Vitesses et procédures statistiques minimax dans des problèmes d'estimation et des tests d'hypothèses

Ghislaine Gayraud

Mémoire de synthèse de l'Université de Rouen
LMRS UMR-6085

Spécialité : **Statistique Mathématique**

présentée par

Ghislaine Gayraud

pour obtenir l'HABILITATION À DIRIGER DES RECHERCHES

Vitesses et procédures statistiques minimax dans des problèmes d'estimation et de tests d'hypothèses.

Soutenue le 6 décembre 2007 devant le jury composé de :

Monsieur	Dominique Fourdrinier	Président
Monsieur	Gérard Kerkyacharian	Examinateur
Madame	Béatrice Laurent-Bonneau	Rapporteur
Monsieur	Serge Pergamenchtchikov	Examinateur
Monsieur	Alexandre Tsybakov	Examinateur

Remerciements

Je tiens à remercier Alexandre Tsybakov, qui a été mon directeur de thèse. Je le remercie pour le soutien constant qu'il m'a toujours témoigné et pour sa très grande culture statistique qu'il n'hésite pas à faire partager.

Je remercie les rapporteurs Eduard Belitser, Yuri Ingster et Béatrice Laurent d'avoir accepté de référer l'ensemble de mes travaux. Merci pour leurs encouragements au travers de leur rapport.

Je remercie Dominique Fourdrinier d'avoir accepté de faire partie de mon jury et d'en être le président. Je remercie également Gérard Kerkyacharian et Serge Pergamenchtchikov pour m'avoir fait l'honneur de participer à ce jury.

Mes remerciements s'adressent aussi à Emmanuelle Gautherat, Christophe Pouet, Judith Rousseau et Karine Tribouley, avec qui j'ai pris grand plaisir à collaborer.

Je voudrais remercier mes collègues de Rouen, qu'ils soient du laboratoire LMRS ou du département de Mathématiques, avec qui j'ai grand plaisir à discuter ; un merci spécial à Isabelle, Marc, Paul, Patrizia et Vlad pour leur aide constante.

Je tiens également à remercier mes collègues du Laboratoire de Statistique du CREST, avec qui je partage bureaux, pauses déjeuner et séminaires.

Merci enfin à ma tribu parisienne et à ma famille : simplement parce qu'ils sont toujours là.

Table des matières

Chapitre 1

Présentation générale.

Ce texte constitue une synthèse de mes travaux de recherche dont la liste est proposée en fin de mémoire. Seuls les articles [1] et [3] ne seront pas détaillés dans ce rapport puisqu'ils le sont dans mon rapport de thèse. Au cours de ma thèse, je me suis intéressée, dans un cadre non-paramétrique, à la résolution minimax de problèmes d'estimation et de test d'hypothèses, concernant des supports de densité et certaines de leurs fonctionnelles. Quatre des articles que je vais présenter dans ce rapport s'inscrivent dans la continuité de mes travaux de thèse : il s'agit des articles [5], [6], [8] et [9] dans lesquels l'objet d'intérêt s'interprète comme étant *une discontinuité*, qui peut être elle-même caractérisée par un support de densité. Il existe également un lien méthodologique entre mes travaux de thèse et les articles [2], [4] et [7] puisque les problèmes qui y sont exposés sont résolus via l'approche minimax. En revanche, l'article [10] ne peut être directement connecté à mes autres travaux, puisqu'il concerne un problème paramétrique de restitution d'un signal, à partir de données dépendantes. Néanmoins, l'objectif final de notre travail reste la résolution non-paramétrique du problème, via l'approche minimax.

Mes travaux s'articulent autour de trois thématiques : inférence non-paramétrique, inférence Bayésienne non-paramétrique et inférence paramétrique dans un modèle de déconvolution aveugle bruitée.

La thématique statistique non-paramétrique représente la partie la plus conséquente de mes travaux de recherche, qui sera détaillée dans le chapitre 2 de cette synthèse. Dans ce cadre, je me suis intéressée à des problèmes d'estimation et à des problèmes de tests d'hypothèses à partir d'observations indépendantes et identiquement distribuées. Je vais à présent lister brièvement les problèmes que j'ai considérés.

L'estimation adaptative de la norme L_2 au carré de la fonction signal, dans le modèle du bruit blanc Gaussien, a été initiée par Efromovich et Low (1996) (procédures de Lepski), puis étudiée par Laurent et Massart (2000) (sélection de modèles) : ils ont montré que la vitesse optimale d'estimation correspond à la vitesse paramétrique dans le cas régulier et à la vitesse minimax non-paramétrique entachée d'une perte d'efficacité dans le cas irrégulier. Ma contribution sur ce sujet, menée en collaboration avec Karine Tribouley (article [2]), est un théorème de la limite centrale adaptatif pour une estimation construite par ondelettes seuillées. Son avantage est de fournir une vitesse qui s'adapte au cadre régulier ou au cadre irrégulier, sans que l'utilisateur connaisse à l'avance la nature de ce cadre. D'un point de vue pratique, il permet de construire des intervalles de confiance et des procédures de test adaptatives autour/sur la norme L_2 au carré de la fonction signal.

Dans le problème d'estimation d'un point de discontinuité d'une densité inconnue (article [5]), j'obtiens comme vitesse minimax la vitesse de convergence paramétrique : cela correspond au phénomène observé dans les problèmes semi-paramétriques réguliers. La procédure d'estimation proposée est extrêmement simple : elle est construite à partir de différences d'histogrammes. La conséquence pratique de la simplicité de sa construction est qu'elle est aisée à implémenter.

En collaboration avec Alexandre Tsybakov, nous nous sommes intéressés au problème de test minimax d'hypothèses composées, dans un modèle de régression binaire d'images (article [6]). Le cadre de cette étude est très général : l'hypothèse nulle est composite, notre statistique de test asymptotiquement minimax est libre de tout paramètre de régularité. Les hypothèses sous lesquelles la relation de la borne supérieure est valide, sont peu contraignantes.

En collaboration avec Christophe Pouet, nous avons étudié le problème minimax et minimax adaptatif de test d'hypothèses composites, dans un modèle de régression (articles [4] et [7]). Nous obtenons la relation de borne supérieure sans faire d'hypothèse de normalité sur la distribution des erreurs. Dans la version adaptative du problème, il est possible de considérer une hypothèse nulle composée non-paramétrique. Nous avons mis en évidence le lien existant entre le paramètre de régularité de la fonction signal, la régularité des erreurs et le contrôle de la complexité de la famille placée sous l'hypothèse nulle (uniquement pour la version adaptative). Notre statistique de test est asymptotiquement minimax et, dans la version adaptative, est libre de tout paramètre de régularité de la fonction signal. La connaissance a priori d'une condition de régularité sur la fonction signal est une conséquence incontournable lorsque la norme L_2 est choisie afin de séparer l'hypothèse nulle de l'alternative. En effet, cette information a priori permet de contrôler l'approximation de la norme L_2 par une pseudo-norme discrète qui est inhérente à la construction de la statistique de test. Nous avons également établi la relation de la borne inférieure en renforçant les hypothèses concernant la distribution des erreurs.

Mes travaux effectués en statistique Bayésienne non-paramétrique ont été menés en collaboration avec Judith Rousseau (articles [8] et [9]). Ils seront détaillés dans le chapitre 3 de ce rapport. Nous avons considéré le problème d'estimation non-paramétrique d'ensembles à niveau à partir d'observations indépendantes et identiquement distribuées. A notre connaissance, nos résultats sur les propriétés asymptotiques d'estimation Bayésienne d'ensembles à niveau sont les premiers à avoir été obtenus. Le contexte non-paramétrique d'un problème Bayésien exige de préter une attention particulière au choix de la loi a priori. Certaines lois a priori peuvent conduire à la non-consistance de la loi a posteriori. Le choix d'une famille de lois a priori *objectives* (i.e. ayant des bonnes propriétés d'objectivité) devient alors nécessaire et la consistance de la loi a posteriori est un minimum à garantir. Nous avons donc établi la consistance de la loi a posteriori pour une famille assez générale de lois a priori. Ce résultat entraîne la consistance de l'estimateur Bayésien non-paramétrique d'ensembles à niveau. Afin de mesurer les performances de l'estimateur Bayésien d'ensembles à niveau, nous nous sommes ensuite intéressées à déterminer sa vitesse de convergence. Comme pour la consistance, nous avons déterminé les vitesses de convergence de l'estimateur Bayésien, pour une large classe de lois a priori. Nous avons également appliqué nos résultats généraux à des lois a priori particulières afin de comparer la performance de l'estimateur Bayésien à celle d'estimateurs fréquentistes. D'un point de vue pratique, l'approche Bayésienne apparaît adéquate puisqu'elle fournit un estimateur explicite qui, dès lors qu'on est capable de simuler directement dans la loi a posteriori, est facilement implémentable.

Un unique article (article [10]) se situe dans la troisième thématique : il correspond à un travail mené en collaboration avec Emmanuelle Gautherat et sera détaillé au chapitre 4. Il s'agit, dans un modèle de déconvolution aveugle bruitée, de restituer la loi du signal entrant. L'hypothèse cruciale dans ce modèle concerne la loi du processus des innovations : elle est discrète et possède un nombre fini de valeurs possibles. L'exploitation de cette information a priori permet d'établir, dans un contexte très général (observations non-indépendantes, processus de sortie non nécessairement causal), une caractérisation du niveau de bruit et du filtre inverse (Gassiat, Gautherat, 1998, 1999). Pour ce problème et pour des données complexes, nous avons proposé une nouvelle procédure d'estimation du niveau de bruit et du filtre inverse. La nouveauté se situe dans l'expression du critère empirique qui est explicite, et permet de construire nos estimateurs à partir de zéros de ce critère. Nous avons établi la consistance des estimateurs, ainsi que leur distribution asymptotique. D'un point de vue pratique, cette méthode d'estimation présente l'avantage d'être facilement implémentable.

Chapitre 2

Statistique non-paramétrique.

Dans un cadre non-paramétrique, je me suis intéressée à la détermination de vitesses *optimales* dans des problèmes d'estimation et de test d'hypothèses. La notion *d'optimalité* est ici relative au principe de minimaxité.

Ce chapitre se décompose en une introduction générale sur le principe de minimaxité dans un contexte non-paramétrique, une partie détaillant les problèmes d'estimation auxquels je me suis intéressée, une partie décrivant les problèmes de tests d'hypothèses que j'ai étudiés, et une partie présentant mes projets et travaux actuels de recherche.

