.3.18). Nous savons que plus les

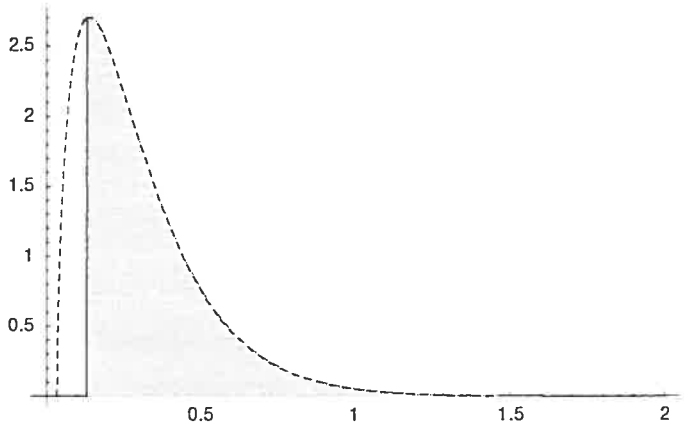


FIGURE 2.5. Illustration du troisième test : approximation de la densité de Q sous H_0 (pointillés) et résultat du troisième test (région ombragée).

fonctions estimées $\widehat{g}_1$ et $\widehat{g}_2$ sont éloignées, plus la distance $d(\widehat{\gamma}_1, \widehat{\gamma}_2)$ est grande

et donc plus la quantité calculée en (2.3.18) est petite. Ainsi, nous rejetons H_0 lorsque la probabilité obtenue en (2.3.18) est faible.

Le comportement de ces tests sera étudié au chapitre 3. Pour le moment, présentons une dernière méthode nous permettant de confronter les hypothèses H_0 et H_a exposées en (2.3.15). Il s'agit d'intervalles de confiance simultanés pour la différence entre les fonctions g_1 et g_2.

2.4. Test basé sur le concept d'intervalles de confiance simultanés

Nous désirons obtenir des intervalles de confiance simultanés pour la différence entre les deux fonctions estimées de façon à avoir un niveau de confiance global égal à $1-\alpha$. Si $M = 2K_0 + 1 + \sum_{j=1}^{L}(2K_j+1)$, dénotons la base de fonctions utilisée par

$$\begin{aligned}\boldsymbol{b}(x) &= (\phi_{-K_0}(x),\ \ldots\ ,\ \phi_{K_0}(x),\ \psi_{0,-K_0}(x),\ \ldots\ ,\ \psi_{L,K_L}(x))^{\top} \\ &= (b_1(x),\ \ldots\ ,\ b_{2K_0+1}(x),\ b_{2K_0+2}(x),\ \ldots\ ,\ b_M(x))^{\top}\end{aligned}$$

et le vecteur des différences entre les coefficients estimés pour les deux échantillons par

$$\begin{aligned}\widehat{\boldsymbol{\theta}} &= \widehat{\boldsymbol{\gamma}}_1 - \widehat{\boldsymbol{\gamma}}_2 \\ &= \big((\widehat{\alpha}_{(1)-K_0} - \widehat{\alpha}_{(2)-K_0}),\ \ldots\ ,\ (\widehat{\alpha}_{(1)K_0} - \widehat{\alpha}_{(2)K_0}), \\ &\qquad (\widehat{\beta}_{(1)0,-K_0} - \widehat{\beta}_{(2)0,-K_0}),\ \ldots\ ,\ (\widehat{\beta}_{(1)L,K_L} - \widehat{\beta}_{(2)L,K_L})\big)^{\top} \\ &= \left(\widehat{\theta}_1,\ \ldots\ ,\ \widehat{\theta}_{2K_0+1},\ \widehat{\theta}_{2K_0+2},\ \ldots\ ,\ \widehat{\theta}_M\right)^{\top}.\end{aligned}$$

Ainsi, nous avons

$$\begin{aligned}\widehat{g}_1(t) - \widehat{g}_2(t) &= \sum_{|k|\le K_0} (\widehat{\alpha}_{(1)k} - \widehat{\alpha}_{(2)k})\phi_k(t) + \sum_{j\le L}\sum_{|k|\le K_j} (\widehat{\beta}_{(1)j,k} - \widehat{\beta}_{(2)j,k})\psi_{j,k}(t) \\ &= \boldsymbol{b}^{\top}(t)\widehat{\boldsymbol{\theta}}.\end{aligned}$$

Pour une valeur fixée de t, disons t_0,

$$\widehat{g}_1(t_0) - \widehat{g}_2(t_0) = \boldsymbol{b}^\top(t_0)\widehat{\boldsymbol{\theta}}.$$

Étant donné que les fonctions g_1 et g_2 sont estimées sur leurs domaines respectifs $[t_{11}, t_{1n_1}]$ et $[t_{21}, t_{2n_2}]$, les intervalles de confiance sont calculés pour des valeurs $\boldsymbol{u} = \{u_1, \ldots, u_{n'}\}$, $n' \leq n_1 + n_2$, correspondant aux observations confondues et ordonnées des deux échantillons satisfaisant

$$u_1 < u_2 < \ldots < u_{n'},$$

$$\begin{aligned} u_1 &= \max\left\{\min_{i=1,\ldots,n_1}\{t_{1i}\}, \min_{j=1,\ldots,n_2}\{t_{2j}\}\right\}, \\ u_{n'} &= \min\left\{\max_{i=1,\ldots,n_1}\{t_{1i}\}, \max_{j=1,\ldots,n_2}\{t_{2j}\}\right\}. \end{aligned}$$

Notons que nous aurions pu choisir de calculer les intervalles de confiance aux points d'abscisses $\boldsymbol{u'} = \{u'_1, \ldots, u'_n\}$, $n = n_1 + n_2$, tels que

$$u'_1 < u'_2 < \ldots < u'_n,$$

$$u'_{\min} = \min_{\substack{i=1,\ldots,n_1 \\ j=1,\ldots,n_2}} \{t_{1i}, t_{2j}\},$$

$$u'_{\max} = \max_{\substack{i=1,\ldots,n_1 \\ j=1,\ldots,n_2}} \{t_{1i}, t_{2j}\},$$

$$u'_i \sim \mathcal{U}(u'_{\min}, u'_{\max}), \quad i = 1, \ldots, n.$$

Toutefois, en procédant ainsi, nous risquons de nous retrouver dans l'une des situations suivantes

$$\begin{aligned} u'_{\min} &< \min_{i=1,\ldots,n_1}\{t_{1i}\} \quad \text{et/ou} \quad u'_{\max} > \max_{i=1,\ldots,n_1}\{t_{1i}\}, \\ u'_{\min} &< \min_{j=1,\ldots,n_2}\{t_{2j}\} \quad \text{et/ou} \quad u'_{\max} > \max_{j=1,\ldots,n_2}\{t_{2j}\}. \end{aligned}$$

Cela nous ramène à un cas d'intervalle de confiance de type prévisionnel pour lequel nous n'avons pas d'intérêt. De plus, ces situations font en sorte que la

variance des estimateurs des coefficients devient si élevée que les fonctions estimées n'ont plus de sens. Pour ces raisons, il nous semble plus adéquat de calculer les intervalles de confiance aux points d'abscisses $\{u_1, \ldots, u_{n'}\}$.

D'après Seber (1977, section 5.1.1 c), l'intervalle de confiance pour toute combinaison linéaire $\boldsymbol{b}^\top(u_i)\boldsymbol{\theta}$ est donné par

$$\boldsymbol{b}^\top(u_i)\widehat{\boldsymbol{\theta}} \pm \sqrt{d\mathcal{F}^{\alpha}_{d,n-p}\widehat{\mathbb{V}\mathrm{ar}}\left(\boldsymbol{b}^\top(u_i)\boldsymbol{\theta}\right)}, \tag{2.4.1}$$

où $\mathcal{F}^{\alpha}_{d,n-p}$ représente le $(1-\alpha)^{\mathrm{e}}$ quantile d'une distribution de Fisher avec d et $(n-p)$ degrés de liberté, $\alpha = 0,05$, $d = p = M, n = n_1 + n_2$ et $i = 1, \ldots n'$. Puisque

$$\widehat{\mathbb{V}\mathrm{ar}}\left(\boldsymbol{b}^\top(u_i)\boldsymbol{\theta}\right) = \boldsymbol{b}^\top(u_i)\boldsymbol{\Sigma}\boldsymbol{b}(u_i),$$

alors l'équation (2.4.1) devient

$$\boldsymbol{b}^\top(u_i)\widehat{\boldsymbol{\theta}} \pm \sqrt{d\mathcal{F}^{\alpha}_{d,n-p}\boldsymbol{b}^\top(u_i)\boldsymbol{\Sigma}\boldsymbol{b}(u_i)}. \tag{2.4.2}$$

Plus la proportion d'intervalles de confiance incluant la valeur 0 sera élevée, plus nous pencherons en faveur de H_0. Le comportement de ce test sera étudié au chapitre 3.

Ceci clôture le second chapitre du présent mémoire. Le chapitre à venir décrit l'analyse de simulations effectuée et présente les divers résultats obtenus. De plus, afin de montrer à quel type de problème se prête le travail présenté jusqu'ici, nous appliquons les méthodes développées à un jeu de données réelles.

Chapitre 3

ÉTUDE DE SIMULATION ET EXEMPLE

Ce troisième et dernier chapitre comporte deux objectifs. Tout d'abord, nous désirons étudier le comportement des tests proposés au chapitre précédent, puis nous voulons les appliquer à une situation réelle. Ainsi, la première section est consacrée à une étude de simulation, puis la seconde à la présentation d'un exemple utilisant des données réelles observées.

3.1. ÉTUDE DE SIMULATION

Décrivons le contexte de simulation dans lequel nous nous situons. Premièrement, les différentes paires de fonctions g_1 et g_2 utilisées sont présentées au tableau 3.1 et correspondent à celles d'abord utilisées par Kulasekera (1995). Au total, nous avons six paires de fonctions égales et dix paires de fonctions différentes.

TABLEAU 3.1. Paires de fonctions utilisées pour l'étude de simulation.

$g_1(t)$	$g_2(t)$			
$\sqrt{t}$	$\sqrt{t}$	t^2	$\sqrt{t}+t$	$\sqrt{t}+1$
t^2	t^2	t^2+t		
$\cos(\pi t)$	$\cos(\pi t)$	$\cos(\pi t)+t$	$\cos(\pi t)+1$	
$\cos(2\pi t)$	$\cos(2\pi t)$	$\cos(2\pi t)+t$	$\cos(2\pi t)+1$	
$\cos(4\pi t)$	$\cos(4\pi t)$	$\sin(4\pi t)$		
$\cos^2(2\pi t)$	$\cos^2(2\pi t)$	$\sin^2(2\pi t)$		

Les tailles échantillonnales des deux jeux de données sont considérées égales et les valeurs retenues pour celles-ci sont $n = n_1 = n_2 = 20, 30$ et 50. Les observations $\{t_{1i}, i = 1, \ldots, n_1\}$ et $\{t_{2i}, i = 1, \ldots, n_2\}$ sont générées uniformément sur l'intervalle unitaire $[0, 1]$. Quant à elles, les erreurs sont distribuées selon une loi

normale de moyenne 0 et d'écart type égal à 0,1. Ensuite, pour $\ell = 1, 2$, nous calculons

$$y_{\ell i} = g_\ell(t_{\ell i}) + \varepsilon_{\ell i}, \qquad i = 1, 2, \ldots, n.$$

Comme nous l'avons déjà mentionné auparavant, l'estimation des fonctions se fait à l'aide de la base d'ondelettes de Daubechies et l'ordre sélectionné est 2. Un tel choix pour l'ordre nous permet d'obtenir une modélisation satisfaisante des fonctions pour les différentes tailles échantillonnales retenues.

Tout juste avant de présenter les résultats obtenus, précisons que cette étude de simulation a été entièrement réalisée à l'aide du logiciel *Mathematica 5.0*. La programmation complète et commentée est disponible à l'Annexe B.

3.1.1. Résultats

Les résultats de l'étude sont présentés sous forme de tableaux séparés pour chaque test. Les étapes suivantes mènent à l'obtention de tels tableaux. Pour chacune des six paires de fonctions égales et pour $\ell = 1, 2$,

1) générer des observations $t_{\ell i}$, $i = 1, \ldots, n$;
2) générer des erreurs $\varepsilon_{\ell i}$, $i = 1, \ldots, n$;
3) calculer $y_{\ell i}$, $i = 1, \ldots, n$;
4) modéliser les fonctions à l'aide de la base d'ondelettes de Daubechies d'ordre 2;
5) calculer l'approximation d'Imhof pour la distribution de la distance (observée et sous H_0) entre les fonctions;
6) calculer l'approximation par point de selle de la distribution de la distance (observée et sous H_0) entre les fonctions;
7) obtenir la valeur des trois premiers tests selon les deux méthodes d'approximation;
8) calculer les intervalles de confiance simultanés et trouver la proportion de ceux-ci qui incluent 0.

Comme mentionné à la fin de la section 2.2, nous cherchons une alternative au test de Bayes. Une façon de procéder est la suivante. En répétant les étapes ci-haut

1000 fois, nous sommes en mesure de trouver le seuil correspondant à chaque test au niveau 5% en prenant le 95^{e} percentile. Afin de trouver la puissance correspondante des tests, nous suivons à nouveau ces 8 étapes et les répétons 1000 fois, mais en utilisant une paire de fonctions différentes telle que $g_1(t)$ correspond aux fonctions égales d'abord utilisées (voir tableau 3.1). Par exemple, pour trouver les puissances des tests associés à la fonction $\cos(4\pi t)$, nous prenons $g_1(t) = \cos(4\pi t)$ et $g_2(t) = \sin(4\pi t)$.

Soulignons que cette façon de procéder nous permet d'obtenir un seuil pour chaque test qui est propre à la paire de fonctions utilisées, soit un *seuil individuel*, auquel correspond une *puissance individuelle*. Nous aimerions maintenant déterminer un seuil pour chaque test qui soit valide pour toute paire de fonctions, c'est-à-dire un *seuil commun*. Pour ce faire, voici comment nous procédons. Pour chaque test et pour chaque valeur de n,

1) calculer $\text{seuil}_{\text{min}} = \min_{i=1,\ldots,6}(\text{seuil}_i)$ et $\text{seuil}_{\text{max}} = \max_{i=1,\ldots,6}(\text{seuil}_i)$ où les seuil_i, $i = 1,\ldots,6$ correspondent aux seuils individuels obtenus pour les six paires de fonctions égales ;

2) définir le vecteur

$$\textbf{seuil}_{\text{possible}} = (\text{seuil}_{\text{min}},\ \text{seuil}_{\text{min}} + 0{,}01,\ \ldots\ ,\ \text{seuil}_{\text{max}})$$

composé, au total, de $((\text{seuil}_{\text{max}} - \text{seuil}_{\text{min}})/0{,}01)$ seuils communs possibles ;

3) construire une matrice dont les éléments, dénotés par mat_{ij}, $i = 1,\ldots,6$ et $j = 1,\ldots,((\text{seuil}_{\text{max}} - \text{seuil}_{\text{min}})/0{,}01)$, sont définis comme suit : pour les six paires de fonctions égales, calculer les niveaux correspondants à chacun des seuils communs possibles ;

4) construire la matrice composée des éléments $\text{diff}_{ij} = |\text{mat}_{ij} - 50|$;

5) définir le vecteur **med** ayant pour éléments $\text{med}_j = \underset{i=1,\ldots,6}{\text{médiane}}(\text{diff}_{ij})$;

6) choisir, parmi les seuils communs possibles, celui correspondant à l'élément minimum du vecteur **med**.

Une fois les seuils communs déterminés pour chacun des tests et pour chaque valeur de n, nous pouvons trouver les puissances communes leur étant associées.

Commentons maintenant les tableaux 3.2 à 3.7 comportant les divers résultats obtenus pour les tests basés sur l'approximation de la distance entre les fonctions.

TABLEAU 3.2. Niveaux empiriques ($\times 1000$) pour le premier test (seuils communs pour l'approximation d'Imhof : 0,196 si $n = 20$, 0,137 si $n = 30$ et 0,0997 si $n = 50$; seuils communs pour l'approximation par point de selle : 0,206 si $n = 20$, 0,146 si $n = 30$ et 0,129 si $n = 50$).

		Approximation d'Imhof			Approximation par point de selle		
$g_1(t) = g_2(t)$	n	niveau	seuil	niveau commun	niveau	seuil	niveau commun
$\sqrt{t}$	20	48	0,129	8	49	0,134	13
	30	50	0,139	55	49	0,154	55
	50	50	0,128	177	49	0,140	75
t^2	20	50	0,188	45	49	0,197	35
	30	50	0,178	138	49	0,169	85
	50	49	0,167	318	49	0,228	184
$\cos(\pi t)$	20	50	0,105	11	50	0,134	11
	30	48	0,124	33	49	0,143	44
	50	49	0,111	87	48	0,117	23
$\cos(2\pi t)$	20	49	0,135	34	50	0,167	33
	30	50	0,078	12	51	0,173	115
	50	50	0,076	13	49	0,049	9
$\cos(4\pi t)$	20	48	0,334	142	49	0,333	137
	30	50	0,172	72	49	0,162	62
	50	48	0,064	17	49	0,064	11
$\cos^2(2\pi t)$	20	51	0,222	68	49	0,244	68
	30	49	0,099	25	50	0,144	48
	50	50	0,064	14	47	0,158	80

Précisons que les tableaux 3.2 et 3.3 présentent respectivement les niveaux et les puissances empiriques obtenues pour le premier test. Pour leur part, les tableaux 3.4 et 3.5 contiennent cette même information, mais pour le deuxième test. De même, les tableaux 3.6 et 3.7 se rapportent au troisième test. Mentionnons que les résultats de l'approximation d'Imhof pour le troisième test ne sont pas présentés aux tableaux 3.6 et 3.7, puisque la quantité $d(\widehat{\gamma}_1, \widehat{\gamma}_2)$ (voir équation (2.3.18)) intervenant dans le calcul du test est souvent plus petite que la borne inférieure définie pour l'approximation d'Imhof.

De façon générale, les puissances individuelles correspondant à un niveau de 5% pour chaque fonction sont supérieures à 90%, donc satisfaisantes. Cependant, quelques exceptions surviennent. Par exemple, à partir des tableaux 3.3, 3.5 et 3.7, nous constatons que les fonctions $g_1(t) = \cos(2\pi t)$ et $g_2(t) = \cos(2\pi t) + t$

TABLEAU 3.3. Puissances empiriques ($\times 1000$) pour le premier test (les puissances communes sont celles associées aux seuils et niveaux communs du tableau 3.2).

			Approximation d'Imhof		Approximation par point de selle	
$g_1(t)$	$g_2(t)$	n	puissance	puissance commune	puissance	puissance commune
$\sqrt{t}$	t^2	20	947	938	949	940
		30	954	968	958	971
		50	971	991	970	989
$\sqrt{t}$	$g_1(t)+t$	20	951	943	950	939
		30	953	954	954	958
		50	966	970	967	968
$\sqrt{t}$	$g_1(t)+1$	20	981	972	979	970
		30	985	985	982	983
		50	988	994	987	989
t^2	$g_1(t)+t$	20	975	973	973	971
		30	985	992	978	979
		50	986	993	990	995
$\cos(\pi t)$	$g_1(t)+t$	20	936	809	937	785
		30	948	944	946	943
		50	952	961	959	954
$\cos(\pi t)$	$g_1(t)+1$	20	990	983	989	987
		30	990	988	989	988
		50	997	998	996	996
$\cos(2\pi t)$	$g_1(t)+t$	20	402	179	217	139
		30	707	600	460	592
		50	979	964	980	933
$\cos(2\pi t)$	$g_1(t)+1$	20	950	912	938	906
		30	965	960	956	967
		50	974	967	989	960
$\cos(4\pi t)$	$\sin(4\pi t)$	20	216	432	221	422
		30	556	696	605	672
		50	928	881	924	823
$\cos^2(2\pi t)$	$\sin^2(2\pi t)$	20	889	911	851	897
		30	974	956	958	957
		50	983	975	965	969

mènent à de faibles puissances lorsque la taille échantillonnale considérée est inférieure à $n = 50$. Cela s'explique par le fait que la modélisation par ondelettes ne parvient pas à bien distinguer ces fonctions lorsqu'il y a peu d'observations. Globalement, les fonctions estimées sont plus rapprochées que les vraies fonctions et cela est amplifié lorsque $t < 0{,}5$. Référons nous à la figure 3.1 afin de

TABLEAU 3.4. Niveaux empiriques ($\times 1000$) pour le deuxième test (seuils communs pour l'approximation d'Imhof : 0,562 si $n = 20$, 0,647 si $n = 30$ et 0,657 si $n = 50$; seuils communs pour l'approximation par point de selle : 0,532 si $n = 20$, 0,605 si $n = 30$ et 0,624 si $n = 50$).

		Approximation d'Imhof			Approximation par point de selle		
$g_1(t) = g_2(t)$	n	niveau	seuil	niveau commun	niveau	seuil	niveau commun
$\sqrt{t}$	20	51	0,647	4	49	0,614	7
	30	50	0,636	67	50	0,601	57
	50	50	0,649	72	50	0,617	66
t^2	20	50	0,580	33	48	0,551	32
	30	51	0,597	168	49	0,561	142
	50	49	0,609	220	50	0,559	328
$\cos(\pi t)$	20	50	0,663	11	51	0,636	11
	30	49	0,652	46	50	0,614	40
	50	53	0,667	25	51	0,636	29
$\cos(2\pi t)$	20	50	0,622	36	50	0,592	37
	30	49	0,692	14	50	0,619	35
	50	51	0,701	6	51	0,682	7
$\cos(4\pi t)$	20	49	0,434	136	50	0,405	133
	30	49	0,570	92	50	0,555	74
	50	51	0,695	28	49	0,681	20
$\cos^2(2\pi t)$	20	49	0,538	64	50	0,506	69
	30	51	0,671	32	49	0,628	33
	50	50	0,713	10	51	0,641	26

mieux visualiser cette explication. Sur cette figure apparaissent, pour $n = 20$, les données simulées, ainsi que l'estimation par ondelettes correspondante pour chaque échantillon. Les vraies fonctions $g_1 = \cos(2\pi t)$ et $g_2 = \cos(2\pi t) + t$ sont également sur le graphique. Toujours à partir des tableaux 3.3, 3.5 et 3.7, nous remarquons aussi que les fonctions $g_1(t) = \cos(4\pi t)$ et $g_2(t) = \sin(4\pi t)$ mènent à de faibles puissances lorsque $n = 20, 30$. L'explication provient des entrecroisements fréquents des fonctions sur le domaine $[0, 1]$. Pour le moment, mentionnons seulement que la portion supérieure de la figure 3.6 illustre cette explication et que de plus amples détails à ce sujet seront fournis un peu plus tard dans ce chapitre.

Faisons également la remarque que les puissances individuelles augmentent habituellement en fonction de n. Effectivement, plus n augmente, plus précises

TABLEAU 3.5. Puissances empiriques ($\times 1000$) pour le deuxième test (les puissances communes sont celles associées aux seuils et niveaux communs du tableau 3.4).

			Approximation d'Imhof		Approximation par point de selle	
$g_1(t)$	$g_2(t)$	n	puissance	puissance commune	puissance	puissance commune
$\sqrt{t}$	t^2	20	947	934	937	922
		30	953	973	946	971
		50	979	988	969	986
$\sqrt{t}$	$g_1(t)+t$	20	946	929	942	918
		30	962	963	951	953
		50	968	969	962	963
$\sqrt{t}$	$g_1(t)+1$	20	965	958	961	957
		30	981	982	976	976
		50	982	983	980	981
t^2	$g_1(t)+t$	20	978	977	975	970
		30	987	989	984	987
		50	988	989	988	990
$\cos(\pi t)$	$g_1(t)+t$	20	935	770	928	767
		30	950	948	940	937
		50	968	966	964	962
$\cos(\pi t)$	$g_1(t)+1$	20	983	978	983	978
		30	990	990	990	990
		50	996	996	995	995
$\cos(2\pi t)$	$g_1(t)+t$	20	259	107	276	113
		30	776	631	676	625
		50	794	763	774	683
$\cos(2\pi t)$	$g_1(t)+1$	20	863	810	861	812
		30	892	885	883	882
		50	927	915	924	913
$\cos(4\pi t)$	$\sin(4\pi t)$	20	210	398	214	394
		30	417	738	479	675
		50	876	828	875	806
$\cos^2(2\pi t)$	$\sin^2(2\pi t)$	20	877	904	879	907
		30	926	920	920	915
		50	938	924	923	921

sont les approximations de la distance entre les fonctions et meilleurs sont les tests basés sur ces approximations.

En ce qui concerne la détermination d'un seuil commun pour toutes les fonctions, les niveaux communs obtenus sont instables. Par exemple, le seuil commun obtenu pour le premier test avec l'approximation d'Imhof pour $n = 20$ mène à

TABLEAU 3.6. Niveaux empiriques ($\times 1000$) pour le troisième test (seuils communs pour l'approximation par point de selle : 0,973 si $n = 20$, 0,99 si $n = 30$ et 0,994 si $n = 50$).

		Approximation d'Imhof			Approximation par point de selle		
$g_1(t) = g_2(t)$	n	niveau	seuil	niveau commun	niveau	seuil	niveau commun
$\sqrt{t}$	20	-	-	-	51	0,989	17
	30	-	-	-	44	0,993	29
	50	-	-	-	50	0,995	39
t^2	20	-	-	-	50	0,983	24
	30	-	-	-	50	0,990	50
	50	-	-	-	47	0,995	40
$\cos(\pi t)$	20	-	-	-	48	0,986	27
	30	-	-	-	50	0,992	36
	50	-	-	-	52	0,995	42
$\cos(2\pi t)$	20	-	-	-	50	0,938	74
	30	-	-	-	47	0,984	75
	50	-	-	-	44	0,991	68
$\cos(4\pi t)$	20	-	-	-	49	0,838	168
	30	-	-	-	48	0,823	194
	50	-	-	-	50	0,891	211
$\cos^2(2\pi t)$	20	-	-	-	49	0,914	115
	30	-	-	-	49	0,972	92
	50	-	-	-	47	0,985	97

des niveaux communs qui varient de 1% à 14%. Cela provient du fait que les seuils individuels des différentes fonctions sont souvent eux-mêmes très variables. Les fonctions les plus problématiques semblent être $\cos(4\pi t)$ et $\cos^2(2\pi t)$ dont les seuils individuels sont souvent éloignés de ceux obtenus pour les autres fonctions. Conséquemment, pour les tests basés sur l'une ou l'autre des approximations de la distribution de la distance entre g_1 et g_2, l'intention que nous avions de déterminer un seuil commun pour toute paire de fonctions ne fonctionne pas. Malgré cela, mentionnons tout de même que les puissances communes qui correspondent aux niveaux communs obtenus sont majoritairement bonnes. Les paires de fonctions faisant exception sont $\cos(4\pi t)$ et $\sin(4\pi t)$ puis $\cos(2\pi t)$ et $\cos(2\pi t) + t$ qui mènent à de faibles puissances communes lorsque $n < 50$.

À présent, commentons le tableau 3.8 qui comporte les résultats du test faisant intervenir les intervalles de confiance simultanés. Globalement, les résultats

TABLEAU 3.7. Puissances empiriques (×1000) pour le troisième test (les puissances communes sont celles associées aux seuils et niveaux communs du tableau 3.6).

$g_1(t)$	$g_2(t)$	n	Approximation d'Imhof		Approximation par point de selle	
			puissance	puissance commune	puissance	puissance commune
$\sqrt{t}$	t^2	20	-	-	988	974
		30	-	-	997	997
		50	-	-	998	998
$\sqrt{t}$	$g_1(t)+t$	20	-	-	964	959
		30	-	-	988	985
		50	-	-	992	990
$\sqrt{t}$	$g_1(t)+1$	20	-	-	993	987
		30	-	-	997	996
		50	-	-	999	999
t^2	$g_1(t)+t$	20	-	-	991	991
		30	-	-	993	993
		50	-	-	994	994
$\cos(\pi t)$	$g_1(t)+t$	20	-	-	983	973
		30	-	-	994	994
		50	-	-	998	998
$\cos(\pi t)$	$g_1(t)+1$	20	-	-	995	992
		30	-	-	999	999
		50	-	-	999	999
$\cos(2\pi t)$	$g_1(t)+t$	20	-	-	394	428
		30	-	-	801	876
		50	-	-	864	870
$\cos(2\pi t)$	$g_1(t)+1$	20	-	-	920	950
		30	-	-	932	936
		50	-	-	940	947
$\cos(4\pi t)$	$\sin(4\pi t)$	20	-	-	302	442
		30	-	-	656	873
		50	-	-	809	996
$\cos^2(2\pi t)$	$\sin^2(2\pi t)$	20	-	-	649	910
		30	-	-	970	980
		50	-	-	973	981

sont très satisfaisants. Le seuil commun utilisé pour ce test correspond à 97,5%. Cela signifie que si la proportion d'intervalles de confiance qui incluent la valeur 0 est inférieure à 97,5%, alors nous rejetons H_0, l'hypothèse d'égalité des fonctions. Notons que ce seuil commun a été choisi de façon à ce que qu'il soit valide pour chaque taille échantillonnale retenue, c'est-à-dire pour $n = 20, 30$ et 50. Les niveaux empiriques sont assez stables et se situent près de 1%. Les puissances correspondantes sont toutes très élevées, soit supérieures à 96% et elles augmen-

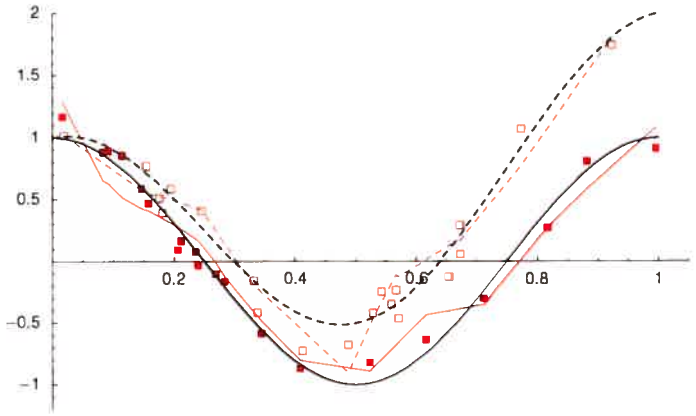


FIGURE 3.1. Paire de fonctions problématique lorsque $n = 20$: $g_1(t) = \cos(2\pi t)$ (trait plein noir), $g_2(t) = \cos(2\pi t) + t$ (pointillés noirs), y_1 (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_2 (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges).

tent avec n. De plus, contrairement à ce que nous avions remarqué précédemment aux tableaux 3.3, 3.5 et 3.7, les puissances pour les paires de fonctions $g_1(t) = \cos(2\pi t)$, $g_2(t) = \cos(2\pi t) + t$ et $g_1(t) = \cos(4\pi t)$, $g_2(t) = \sin(4\pi t)$ sont ici très satisfaisantes.

Tout juste avant de passer à la comparaison de ces résultats avec ceux obtenus par Angers (2003), clôturons la présente sous-section en fournissant quelques illustrations graphiques. Nous avons d'abord retenu un exemple de fonctions égales, soit $g_1(t) = g_2(t) = \sqrt{t}$ et ensuite deux autres exemples de fonctions différentes, soit $g_1(t) = \sqrt{t}$ et $g_2(t) = \sqrt{t} + 1$, puis $g_1(t) = \cos(4\pi t)$ et $g_2(t) = \sin(4\pi t)$. La valeur des tailles échantillonnales considérée pour ces exemples est $n = 30$.

Débutons avec la paire de fonctions égales. Pour une des 1000 réalisations effectuées, nous obtenons les figures 3.2 et 3.3. Plus précisément, dans la partie supérieure de la figure 3.2, nous pouvons constater l'allure des données simulées, ainsi que l'estimation par ondelettes correspondante pour chaque échantillon. Les vraies fonctions g_1 et g_2 apparaissent également sur le graphique. Au bas de cette même figure, nous avons les différences $\widehat{g}_1(t) - \widehat{g}_2(t)$, ainsi que les intervalles de confiance simultanés pour la différence entre les fonctions $\widehat{g}_1$ et $\widehat{g}_2$. Notons qu'afin d'empêcher tout chevauchement avec les fonctions $\widehat{g}_1$ et $\widehat{g}_2$, nous avons choisi d'abaisser l'axe des t du graphique des différences $\widehat{g}_1(t) - \widehat{g}_2(t)$ et des intervalles

TABLEAU 3.8. Niveaux communs et puissances communes empiriques ($\times 1000$) pour le quatrième test.

Fonctions	n	niveau/puissance communs	Fonctions	n	niveau/puissance communs
$g_1(t) = g_2(t) = \sqrt{t}$	20	6	$g_1(t) = \sqrt{t}\,;\ g_2(t) = g_1(t) + 1$	20	992
	30	5		30	998
	50	6		50	1000
$g_1(t) = g_2(t) = t^2$	20	10	$g_1(t) = t^2\,;\ g_2(t) = t$	20	982
	30	7		30	990
	50	5		50	1000
$g_1(t) = g_2(t) = \cos(\pi t)$	20	9	$g_1(t) = \cos(\pi t)\,;\ g_2(t) = g_1(t) + t$	20	986
	30	9		30	995
	50	8		50	1000
$g_1(t) = g_2(t) = \cos(2\pi t)$	20	12	$g_1(t) = \cos(\pi t)\,;\ g_2(t) = g_1(t) + 1$	20	990
	30	13		30	996
	50	11		50	1000
$g_1(t) = g_2(t) = \cos(4\pi t)$	20	16	$g_1(t) = \cos(2\pi t)\,;\ g_2(t) = g_1(t) + t$	20	971
	30	14		30	987
	50	9		50	1000
$g_1(t) = g_2(t) = \cos^2(2\pi t)$	20	10	$g_1(t) = \cos(2\pi t)\,;\ g_2(t) = g_1(t) + 1$	20	990
	30	11		30	998
	50	10		50	1000
$g_1(t) = \sqrt{t}\,;\ g_2(t) = t^2$	20	980	$g_1(t) = \cos(4\pi t)\,;\ g_2(t) = \sin(4\pi t)$	20	956
	30	989		30	990
	50	1000		50	999
$g_1(t) = \sqrt{t}\,;\ g_2(t) = g_1(t) + t$	20	992	$g_1(t) = \cos^2(2\pi t)\,;\ g_2(t) = \sin^2(2\pi t)$	20	966
	30	994		30	987
	50	1000		50	1000

de confiance de 1 unité (ligne horizontale noire et pointillée). Ainsi, cela explique pourquoi les intervalles sont centrés en -1, plutôt qu'en 0. Pour sa part, la figure 3.3 illustre les approximations d'Imhof et par point de selle pour la densité de la distance observée et pour celle sous H_0. Les relations à établir entre les tests proposés et ces figures sont les suivantes. D'abord, la figure 3.3 montre une densité observée très proche de celle sous H_0 et ce, pour les deux méthodes d'approximations. De plus, remarquons qu'à la figure 3.2, tous les intervalles de confiance contiennent la valeur -1, c'est-à-dire l'équivalent du 0. Ces constatations reflètent bien la proximité des fonctions estimées.