2.1 Une introduction.

Les techniques non-paramétriques apparaissent naturellement lorsque le statisticien n'a pas d'idée a priori sur la forme du modèle de départ. La statistique non-paramétrique s'est intensément développée à partir des années 60-70. Il existe notamment une littérature abondante sur le problème d'estimation de fonctions telles que la fonction signal (modèle de régression ou modèle du bruit blanc Gaussien) ou encore la fonction de densité. Pour les problèmes d'estimation, s'est alors posée la question d'existence d'estimateur *optimal* parmi tous les estimateurs possibles, pour un paramètre d'intérêt appartenant à une classe fonctionnelle. La minimaxité est un critère d'optimalité qui est parfois contesté parce que trop pessimiste (au sens où il fournit le meilleur estimateur pour la plus mauvaise valeur du paramètre). Néanmoins, cela reste un des seuls critères en statistique non-paramétrique rendant possible la comparaison des procédures d'estimation.

Dans la suite de ce paragraphe, f désigne le paramètre d'intérêt qui appartient *naturellement* à un espace de dimension non finie ; par exemple, si f désigne une densité de probabilité définie sur $\mathbb{R}$ alors f appartient naturellement à l'espace des densités définies sur $\mathbb{R}$. Notons également n le nombre de données observées et $R_n(\hat{f}, f)$ le risque de l'estimateur $\hat{f}$ pour l'estimation de f (l'estimateur $\hat{f}$ est une fonction mesurable des observations). Le risque minimax est alors défini à partir du risque maximal $R_n(\hat{f})$ d'un estimateur $\hat{f}$, par :

$$\begin{aligned} R(\bar{f}) &= \inf_{\hat{f}} R_n(\hat{f}), \\ &\text{avec} \\ R_n(\hat{f}) &= \sup_f R_n(\hat{f}, f) = \sup_f \mathbb{E}_f[l(r_n^{-1} d(f, \hat{f}))], \end{aligned} \tag{2.1}$$

où $\inf_{\hat{f}}$ désigne l'infimum sur tous les estimateurs possibles de f, l est une fonction de perte ($l : \mathbb{R}^+ \to \mathbb{R}^+$), d est une distance entre éléments f, et r_n est une suite de nombres positifs qui décroît vers 0 avec n.

Farrel (1967) et Birgé (1983) ont montré que pour n'importe quel estimateur $\hat{f}$, il existe un élément f pour lequel, le risque maximal ne tend pas asymptotiquement vers 0. Pour remédier à ce mauvais comportement, une approche minimax *alternative* (a contrario de l'approche minimax classique) s'est développée à partir des années 70. Cela consiste à restreindre l'espace auquel appartient naturellement le paramètre d'intérêt, à des boules de classes fonctionnelles. L'approche minimax nécessite alors de spécifier au préalable :

- $\mathcal{B}$, la boule fonctionnelle à laquelle f appartient, qui dépend de M (le rayon de la boule) et de s (un indice de régularité de f),
- d, la distance utilisée pour le calcul du risque maximal (2.1).

Etant données $\mathcal{B}$ et d, la résolution consiste à déterminer r_n *la vitesse minimax de convergence* et à construire $\hat{f}$, *un estimateur asymptotiquement minimax* de f pour lesquels il existe deux constantes positives c_1 et c_2 telles que :

$$\lim \inf_{n \to +\infty} \inf_{\hat{f}} \sup_{f \in \mathcal{B}} \mathbb{E}_f[l(r_n^{-1} d(f, \hat{f}))] \geq c_1, \tag{2.2}$$

$$\lim \sup_{n \to +\infty} \sup_{f \in \mathcal{B}} \mathbb{E}_f[l(r_n^{-1} d(f, \hat{f}))] \leq c_2, \tag{2.3}$$

où $\inf_{\hat{f}}$ désigne l'infimum sur tous les estimateurs possibles de f et l est une fonction de perte. La relation (2.2) garantit qu'aucun estimateur ne puisse avoir une vitesse de convergence plus rapide que r_n. La relation (2.3) assure à l'estimateur $\hat{f}$ d'atteindre, à une constante près, la vitesse minimax r_n. Un choix de modèles sous-jacents (fonction signal, densité de probabilité, densité spectrale) combinés aux facteurs $\mathcal{B}$ et d a engendré un nombre important de publications. Citons en quelques unes : Pinsker (1980) a étudié le comportement asymptotique exact du risque minimax dans le modèle de bruit blanc Gaussien, pour d égale à la norme L_2 au carré et $\mathcal{B}$ un ellipsoïde de Sobolev, l'article de Nussbaum (1985) concerne l'efficacité asymptotique exacte dans un modèle de la régression pour des ellipsoïdes de Sobolev, Donoho et Johnstone (1994a) ont considéré une classe d'ellipsoïdes pour la fonction signal dans un modèle de régression et Golubev, Levit et Tsybakov (1996) ont étudié le problème d'efficacité asymptotique pour l'estimation du signal dans un modèle de bruit blanc Gaussien et pour une classe de fonctions analytiques. Une revue des résultats obtenus en estimation minimax non-paramétrique est donnée dans Tsybakov (2004).

C'est sous ce formalisme que je me suis attachée à résoudre le problème d'estimation d'un point de discontinuité d'une densité de probabilité inconnue quand cette dernière est une fonction de Lipschitz (article [5]).

Au cours des années 90, les problèmes d'estimation minimax non-paramétrique ont été explorés dans un cadre adaptatif. Il s'agit dans ce cadre de résoudre le problème minimax d'estimation en supposant la régularité du paramètre d'intérêt s inconnue, et donc de définir une procédure d'estimation dépendant exclusivement des données. La justification du cadre adaptatif n'est pas uniquement théorique ; en effet, supposer s inconnue apparaît plus réaliste d'un point de vue pratique puisque le paramètre d'intérêt est lui-même inconnu. Le cadre adaptatif permet de gagner en généralité, de fournir des estimateurs qui soient libres de toute hypothèse de régularité concernant le paramètre d'intérêt. La contre-partie est qu'il engendre parfois une perte d'efficacité dans la vitesse de convergence minimax. La théorie de l'estimation adaptative a connu un essor considérable de part l'apparition de nouvelles méthodes d'estimation avec notamment les travaux de Lepskii (1990, 1991) (procédures de Lepskii), de Donoho et Johnstone (1994b, 1995, 1996), Donoho, Johnstone, Kerkyacharian et Picard (1995, 1996), Härdle, Kerkyacharian, Picard et Tsybakov (1998) (ondelettes seuillées), Birgé et Massart (1997) et Barron, Birgé et Massart (1999) (sélection de modèles), Loubes et van de Geer (2002) (fonction de contraste pénalisée). Des procédures d'estimation adaptatives atteignant asymptotiquement le risque minimax exact sur toute une famille de classes de fonctions ont été proposées par Efroimovich et Pinsker (1984), Golubev (1987), Golubev et Nussbaum (1992), Cavalier et Tsybakov (2001) pour la fonction signal et par Golubev (1992) pour la densité.

C'est dans le cadre adaptatif que j'ai établi avec Karine Tribouley, la normalité asymptotique de la norme L_2 au carré de la fonction signal, dans le modèle du bruit blanc Gaussien, et où $\mathcal{B}$ est une boule de Besov (article [2]).

La résolution des problèmes de tests d'hypothèses via l'approche minimax s'est développée au début des années 90. Le principe de minimaxité permet de garantir une *distance de séparation minimale* entre l'hypothèse nulle H_0 et l'hypothèse alternative H_1 pour laquelle, les risques de première espèce et de seconde espèce restent bornés par des constantes fixées à l'avance. Cette distance de séparation minimale est appelée *vitesse minimax de test*.

Les travaux concernant la résolution minimax de tests d'hypothèses non-paramétriques sont postérieurs à l'emergence des études menées sur les problèmes d'estimation minimax non-paramétriques. Ces deux types de problèmes étant liés, il est naturel de vouloir les comparer. Cette comparaison n'aboutit pas à une conclusion unique : le choix de la métrique d conduit à des conclusions différentes. A titre d'exemple, si d correspond à la norme L_1, la vitesse de test minimax coïncide avec la vitesse minimax d'estimation du paramètre f ; en revanche si d désigne la norme L_2, la vitesse minimax de test est plus rapide que la vitesse minimax d'estimation du paramètre f et elle coïncide avec la vitesse minimax d'estimation de la norme L_2 de f (Lepskii, Nemirovskii et Spokoiny, 1999). Les travaux pionniers sur le sujet sont dus à Ingster (1982) et une bibliographie des différents résultats établis dans les problèmes de tests minimax d'hypothèses est disponible dans Ingster (1993) et dans Ingster et Suslina (2002).

Désignons à nouveau par f le paramètre d'intérêt, par $\mathcal{B}$ la boule fonctionnelle à laquelle f appartient, par d la distance utilisée entre les éléments de $\mathcal{B}$ et par n le nombre d'observations. Notons également Δ_n, un test statistique (fonction mesurable des observations à valeurs dans $\{0,1\}$, telle que si $\Delta_n = 1$, alors on rejette H_0 et si $\Delta_n = 0$, alors on accepte H_0). A partir des observations, nous considérons le problème de test suivant :

$$H_0 \ : \ f \in \mathcal{F}_0 \subset \mathcal{B} \quad \text{contre} \quad H_1 \ : \ f \in \Lambda_n(v_n) = \{f \in \mathcal{B} \ : \ \inf_{f_0 \in \mathcal{F}_0} d(f, f_0) \geq v_n\}, \tag{2.4}$$

où v_n est une suite de nombres positifs qui décroît vers 0 avec n et $\mathcal{F}_0$ est une classe d'éléments f entièrement spécifiée. Si la classe $\mathcal{F}_0$ est réduite à un seul élément, alors c'est une hypothèse nulle simple. Si $\mathcal{F}_0$ est une classe paramétrique ou bien une classe plus riche (cela nécessite un contrôle de sa complexité) alors c'est une hypothèse nulle composée. L'alternative $\Lambda_n(v_n)$ est caractérisée par trois facteurs : la boule $\mathcal{B}$, la distance d entre éléments f et v_n. La suite v_n correspond à la distance de séparation minimale permettant de tester H_0 contre H_1, avec des erreurs de première espèce et de seconde espèce bornées. Plus précisément, étant données $\mathcal{B}$, $\mathcal{F}_0$ et d, la résolution consiste à déterminer v_n, *la vitesse minimax de test* et à construire $\hat{\Delta}_n$, *un test statistique asymptotiquement de niveau α fixé dans* $(0,1)$ telles que pour tout α_1 et α_2 fixés dans $(0,1)$, il existe deux constantes positives a et A $(a \leq A)$ vérifiant :

$$\liminf_{n \to +\infty} \inf_{\Delta_n} \sup_{f \in \Lambda_n(a v_n)} \mathbb{P}_f[\Delta_n = 0] \geq \alpha_1, \tag{2.5}$$

$$\limsup_{n \to +\infty} \sup_{f \in \Lambda_n(A v_n)} \mathbb{P}_f[\hat{\Delta}_n = 0] \leq \alpha_2, \tag{2.6}$$

où $\inf\limits_{\Delta_n}$ est l'infimum pris sur l'ensemble des tests statistiques de niveau asymptotique α. La relation (2.5) garantit que les hypothèses H_0 et H_1 ne peuvent être *distinguées* favorablement si la distance les séparant est d'un ordre inférieur à v_n. La relation (2.6) assure à la statistique $\hat{\Delta}_n$ d'atteindre la vitesse minimax de test v_n.