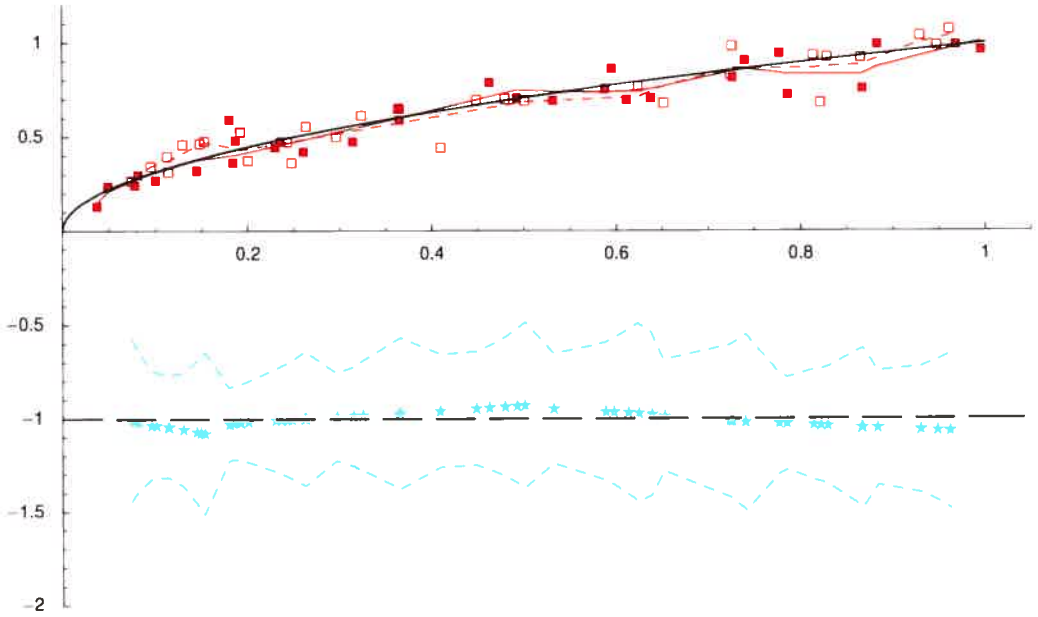


FIGURE 3.2. Graphique de $g_1(t) = g_2(t) = \sqrt{t}$ (trait plein noir), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

Passons maintenant au premier exemple de fonctions différentes et référons nous aux figures 3.4 et 3.5. Notons que toutes les explications fournies précédemment pour les figures 3.2 et 3.3 s'appliquent ici. Comparativement à ce que nous avons constaté auparavant à la figure 3.3, nous remarquons qu'à la figure 3.5, l'approximation de la densité observée est beaucoup plus éloignée de celle sous

H_0. Cette observation coïncide avec nos attentes, c'est-à-dire plus les fonctions g_1 et g_2 sont distantes, plus la densité observée est éloignée de celle sous H_0. Pour leur part, les intervalles de confiance simultanés excluent tous la valeur -1, c'est-à-dire l'équivalent du 0. Cela confirme la disparité les fonctions g_1 et g_2.

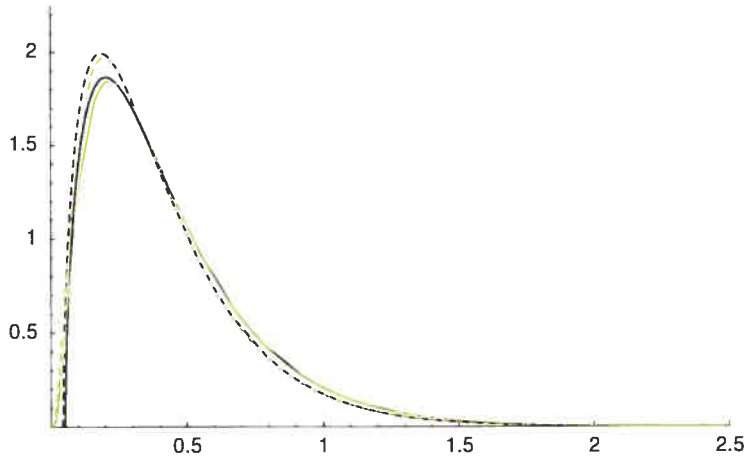


FIGURE 3.3. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

Poursuivons avec le second exemple de fonctions différentes en nous concentrant sur les figures 3.6 et 3.7. D'abord, en comparant la figure 3.7 à la figure 3.5, notons que l'approximation de la densité observée et celle sous H_0 sont plus rapprochées. Cela provient du fait que contrairement aux fonctions $g_1(t) = \sqrt{t}$ et $g_2(t) = \sqrt{t} + 1$ qui ne s'entrecroisent jamais sur le domaine $[0, 1]$, les fonctions $g_1(t) = \cos(4\pi t)$ et $g_2(t) = \sin(4\pi t)$ s'entrecroisent à quatre reprises. L'éloignement entre les fonctions n'est donc pas constant et tend à être nul pour certaines valeurs de t. La figure 3.6 illustre clairement l'impact de ces entrecroisements. Afin d'éviter toute confusion, précisons que l'axe des t du graphique des différences $\widehat{g}_1(t) - \widehat{g}_2(t)$ et des intervalles de confiance a été abaissé de 4 unités (ligne horizontale noire et pointillée) afin d'empêcher tout chevauchement avec les fonctions $\widehat{g}_1$ et $\widehat{g}_2$. Nous pouvons effectivement voir que dans un voisinage de t où les fonctions s'entrecroisent, la différence $\widehat{g}_1(t) - \widehat{g}_2(t)$ est quasiment nulle et les intervalles de confiance correspondants incluent -4, soit l'équivalent du 0.

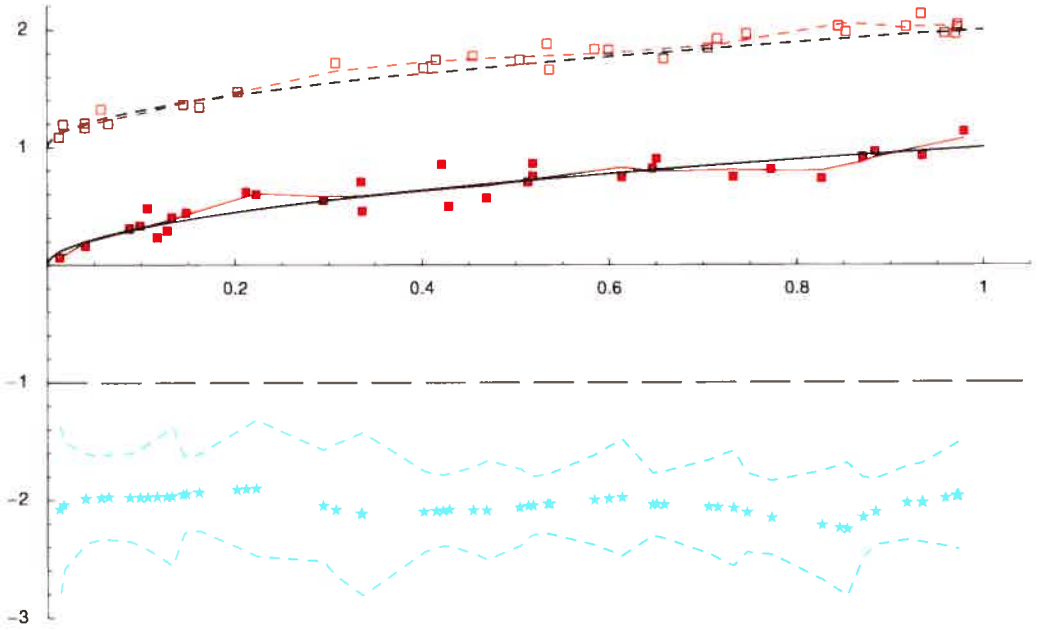


FIGURE 3.4. Graphique de $g_1(t) = \sqrt{t}$ (trait plein noir), $g_2(t) = \sqrt{t} + 1$ (pointillés noirs), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (·) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

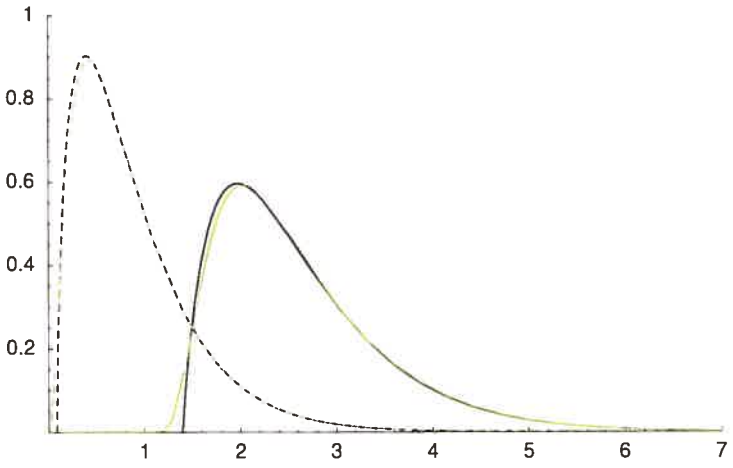


FIGURE 3.5. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

Ceci met fin aux exemples illustratifs de paires de fonctions commentés en détail. Prenez note que des illustrations supplémentaires pour les treize paires de fonctions restantes sont fournies à l'Annexe A.

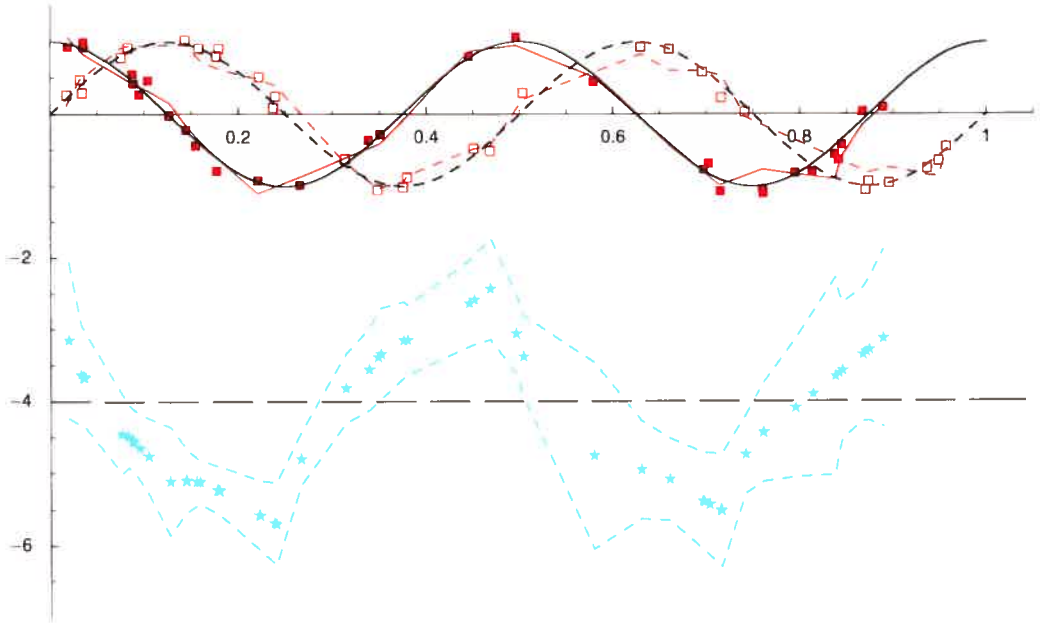


FIGURE 3.6. Graphique de $g_1(t) = \cos(4\pi t)$ (trait plein noir), $g_2(t) = \sin(4\pi t)$ (pointillés noirs), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

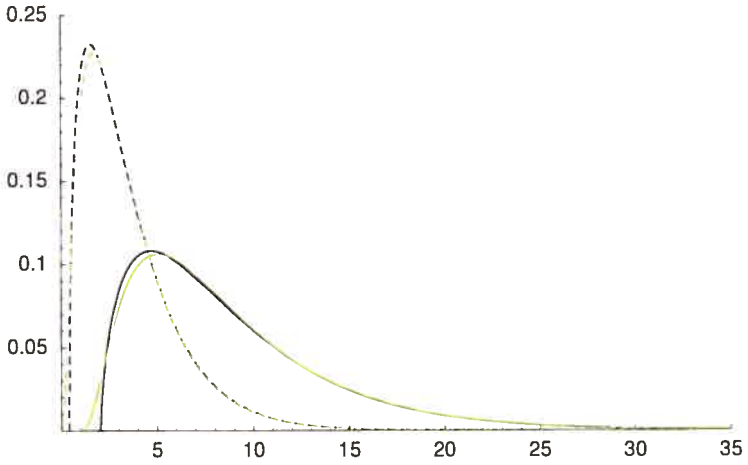


FIGURE 3.7. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

3.1.2. Comparaison avec Angers (2003)

La présente sous-section vise à comparer les résultats des tableaux 3.2 à 3.7 avec ceux des tableaux 3.9 et 3.10 correspondant à ceux obtenus par Angers (2003). Tout d'abord, précisons que le contexte de simulation des deux études est exactement le même. De plus, la comparaison des résultats est directe puisque les fonctions utilisées sont également les mêmes.

TABLEAU 3.9. Niveaux individuels empiriques ($\times 1000$) du test basé sur l'approximation d'Edgeworth (voir Angers, 2003).

$g_1(t) = g_2(t)$	n	niveau	seuil
$\sqrt{t}$	20	53	0.0625
	30	49	0.0825
	50	48	0.1000
t^2	20	52	0.0900
	30	51	0.1200
	50	50	0.1325
$\cos(\pi t)$	20	51	0.0900
	30	46	0.0425
	50	50	0.0725
$\cos(2\pi t)$	20	51	0.0900
	30	51	0.0425
	50	53	0.0525
$\cos(4\pi t)$	20	50	0.0725
	30	50	0.1025
	50	49	0.0325
$\cos^2(2\pi t)$	20	51	0.0750
	30	50	0.1900
	50	49	0.0450

Parmi les tests développés par Angers (2003), nous n'utilisons pas celui reposant sur le facteur de Bayes pour la comparaison puisque nous savons, d'après l'article, que ce dernier est trop conservateur. Le test auquel nous nous rapportons pour la comparaison est celui reposant sur une approximation de la distribution de ξ. Comme nous l'avons déjà mentionné au chapitre 2, la méthode d'approximation utilisée dans Angers (2003) est celle d'Edgeworth. L'expression pour l'approximation de la fonction de répartition $F(\xi|\boldsymbol{y}_1, \boldsymbol{y}_2)$ est décrite par l'équation (2.3.5) et

le test est donné par

$$\sup_{t\in T}\left[F_0(t)-F_1(t)\right], \tag{3.1.1}$$

où $F_i(t)$ dénote la fonction de répartition de ξ sous l'hypothèse H_i, $i=0,1$. Ainsi, lorsque la valeur obtenue pour l'équation (3.1.1) est élevée, l'hypothèse H_0 est rejetée.

TABLEAU 3.10. Puissances individuelles empiriques ($\times 1000$) du test basé sur l'approximation d'Edgeworth au niveau 5% (voir Angers, 2003).

Fonctions	n	puissance
$g_1(t)=\sqrt{t}; g_2(t)=t^2$	20	981
	30	998
	50	992
$g_1(t)=\sqrt{t}; g_2(t)=g_1(t)+t$	20	976
	30	990
	50	1000
$g_1(t)=\sqrt{t}; g_2(t)=g_1(t)+1$	20	991
	30	1000
	50	1000
$g_1(t)=t^2; g_2(t)=g_1(t)+t$	20	984
	30	996
	50	1000
$g_1(t)=\cos(\pi t); g_2(t)=g_1(t)+t$	20	976
	30	995
	50	992
$g_1(t)=\cos(\pi t); g_2(t)=g_1(t)+1$	20	987
	30	1000
	50	1000
$g_1(t)=\cos(2\pi t); g_2(t)=g_1(t)+t$	20	743
	30	975
	50	1000
$g_1(t)=\cos(2\pi t); g_2(t)=g_1(t)+1$	20	743
	30	975
	50	992
$g_1(t)=\cos(4\pi t); g_2(t)=\sin(4\pi t)$	20	831
	30	899
	50	999
$g_1(t)=\cos^2(2\pi t); g_2(t)=\sin^2(2\pi t)$	20	924
	30	994
	50	1000

En comparant les résultats des divers tests, voici ce que nous concluons. Lorsque $n = 20$, les seuils individuels obtenus pour le test développé par Angers (2003) sont moins variables que ceux des trois tests du présent mémoire. Toutefois, si $n = 30$, l'inverse se produit ; les seuils du deuxième test avec l'approximation par point de selle sont très peu variables comparativement à ceux obtenus par Angers (2003). Enfin, lorsque $n = 50$, tous les tests mènent à des seuils de variabilité similaire. En ce qui a trait aux puissances individuelles obtenues, mentionnons qu'elles sont semblables pour la majorité des paires de fonctions étudiées. Cependant, les puissances obtenues par Angers (2003) pour les paires de fonctions $g_1(t) = \cos(2\pi t)$, $g_2(t) = \cos(2\pi t) + t$ et $g_1(t) = \cos(4\pi t)$, $g_2(t) = \sin(4\pi t)$ sont nettement plus élevées.

3.2. Exemple avec données réelles

L'objectif de cette dernière sous-section est d'appliquer la théorie à un contexte de données réelles. Le jeu de données que nous avons sélectionné a été obtenu suite à une étude portant sur l'utilisation de l'éthanol comme carburant dans un moteur d'automobile expérimental à un cylindre (voir Cleveland, 1993). Notons que ces données figurent également parmi celles disponibles dans le logiciel *S-Plus*.

La variable réponse y correspond à la concentration de monoxyde d'azote et de dioxyde d'azote dans l'échappement du moteur. Elle est mesurée pour divers rapports d'équivalence t et rapports de compression du moteur. À titre informatif, le rapport d'équivalence du moteur constitue une mesure de la richesse du mélange en air et en éthanol dans le carburant utilisé.

Ainsi, l'expérience consiste à déterminer l'impact de différentes configurations du moteur sur la concentration d'oxydes d'azote dans l'échappement. La motivation des chercheurs provient du fait que les oxydes d'azote sont des substances polluantes pour l'environnement et potentiellement dangereuses pour la santé humaine.

Étant donné que le rapport de compression ne prend que quelques valeurs, les observations ont été séparées en deux catégories. Lorsque le rapport de compression est inférieur à 10, nous considérons qu'il est « faible ». Autrement, il

est supérieur ou égal à 10 et nous disons qu'il est « élevé ». Ces catégories nous permettent ainsi de définir deux échantillons. Pour chacun d'entre eux, nous modélisons y en fonction de t et comparons les fonctions obtenues.

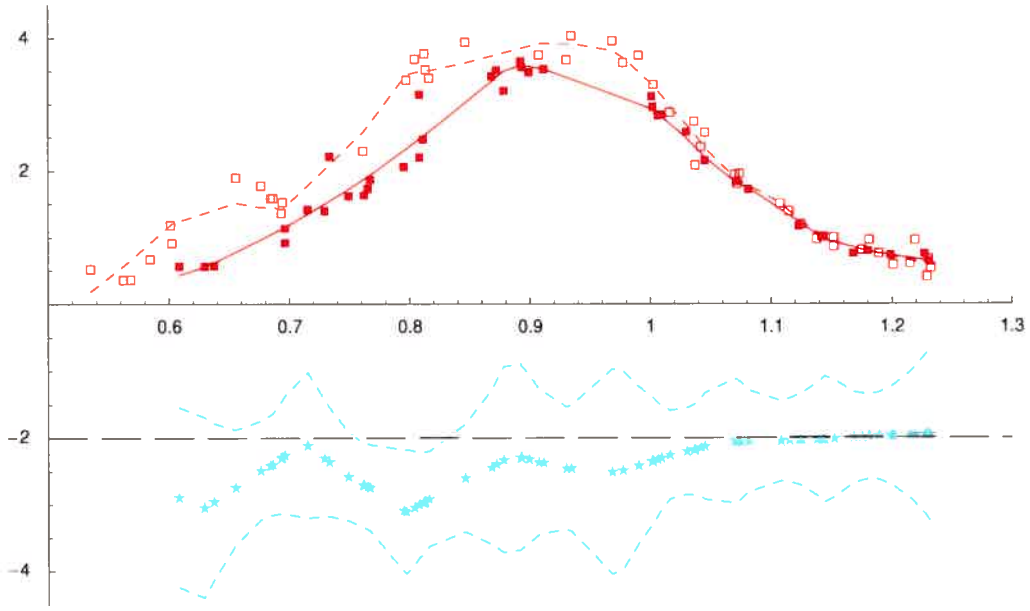


FIGURE 3.8. Concentration d'oxydes d'azote (y) en fonction du rapport d'équivalence (t) : observations pour le premier échantillon (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), observations pour le second échantillon (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

Le premier échantillon formé des données pour lesquelles le rapport de compression est « faible » comporte 39 observations. Le second échantillon possède pour sa part 49 observations. La portion supérieure de la figure 3.8 illustre ces observations ainsi que l'estimation par ondelettes correspondante. À première vue, les fonctions apparaissent différentes, surtout lorsque le rapport d'équivalence est inférieur à 1. Le résultat du test basé sur le concept d'intervalles de confiance simultanés pour la différence entre les fonctions est présenté au bas de cette même figure. Pour éviter tout chevauchement avec les fonctions estimées, l'axe des t du graphique des différences $\widehat{g}_1(t) - \widehat{g}_2(t)$ et des intervalles de confiance a été abaissé de 2 unités (ligne horizontale noire et pointillée). Nous remarquons que les intervalles de confiance qui excluent la valeur -2, c'est-à-dire l'équivalent du 0,

correspondent à ceux dont le rapport d'équivalence est situé entre 0,75 et 0,825. Pour toute autre valeur du rapport d'équivalence, les intervalles calculés incluent la valeur -2; soit parce que les fonctions estimées sont similaires (ex. $t > 1$), ou alors parce qu'il n'y a pas assez d'observations provenant des deux échantillons pour que la différence entre les fonctions estimées soit détectée (ex. $0{,}9 \leq t \leq 1$). Au total, 84% des intervalles calculés contiennent la valeur -2. Comme cette proportion est inférieure au seuil 97,5%, nous concluons que les fonctions des échantillons diffèrent. Ce résultat coïncide avec celui obtenu par Angers (2003) au moyen d'une approche bootstrap.

CONCLUSION

Faisant suite à l'article de Angers (2003), le but de ce mémoire était de développer de nouvelles techniques nous permettant de déterminer si deux fonctions sont égales ou non. Dans un premier temps, nous avons décomposé les fonctions à l'aide de la base d'ondelettes de Daubechies et nous avons aussi proposé des densités *a priori* pour les coefficients d'ondelettes. Par la suite, nous avons approximé la distribution de la distance entre les fonctions, basée sur la notion de distance euclidienne, à l'aide d'un mélange de distributions khi-deux non centrées. Basés sur ces approximations, trois tests ont été suggérés. Par la suite, nous avons également présenté un dernier test reposant sur le concept d'intervalles de confiance simultanés.

Les résultats obtenus au chapitre 3 confirment que les trois premiers tests proposés ne nous permettent pas de trouver un seuil commun pour toute paire de fonctions. En effet, nous avons remarqué que pour des fonctions différentes qui s'entrecroisent à quelques reprises sur le domaine, par exemple $g_1(t) = \cos(4\pi t)$ et $g_2(t) = \sin(4\pi t)$, le seuil tend à être éloigné de ceux obtenus lorsque les fonctions ne s'entrecroisent pas. Conséquemment, l'intention que nous avions de déterminer un seuil commun pour toute paire de fonctions devient illusoire. Soulignons qu'une telle problématique survient également dans la contribution de Angers (2003). Afin de pallier à celle-ci, l'auteur recours à une méthode de réechantillonnage pour confirmer ou réfuter l'égalité des deux courbes. Dans ce mémoire, la méthode recommandée pour la comparaison de deux courbes empiriques est celle faisant intervenir les intervalles de confiance simultanés pour la différence entre les fonctions.

En terminant, énumérons quelques pistes de recherche. Tout d'abord, les tests présentés dans le présent mémoire sont construits de manière à déterminer si les fonctions sont égales ou non sur la totalité de leurs domaines de définition. Toutefois, certaines situations sont telles que nous ne nous intéressons pas au domaine au complet, mais plutôt à une portion de ce celui-ci. Citons l'exemple du chercheur qui désire comparer le rendement de deux médicaments. Il sait déjà qu'au début de l'expérience, les médicaments auront un effet similaire. Il sait aussi qu'à la toute fin de l'expérience, l'effet des médicaments sera complètement atténué. Cependant, il est intéressé à comparer le rendement des médicaments entre ces périodes.

Enfin, une autre piste de recherche à mentionner provient du fait que la théorie présentée dans ce mémoire permet de comparer deux fonctions de régression à la fois. Celle-ci pourrait éventuellement être élargie de façon à pouvoir comparer plusieurs fonctions simultanément.

Annexe A

GRAPHIQUES DES PAIRES DE FONCTIONS

Cette première annexe présente les graphiques des différentes paires de fonctions utilisées dans l'étude de simulation décrite au chapitre 3. Pour obtenir une explication détaillée de telles figures, consultez la sous-section 3.1.1. Pour le moment, rappelons simplement que les tailles échantillonnales retenues correspondent à $n_1 = n_2 = n = 30$ et que la modélisation des fonctions est réalisée à l'aide de la base d'ondelettes de Daubechies d'ordre 2.

A.1. Paires de fonctions égales

A.1.1. Fonctions $g_1(t) = g_2(t) = t^2$

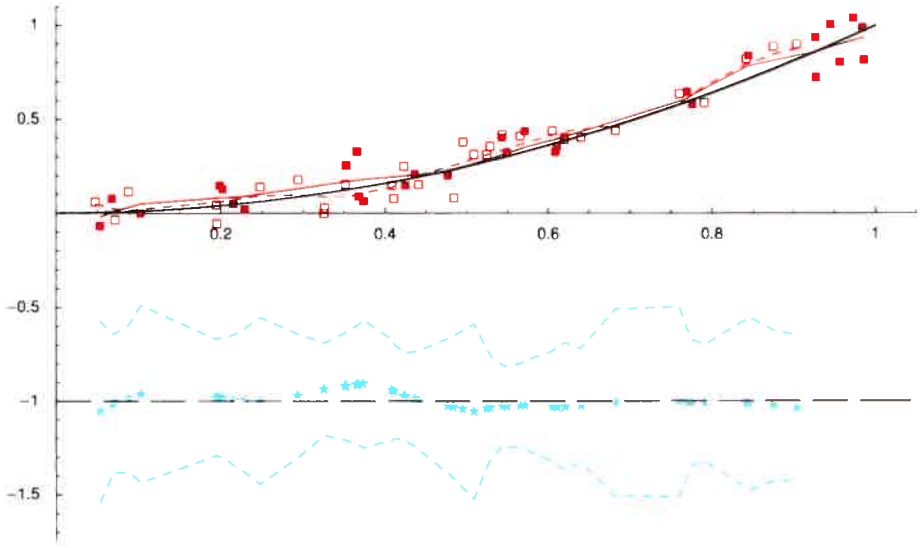


FIGURE A.1. Graphique de $g_1(t) = g_2(t) = t^2$ (trait plein noir), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

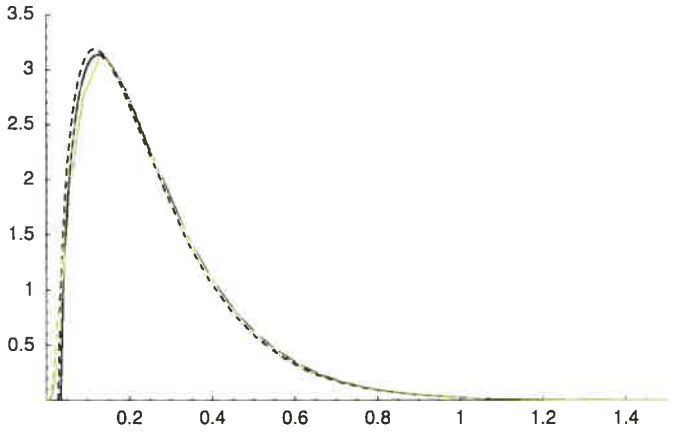


FIGURE A.2. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

A.1.2. Fonctions $g_1(t) = g_2(t) = \cos(\pi t)$

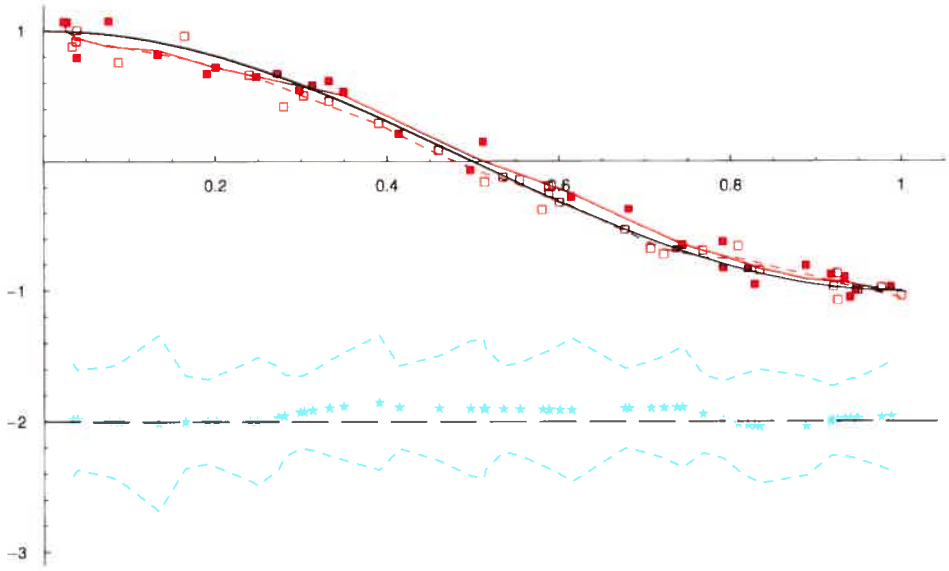


FIGURE A.3. Graphique de $g_1(t) = g_2(t) = \cos(\pi t)$ (trait plein noir), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

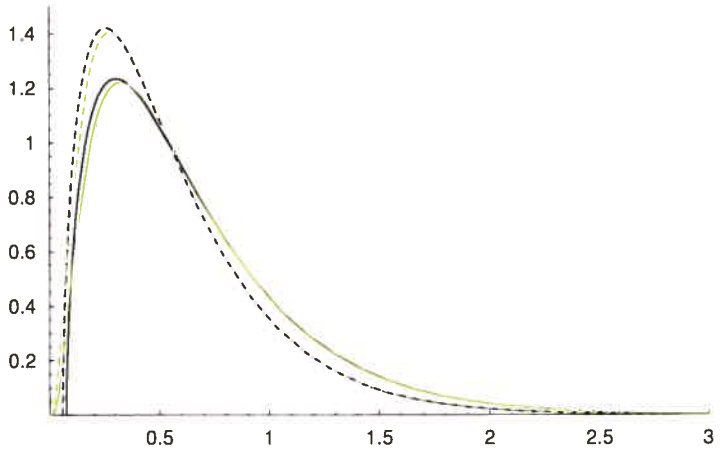


FIGURE A.4. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

A.1.3. Fonctions $g_1(t) = g_2(t) = \cos(2\pi t)$

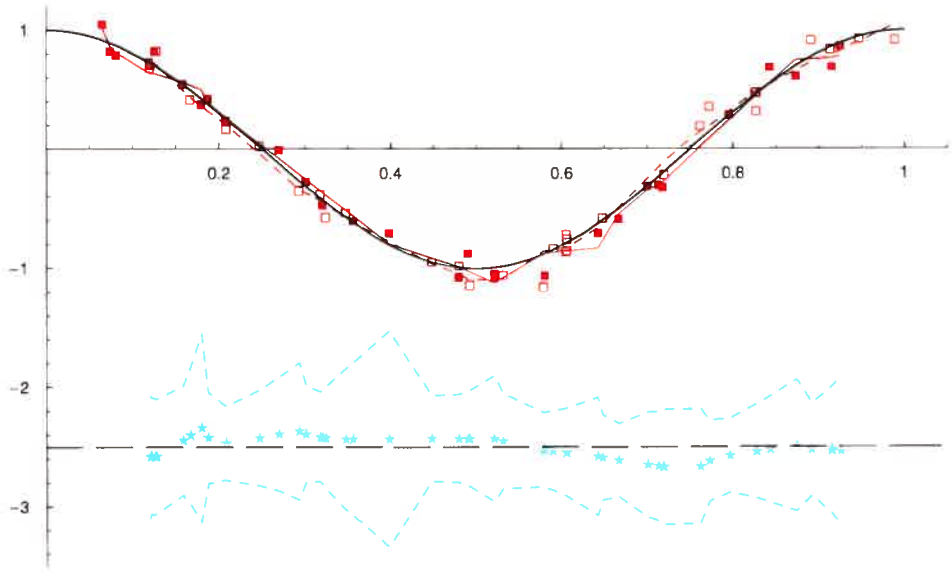


FIGURE A.5. Graphique de $g_1(t) = g_2(t) = \cos(2\pi t)$ (trait plein noir), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

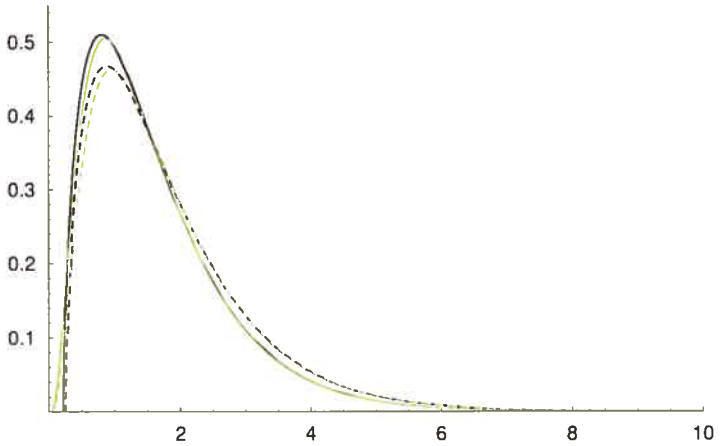


FIGURE A.6. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

A.1.4. Fonctions $g_1(t) = g_2(t) = \cos(4\pi t)$

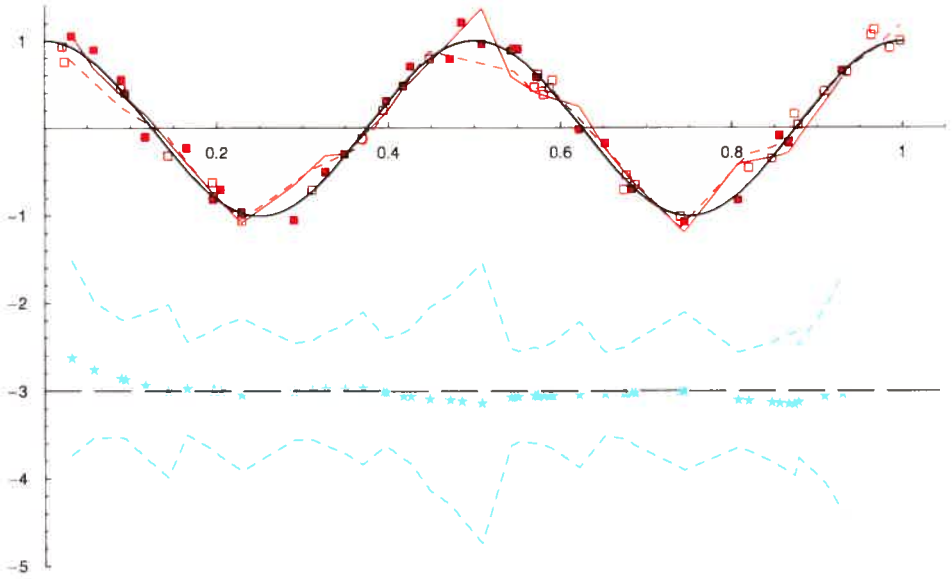


FIGURE A.7. Graphique de $g_1(t) = g_2(t) = \cos(4\pi t)$ (trait plein noir), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (·) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