C'est sous ce formalisme que je me suis attachée à résoudre le problème de test d'hypothèses composées (nulle et alternative) sur le contour d'images dans un modèle binaire d'images, avec $\mathcal{B}$ définie par une classe d'ensembles dont la frontière a une longueur finie, et avec d égale à la mesure de Lebesgue de la différence symétrique entre deux ensembles (article [6], en collaboration avec Alexandre Tsybakov). Il en est de même pour la résolution du problème de test d'hypothèses composées (nulle et alternative) sur la fonction signal, dans un modèle de régression discrète, avec $\mathcal{B}$ définie par une boule de Hölder, et avec d égale à la norme L_2 (article [4], en collaboration avec Christophe Pouet).

Comme pour l'estimation, les problèmes minimax de test d'hypothèses ont été étudiés dans un cadre adaptatif. Cela permet en particulier de fournir des procédures de test qui sont libres de tout paramètre de régularité. D'un point de vue pratique, le gain est incontestable, mais comme pour l'estimation, le cadre adaptatif engendre une perte d'efficacité dans la vitesse minimax de test (article [7] en collaboration avec Christophe Pouet). On peut citer sur cette thématique les travaux de Spokoiny (1996, 1998), Horowitz et Spokoiny (2001), Guerre et Lavergne (2002), Ingster et Suslina (2002) et Baraud, Huet et Laurent (2003).

2.2 Problèmes d'estimation.

2.2.1 L'intégrale de la fonction signal au carré.

Cette partie décrit l'article [2], écrit en collaboration avec Karine Tribouley. Nous nous sommes intéressées à l'estimation adaptative de $\theta = \int_0^1 f^2(t)dt$ où f est la fonction signal dans le modèle du bruit blanc Gaussien. Le processus $Y_n(t)$ est observé sur $[0,1]$ et il généré par l'équation :

$$dY_n(t) = f(t) + n^{-1/2}dB(t), \quad t \in [0,1], \quad n \in \mathbb{N}^\star \tag{2.7}$$

où $B(\cdot)$ est un mouvement Brownien. La fonction f est supposée appartenir à une boule de Besov $\mathcal{B} = B_{s,2,\infty}(M)$, avec $s > 0$ et $M > 0$. La performance des estimateurs est mesurée par le risque quadratique autement dit, d est la distance Euclidienne dans $\mathbb{R}$, et l est la perte au carré.

L'estimation de θ ou plus généralement de fonctionnelles intégrales sur f a été abondamment étudiée au cours des vingt dernières années. Le papier pionnier sur le sujet est attribué à Bickel et Ritov (1988) dans le modèle de densité lorsque cette dernière est Höldérienne (estimation à noyau). Laurent (1996) a également étudié ce problème d'estimation pour une densité appartenant à une classe d'ellipsoïdes. Bickel et Ritov (1988), puis Birgé et Massart (1995) ont établi les bornes inférieures, pour le probléme d'estimation de fonctionnelles intégrales de la densité. L'estimation de l'intégrale de la fonction signal au carré dans le modèle du bruit blanc Gaussien a été traitée dans Ibragimov, Khas'minskii et Nemirovskii (1986) et dans Donoho et Nussbaum (1990) pour une classe d'ellipsoïdes, et dans Lepskii, Nemirovskii et Spokoiny (1999) (cadre plus général puisque estimation de $\|f\|_{L_r}$, pour $r \geq 1$) pour une classe de Hölder. La version adaptative du problème dans le modèle de la densité a été étudiée par Laurent (2005) pour des boules de Besov. L'estimation adaptative de l'intégrale du carré du signal dans le modèle du bruit blanc Gaussien a été traitée par Efromovich et Low (1996) (estimation basée sur les procédures de Lepskii) et par Laurent et Massart (2000) (estimation construite par sélection de modèles).

Lorsque la régularité s de la fonction signal est connue, il est à présent bien connu que dans le cas régulier ($s > 1/4$), la vitesse minimax d'estimation de θ est égale à la vitesse paramétrique ($\tilde{r}_n = n^{-1/2}$) tandis que dans le cas irrégulier ($s \leq 1/4$), la vitesse minimax est $\tilde{r}_n = n^{-4s/(4s+1)}$. Pour $s > 1/4$, le résultat est encore plus précis puisque la constante c_1 dans la relation (2.2) est explicite, et est égale à $c_1 = 4\theta$.

Dans le cas adaptatif i.e. s non connue, Efromovich et Low (1996) et Laurent et Massart (2000) ont proposé un estimateur adaptatif qui atteint la vitesse minimax adaptative suivante :

$$r_n \quad = \quad \begin{cases} n^{-1/2} & \text{quand} \quad s > 1/4, \\ (n\sqrt{1/\log(n)})^{-4s/(4s+1)} & \text{quand} \quad s \leq 1/4. \end{cases} \tag{2.8}$$

L'adaptation engendre donc dans le cas irrégulier une perte d'efficacité de l'ordre d'un terme logarithmique.

Dans notre article, nous proposons un estimateur adaptatif construit sur des méthodes d'ondelettes seuillées qui atteint la vitesse minimax d'estimation définie par (2.8). Le résultat principal de notre article est un théorème de la limite centrale adaptatif dans lequel la vitesse est indépendante de s et elle est aléatoire. Ce résultat fournit des intervalles de confiance adaptatifs autour de θ et permet de résoudre des problèmes de test sur θ. Signalons que notre résultat est du même type que celui de Picard et Tribouley (2000), qui obtiennent un théorème de la limite centrale adaptatif pour la fonction signal en un point fixé, dans les modèles de régression discrète et de bruit blanc Gaussien.

Procédures minimax d'estimation $\hat{\theta}_n$.

Notons φ la fonction d'échelle associée à l'ondelette ψ. Ces fonctions ont un support compact et le q-ième moment de la fonction ψ s'annule pour $q = 1, \ldots, N$, où N est une constante positive fixée à l'avance. Pour tout niveau de résolution $j_0 > 0$ (j_0 entier), on transforme le modèle initial (2.7) en une suite de modèles par niveau

$$\hat\alpha_{j_0k} = \alpha_{j_0k} + n^{-1/2}\zeta_{j_0k}, \qquad k = 0,\dots,2^{j_0},$$
$$\hat\beta_{jk} = \beta_{jk} + n^{-1/2}\epsilon_{jk}, \qquad j = j_0,\dots,\ k = 0,\dots,2^{j},$$

où
$$\hat\beta_{jk} = \int \psi_{jk}dY_n(t),\ \beta_{jk} = \int \psi_{jk}f,\ \hat\alpha_{j_0k} = \int \phi_{j_0k}dY_n(t),\ \alpha_{j_0k} = \int \phi_{j_0k}f,$$
et $\{\epsilon_{jk} = \int \psi_{jk}dB,\ \zeta_{j_0l} = \int \phi_{j_0l}dB\}_{j,k,l}$ est une famille de variables indépendantes Gaussiennes centrées réduites.

Lorsque s est connue, l'estimateur $\tilde\theta_n = \sum_k(\hat\alpha^2_{j_{s,n}k} - \frac{1}{n})$ qui atteint la vitesse minimax $\tilde r_n$, est construit à partir du niveau $j_{s,n}$. Ce niveau dépend fortement de s et il est tel que $2^{j_{s,n}} = 2^{j_{s,n,1}} = n^{2/(1+4s)}$ quand $s \le 1/4$ et $2^{j_{s,n}} = 2^{j_{s,n,2}} = n(\log(n))^{-\gamma}$, pour un certain γ positif quand $s > 1/4$. Dans le cas adaptatif, l'estimateur doit être indépendant de s. La caractérisation suivante "$f \in B_{s,2,\infty}(M) \Longrightarrow \forall j,\quad \sum_k \beta^2_{jk} \le M^2 2^{-2js}$", nous a conduit à considérer le plus petit seuil t_{jn} indépendant de s pour lequel on a $\forall j \ge j_s^\star,\quad \sum_k \beta^2_{jk} \le t_{jn}$. La quantité t_{jn} correspond à un log près à $2^{-2j_{s,n,1}s}$, plus précisément, $t_{jn} = \kappa^2\left(\frac{2^j\log(n)}{n^2}\right)^{1/2}$, pour une certaine constante positive κ. Le niveau j_s^* réalise à une constante près la balance entre $2^{-2j_s^*s}$ et t_{jn} (la quantité j_s^* dépend de s et corrrespond à une légère modification de $j_{s,n,1}$). L'estimateur qui en découle est un estimateur seuillé défini par :

$$\hat\theta_n = \sum_k(\hat\alpha^2_{j_0k} - \frac{1}{n}) + \sum_{j_0}^{j_1}\sum_k(\hat\beta^2_{jk} - \frac{1}{n})1\!\!1_{\{\sum_k(\hat\beta^2_{jk}-\frac{1}{n})\ge t_{jn}\}}, \tag{2.9}$$

où $2^{j_0} = n/(\log n)^\gamma$, et $2^{j_1} = n$ pour un certain $\gamma > 0$.
Bien que l'estimateur $\hat\theta_n$ soit adaptatif, il subsiste dans la décomposition de l'erreur $\theta - \hat\theta_n$ un terme de biais aléatoire B_n. C'est la présence de B_n qui engendre, pour l'adaptation, la perte d'un facteur logarithmique dans la vitesse minimax d'estimation. Le terme B_n étant dominant dans le cas irrégulier, il va falloir l'estimer. Il est défini par :

$$B_n = \sum_{j_0}^{j_1}\sum_k \beta^2_{jk}1\!\!1_{\{\sum_k(\hat\beta^2_{jk}-\frac{1}{n})<t_{j,n}\}}1\!\!1_{\{\sum_k\beta^2_{jk}<2t_{j,n}\}}.$$

L'estimation de B_n nécessite l'ajout d'une hypothèse, notée $[H_s(M,\kappa)]$. Elle consiste à contraindre les fonctions irrégulières ($s < 1/4$) "à avoir *une irrégularité régulière*" (une hypothèse similaire est considérée dans Picard, Tribouley (2000)). Soulignons que $[H_s(M,\kappa)]$ est moins restrictive qu'une hypothèse concernant l'existence d'une borne inférieure pour la régularité s (ce qui est classiquement supposé dans un contexte adaptatif).