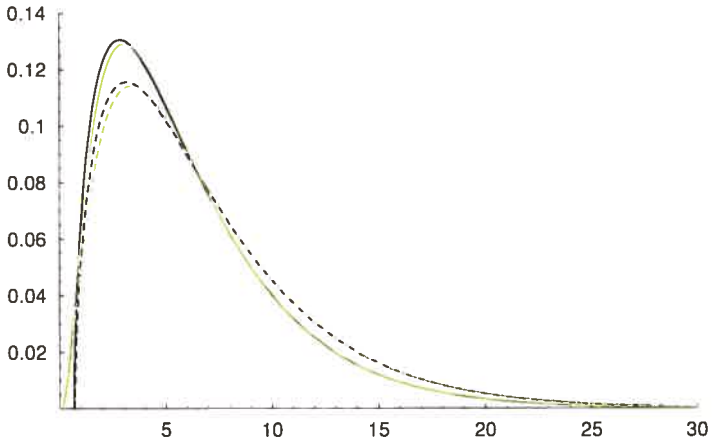


FIGURE A.8. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

A.1.5. Fonctions $g_1(t) = g_2(t) = \cos^2(2\pi t)$

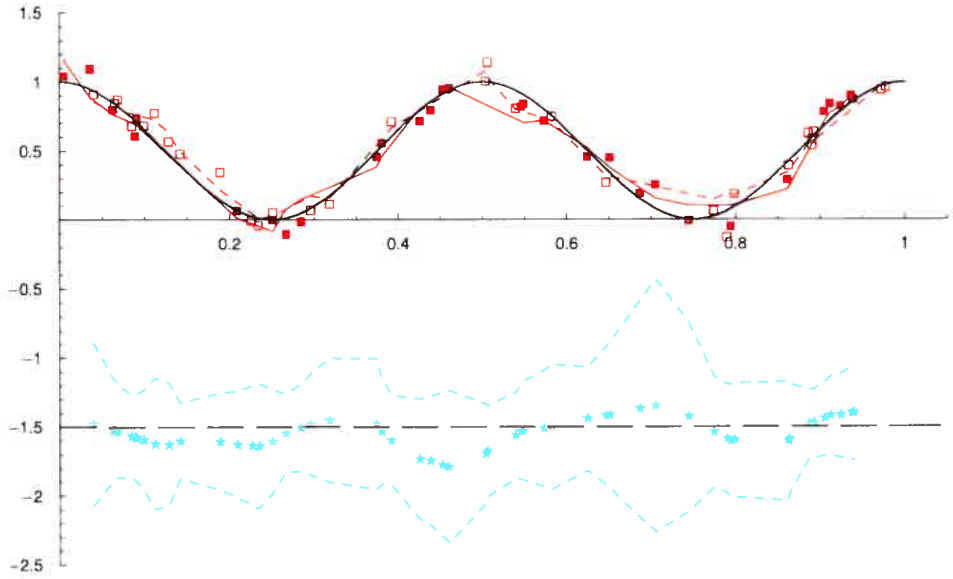


FIGURE A.9. Graphique de $g_1(t) = g_2(t) = \cos^2(2\pi t)$ (trait plein noir), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

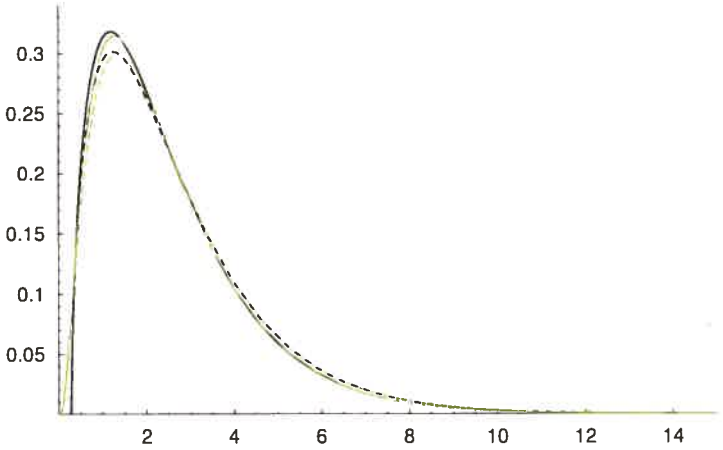


FIGURE A.10. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

A.2. Paires de fonctions différentes

A.2.1. Fonctions $g_1(t) = \sqrt{t}$ et $g_2(t) = t^2$

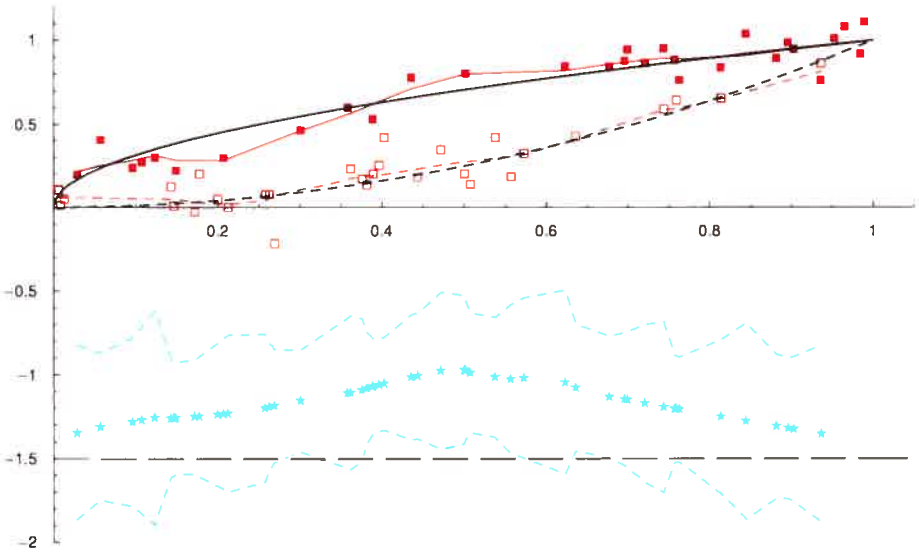


FIGURE A.11. Graphique de $g_1(t) = \sqrt{t}$ (trait plein noir), $g_2(t) = t^2$ (pointillés noirs), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (·) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

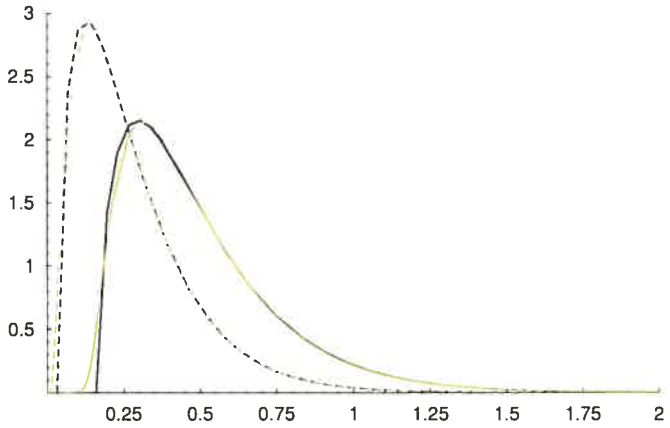


FIGURE A.12. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

A.2.2. Fonctions $g_1(t) = \sqrt{t}$ et $g_2(t) = \sqrt{t} + t$

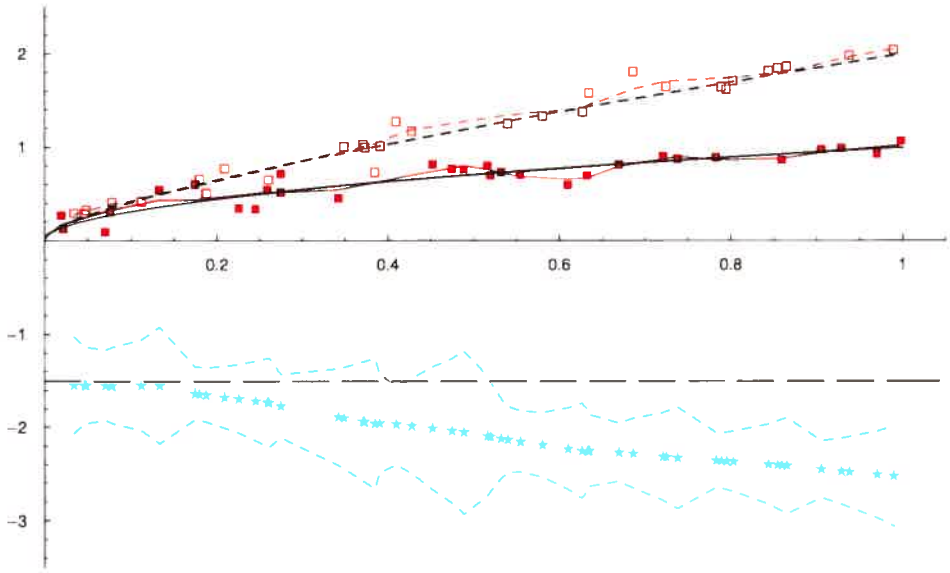


FIGURE A.13. Graphique de $g_1(t) = \sqrt{t}$ (trait plein noir), $g_2(t) = \sqrt{t} + t$ (pointillés noirs), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

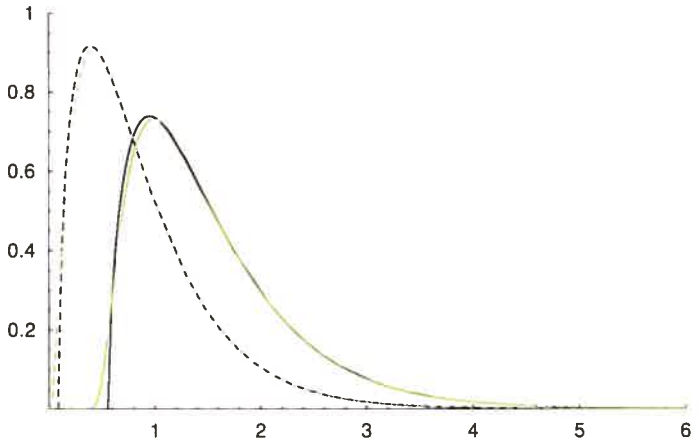


FIGURE A.14. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

A.2.3. Fonctions $g_1(t) = t^2$ et $g_2(t) = t^2 + t$

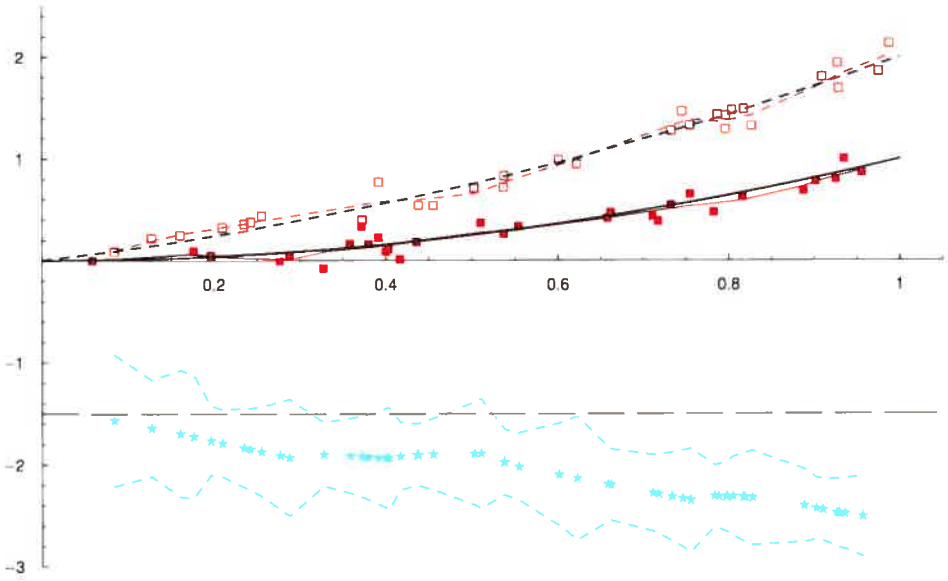


FIGURE A.15. Graphique de $g_1(t) = t^2$ (trait plein noir), $g_2(t) = t^2 + t$ (pointillés noirs), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\tilde{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

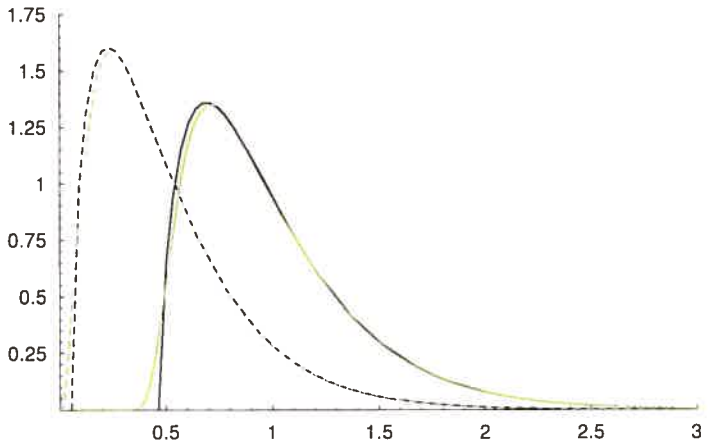


FIGURE A.16. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

A.2.4. Fonctions $g_1(t) = \cos(\pi t)$ et $g_2(t) = \cos(\pi t) + t$

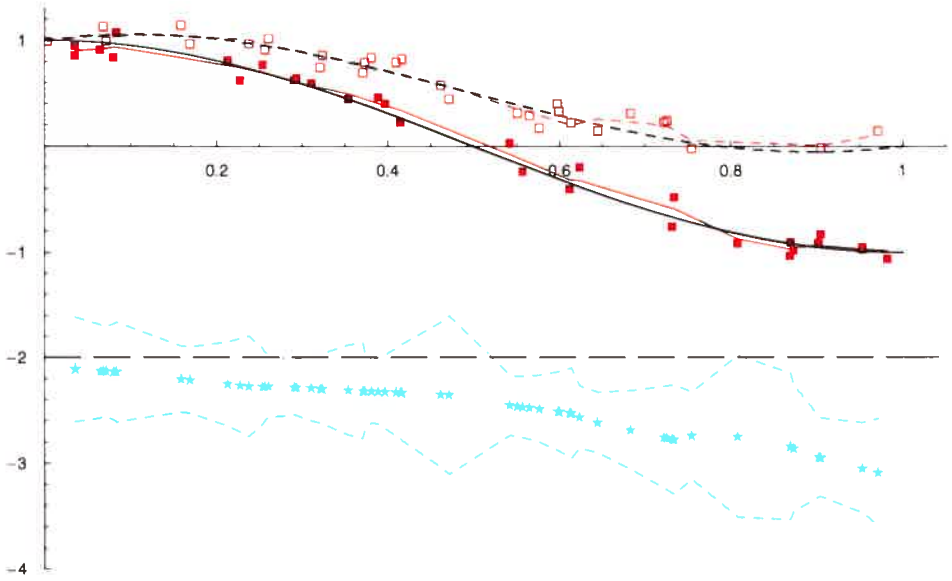


FIGURE A.17. Graphique de $g_1(t) = \cos(\pi t)$ (trait plein noir), $g_2(t) = \cos(\pi t) + t$ (pointillés noirs), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

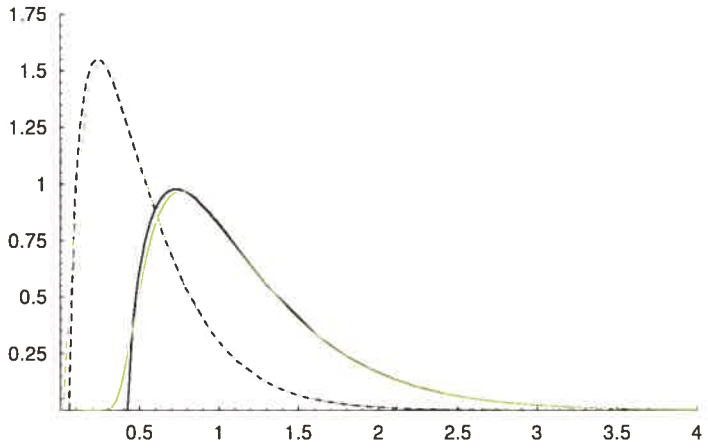


FIGURE A.18. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

A.2.5. Fonctions $g_1(t) = \cos(\pi t)$ et $g_2(t) = \cos(\pi t) + 1$

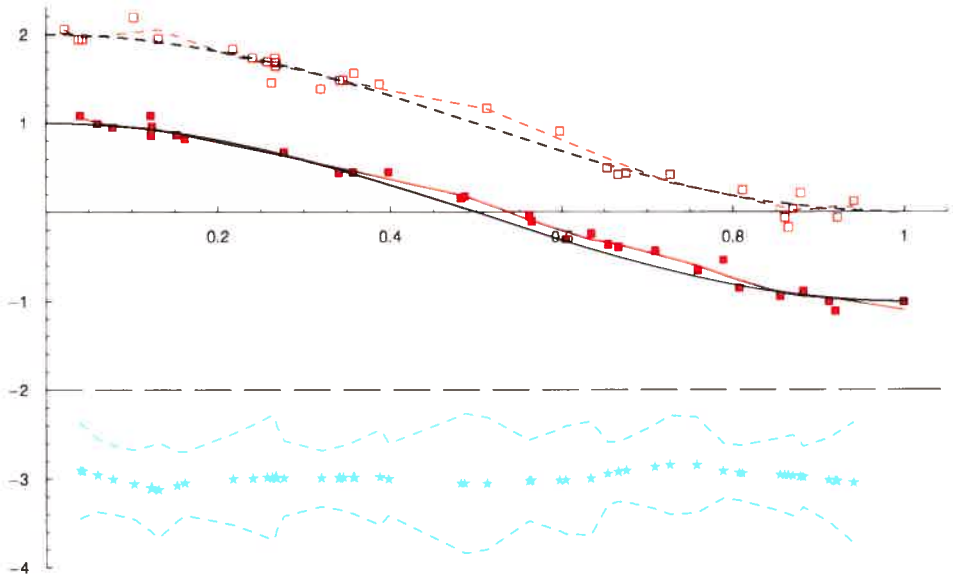


FIGURE A.19. Graphique de $g_1(t) = \cos(\pi t)$ (trait plein noir), $g_2(t) = \cos(\pi t) + 1$ (pointillés noirs), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