$[H_s(M,\kappa)]$: il existe une suite $\rho_n > 0$ telle que pour tout $n \geq 2$

$$\exists j, \quad j_s^* - \rho_n \leq j \leq j_s^* \ : \quad \sum_k \beta_{jk}^2 \geq 4M^2 2^{-2j_s^* s}.$$

Le niveau j_s^* étant dépendant de s, nous allons l'estimer par $\hat{\jmath}$ qui en probabilité sera inférieur à j_s^*. Pour ce faire, l'échantillon initial est partitionné en deux sous-échantillons de même taille qui seront repérés par "(1)" pour le premier sous-échantillon, et par "(2)" pour le second sous-échantillon. L'estimateur $\hat{\theta}_n$ défini par (2.9) devient alors $\hat{\theta}_n = \hat{\theta}_n(1)$: il dépend donc uniquement du sous-échantillon (1). L'estimateur $\hat{\jmath} = \hat{\jmath}(2)$ est alors défini par :

$$\hat{\jmath} = \hat{\jmath}(2) = \begin{cases} j_0 \quad \text{si aucun} \quad j \in \{j_0, \ldots, j_1\}, \quad \text{ne satisfait} \quad \sum_k (\hat{\beta}_{jk}(2)^2 - \frac{1}{n}) \geq t_{j,n}/4, \\ \sup\{j \in \{j_0, \ldots, j_1\} \ : \ \sum_k (\hat{\beta}_{jk}(2)^2 - \frac{1}{n}) \geq t_{j,n}/4\}. \end{cases}$$

Soulignons que dans le cas régulier, l'estimation de B_n est inutile ; dans ce cas, $\hat{\jmath}(2)$ est, en probabilité, presque égal à j_0. Nous pouvons alors estimer B_n par :

$$\hat{B}_n = \sum_{\hat{\jmath}(2) \leq j \leq \hat{\jmath}(2)+\tau_n} \sum_k \left(\hat{\beta}_{jk}(1)^2 - \frac{1}{n} \right) \mathbb{1}_{\hat{\jmath}(2) \neq j_0},$$

pour une certaine suite τ_n de réels positifs.

Théorème de la limite centrale adaptatif.

Soient $\kappa^2 > 32$, $s \in]0, N]$ et $M > 0$. Nous supposons que $f \in B_{s,2,\infty}(M)$ et que l'hypothèse $[H_s(M,\kappa)]$ est satisfaite dans le cas $s \leq 1/4$. Si τ_n est tel que

$$2 \sup(\rho_n, \log_2 \log n) < \tau_n < 2(\kappa^2/16)^2 \log_2 n,$$

alors

$$d(n) \left(\hat{\theta}_n(1) - \theta + \hat{B}_n \right) \xrightarrow{\mathcal{D}}_{n \to +\infty} \mathcal{N}(0,1), \tag{2.10}$$

où la vitesse aléatoire est donnée par

$$d(n) = \left(\sum_k (\frac{4\alpha_{j_0 k}^2}{n} + \frac{2}{n^2}) + \sum_{\hat{\jmath}(2) \leq j \leq \hat{\jmath}(2)+\tau_n} \sum_k (\frac{4\beta_{jk}^2}{n} + \frac{2}{n^2}) \ \mathbb{1}_{\{\hat{\jmath}(2) \neq j_0\}} \right)^{-1/2}.$$

D'un point de vue pratique, la relation (2.10) est inutilisable car dépendante de quantités inconnues. Nous allons donc estimer $d(n)$ à partir des quantités suivantes :

$$\begin{aligned}
\hat{w}(n) &= \hat{v}_0(n) + \hat{c}(n) 1_{\{\hat{\jmath}(2) \neq j_0\}} \\
&\text{où} \\
\hat{v}_0(n) &= \frac{4}{n} \sum_k (\hat{\alpha}_{j_0 k}(1)^2 - \frac{1}{n}), \\
\hat{c}(n) &= \frac{4}{n} \sum_{\hat{\jmath}(2)}^{\hat{\jmath}(2)+\tau_n} \sum_k (\hat{\beta}_{jk}(1)^2 - \frac{1}{n}),
\end{aligned}$$

sont des estimateurs du carré de l'inverse des vitesses de convergence du cas régulier et du cas irrégulier. Nous montrons que $\hat{w}(n) = d(n)^{-2}(1 + o_P(1))$.

COMMENTAIRES.

- Si $\Phi^{-1}(\alpha)$ est le α-quantile d'une normale centrée réduite, notre résultat fournit un intervalle de confiance asymptotiquement de niveau α autour de θ, qui a la forme :

$$\left[\hat{\theta}_n(1) + \hat{B}_n - \hat{w}(n)^{1/2}\Phi^{-1}(\alpha/2),\ \hat{\theta}_n(1) + \hat{B}_n + \hat{w}(n)^{1/2}\Phi^{-1}(\alpha/2)\right]$$

- Notre résultat permet de résoudre les tests paramétriques du type : $H_0 \ : \ \theta = \theta_0$ contre $H_1 \ : \ \theta = \theta_1$, avec $\theta_0 < \theta_1$ fixés. Tester H_0 contre H_1, avec un risque de première espèce asymptotiquement égal à α, est possible en prenant un test dont la région critique est :

$$\mathcal{R} = \{\hat{\theta}_n(1) + \hat{B}_n \ > \theta_0 + \hat{w}(n)^{1/2}\Phi^{-1}(1-\alpha)\}.$$

  La puissance est alors égale à $\eta_n = 1 - \Phi((\theta_0 - \theta_1)\hat{w}(n)^{-1/2} + \Phi^{-1}(1-\alpha))$.
- Je n'ai pas présenté tous les résultats que nous avons obtenus dans cet article. Nous avons également montré que l'estimateur adaptatif $\hat{\theta}_n(1)$ atteint la vitesse minimax adaptative définie par (2.8) et que dans le cas régulier, notre estimateur est asymptotiquement efficace.

2.2.2 Point de discontinuité d'une densité de probabilité.

Le statisticien observe simultanément $(X_1, \ldots, X_n)$, n variables aléatoires réelles i.i.d., admettant une densité de probabilité f inconnue. La densité f est définie sur $[0, 1]$ et elle admet un unique point de discontinuité θ.

Je me suis intéressée à déterminer la vitesse minimax dans le problème d'estimation de θ (article [5]). J'ai donc établi les relations (2.2) et (2.3) dans lequelles :

– $\mathcal{B} = \{f$ densités définies sur $[0,1]$: f admet un unique point de discontinuité $\theta \in (h, 1-h)$ avec h fixé dans $(0,1/2)$, la taille du saut en θ est au moins égale à $a > 0$, f est continue à droite, f est bornée inférieurement par m et supérieurement par M, f est Lipschitz en dehors de θ avec L pour constante de Lipschitz$\}$ (les constantes h, a, m et M sont données)
– le risque intervenant dans les relations (2.2) et (2.3) pour un estimateur $\hat{\theta}_n$ est :

$$\mathbb{E}_f[l(r_n^{-1}d(\theta, \hat{\theta}_n))] \quad = \quad \mathbb{E}_f[\mathbb{1}_{\{r_n^{-1}|\theta-\hat{\theta}_n| \geq C\}}], \tag{2.11}$$

où C est une constante positive.

L'estimation minimax d'un point de discontinuité à partir de données uni-dimensionnelles a été étudié par Korostelev (1987), dans le modèle du bruit blanc Gaussien, et pour une classe de fonctions similaire à la classe $\mathcal{B}$ définie ci-dessus. Korostelev (1987) a montré que la vitesse minimax de convergence est la même que la vitesse minimax obtenue par Ibragimov et Khas'minskii (1981), dans une formulation paramétrique du problème. La vitesse minimax dans ce cas est $r_n = n^{-1}$. Raimondo (1998) a considéré l'estimation d'un point de discontinuité de la α-ième dérivée de la fonction de régression, qui satisfait la condition de Lipschitz en dehors du point de discontinuité ($\alpha \geq 0$). Sa vitesse de convergence n'est pas optimale (elle l'est uniquement lorsque $\alpha < 1/2$). C'est Neumann (1997) dans un modéle de déconvolution de la densité qui a déterminé la vitesse minimax d'estimation d'un point de discontinuité : il s'avère que la vitesse minimax est différente selon que α est supérieur ou inférieur à 1/2. Goldenshluger, Tsybakov et Zeevi (2006) ont généralisé l'étude de Neumann (1997) en établissant la vitesse minimax d'estimation d'un point de discontinuité dans un problème inverse bruité défini à partir d'un noyau de convolution : cette vitesse dépend de la régularité que possède la fonction en dehors de son point de discontinuité et du comportement de la queue de la transformée de Fourier du noyau de convolution. Chu et Cheng (1996) ont proposé des estimateurs de p points de discontinuité d'une densité, pour une classe de densités Lipschitziennes en dehors des points de discontinuité, et lorsque les points de discontinuité sont supposés être distants d'au moins une constante strictement positive. Leur méthode d'estimation consiste à maximiser la valeur absolue d'une difference d'estimateurs à noyau de la densité mais elle ne permet pas d'atteindre la vitesse minimax.

Dans l'article [5], nous avons proposé un estimateur asymptotiquement minimax du point de discontinuité qui, de part sa construction, est simple à implémenter.

Procédures minimax d'estimation $\hat{\theta}_n$.