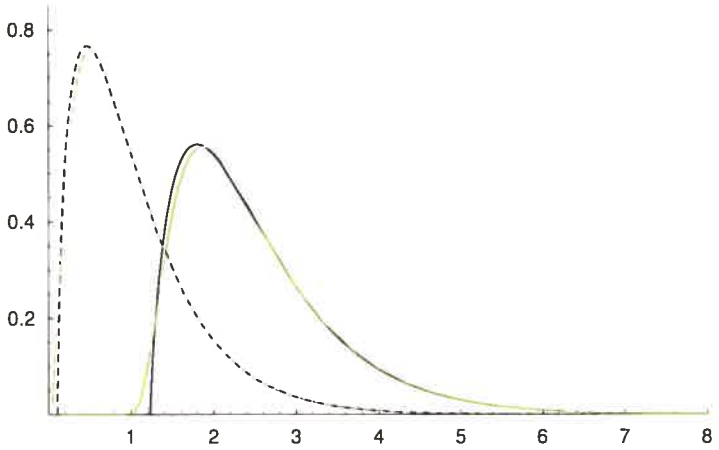


FIGURE A.20. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

A.2.6. Fonctions $g_1(t) = \cos(2\pi t)$ et $g_2(t) = \cos(2\pi t) + t$

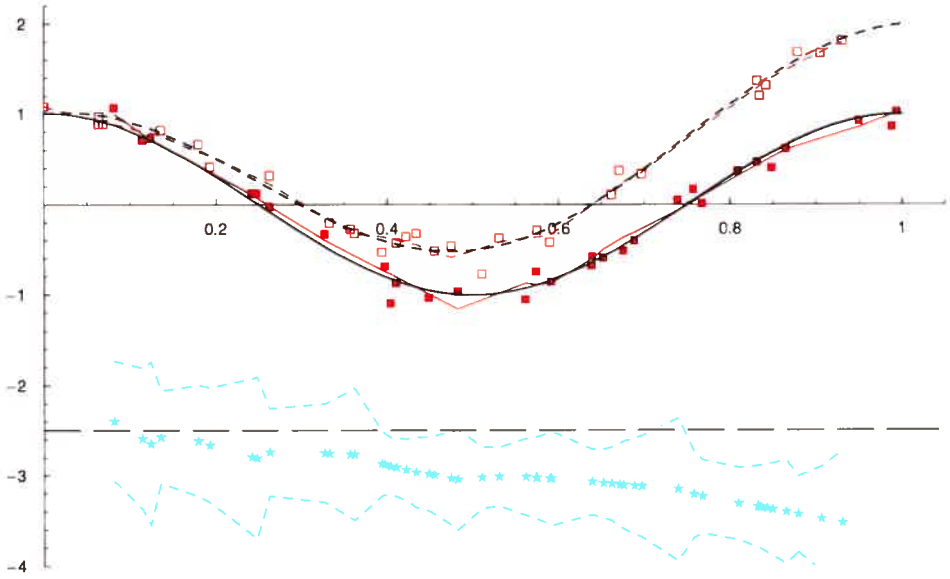


FIGURE A.21. Graphique de $g_1(t) = \cos(2\pi t)$ (trait plein noir), $g_2(t) = \cos(2\pi t) + t$ (pointillés noirs), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

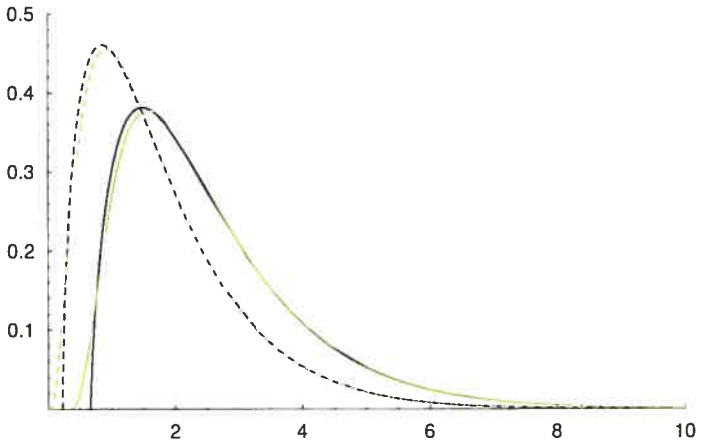


FIGURE A.22. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

A.2.7. Fonctions $g_1(t) = \cos(2\pi t)$ et $g_2(t) = \cos(2\pi t) + 1$

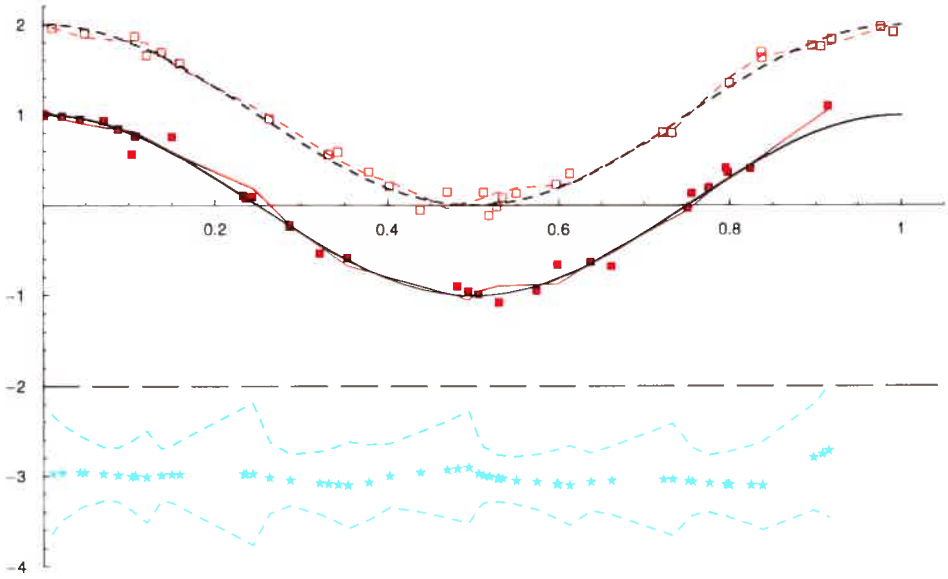


FIGURE A.23. Graphique de $g_1(t) = \cos(2\pi t)$ (trait plein noir), $g_2(t) = \cos(2\pi t) + 1$ (pointillés noirs), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (⋆) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

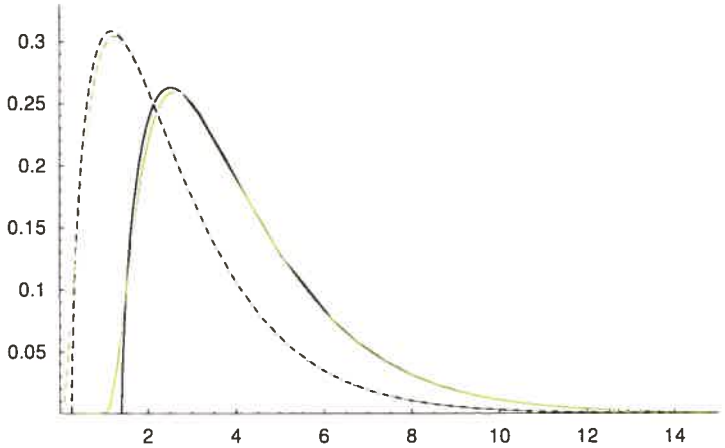


FIGURE A.24. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

A.2.8. Fonctions $g_1(t) = \cos^2(2\pi t)$ et $g_2(t) = \sin^2(2\pi t)$

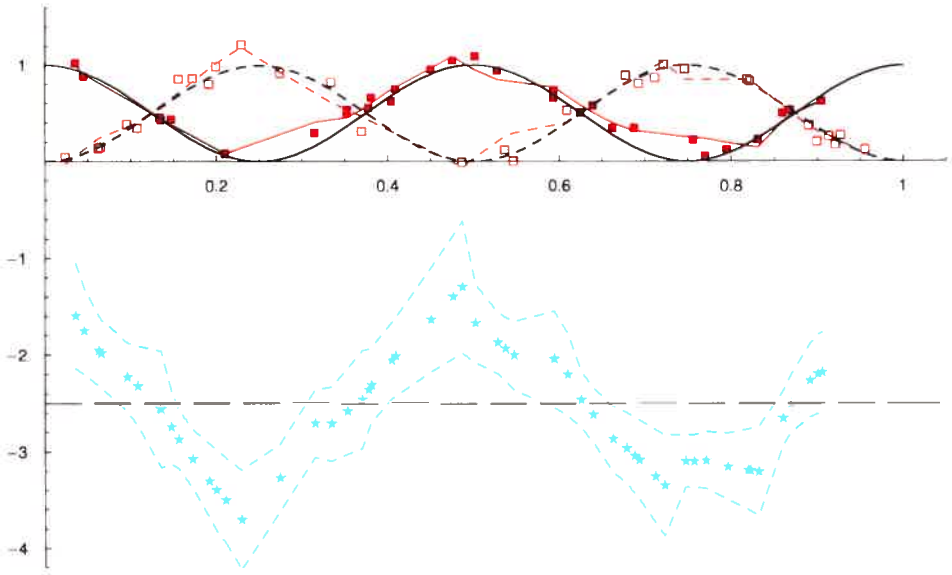


FIGURE A.25. Graphique de $g_1(t) = \cos^2(2\pi t)$ (trait plein noir), $g_2(t) = \sin^2(2\pi t)$ (pointillés noirs), y_{1i} (■), $\hat{y}_1 = \hat{g}_1(t_1)$ (trait plein rouge), y_{2i} (□), $\hat{y}_2 = \hat{g}_2(t_2)$ (pointillés rouges), $\hat{g}_1(t) - \hat{g}_2(t)$ (·) et bornes inférieures et supérieures des intervalles de confiance simultanés pour la différence entre les fonctions (pointillés bleus).

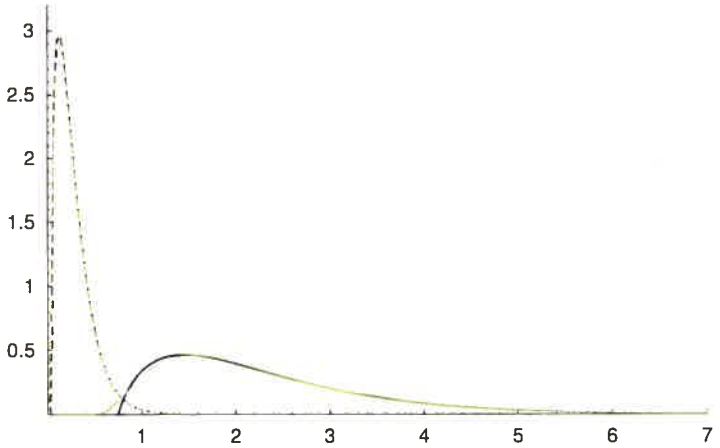


FIGURE A.26. Approximation d'Imhof de la distribution de la distance : observée (trait plein noir) et sous H_0 (pointillés noirs). Approximation par point de selle de la distribution de la distance : observée (trait plein vert) et sous H_0 (pointillés verts).

Annexe B

PROGRAMMATION

Cette seconde annexe contient l'ensemble des programmes utilisés pour la génération des résultats présentés dans ce mémoire. Précisons que le logiciel utilisé est *Mathematica 5.0*.

Dans l'ordre, nous présentons les programmes pour la modélisation par base d'ondelettes, l'approximation de la distribution de ξ selon la méthode d'Imhof et celle par point de selle, puis finalement celui pour le calcul des intervalles de confiance simultanés. Précisons que les différentes « librairies » à inclure sont les suivantes :

1) LinearAlgebra « MatrixManipulation » ;

2) Statistics « ContinuousDistributions » ;

3) Statistics « DataManipulation » ;

4) Statistics « DescriptiveStatistics.

B.1. MODÉLISATION PAR BASE D'ONDELETTES

```
(* Filtres d'ondelettes de Daubechies d'ordre 2 *)

Daub = {0.4829629, 0.8365163, 0.2241439, -0.1294095};


(* Algorithme pyramidal de Daubechies-Lagarias permettant de calculer les
   fonctions phi(x) et psi(x). *)

dyad[x_, n_] := RealDigits[x, 2, n, -1][[1]];
T0[fil_] := Module[{n, domain},
                   n = Length[fil]; domain = n - 1;
```

```
                    Table[If[(2*i - j > n || 2*i - j < 1),
                            0, Sqrt[2]*fil[[2*i - j]]],
                          {i, 1, domain}, {j, 1, domain}]];
T1[fil_] := Module[{n, domain},
                    n = Length[fil]; domain = n - 1;
                    Table[If[(2*i - j + 1 > n || 2*i - j + 1 < 1),
                            0, Sqrt[2] fil[[2*i - j + 1]]],
                          {i, 1, domain}, {j, 1, domain}]];

PhiD[x_, fil_, n_] := Module[{daun,Cn,int,dec,dy,tt,prodmat,onevec,phis},
                              daun = Length[fil]/2; Cn = Length[fil] - 1;
                              int = Floor[x]; dec = x - int;
                              dy = dyad[dec, n]; tt = {T0[fil], T1[fil]};
                              prodmat = IdentityMatrix[Cn];
                              For[i = 1, i <= n, i++,
                                  prodmat = prodmat.tt[[dy[[i]] + 1]]];
                                  onevec = Table[1, {i, 1, Cn}];
                                  phis = prodmat.onevec/Cn;
                                  phis[[int + 1]]];
PhiD[x_, fil_, n_] := 0 /; (x <= 0 || x >= Length[fil] - 1);

PsiD[x_, fil_, n_] := Module[{daun, Cn, xx, int, dec, dy, tt, prodmat,
                               onevec, v, mirror, u},
                             daun = Length[fil]/2; Cn = Length[fil] - 1;
                             xx = 2*x; int = Floor[xx]; dec = xx - int;
                             dy = dyad[dec, n]; tt = {T0[fil], T1[fil]};
                             prodmat = IdentityMatrix[Cn];
                             For[i = 1, i <= n, i++,
                                 prodmat = prodmat.tt[[dy[[i]] + 1]]];
                                 onevec = Table[1, {i, 1, Cn}];
                                 v = prodmat.onevec/Cn;
                                 mirror = Flatten[Table[((-1)^i)*fil[[i]],
                                                        {i, 1, Length[fil]}]];
                                 u = Table[If[(i + 1 - int > 0 &&
                                              i + 1 - int < Cn + 2),
                                              mirror[[i + 1 - int]], 0],
                                              {i, 1, Cn}];
                                           Sqrt[2]*Sum[v[[i]]*u[[i]],
                                           {i, 1, Cn}]];
PsiD[x_, fil_, n_] := 0 /; (x <= 1 - Length[fil]/2 || x >= Length[fil]/2);

p = Length[Daub]/2;
Phi[x_] := PhiD[x, Daub, 25];
Psi[x_] := PsiD[x, Daub, 25];


(* Entrée des données réelles *)

xtmp1 = {1.001, 1.231, 1.123, 0.696, 0.808, 1.071, 1.009, 1.142, 0.767, 1.006,
         0.893, 1.081, 0.868, 0.762, 1.144, 1.045, 0.637, 0.733, 0.715, 0.872,
         0.765, 0.878, 0.811, 0.729, 0.911, 0.808, 1.168, 0.749, 0.892, 1.002,
         0.696, 1.199, 1.030, 0.899, 1.227, 1.180, 0.795, 0.629, 0.608};

ylist1 = {3.120, 0.638, 1.170, 0.926, 3.148, 1.836, 2.845, 1.013, 1.869, 2.836,
```

```
          3.567, 1.719, 3.423, 1.634, 1.021, 2.157, 0.571, 2.219, 1.419, 3.519,
          1.732, 3.206, 2.471, 1.397, 3.536, 2.202, 0.756, 1.620, 3.656, 2.964,
          1.139, 0.727, 2.581, 3.488, 0.754, 0.797, 2.064, 0.561, 0.563};

xtmp2 = {0.907, 0.761, 1.108, 1.016, 1.189, 1.042, 1.215, 0.930, 1.152, 1.138,
         0.601, 0.686, 1.072, 1.074, 0.934, 1.229, 1.175, 0.568, 0.977, 1.152,
         0.693, 1.232, 1.036, 1.125, 0.797, 1.115, 1.070, 1.219, 0.676, 1.045,
         0.968, 0.846, 0.684, 0.812, 1.230, 0.804, 0.813, 1.002, 0.602, 0.694,
         0.816, 1.037, 1.181, 0.990, 1.201, 0.584, 0.562, 0.535, 0.655};

ylist2 = {3.741, 2.295, 1.498, 2.881, 0.760, 2.358, 0.606, 3.669, 1.000, 0.981,
          1.192, 1.590, 1.806, 1.962, 4.028, 0.414, 0.812, 0.374, 3.623, 0.866,
          1.369, 0.542, 2.739, 1.200, 3.361, 1.390, 1.947, 0.962, 1.777, 2.571,
          3.952, 3.931, 1.587, 3.760, 0.672, 3.677, 3.517, 3.290, 0.923, 1.527,
          3.388, 2.085, 0.966, 3.732, 0.586, 0.678, 0.370, 0.530, 1.900};


(* Ramener les observations générées sur le domaine [0, 1] si cela n'est pas
   déjà ainsi *)

xmin = Min[xtmp1, xtmp2];
xmax = Max[xtmp1, xtmp2];
xlist1 = (xtmp1 - xmin)/(xmax - xmin);
xlist2 = (xtmp2 - xmin)/(xmax - xmin);
nobs1 = Length[xlist1];
nobs2 = Length[xlist2];
nobs = Min[nobs1, nobs2];


(* Construction d'une matrice diagonale dont les éléments sont les facteurs
   2^(j/2) ayant étés omis dans la définition des psi_(j,k)(x) *)

s = 1;
tmp = 1;
bd = 1 + nobs/2 - 2*p;
For[i = 0, i < Floor[Log[2, nobs]], i++,
    {tmp = p*i + 2^i; If[tmp >= bd, Break[]]}];
j0 = i - 1;
j0 = Max[i - 1, 1];
indPhi = Table[i, {i, -2*p + 2, 0}];
indPsik = Flatten[Table[i, {j, 0, j0}, {i, 1 - p, 2^j + p - 2}]];
indPsij = Flatten[Table[j, {j, 0, j0}, {i, 1 - p, 2^j + p - 2}]];
power2 = 2^Flatten[{Table[0, {i, 1, Length[indPhi]}], (-indPsij)}];
powerMat = DiagonalMatrix[power2];


(* Matrice de variance-covariance a priori des coefficients alpha et beta *)

gam1 = Table[1, {i, 1, Length[indPhi]}];
gam2 = Table[2^(-indPsij[[i]]*(s + 0.5)), {i, 1, Length[indPsij]}];
Gam = DiagonalMatrix[Flatten[Append[gam1, gam2]]];


(* Matrice de variance-covariance Qn *)
```

```
QnDist[x_, y_] := Module[{kmin, kmax, qtmp, j},
                         qtmp = 0;
                         For[i = 1, i <= 10, i++,
                             {j = j0 + i; kmin = Ceiling[2^j*Max[x, y] - p];
                              kmax = Floor[2^j*Min[x, y] + p - 1];
                              argi = Table[2^j*x - k, {k, kmin, kmax}];
                              argj = Table[2^j*y - k, {k, kmin, kmax}];
                              qtmp = qtmp +
                              Sum[(Psi[argi[[ii]]]*Psi[argj[[ii]]])/2^(2*j*s),
                                  {ii, 1, kmax - kmin + 1}];}];
                         Return[qtmp]];

Qntrian1 = Table[QnDist[xlist1[[i]], xlist1[[j]]],
                 {j, 1, nobs1}, {i, j, nobs1}];
Qn1 = Table[If[j >= i, Qntrian1[[i]][[j - i + 1]], Qntrian1[[j]][[i - j + 1]]],
            {j, 1, nobs1}, {i, 1, nobs1}];

Qntrian2 = Table[QnDist[xlist2[[i]], xlist2[[j]]],
                 {j, 1, nobs2}, {i, j, nobs2}];
Qn2 = Table[If[j >= i, Qntrian2[[i]][[j - i + 1]], Qntrian2[[j]][[i - j + 1]]],
            {j, 1, nobs2}, {i, 1, nobs2}];


(* Matrice comportant les fonctions de la base d'ondelettes évaluées
   aux abscisses *)

XPhi1 = Table[Phi[xlist1[[ii]] - indPhi[[jj]]],
              {ii, 1, nobs1}, {jj, 1, Length[indPhi]}];
XPsi1 = Table[Psi[2^indPsij[[jj]]*xlist1[[ii]] - indPsik[[jj]]],
              {ii, 1, nobs1}, {jj, 1, Length[indPsij]}];
xmat1 = AppendRows[XPhi1, XPsi1];

XPhi2 = Table[Phi[xlist2[[ii]] - indPhi[[jj]]],
              {ii, 1, nobs2}, {jj, 1, Length[indPhi]}];
XPsi2 = Table[Psi[2^indPsij[[jj]]*xlist2[[ii]] - indPsik[[jj]]],
              {ii, 1, nobs2}, {jj, 1, Length[indPsij]}];
xmat2 = AppendRows[XPhi2, XPsi2];


(* Décomposition spectrale *)

xtx1 = Transpose[xmat1].xmat1;
mat11 = xmat1.Gam.Transpose[xmat1] + Qn1;
{v11, v21, v31} = SingularValueDecomposition[mat11, Tolerance -> 0];
dmat1 = Table[v21[[i, i]], {i, 1, Length[v21]}];
hmat1 = v11;
svec1 = Transpose[hmat1].ylist1;

xtx2 = Transpose[xmat2].xmat2;
mat12 = xmat2.Gam.Transpose[xmat2] + Qn2;
{v12, v22, v32} = SingularValueDecomposition[mat12, Tolerance -> 0];
dmat2 = Table[v22[[i, i]], {i, 1, Length[v22]}];
hmat2 = v12;
```

```
svec2 = Transpose[hmat2].ylist2;


(* Estimateurs pour sigma et tau *)

LogM[ss_, dd_, v_] := Module[{vv},
                              vv = Abs[v];
                              LM = 0.5*Length[ss]*Log[Mean[ss^2/(vv + dd)]] +
                                   0.5*Sum[Log[vv + dd[[i]]],
                                           {i, 1, Length[ss]}] +
                                   0.5*Length[ss];
                              LM];

v1low = 0.05; v1up = 1; v2low = 0.09; v2up = 1;

Opt1 = FindMinimum[LogM[svec1, dmat1, v], {v, 0.5*(v1low + v1up)},
                   MaxIterations -> 100];
v1hat = Abs[v] /. Opt1[[2]];
tau1hat = Sum[svec1[[i]]^2/(v1hat + dmat1[[i]]), {i, 1, nobs1}]/nobs1;

Opt2 = FindMinimum[LogM[svec2, dmat2, v], {v, 0.5*(v2low + v2up)},
                   MaxIterations -> 100];
v2hat = Abs[v] /. Opt2[[2]];
tau2hat = Sum[svec2[[i]]^2/(v2hat + dmat2[[i]]), {i, 1, nobs2}]/nobs2;


(* Vecteur moyen et matrice de variance-covariance pour la distribution
   a posteriori des coefficients alpha et beta *)

amat1 = Gam.Transpose[xmat1].hmat1.DiagonalMatrix[1/(v1hat + dmat1)].
        Transpose[hmat1];
beta1 = amat1.ylist1;
Var1 = tau1hat*(Gam - amat1.xmat1.Gam);

amat2 = Gam.Transpose[xmat2].hmat2.DiagonalMatrix[1/(v2hat + dmat2)].
        Transpose[hmat2];
beta2 = amat2.ylist2;
Var2 = tau2hat*(Gam - amat2.xmat2.Gam);


(* Vecteur moyen et matrice de variance-covariance pour la distribution
   a posteriori de la différence entre les coefficients alpha et beta *)

Moyvec = (beta1 - beta2);
Bmat = Var1 + Var2;


(* Distance observée entre les deux fonctions *)

dist = Moyvec.powerMat.Moyvec;


(* Valeurs prédites *)
```

```
yhat1 = xmat1.beta1;
yhat2 = xmat2.beta2;


(* Calcul des données sous l'hypothèse que les fonctions sont égales *)

y1H0 = Table[0, {i, 1, nobs1}];
For[ii = 1, ii <= nobs1, ii++,
    {diff = Min[Select[Abs[xlist2 - xlist1[[ii]]], #1 > 0 &]];
     ytmp = Select[Table[If[Abs[xlist2[[k]] - xlist1[[ii]]] <= 2*diff,
                            ylist2[[k]]], {k, 1, nobs2}], NumberQ[#1] &];
     If[Length[ytmp] > 0,
        y1H0[[ii]] = (ylist1[[ii]] + Mean[ytmp])/2,
        y1H0[[ii]] = ylist1[[ii]]];}];

y2H0 = Table[0, {i, 1, nobs2}];
For[ii = 1, ii <= nobs2, ii++,
    {diff = Min[Select[Abs[xlist1 - xlist2[[ii]]], #1 > 0 &]];
     ytmp = Select[Table[If[Abs[xlist1[[k]] - xlist2[[ii]]] <= 2*diff,
                            ylist1[[k]]], {k, 1,nobs1}], NumberQ[#1] &];
     If[Length[ytmp] > 0,
        y2H0[[ii]] = (ylist2[[ii]] + Mean[ytmp])/2,
        y2H0[[ii]] = ylist2[[ii]]];}];


(* Estimateurs pour sigma et tau calculés à l'aide des données sous H0 *)

svec1H0 = Transpose[hmat1].y1H0;
Opt1H0 = FindMinimum[LogM[svec1H0, dmat1, v], {v, 0.5*(v1low + v1up)},
                     MaxIterations -> 100];
v1hatH0 = Abs[v] /. Opt1H0[[2]];
tau1hatH0 = Sum[svec1H0[[i]]^2/(v1hatH0 + dmat1[[i]]), {i, 1, nobs1}]/nobs1;

svec2H0 = Transpose[hmat2].y2H0;
Opt2H0 = FindMinimum[LogM[svec2H0, dmat2, v], {v, 0.5*(v2low + v2up)},
                     MaxIterations -> 100];
v2hatH0 = Abs[v] /. Opt2H0[[2]];
tau2hatH0 = Sum[svec2H0[[i]]^2/(v2hatH0 + dmat2[[i]]), {i, 1, nobs2}]/nobs2;


(* Vecteur moyen et matrice de variance-covariance pour la distribution
   a posteriori des coefficients alpha et beta sous H0 *)

amat1H0 = Gam.Transpose[xmat1].hmat1.DiagonalMatrix[1/(v1hatH0 + dmat1)].
          Transpose[hmat1];
beta1H0 = amat1H0.y1H0;
Var1H0 = tau1hatH0*(Gam - amat1H0.xmat1.Gam);


amat2H0 = Gam.Transpose[xmat2].hmat2.DiagonalMatrix[1/(v2hatH0 + dmat2)].
          Transpose[hmat2];
beta2H0 = amat2H0.y2H0;
Var2H0 = tau2hatH0*(Gam - amat2H0.xmat2.Gam);
```

```
(* Vecteur moyen et matrice de variance-covariance pour la distribution
   a posteriori de la différence entre les coefficients alpha et beta
   sous H0 *)

MoyvecH0 = beta1H0 - beta2H0;
BmatH0 = Var1H0 + Var2H0;


(* Valeurs prédites sous H0 *)

yhat1H0 = xmat1.beta1H0;
yhat2H0 = xmat2.beta2H0;
```

B.2. Approximation d'Imhof

```
(* 1. Distribution et densité de Q observée *)
(*******************************************)

(* Rendre la matrice Bmat parfaitement symétrique *)

mat = Bmat;
New1 = Table[If[i < j, mat[[i, j]], 0], {i, 1, Length[Moyvec]},
             {j, 1, Length[Moyvec]}];
New2 = Transpose[New1];
new3 = Table[mat[[i, i]], {i, 1, Length[Moyvec]}];
Var = New1 + New2 + DiagonalMatrix[new3];


(* Trouver les valeurs propres distinctes *)

ValprTemp = Eigenvalues[powerMat.Var];
Valpr = Transpose[Reverse[Frequencies[ValprTemp]]][[2]];


(* Ordre de multiplicité des valeurs propres *)

hrtemp = Transpose[Reverse[Frequencies[ValprTemp]]][[1]];


(* Calcul des paramètres de non-centralité *)

u = CholeskyDecomposition[Var];
l = Transpose[u];

VectprTemp = Eigenvectors[u.powerMat.l];
norme = Table[Sum[VectprTemp[[i, j]]^2, {j, 1, Length[ValprTemp]}],
              {i, 1, Length[ValprTemp]}];
Vectpr = Transpose[VectprTemp/Sqrt[norme]];

deltatemp = Transpose[Vectpr].Inverse[l].Moyvec;
deltatemp2 = deltatemp^2;
top = Flatten[{{0}, Table[Sum[hrtemp[[j]], {j, 1, i}],
```

```
                              {i, 1, Length[hrtemp]}]}];
delta = Table[Sum[deltatemp2[[i + top[[j]]]], {i, 1, hrtemp[[j]]}],
              {j, 1, Length[hrtemp]}];


(* Calcul des quantités intervenant dans l'approximation d'Imhof *)

e1 = Sum[Valpr[[i]]*(hrtemp[[i]] + delta[[i]]), {i, Length[Valpr]}];
e2 = Sum[Valpr[[i]]^2*(hrtemp[[i]] + 2*delta[[i]]), {i, Length[Valpr]}];
e3 = Sum[Valpr[[i]]^3*(hrtemp[[i]] + 3*delta[[i]]), {i, Length[Valpr]}];
hprime = e2^3/e3^2;


(* Expressions pour l'approximation de la distribution et la densité de Q *)

chidist = ChiSquareDistribution[hprime];
f[x_] := (x - e1)*(hprime/e2)^(1/2) + hprime;



(* 2. Distribution et densité de Q sous H0 *)
(*****************************************)

(* Voir les commentaires ci - haut *)

matH0 = BmatH0;
New1H0 = Table[If[i < j, matH0[[i, j]], 0], {i, 1, Length[MoyvecH0]},
               {j, 1, Length[MoyvecH0]}];
New2H0 = Transpose[New1H0];
new3H0 = Table[matH0[[i, i]], {i, 1, Length[MoyvecH0]}];
VarH0 = New1H0 + New2H0 + DiagonalMatrix[new3H0];


ValprTempH0 = Eigenvalues[powerMat.VarH0];
ValprH0 = Transpose[Reverse[Frequencies[ValprTempH0]]][[2]];


hrtempH0 = Transpose[Reverse[Frequencies[ValprTempH0]]][[1]];


uH0 = CholeskyDecomposition[VarH0];
lH0 = Transpose[uH0];

VectprTempH0 = Eigenvectors[uH0.powerMat.lH0];
normeH0 = Table[Sum[VectprTempH0[[i, j]]^2, {j, 1, Length[ValprTempH0]}],
                {i, 1, Length[ValprTempH0]}];
VectprH0 = Transpose[VectprTempH0/Sqrt[normeH0]];

deltatempH0 = Transpose[VectprH0].Inverse[lH0].MoyvecH0;
deltatemp2H0 = deltatempH0^2;
topH0 = Flatten[{{0}, Table[Sum[hrtempH0[[j]], {j, 1, i}],
                            {i, 1, Length[hrtempH0]}]}];
deltaH0 = Table[Sum[deltatemp2H0[[i + topH0[[j]]]], {i, 1, hrtempH0[[j]]}],
                {j, 1, Length[hrtempH0]}];
```

```
e1H0 = Sum[ValprH0[[i]]*(hrtempH0[[i]] + deltaH0[[i]]), {i, Length[ValprH0]}];
e2H0 = Sum[ValprH0[[i]]^2*(hrtempH0[[i]] + 2*deltaH0[[i]]),
          {i, Length[ValprH0]}];
e3H0 = Sum[ValprH0[[i]]^3*(hrtempH0[[i]] + 3*deltaH0[[i]]),
          {i, Length[ValprH0]}];
hprimeH0 = e2H0^3/e3H0^2;


chidistH0 = ChiSquareDistribution[hprimeH0];
fH0[x_] := (x - e1H0)*(hprimeH0/e2H0)^(1/2) + hprimeH0;



(* 3. Tests basés sur l'approximation de la distribution de Q *)
(*************************************************************)

(* Calcul du mode de l'approximation de la densité de Q observée et sous H0 *)

modei = z /. Minimize[-f'[z]*PDF[chidist, f[z]], z > e1 - (e2^2/e3), z][[2]];
modeiH0 = z /. Minimize[-fH0'[z]*PDF[chidistH0, fH0[z]],
                        z > e1H0 - (e2H0^2/e3H0), z][[2]];


(* Calcul des trois tests *)

cr1Im = 1 - CDF[chidistH0, fH0[modei]];
cr2Im = CDF[chidistH0, fH0[modei]] - CDF[chidistH0, fH0[modeiH0]];
cr3Im = CDF[chidistH0, fH0[dist]];
```