Notre procédure d'estimation requière la construction d'un estimateur préliminaire $\tilde{\theta}_n$. Notons $\tilde{N} = \tau_n^{-1} = [n/\log n]$ et $r_n = 1/n$ ($[\cdot]$ désigne la partie entière). Partitionons $[0,1]$ en $\tilde{N}$ intervalles réguliers $[\tilde{b}_{k-1}, \tilde{b}_k]$ $\forall k \in \{1, \ldots, \tilde{N}-1\}$, avec $0 = \tilde{b}_0 < \tilde{b}_1 < \ldots < \tilde{b}_{\tilde{N}} =$

1, de longueur τ_n. Pour tout $k \in \{2, \ldots, \tilde{N}\}$, nous considérons $\tilde{Y}_k = \tilde{f}_{n,k} - \tilde{f}_{n,k-1}$, où $\tilde{f}_{n,k}$ est l'histogramme de $f(x)$ pour n'importe quel point $x \in [\tilde{b}_{k-1}, \tilde{b}_k)$, respectivement de $[\tilde{b}_{\tilde{N}-1}, \tilde{b}_{\tilde{N}}]$, quand $k = \tilde{N}$. Notons qu'une différence entre $\tilde{f}_{n,k}$ et $\tilde{f}_{n,k-1}$ suffisamment grande, indique probablement la présence de θ dans l'intervalle $[\tilde{b}_{k-2}, \tilde{b}_k)$. Ainsi, $\tilde{\theta}_n$ est défini par :

$$\tilde{\theta}_n = \tilde{b}_{k^\star}, \quad \text{où} \quad k^\star = \begin{cases} \min\{2 \leq k < \tilde{N} : |\tilde{Y}_k| \geq ca\}, & \text{où } 0 < c \leq 1 \text{ est une constante fixée} \\ \tilde{N}, & \text{sinon.} \end{cases}$$

La deuxième étape de la procédure d'estimation consiste à découper $[\tilde{b}_{k^\star-2}, \tilde{b}_{k^\star}]$ en $N = [2\tau_n r_n^{-1}]$ intervalles réguliers de manière à localiser plus précisément θ. La nouvelle partition est formée par les noeuds $\tilde{b}_{k^\star-2} = b_0 < b_1 < \ldots < b_N = \tilde{b}_{k^\star}$ équi-espacés de longueur de l'ordre de r_n. Pour tout $j \in \{1, \ldots, N\}$, nous considérons $Y_j = \dfrac{1}{j}\displaystyle\sum_{k=1}^{j} |f_{n,k} - \tilde{f}_{n,k^\star-2}|$, où $f_{n,k}$ est l'histogramme de $f(x)$ pour n'importe quel point $x \in [b_{k-1}, b_k)$, respectivement de $[b_{N-1}, b_N]$ lorsque $k = N$. L'estimateur $\hat{\theta}_n$ est alors défini par :

$$\hat{\theta}_n = b_{j^\star}, \quad \text{où} \quad j^\star = \begin{cases} \min\{j \in \{1, \ldots, N\} : Y_j \geq c'a\}, & \text{avec } 0 < c' \leq 1 \text{ fixée} \\ N, & \text{sinon.} \end{cases} \tag{2.12}$$

Une grande valeur de Y_j indique probablement la présence de θ dans $[b_{j-1}, b_j]$.

Résultats et commentaires.

Nous considérons le risque défini par l'équation (2.11), qui correspond à la probabilité que la distance renormalisée entre l'estimateur et la quantité estimée dépasse un seuil $C > 0$. Nos résultats s'énoncent comme suit. Rappelons que $r_n = n^{-1}$, alors

1. il existe une constante $C > 0$ telle que

$$\liminf_{n\to\infty} \inf_{\theta_n} \sup_{f\in\mathcal{B}} \mathbb{P}_f(r_n^{-1}|\theta - \theta_n| \geq C) \geq p_0,$$

   pour p_0 une constante strictement positive et où $\inf_{\theta_n}$ désigne l'infimum sur l'ensemble des estimateurs possibles de θ,

2. la famille des estimateurs $\hat{\theta}_n$ définie par la relation (2.12) est telle que

$$\lim_{C\to\infty} \limsup_{n\to\infty} \sup_{f\in\mathcal{B}} \mathbb{P}_f(r_n^{-1}|\hat{\theta}_n - \theta| > C) = 0.$$

Cela me conduit à faire quelques remarques.

– La vitesse minimax que nous obtenons dans le contexte non-paramétrique est identique à celle obtenue dans le cadre paramétrique d'estimation d'un point de discontinuité (Ibragimov et Khas'minskii, 1981). Ce résultat est propre aux problèmes d'estimation semi-paramétriques réguliers.
– Nous avons également établi la vitesse de convergence de l'estimateur préliminaire $\tilde{\theta}_n$. Il atteint la vitesse de convergence $\log(n)n^{-1}$, il est par conséquent "presque" minimax.
– Notre procédure d'estimation est relativement simple à implémenter et elle permet d'atteindre la vitesse optimale. Nous aurions pu construire nos estimateurs à partir d'outils plus sophistiqués, comme les estimateurs à noyau : cela aurait conduit à la même vitesse de convergence.
– Nous aurions pu résoudre le problème d'estimation d'un point de discontinuité pour une fonction de perte du type $w(u) = |u|^d$, où d est un entier strictement positif. Dans ce cas, notre estimateur n'aurait pas permis d'atteindre la vitesse minimax r_n. Sa vitesse aurait été de l'ordre de $\log(n)n^{-1}$ (induite par l'étape préliminaire d'estimation).
– Nous avons étudié uniquement le cas uni-dimensionnel. Contrairement au cas univarié, la notion de discontinuité d'une densité dans un cadre multi-dimensionnel, est à formuler explicitement (elle peut être déclinée sous différentes formes). A titre d'exemple, Freidlin et Korostelev (1995) ont considéré des domaines G_t grandissant dans $\mathbb{R}^2$ au cours du temps dans un modèle continu d'images binaires. Ils ont considéré la discontinuité comme étant l'instant inconnu τ à partir duquel le domaine grandit. Un autre type de discontinuité caractérisée par un saut dans la fonction signal d'un modèle binaire continu ou discret d'images a été considéré par Khas'minskii et Lebedev (1990), et par Korostelev et Tsybakov (1993). Korostelev et Tsybakov (1993) ont étudié l'estimation minimax d'ensembles dans un modèle de régression binaire en considérant comme déterministes ou aléatoires les variables d'entrée et pour une classe d'ensembles réguliers. La discontinuité correspond à la frontière de ces ensembles et c'est l'ensemble tout entier qu'ils estiment. Tsybakov (1994) a généralisé le problème d'estimation d'une discontinuité en multivarié au travers de l'estimation d'ensembles G en considérant trois modèles distincts : modèle de régression sur la segmentation d'images (G correspond à une image), saut d'une densité supposée inconnue (G désigne le support de la densité), processus de Poisson dont l'intensité est strictement positive sur un ensemble compact G et nulle en dehors de G. Pour ces modèles et des classes d'ensembles réguliers, Tsybakov (1994) a obtenu les vitesses minimax d'estimation pour la distance de Hausdorff et la mesure de Lebesgue de la différence symétrique.

2.3 Problèmes de test d'hypothèses composées.

Il me faut souligner qu'une des difficultés principales dans la résolution minimax de problèmes de test d'hypothèses dans le cas d'hypothèse nulle composée, réside dans l'obtention de la borne inférieure (2.5). Cette difficulté est inhérente à la construction d'une famille paramétrique qui doit être d'une part, incluse dans $\mathcal{B}$ et d'autre part, séparée de tout élément appartenant à $\mathcal{F}_0$ ($\mathcal{F}_0$ caractérise l'hypothèse nulle). L'obtention de la borne inférieure requière l'existence d'au moins un élément f_0 dans $\mathcal{F}_0$ vérifiant des conditions supplémentaires. Nous nommerons par la suite abusivement l'élément f_0 comme étant *l'élément plus régulier de* $\mathcal{F}_0$. C'est autour de f_0 que sera construite une famille paramétrique randomisée incluse dans $\mathcal{B}$ et séparée de tout élément de $\mathcal{F}_0$.

2.3.1 Modèle de régression binaire d'images.

En collaboration avec Alexandre Tsybakov, nous nous sommes intéressés dans l'article [6] à un problème de test d'hypothèses, dans un modèle de régression binaire d'images. Nous observons $Y_1, \ldots, Y_n$, n variables aléatoires réelles i.i.d. générées par l'équation suivante :

$$Y_i = \mathbb{1}_{x_i \in G} + \xi_i, \quad i \in \{1, \ldots, n\},$$

où $G \subset [0,1]^2$ est le paramètre d'intérêt, les ξ_i sont des variables i.i.d. centrées de variance finie σ^2, et les x_i sont déterministes et répartis régulièrement sur la grille $\Gamma \subset K = [0,1] \times [0,1]$, avec $\Gamma = \{x = (\frac{j}{\sqrt{n}}, \frac{k}{\sqrt{n}}),\ 1 \leq j,\ k \leq \sqrt{n},\ j,\ k\ \text{ entiers}\}$.

Nous avons considéré le problème de test minimax défini par (2.4) pour lequel,

$$\begin{aligned}
\mathcal{B} &= \mathcal{G}(\epsilon_0, L_0),\ \epsilon_0 > 0, L_0 > 0 \text{ fixés},\\[2ex]
\mathcal{G}(\epsilon_0, L_0) &= \{G\ :\ \forall \epsilon \in (0, \epsilon_0),\ \text{ les } \epsilon\text{-voisinages pour la norme Euclidienne de } \partial G \cap int K\\
&\qquad \text{ont une mesure de Lebesgue inférieure à } \ L_0 \epsilon,\ \text{où } \partial G \text{ est la frontière de } G\ \},\\[2ex]
\mathcal{F}_0 &= \mathcal{G}_\Theta = \{G_\theta \in \mathcal{G}(\epsilon_0, L_0), \theta \in \Theta\},\ \Theta \subset \mathbb{R}^k,\ k \in \mathbb{N}^\star,\\
&\qquad \text{et } \ \forall \theta \in \Theta,\ G_\theta \text{ est un ensemble fermé de } K,\\[2ex]
d(G_1, G_2) &= \mu(G_1 \Delta G_2) \text{ où } \mu \text{ est la mesure de Lebesgue sur } \mathbb{R}^2 \text{ et } G_1 \Delta G_2\\
&\qquad \text{est la différence symétrique entre } G_1 \text{ et } G_2.
\end{aligned}$$

Soulignons qu'un ensemble G est dans $\mathcal{G}(\epsilon_0, L_0)$ pour un certain $\epsilon_0 > 0$ si ∂G est une courbe rectifiable de longueur inférieure à une constante $L' = L'(\epsilon_0, L_0)$. Par exemple, ∂G

peut être le graphe d'une fonction de Lipschitz définie sur $[0,1]$ et à valeurs dans $(0,1)$. Nous considérons uniquement les ensembles appartenant à $\mathcal{G}(\epsilon_0, L_0)$ afin d'exclure les cas pathologiques d'ensembles qui ont une frontière très oscillante ou qui ont une longueur de frontière infinie.