B.3. Approximation par point de selle

```
(* 1. Distribution et densité de Q observée *)
(*******************************************)

(* Fonction génératrice des cumulants et dérivées *)

cgf[biz_] := (-1/2)*Sum[hrtemp[[i]]*Log[1 - 2*biz*Valpr[[i]]],
                        {i, 1, Length[Valpr]}] +
             Sum[(biz*delta[[i]]*Valpr[[i]])/(1 - 2*biz*Valpr[[i]]),
                 {i, 1, Length[Valpr]}];
der1[biz_] := cgf'[biz];
der2[biz_] := cgf''[biz];
der3[biz_] := cgf'''[biz];


(* Fonctions w et v ainsi que leurs dérivées *)

fctw[psi_, t_] := Sign[psi]*(2*(psi*t - cgf[psi]))^(1/2);
```

```
derfctw[psi_, t_] := Sign[psi]/Sqrt[2]*psi/Sqrt[psi*t - cgf[psi]];

fctu[psi_, t_] := psi*(der2[psi])^(1/2);
derfctu[psi_, t_] := (der2[psi])^(-1/2) + psi*der3[psi]/(2*Sqrt[der2[psi]^3]);


(* Expressions pour l'approximation de la distribution et la densité de Q *)

saddist = NormalDistribution[0, 1];
sol[psi_, t_] := fctw[psi, t] + Log[fctu[psi, t]/fctw[psi, t]]/fctw[psi, t];
dersol[psi_, t_] := derfctw[psi, t]*
                    (1 - (Log[fctu[psi, t]/fctw[psi, t]] + 1)/(fctw[psi, t]^2))
                    + (fctu[psi, t]*fctw[psi,t])^(-1)*derfctu[psi, t];
sadd[psi_, t_] := CDF[saddist, sol[psi, t]];
sad[psi_, t_] := dersol[psi, t]*PDF[saddist, sol[psi, t]];


(* 2. Distribution et densité de Q sous H0 *)
(*****************************************)

(* Voir les commentaires ci - haut *)

cgfH0[biz_] := (-1/2)*Sum[hrtempH0[[i]]*Log[1 - 2*biz*ValprH0[[i]]],
                          {i, 1, Length[ValprH0]}] +
               Sum[(biz*deltaH0[[i]]*ValprH0[[i]])/(1 - 2*biz*ValprH0[[i]]),
                   {i, 1, Length[ValprH0]}];
der1H0[biz_] := cgfH0'[biz];
der2H0[biz_] := cgfH0''[biz];
der3H0[biz_] := cgfH0'''[biz];


fctwH0[a_, t_] := Sign[a]*(2*(a*t - cgfH0[a]))^(1/2);
derfctwH0[a_, t_] := Sign[a]/Sqrt[2]*a/Sqrt[a*t - cgfH0[a]];

fctuH0[a_, t_] := a*(der2H0[a])^(1/2);
derfctuH0[a_, t_] := (der2H0[a])^(-1/2) +
                     a*der3H0[a]/(2*Sqrt[der2H0[a]^3]);


solH0[a_, t_] := fctwH0[a, t] + Log[fctuH0[a, t]/fctwH0[a, t]]/fctwH0[a, t];
dersolH0[a_, t_] := derfctwH0[a, t]*
                    (1 - (Log[fctuH0[a, t]/fctwH0[a, t]] + 1)/(fctwH0[a,t]^2))
                    + (fctuH0[a, t]*fctwH0[a, t])^(-1)*derfctuH0[a, t];
saddH0[psi_, t_] := CDF[saddist, solH0[psi, t]];
sadH0[psi_, t_] := dersolH0[psi, t]*PDF[saddist, solH0[psi, t]];



(* 3. Tests basés sur l'approximation de la distribution de Q *)
(************************************************************)

(** PREMIER TEST **)

(* Au lieu de lister les valeurs de q qui nous intéressent et de solutionner
```

```
   cgf'(psi) = q, nous procédons à l'envers. Nous listons les valeurs de psi
   qui correspondent aux valeurs de q qui nous intéressent. Pour ces valeurs
   de psi, nous calculons cgf' = q *)

psitemp1 = Flatten[{Reverse[Table[i*(-100000), {i, 1, 9}]],
                    Reverse[Table[i*(-10000), {i, 1, 9}]],
                    Reverse[Table[i*(-1000), {i, 1, 9}]],
                    Reverse[Table[i*(-10), {i, 1, 90}]], Table[i, {i,-9,-1}],
                    Reverse[Table[i*(-0.01), {i, 1, 90}]],
                    Table[i*0.01,
                          {i, 1, Floor[((1/2)*Min[1/Valpr]-0.01)/0.01]}]}];
qtemp1 = der1[psitemp1];


(* Nous trouvons la plus petite valeur de q pour laquelle l'approximation de
   la fonction de répartition de Q vaut 1. Ensuite, nous conservons les
   valeurs de q inférieures ou égale à celle-ci *)

Do[deb; Maxpsitemp1 = deb; If[CDF[chidist, f[deb]] == 1, Break[]],
   {deb, Ceiling[c1 - (e2^2/e3)], 1000}];
qtemp2 = Table[If[qtemp1[[i]] <= Maxpsitemp1, qtemp1[[i]], 0],
               {i, 1, Length[qtemp1]}];
Valq = Select[qtemp2, # > 0 &];
psitemp2 = Select[Table[If[qtemp1[[i]] <= Maxpsitemp1, i, 0],
                  {i, 1, Length[qtemp1]}], # > 0 &];
Valpsi = psitemp1[[psitemp2]];


(* Calcul de l'approximation par point de selle de la densité de Q pour les
   valeurs de q et de psi trouvées ci-haut *)

DensSad = sad[Valpsi, Valq];


(* Calcul du mode de cette densité *)

maxk = Max[DensSad];
pos = Select[Table[If[DensSad[[i]] == maxk, i, 0],
             {i, 1, Length[Valpsi]}], # > 0 &][[1]];
modek = Valq[[pos]];


(* Évaluation du premier test *)

toto = Solve[der1H0[x] == modek, x];
tempor = Select[x /. toto,
                # \[Element] Reals && # < (1/2)*Min[1/ValprH0] &][[1]];
cr1Ku = 1 - saddH0[tempor, modek];


(** DEUXIÈME TEST **)

(* Les commentaires du PREMIER TEST s'appliquent tous, mais ils font toutefois
   référence aux approximations de la distribuion et de la densité de Q sous
```

```
   H0 *)

psitemp1H0 = Flatten[{Reverse[Table[i*(-100000), {i, 1, 9}]],
                      Reverse[Table[i*(-10000), {i, 1, 9}]],
                      Reverse[Table[i*(-1000), {i, 1, 9}]],
                      Reverse[Table[i*(-10), {i, 1, 90}]], Table[i, {i, -9, -1}],
                      Reverse[Table[i*(-0.01), {i, 1, 90}]],
                      Table[i*0.01,
                            {i, 1, Floor[((1/2)*Min[1/ValprH0] - 0.01)/0.01]}]}];
qtemp1H0 = der1H0[psitemp1H0];


Clear[deb];
Do[deb; Maxpsitemp1H0 = deb; If[CDF[chidistH0, fH0[deb]] == 1, Break[]],
   {deb, Ceiling[c1H0 - (e2H0^2/e3H0)], 1000}];
qtemp2H0 = Table[If[qtemp1H0[[i]] <= Maxpsitemp1H0, qtemp1H0[[i]], 0],
                 {i, 1, Length[qtemp1H0]}];
ValqH0 = Select[qtemp2H0, # > 0 &];
psitemp2H0 = Select[Table[If[qtemp1H0[[i]] <= Maxpsitemp1H0, i, 0],
                    {i, 1, Length[qtemp1H0]}], # > 0 &];
ValpsiH0 = psitemp1H0[[psitemp2H0]];


DensSadH0 = sadH0[ValpsiH0, ValqH0];


maxkH0 = Max[DensSadH0];
posH0 = Select[Table[If[DensSadH0[[i]] == maxkH0, i, 0],
               {i, 1, Length[ValpsiH0]}], # > 0 &][[1]];
modekH0 = ValqH0[[posH0]];


(* Évaluation du deuxième test *)

totoH0 = Solve[der1H0[x] == modekH0, x];
temporH0 = Select[x /. totoH0,
                  # \[Element] Reals && # < (1/2)*Min[1/ValprH0] &][[1]];
cr2Ku = saddH0[tempor, modek] - saddH0[temporH0, modekH0];


(** TROISIÈME TEST **)

(* Évaluation du troisième test *)

totodist = Solve[der1H0[x] == dist, x];
tempordist = Select[x /. totodist,
                    # \[Element] Reals && # < (1/2)*Min[1/ValprH0] &][[1]];
cr3Ku = saddH0[tempordist, dist];
```

B.4. INTERVALLES DE CONFIANCE SIMULTANÉS

```
(* Liste contenant les points d'abscisses où les intervalles de confiance
   seront calculés *)

xlist = Select[Select[Sort[Flatten[Append[xlist1, xlist2]]],
                      # >= Max[{Min[xlist1], Min[xlist2]}] &],
              # <= Min[{Max[xlist1], Max[xlist2]}] &];


(* Vecteur des différences entre les coefficients estimés pour les deux
   échantillons *)

beta = beta1 - beta2;


(* Matrice comportant les fonctions de la base d'ondelettes évaluées aux
   points d'abscisses *)

XPhi = Table[Phi[xlist[[ii]] - indPhi[[jj]]],
             {ii, 1, Length[xlist]}, {jj, 1, Length[indPhi]}];
XPsi = Table[Psi[2^indPsij[[jj]]*xlist[[ii]] - indPsik[[jj]]],
             {ii, 1, Length[xlist]}, {jj, 1, Length[indPsij]}];
xmat = AppendRows[XPhi, XPsi];


(* Calcul des bornes inférieures et supérieures des intervalles de confiance
   simultanés *)

inf = xmat.beta -
      Sqrt[Length[beta]*
           Quantile[FRatioDistribution[Length[beta],
                                       nobs1 + nobs2 - Length[beta]], 0.95]]*
      Sqrt[Table[xmat[[i]].Bmat.xmat[[i]], {i, 1, Length[xlist]}]];
sup = xmat.beta +
      Sqrt[Length[beta]*
           Quantile[FRatioDistribution[Length[beta],
                                       nobs1 + nobs2 - Length[beta]], 0.95]]*
      Sqrt[Table[xmat[[i]].Bmat.xmat[[i]], {i, 1, Length[xlist]}]];
```

BIBLIOGRAPHIE

ABRAMOVICH, F. ET ANGELINI, C. (2003), Testing in mixed-effects FANOVA models, *Rapport technique*, RP-SOR-03-03, Université de Tel Aviv, Israël.

ABRAMOVICH, F. ET SAPATINAS, T. (1999), Bayesian approach to wavelet decomposition and shrinkage, *Lecture Notes in Statistics*, **141**, 33-50.

ALPERT, B. K. (1992), Wavelets and other basis for fast numerical linear algebra, *Wavelets - A tutorial in theory and applications*, ed. Chui C. K., Academic Press, Boston, 181-216.

ANGERS, J.-F. ET DELAMPADY, M. (1992), Hierarchical Bayesian curve fitting and smoothing, *Canadian Journal of Statistics*, **20**, 35-49.

ANGERS, J.-F. ET DELAMPADY, M. (2001), Bayesian nonparametric regression using wavelets, *Sankhyã, Series A*, **63**, 287-308.

ANGERS, J.-F. (2003), Curves comparison using wavelet, *Wavelets and their applications*, Allied Publishers Private Limited, 47-62.

BARNDORFF-NIELSEN, O. E. (1990), Approximate interval probabilities, *Journal of the Royal Statistical Society, Series B*, **52**, 485-496.

BRILLINGER, D. R. (1973), The analysis of time series collected in an experiment design, *Multivariate Analysis III*, ed. Krishnaiah P., Academic Press, New York, 241-256.

BRILLINGER, D. R. (1981), Some aspects of the analysis of evoked response experiments, *Statistics and related topics*, eds. Csörgö M., Dawson D., Rao J., et Saleh A., Amsterdam, 155-168.

CHIPMAN, H. A., KOLACZIK, E. D. ET MCCULLOCH, R. E. (1997), Adaptative Bayesian wavelet shrinkage, *Journal of the American Statistical Association*, **92**, 1413-1421.

CLEVELAND, W. (1993), *Visualizing data*, Hobart Press, New Jersey.

CLYDE, M. A. ET GEORGE, E. I. (1999), Empirical Bayes estimation in wavelet nonparametric regression, *Lecture Notes in Statistics*, **141**, 309-322.

CRAVEN, P. ET WAHBA, G. (1979), Smoothing noisy data with spline functions : estimating the correct degree of smoothing by the method of generalized cross-validation, *Numerische Mathematik*, **31**, 377-403.

DANIELS, H. E. (1954), Saddlepoint approximations in statistics, *Annals of Mathematical Statistics*, **25**, 631-650.

DAUBECHIES, I. (1992), *Ten lectures on wavelets*, SIAM, Philadelphie.

DAUBECHIES, I. ET LAGARIAS, J. (1991), Two-scale difference equations I : existence and global regularity of solutions, *SIAM Journal on Mathematical Analysis*, **22**, 1388-1410.

DAUBECHIES, I. ET LAGARIAS, J. (1992), Two-scale difference equations II : local regularity, infinite products of matrices and fractals, *SIAM Journal on Mathematical Analysis*, **23**, 1031-1079.

DE BOOR, C. (1978), *A practical guide to splines*, Springer-Verlag, New York.

DONOHO, D. L. ET JOHNSTONE, I. M. (1994), Ideal spatial adaptation via wavelet shrinkage, *Biometrika*, **81**, 425-455.

DONOHO, D. L. ET JOHNSTONE, I. M. (1995), Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association*, **90**, 1200-1224.

DONOHO, D. L. ET JOHNSTONE, I. M. (1998), Minimax estimation via wavelet shrinkage, *Annals of Statistics*, **26**, 879-921.

FAN, J. AND LIN, S.-K. (1998), Test of significance when data are curves, *Journal of the American Statistical Association*, **93**, 1007-1021.

FOSTER, D. ET GEORGE, E. (1994), The risk inflation criterion for multiple regression, *Annals of Statistics*, **22**, 1947-1975.

GELMAN, A., CARLIN, J. B., STERN, H. S. ET RUBIN, D. B. (2004), *Bayesian data analysis*, Chapman & Hall, Floride.

GROSSMAN, A. ET MORLET, J. (1984), Decomposition of Hardy functions into square integrable wavelets of constant shape, *SIAM Journal on Mathematical Analysis*, **15**, 723-736.

HALL, P. ET HART, J. D. (1990), Bootstrap test for the difference between means in nonparametric regression, *Journal of the American Statistical Association*, **85**, 1039-1049.

HAAR, A. (1910), Zur theorie der orthogonalen funktionensysteme, *Mathematische Annalen*, **69**, 331-371.

HÄRDLE, W., KERKYACHARIAN, G., PICARD, D. ET TSYBAKOV, A. B. (1998), *Wavelets, approximations, and statistical applications*, Springer-Verlag, New York.

IMHOF, J. P. (1961), Computing the distribution of a quadratic form in normal variables, *Biometrika*, **48**, 419-426.

JOHNSON, N. L. (1959), On an extension of the connexion between Poisson and χ^2 distributions, *Biometrika*, **46**, 352-363.

JOHNSON, N. L. ET KOTZ, S. (1970), *Distributions in statistics : continuous univariate distributions - 2*, Wiley, New York.

KIMELDORF, G. AND WAHBA, G. (1971), Some results on Tchebycheffian spline functions, *Journal of Mathematical Analysis and Applications*, **33**, 82-95.

KLEIN, J. P. ET MOESCHBERGER, M. L. (1997), *Survival analysis : techniques for censored and truncated data*, Springer, New York.

KOUL, H. L. AND SCHICK, A. (1997), Testing for the equality of two nonparametric regression curves, *Journal of Statistical Planning and Inference*, **65**, 293-314.

KULASEKERA, K. B. (1995), Comparison of regression curve using quasi-residuals, *Journal of the American Statistical Association*, **90**, 1085-1093.

KUONEN, D. (1999), Saddlepoint approximations for distributions of quadratic forms in normal variables, *Biometrika*, **86**, 929-935.

LANGE, K. (1998), *Numerical analysis for statisticians*, Springer-Verlag, New York.

LEBLANC, A. (2001), Utilisation des ondelettes de Haar en estimation bayésienne, *Thèse de doctorat*, Université de Montréal, Montréal.

LINDLEY, D. V. ET SMITH, A. F. (1972), Bayes estimates for the linear model, *Journal of the Royal Statistical Society, Series B*, **34**, 1-41.

MALLAT, S. (1989), Multiresolution approximations and wavelet orthonormal bases of $\mathcal{L}^2(\mathbb{R})$, *Transactions of the American Mathematical Society*, **315**, 69-88.

MALLAT, S. (1998), *A wavelet tour of signal processing*, Academic Press, San Diego.

MEYER, Y. (1990), *Ondelettes et opérateurs I : ondelettes*, Hermann, Paris.

MEYER, Y. (1992), *Wavelets and operators*, Cambridge University Press, Cambridge.

MÜLLER, P. ET VIDAKOVIC, B. (1999A), *Bayesian inference in wavelet-based models*, Springer-Verlag, New York.

MÜLLER, P. ET VIDAKOVIC, B. (1999B), MCMC methods in wavelet shrinkage : non-equally spaced regression, density and spectral density estimation, *Lecture Notes in Statistics*, **141**, 187-202.

ODGEN, R.T. (1997), *Essential wavelets for statistical applications and data analysis*, Birkhäuser, Boston.

PEARSON, E. S. (1959), Note on an approximation to the distribution of non-central χ^2, *Biometrika*, **46**, 364.

ROBERT, C. P. (2001), *The Bayesian choice*, 2^e édition, Springer-Verlag, New York.

RUDIN, W. (1987), *Real and complex analysis*, 3^e édition, McGraw-Hill, New York.

SCHOENBERG, I. J. (1964), Spline functions and the problem of graduation, *Proceedings of the National Academy of Sciences of the United States of America*, **52**, 947-950.

SCHUMAKER, L. L. (1981), *Splines functions : basic theory*, Wiley, New York.

SEARLE, S. (1971), *Linear models*, Wiley, New York.

SEARLE, S. (1982), *Matrix algebra useful for statistics*, Wiley, New York.

SEBER, G. A. F. (1977), *Linear regression analysis*, Wiley, New York.

SHUMWAY, R. (1988), *Applied statistical time series analysis*, Prentice-Hall, New Jersey.

SILVERMAN, B. W. (1985), Some aspects of the spline smoothing approach to non-parametric regression curve fitting, *Journal of the Royal Statistical Society, Series B*, **47**, 1-52.

TAYLOR, B. (1715), *Methodus incrementorum directa et inversa*, London.

VAN DER LINDE, A. (1993), A note on smoothing splines as Bayesian estimates, *Statistics and Decisions*, **11**, 61-67.

VIDAKOVIC, B. (1999), *Statistical modeling by wavelets*, John Wiley, New York.

WAHBA, G. (1978), Improper priors, spline smoothing and the problem of guarding against model errors in regression, *Journal of the Royal Statistical Society, Series B*, **40**, 364-372.

WAHBA, G. ET WOLD S. (1975), Fitting splines functions by cross-validation, *Communications in Statistics*, **4**, 1-17.

WEERAHANDI, S. ET ZIDEK, J. V. (1988), Bayesian nonparametric smoothers for regular processes, *Canadian Journal of Statistics*, **16**, 61-74.

WALNUT, D. F. (2002), *An introduction to wavelet analysis*, Birkhäuser, Boston.

YAU, P. ET KOHN, R. (1999), Wavelet nonparametric regression using basis averaging, *Lecture Notes in Statistics*, **141**, 95-108.

ZHAO, L. (2000), Bayesian aspects of some nonparametric problems, *Annals of Statistics*, **28**, 532-552.