Il existe une littérature abondante sur le problème d'estimation de discontinuité dans un modèle d'images : nous avons cité quelques articles sur ce sujet dans la partie 2.2.2. Les premiers travaux sur les problèmes de test minimax d'hypothèses concernant des ensembles sont disponibles dans l'article [3], qui a constitué une partie de ma thèse. Il s'agissait de tester une hypothèse nulle simple contre une alternative non-paramétrique, sur les supports de densité et de certaines de leurs fonctionnelles intégrales, lorsque les ensembles étaient supposés appartenir à la classe des fragments (la classe des fragments est constituée d'ensembles définis sous une frontière, qui est une fonction réelle Höldérienne).

Dans l'article [6], le problème de test auquel nous nous sommes intéressés est différent : le modèle est celui d'une régression binaire d'images et aucune hypothèse a priori de régularité sur l'ensemble n'est imposée.

Test statistique.

Nous considérons $\hat{\theta}_n$ l'estimateur des moindres carrés de θ défini par

$$\sum_{i=1}^{n}(Y_i - \mathbb{1}_{x_i \in G_{\hat{\theta}_n}})^2 = \min_{\theta\in\Theta}\sum_{i=1}^{n}(Y_i - \mathbb{1}_{x_i\in G_\theta})^2. \tag{2.13}$$

Le test statistique est alors défini par

$$\begin{aligned} \Delta_n &= \mathbb{1}_{\{T_n > c_0 n^{-1/2}\}}, && (2.14)\\ &\text{où} \\ c_0 \ \ \text{est le} & \ \ (1-\alpha) - \text{quantile d'une loi normale standard}, \\ T_n &= \frac{1}{n\sqrt{\mu}_4}\sum_{i=1}^{n}(Y_i - \mathbb{1}_{x_i\in G_{\hat{\theta}_n}})^2 - \frac{1}{\sqrt{\mu}_4}\sigma^2, && (2.15)\\ &\text{avec} \\ \mu_4 &= \mathbb{E}(\xi_i^4) - \sigma^4. \end{aligned}$$

Nous admettons que $\mathbb{E}(\xi_i^4)$ existe. Soulignons que l'unicité de $\hat{\theta}_n$ n'est pas requise : notre résultat est valide pour n'importe quel $\hat{\theta}_n$ vérifiant (2.13). D'un point de vue pratique, il n'est pas raisonnable de supposer σ^2 et μ_4 connus. Pour remédier à cela, nous numérotons les points x_i de la grille Γ de sorte que $\|x_i - x_{i-1}\| = 1/\sqrt{n}$, où $\|\cdot\|$ est la norme Euclidienne dans $\mathbb{R}^2$. Nous pouvons alors définir $\hat{\sigma}^2$ et $\hat{\mu}_4$ les estimateurs de σ^2 et de μ_4 par

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{2(n-1)}\sum_{i=2}^{n}(Y_i - Y_{i-1})^2, \\
\hat{\mu}_4 &= \frac{1}{2(n-1)}\sum_{i=2}^{n}(Y_i - Y_{i-1})^4 - \left(\frac{1}{n-1}\sum_{i=2}^{n}(Y_i - Y_{i-1})^2\right)^2.
\end{aligned}$$

Ces estimateurs sont des estimateurs consistants de σ^2 et de μ_4 uniformément sur la classe $\mathcal{G}(\varepsilon_0, L_0)$: la consistance est par conséquent valide sous H_0 et sous H_1. Notons $\hat{T}_n$ la statistique définie par l'expression (2.15) dans laquelle $\hat{\sigma}^2$ et $\hat{\mu}_4$ remplacent σ^2 et μ_4. Le test statistique actualisé $\hat{\Delta}_n$ (qui remplace Δ_n défini par (2.14)) est alors

$$\hat{\Delta}_n = \mathbb{1}_{\hat{T}_n > c_0 n^{-1/2}}. \tag{2.16}$$

Résultats et commentaires.

Nous supposons les hypothèses suivantes :

- [A1] Les ξ_i sont i.i.d., centrées de variance σ^2 finie et elles admettent un moment d'ordre 4.
- [A2] Il existe un $\epsilon_0 > 0$ tel que $\mathcal{G}_\Theta \subset \mathcal{G}(\epsilon_0, L_0)$.
- [A3] La dimension de Vapnik-Chervonenkis V de $\mathcal{G}_\Theta$ est finie.

Sous l'hypothèse [A1], $\hat{\Delta}_n$ défini par (2.16) est asymptotiquement de niveau α.

Sous [A1]-[A3] et si $\exists\ B > 0$ telle que $|\xi_i| < B$ p.s. pour tout $i = 1, \ldots, n$, alors pour n'importe quel $\alpha_1 \in (0,1)$, il existe une constante $A > 0$ telle que la relation (2.6) est satisfaite avec $v_n = n^{-1/2}$ et $\hat{\Delta}_n$ défini par (2.16).
Soulignons que, l'hypothèse additionnelle, qui spécifie que les erreurs sont bornées p.s., est une condition technique. Néanmoins, elle semble relativement naturelle dans un contexte d'analyse d'images.

Nous prouvons la relation (2.5) en utilisant l'inclusion de la classe $\mathcal{G}_\Theta$ dans la classe des fragments, qui est constituée des ensembles G_θ définis par :

$$G_\theta = \{(x,y) \in [0,1]^2 :\ 0 \le y \le g(x,\theta)\},\ \forall \theta \in \Theta,$$

où $g(x,\theta)\ :\ [0,1] \to [0,1]$ est Lipschitzienne, de constante de Lipschitz $L > 0$. Cela garantit que $\mathcal{G}_\Theta \subset \mathcal{G}(\epsilon_0, L_0)$ avec $L_0 = 2\sqrt{L^2+1}$.
L'hypothèse nulle étant composée, il est nécessaire d'avoir l'existence d'au moins un ensemble *plus régulier* G_{θ_0} dans $\mathcal{G}_\Theta$: il existe $\theta_0 \in \Theta$ et des nombres réels $0 \le a_1 \le b_1 \le 1$, $c_1 > 0$, tels que pour tout $n \ge 1$, l'ensemble

$$\{(x,y) : a_1 \le x \le b_1,\ g(x,\theta_0) - \frac{c_1}{\sqrt{n}} \le y \le g(x,\theta_0) + \frac{c_1}{\sqrt{n}}\} \tag{2.17}$$

est inclu dans $[0,1]^2$ et ne contient aucun des points x_i de la grille Γ. Remarquons que toute famille possédant une frontière constante ou plus généralement toute famille ayant une frontière polynomiale vérifie la condition (2.17).
Sous ces hypothèses, et si de plus $g(x,\theta)$ est Lipschitzienne en $\theta \in \Theta$, alors il existe une constante $a > 0$ telle que la relation (2.5) est vérifée avec $v_n = n^{-1/2}$.

- Le cas d'images de dimension $d > 2$ pourrait être envisagé. Il faut cependant noter que dans ce cas, la vitesse minimax de test est $v_n = n^{-1/d}$, qui se dégrade fortement avec la dimension des points x_i, formant la grille régulière.
- Notre test statistique dépend uniquement des données. Pour valider la relation de la borne supérieure, aucune condition de régularité sur les ensembles n'est requise. Une information a priori sur la régularité des ensembles aurait probablement conduit à une vitesse de test meilleure : elle aurait été égale à la vitesse de test dans le problème de test sur les supports de densité (article [3]).
- Une hypothèse nulle non-paramétrique pourrait être considérée : le résultat de vitesse minimax de test serait inchangé à condition que la dimension de Vapnik de la classe sous H_0 soit finie.

2.3.2 Modèle de régression : cadres minimax et minimax adaptatif.

Cette partie regroupe les résultats des articles [4] et [7] écrit en collaboration avec Christophe Pouet. Nous disposons de n variables aléatoires $Y_1, \ldots, Y_n$, i.i.d. issues du modèle de régression discrète :

$$Y_i = f(x_i) + \xi_i,\ i = 1, \ldots, n,$$

où f est le paramètre d'inférence et est définie sur $[0,1]$ et à valeurs réelles, les points x_i sont déterministes, équi-espacés sur $[0,1]$ et numérotés tels que $|x_{i-1} - x_i| = \frac{1}{n}$, $\forall i = 2, \ldots, n$, les $(\xi_i)_{1 \le i \le n}$ sont i.i.d., centrées, de variance inconnue $\sigma^2 > 0$ finie.

Nous avons considéré le problème de test d'hypothèses défini en (2.4) pour lequel,

$$\begin{aligned}
\mathcal{B} &= \Sigma(\beta, C, M), \\
\Sigma(\beta, C, M) &= \left\{ \begin{array}{ll} \{f \in H(\beta, C) \ : \ \|f\|_\infty \leq M\} & \text{si } \frac{1}{4} < \beta \leq 1 \\ \{f \in H(\beta, C) \ : \ \|f\|_\infty \leq M, \ \|f'\|_\infty \leq M\} & \text{si } \beta > 1 \end{array} \right. , \\
H(\beta, C) &= \{f : \text{Hölder, de régularité } \beta \text{ et de constante de Hölder } C \ \}, \\
\mathcal{F}_0 &= \left\{ \begin{array}{ll} \mathcal{F}_\Theta \subset \Sigma(\beta, C, M), \Theta \in \mathbb{R}^d & \text{et cas } \beta \text{ connue} \\ \mathcal{F}_0 \subset \Sigma(\beta, C, M) & \text{et cas adaptatif} \end{array} \right. , \\
d(f_1, f_2) &= \|f_1 - f_2\|_{L_2}, \ \forall \, f_1, f_2 \ \in \ \mathcal{B}.
\end{aligned}$$

Dans le cadre adaptatif, β, la régularité de f, étant inconnue, elle est supposée appartenir à

$$\mathcal{T} \ = \ \{\beta \ : \ 1/4 < \beta_\star \leq \beta \leq \beta^\star\}, \tag{2.18}$$

où $\beta_\star$ et $\beta^\star$ sont fixées. Il s'ensuit que l'alternative est constituée de l' des boules fonctionnelles de Hölder sur l'ensemble $\mathcal{T}$ i.e.

$$H_1 \ : \ f \in \Lambda(v_n) = f \in \cup_{\beta \in \mathcal{T}} \{f \in \Sigma(\beta, C, M) \ : \ \inf_{f_0 \in \mathcal{F}_0} \|f - f_0\|_{L_2} > v_n\}.$$

La condition (2.18) est une hypothèse classique dans le cadre adaptatif (même condition dans Spokoiny, 1996). Dans le cadre adaptatif, l'erreur de seconde espèce s'exprime sous la forme :

$$\sup_{\beta \in \mathcal{T}} \ \sup_{\{f \in \Sigma(\beta, C, M) \, : \, \inf_{f_0 \in \mathcal{F}_0} \|f - f_0\|_{L_2} \geq v_n\}} \mathbb{P}_f[\Delta_n = 0].$$

Nous supposons de plus que la classe fonctionnelle $\mathcal{F}_0$ caractérisant H_0 est incluse dans la classe possédant la plus grande régularité $\Sigma(\beta^\star, C, M)$.

Le problème minimax de test d'hypothèses sur la fonction de régression a fait l'objet de nombreuses publications. Lorsque H_0 est simple (i.e. $H_0 \ : \ f \equiv 0$, pas de signal sous H_0), Ingster (1982), Ermakov (1990) ont étudié le cas d'une alternative définie par un ellipsoïde et séparée de la fonction nulle par la norme L_2. Ingster (1986) a considéré une boule de Hölder distante de la fonction nulle par la norme L_p et en un point fixé. Le cas des ellipsoïdes de $l(p)$, $0 < p \leq +\infty$ a été traité par Ingster (1990) et Suslina (1993). Lepskii (1993) et Lepskii et Tsybakov (2000) se sont attachés à déterminer la distance exacte (valeur de la constante devant la vitesse minimax de test) pour des alternatives Höldériennes séparées de la fonction nulle par la norme sup. Lepskii et Spokoiny (1999) ont étudié le cas d'une boule de Besov séparée de la fonction nulle par la norme L_p, l'extension adaptative se trouve dans Spokoiny (1996) pour $p = 2$ et dans Spokoiny (1998) pour $p \neq 2$. Le cas d'une hypothèse nulle composée a été étudié par Baraud, Huet et Laurent (2003) (cadre adaptatif et résultats obtenus à distance finie) pour des alternatives Höldériennes séparées de l'hypothèse nulle par une distance discrète. Horowitz et Spokoiny (2001) (cadre adaptatif) ont considéré des alternatives Höldériennes séparées d'une classe paramétrique par une distance discrète. Fromont et Lévy-Leduc (2003) (cadre adaptatif et résultats obtenus à distance finie) ont étudié le cas de fonctions de regression périodiques avec une distance discrète séparant une boule de Sobolev de l'hypothèse nulle.

L'apport de notre travail a été de considérer une hypothèse nulle composée. Notre étude permet dans le cadre adaptatif de considérer une hypothèse nulle non-paramétrique, ce qui n'est pas le cas de Horowitz *et al.* (2001) ni de Guerre et Lavergne (2002). Ces derniers ont une hypothèse nulle paramétrique nécessitant l'existence d'un estimateur consistant sous H_0. Notre choix de la norme L_2 séparant H_0 de l'alternative fait apparaître dans le risque de seconde espèce un terme d'approximation de la norme L_2 par une "norme" discrète, qu'il est nécessaire de contrôler. Ce contrôle est possible à condition d'avoir une hypothèse de régularité sur f. En revanche, si nous avions considéré une distance discrète de séparation entre H_0 et H_1 comme dans les travaux de Baraud *et al.* (2003), de Horowitz *et al.* (2001) et de Fromont *et al.* (2003), ce contrôle n'aurait pas été nécessaire. Contrairement aux articles de Baraud *et al.* (2003), de Fromont *et al.* (2003) et de Spokoiny (1996), nous n'avons pas fait d'hypothèse de normalité sur la loi des erreurs pour l'obtention de la borne supérieure (relation (2.6)). Nous montrons que les hypothèses sur la loi des erreurs sont inversement restrictives à la régularité β de la fonction f. Enfin, nous avons traité le problème en déroulant complètement la méthode minimax (Spokoiny (1996) le fait également) c'est à dire en établissant la relation de la borne supérieure (2.6) et la relation de la borne inférieure (2.5). La validité de la relation de la borne inférieure (2.5) a nécessité de renforcer les hypothèses (en particulier celles qui concernent la loi des erreurs).

Procédures minimax de test.

β connue.
Fixons β dans $\mathcal{T}$. Pour une certaine fonction f_0 dans $\mathcal{F}_0$, nous considérons

$$\tilde{T}_{n,\beta,f_0} = \frac{1}{\sqrt{\tilde{m}_\beta}} \sum_{k=1}^{\tilde{m}_\beta} \frac{\tilde{m}_\beta}{n\hat{\sigma}_n^2} \sum_{\substack{i \in I_k,\, j \in I_k \\ i \neq j}} (Y_i - f_0(x_i))(Y_j - f_0(x_j)), \tag{2.19}$$

où $\tilde{m}_\beta = n^{2/(4\beta+1)}$, $\hat{\sigma}_n^2 = \frac{1}{2(n-1)}\sum_{i=2}^n (Y_i - Y_{i-1})^2$ est un estimateur de σ^2, $I_k = \{i : x_i \in A_k\}$, avec $A_k = [\frac{k-1}{\tilde{m}_\beta}, \frac{k}{\tilde{m}_\beta}[,\ \forall k \in \{1, \dots, \tilde{m}_\beta - 1\}$ et $A_{\tilde{m}_\beta} = [\frac{\tilde{m}_\beta - 1}{\tilde{m}_\beta}, 1]$. Finalement, le test statistique est défini par

$$\tilde{\Delta}_{n,\beta} = 1\!\mathrm{I}_{\inf_{f \in \mathcal{F}_0} \tilde{T}_{n,\beta,f_0} > c_0}, \tag{2.20}$$

où $c_0 = \varphi^{-1}(1-\alpha)$ est le $(1-\alpha)$-quantile d'une loi normale standard.

Idée de la construction de la statistique $\tilde{T}_{n,\beta,f_0}$ définie par (2.19) :
La statistique $J_n = \hat{\sigma}_n^2 \frac{\sqrt{\tilde{m}_\beta}}{n} \tilde{T}_{n,\beta,f_0}$ estime la quantité $\|f - f_0\|_{L_2}^2$, qui est supérieure à v_n^2 sous H_1. Sous H_1, J_n doit alors être supérieure à v_n^2. Or, J_n se décompose en un terme de biais, supérieur à $v_n^2 - v_n \tilde{m}_\beta^{-\beta}$ sous H_1, et en deux termes stochastiques de l'ordre de $\frac{\sqrt{\tilde{m}_\beta}}{n}$, respectivement de $\frac{\|f - f_0\|_{L_2}}{\sqrt{n}}$. Equilibrer ces termes sous la condition $J_n > v_n^2$ conduit à $v_n = O(\tilde{m}_\beta^{-\beta})$ et à $\tilde{m}_\beta = n^{2/(4\beta+1)}$.

β inconnue.
Dans le cas de $\beta \in \mathcal{T}$ inconnue, et pour une fonction f_0 in $\mathcal{F}_0$, nous considérons

$$T_{n,\beta,f_0} = \frac{1}{\sqrt{m_\beta}} \sum_{k=1}^{m_\beta} \frac{m_\beta}{n\hat{\sigma}_n^2} \sum_{\substack{i \in I_k,\, j \in I_k \\ i \neq j}} (Y_i - f_0(x_i))(Y_j - f_0(x_j)),$$

où $m_\beta = (t_n^{-1} n)^{2/(4\beta+1)}$ avec $t_n = \sqrt{\log(\log n)}$, $\hat{\sigma}_n^2 = \frac{1}{2(n-1)}\sum_{i=2}^n (Y_i - Y_{i-1})^2$ est un estimateur de σ^2 et $I_k = \{i : x_i \in A_k\}$, avec $A_k = [\frac{k-1}{m_\beta}, \frac{k}{m_\beta}[,\ \forall k \in \{1, \dots, m_\beta - 1\}$ et $A_{m_\beta} = [\frac{m_\beta - 1}{m_\beta}, 1]$. La quantité m_β est supposée être entière, si tel n'est pas le cas, m_β est égale à la partie entière de $(t_n^{-1} n)^{2/(4\beta+1)}$.

On peut montrer que n'importe quelles régularités β_1 et β_2 dans $\mathcal{T}$ distantes d'au plus d'un ordre de $\frac{1}{\log n}$, produisent le même m_{β_1} (la distance séparant β_1 et β_2 dans $\mathcal{T}$ est la distance Euclidienne). Ainsi, on transforme la fourchette d'adaptation initiale $\mathcal{T}$ en une fourchette $\mathcal{M}$ définie par $\mathcal{M} = \{m_\beta \ : \ m_{\beta^\star} \le m_\beta \le m_{\beta_\star}\}$ dans laquelle deux m_{β_j} et $m_{\beta_{j+1}}$ consécutifs correspondent à deux régularités β_j et β_{j+1} distantes d'au moins $c\frac{1}{\log n}$, où c est une constante positive adaptative qui rend m_{β_j} et $m_{\beta_{j+1}}$ entiers, et qui dépend de β_* et de β^*. Il existe alors une bijection entre $\mathcal{M}$ et un sous-ensemble $\mathcal{T}^* \subset \mathcal{T}$ de $\beta \in \mathcal{T}$ contenant des régularités distantes deux à deux d'au moins $c\frac{1}{\log n}$. Les régularités $\beta_\star$ et $\beta^\star$ étant fixées, le cardinal de l'ensemble $\mathcal{M}$ est de l'ordre d'un $\log n$ quand n est suffisamment grand. Le test statistique adaptatif est alors défini par :

$$\Delta_n = 1\!\mathrm{I}_{\sup_{m_\beta \in \mathcal{M}} \inf_{f_0 \in \mathcal{F}_0} T_{n,\beta,f_0} \geq \rho_n} = 1\!\mathrm{I}_{T_n \geq \rho_n}, \tag{2.21}$$

où $\rho_n = \sqrt{8 \log(\log n)}$.

L'idée de cette procédure de test est qu'elle rejette H_0 dès qu'un des tests $\Delta_{n,\beta} = 1\!\mathrm{I}_{\inf_{f_0 \in \mathcal{F}_0} T_{n,\beta,f_0} \geq \rho_n}$ pour $\beta \in \mathcal{T}^*$ rejette H_0. Sous H_0, T_n se comporte asymptotiquement comme le supremum sur $\mathcal{T}^*$ d'une variable aléatoire Gaussienne standard, dont la distribution dégénère au-delà de $\sqrt{8 \log(card(\mathcal{T}^*))} = \sqrt{8 \log(\log n)}$ (via l'inégalité de Berry-Essen). C'est ici qu'apparaît la perte $\sqrt{\log(\log n)}$ dans la vitesse minimax adaptative de test $v_{nt_n^{-1}} = (t_n^{-1} n)^{-2\beta/(4\beta+1)}$.

RÉSULTATS.

Vitesse minimax.
Nous avons fait les hypothèses suivantes :

- [A1] L'ensemble Θ est un compact de $\mathbb{R}^d$ et $\sup_{0 \leq x \leq 1} |f_\theta(x) - f_{\theta'}(x)| \leq Q \|\theta - \theta'\|^\nu$, $\forall \theta,\ \theta' \in \Theta$, où $\nu > 0$ et $Q > 0$ sont des constantes et $\|\cdot\|$ est la norme Euclidienne dans $\mathbb{R}^d$.
- [A2] Les $(\xi_i)_{1 \leq i \leq n}$ sont i.i.d. centrées de variance σ^2 finie ; elles admettent une densité de probabilité g strictement positive sur tout $\mathbb{R}$.
- [A3.Sup] $\mathbb{P}[|\xi_i| \geq y) = O(y^{-\delta})$, $\forall i \in \{1, \ldots, n\}$, et pour un $\delta > (4\beta+1)/(2\beta)$.
- [A3.Inf] La densité g est deux fois différentiable et $\exists C_1 > 0\ : \forall y\ ;\ |y| \leq C_1$, on a

$$\begin{aligned}
&|\int \frac{(g'(x))^2}{g(x)} dx| \leq \kappa_1,\\
&|\int \frac{1}{g(x)} (\int_0^1 g^{(2)}(x+ty)(1-t)dt)^2 dx| \leq \kappa_2,\\
&|\int \frac{g'(x)}{g(x)} (\int_0^1 g^{(2)}(x+ty)(1-t)dt)^2 dx| \leq \kappa_3,\\
&|\int \frac{1}{g(x)} (\int_0^1 g^{(2)}(x+ty)(1-t)dt)(\int_0^1 g^{(2)}(x-ty)(1-t)dt) dx| \leq \kappa_4,
\end{aligned}$$

  avec κ_1, κ_2, κ_3, κ_4, des constantes positives.

Nous avons obtenu les résultats suivants :
Sous l'hypothèse [A2], le test statistique $\tilde{\Delta}_{n,\beta}$ défini par (2.20) est asymptotiquement de niveau α.
Sous les hypothèses [A1]-[A3.Sup] et pour tout $\alpha_1 \in (0,1)$, il existe une constante $A > 0$ telle que la relation (2.6) soit valide avec le test statistique $\tilde{\Delta}_{n,\beta}$, défini par (2.20). La vitesse minimax de test est $v_n = n^{-2\beta/(4\beta+1)}$.

Comme je l'ai déjà mentionné, le cas d'hypothèse nulle composée nécessite l'existence d'une fonction *plus régulière* f_{θ_0} dans $\mathcal{F}_\Theta$.
Sous les hypothèses [A1]-[A3.Inf], pour $v_n = n^{-2\beta/(4\beta+1)}$ et si il existe une fonction $f_{\theta_0} \in \mathcal{F}_\Theta$ telle que $f_{\theta_0} \in \Sigma(\beta, C', M')$, avec $C' < L$ et $M' < M$, alors pour tout $\alpha_2 \in (0,1)$, il existe une constante $a > 0$ telle que la relation (2.5) soit satisfaite.

Quelques commentaires sur ces résultats :

- Dans le cas des fonctions Höldériennes, la vitesse minimax de test est meilleure que la vitesse minimax d'estimation de la fonction de régression. La vitesse minimax de test correspond à la vitesse minimax d'estimation de la norme L_2 d'une fonction de régression (Cf Lepskii *et al.* (1999) dans le modèle du bruit blanc Gaussien). Par ailleurs, la vitesse minimax de test est identique à celle qui apparaît dans le cas d'une hypothèse nulle simple.
  L'hypothèse [A1] pourrait être remplacée par le contrôle d'une entropie sur la classe $\mathcal{F}_\Theta$; ainsi il serait permis de considérer une classe de fonctions non-paramétrique sous H_0. Horowitz et Spokoiny (2001) ont obtenu un résultat adaptatif sous une hypothèse nulle paramétrique en imposant l'existence d'un estimateur consistant de θ (c'est une hypothèse sur θ beaucoup plus forte que [A1]).
- La difficulté de la preuve de la borne supérieure se situe dans l'approximation de $\|f - f_\theta\|^2_{L_2}$ par une somme discrète de termes $(f - f_\theta)^2$ calculés aux points x_i, $i = 1, \ldots, n$. Le contrôle de cette approximation est effectué via une inégalité due à Ingster (1993), qui nécessite d'imposer à f une régularité.
- L'hypothèse [A3.Sup] implique l'existence d'un moment d'ordre k des variables aléatoires ξ_i. L'ordre du moment existant est lié à la régularité β de la fonction de régression. Soulignons qu'un β tendant vers l'infini requière l'existence d'un moment d'ordre $k = 2 + \epsilon$ ($\epsilon > 0$ aussi petit que l'on veut). A l'inverse, un β proche de $1/4$ conduit à l'existence d'un moment d'ordre $k = 4 + \epsilon$ ($\epsilon > 0$ aussi petit que l'on veut).
- Nous avons obtenu la relation de la borne inférieure dans un cadre plus général que celui du modèle de régression Gaussienne. En particulier, l'hypothèse [A3.Inf] est valide pour un mélange de distributions Gaussiennes.

Vitesse minimax adaptative.
Nous avons fait les hypothèses suivantes :

- [B1.Sup] Les variables aléatoires $\xi_1, \ldots, \xi_n$ sont i.i.d., centrées, de variance finie $\sigma^2 > 0$.
- [B1.Inf] Les $\xi_1, \ldots, \xi_n$ sont des variables aléatoires i.i.d., Gaussiennes, centrées et de variance finie $\sigma^2 > 0$.
- [B2.Sup] $\lim\limits_{y \to +\infty} \mathbb{P}\left(|\xi_i| \geq y\right) y^{\frac{1}{\kappa}(2+\frac{1}{2\beta_\star})\wedge 4} = 0$, $\forall i \in \{1, \ldots, n\}$, avec $0 < \kappa < 1$.
- [B3] La δ–entropie de la classe $\mathcal{F}_0$ calculée avec la norme sup, est bornée par δ^{-r},

où r est inférieur à $\dfrac{2\beta_\star(1-\kappa)(1+4\beta^\star)}{(16+8\kappa)\beta_\star\beta^\star+2(2\beta^\star+\kappa\beta_\star)}$.
Nous avons obtenu les résultats suivants :
Sous l'hypothèse [B1.Sup], le test statistique Δ_n défini par (2.21) vérifie quand n tend vers l'infini,

$$\sup_{f_0\in\mathcal{F}_0} \mathbb{P}_{f_0}(\Delta_n = 1) = o_n(1). \tag{2.22}$$

Sous les hypothèses [B1.Sup], [B2.Sup] et [B3], il existe une constante $A > 0$ suffisamment grande telle que quand n tend vers l'infini,

$$\sup_{\beta\in\mathcal{T}} \sup_{f\in\Lambda(Av_{nt_n^{-1}})} \mathbb{P}_f(\Delta_n = 0) = o_n(1), \tag{2.23}$$

où Δ_n est défini par (2.21) et où $t_n = \sqrt{\log(\log n)}$ est la perte due à l'adaptation. La vitesse minimax adaptative de test est $v_{nt_n^{-1}} = (nt_n^{-1})^{-2\beta/(4\beta+1)}$.

La relation de la borne inférieure est établie dans la cas Gaussien.
Sous les hypothèses [B1.Inf] et [B3], pour $v_{nt_n^{-1}} = (nt_n^{-1})^{-2\beta/(4\beta+1)}$ et si il existe une fonction $f_0 \in \mathcal{F}_0$ telle que $f_0 \in \Sigma(\beta, C', M')$ pour tout $\beta \in \mathcal{T}$, avec $C' < C$ et $M' < M$, alors pour tout $\alpha_2 \in (0,1)$, il existe une constante $a > 0$ telle que

$$\inf_{\tilde{\Delta}_n} \sup_{\beta\in\mathcal{T}} \sup_{f\in\Lambda(av_{nt_n^{-1}})} \mathbb{P}_f(\tilde{\Delta}_n = 0) \geq \alpha_2 + o_n(1), \tag{2.24}$$

où $\inf_{\tilde{\Delta}_n}$ est l'infimum pris sur l'ensemble des tests statistiques asymptotiquement de niveau α, avec $\alpha \in (0,1)$ fixé à l'avance.

Quelques commentaires sur ces résultats :
- Les résultats (2.22) et (2.23) prouvent que notre test statistique atteint asymptotiquement n'importe quelles erreurs de première et de seconde espèce.
- Le test statistique (2.21) étant défini par l'indicatrice d'un supremum, son utilisation pratique semble être difficile. Toutefois, sur des classes $\mathcal{F}_0$ simples, comme par exemple des classes paramétriques, son implémetation peut s'effectuer relativement facilement.
- Notre travail met en évidence le lien existant entre le paramètre de régularité β, le contrôle de l'entropie de la classe $\mathcal{F}_0$ et la queue de distribution des erreurs. En effet, plus la classe $\mathcal{F}_0$ est riche, plus les queues de distribution des erreurs doivent être légères. Le choix de $\beta_\star$ est plus important que celui de $\beta^\star$: lorsque $\beta_\star$ est grand, la classe $\mathcal{F}_0$ peut être indifféremment une classe riche ou pauvre. En revanche, au plus

$\beta_\star$ est petit, avec $\beta^\star$ fonction de $\beta_\star$, au plus la classe $\mathcal{F}_0$ devra être pauvre. Il est impossible de fournir une paire $(\beta_\star, \beta^\star)$ pour une classe $\mathcal{F}_0$ riche et une distribution des erreurs à queues lourdes : dans cette situation, l'hypothèse [B3] n'est jamais satisfaite. En revanche, [B3] est toujours satisfaite pour une classe $\mathcal{F}_0$ pauvre et pour des queues légères de la distribution des erreurs.
- La relation de la borne inférieure (2.24) est établie dans le cas Gaussien, comme dans Spokoiny (1996). La preuve de (2.24) repose sur la construction d'une famille paramétrique aléatoire autour de f_0, l'élément *plus régulier* dans $\mathcal{F}_0$. Spokoiny (1996) propose une famille paramétrique aléatoire construite, via les ondelettes, à partir de niveaux de résolution orthogonaux. Cela garantit l'indépendance entre les paquets de variables qui sont associés à différentes valeurs du paramètre de régularité. Dans notre cas, nous n'avons pas l'indépendance entre les variables associées à des régularités distinctes. Pour contourner cette difficulté, nous avons partitionné l'ensemble des régularités afin de former des paquets de variables, pour lequelles la dépendance est contrôlée.

2.4 Perspectives.

- Approche maxiset. L'approche maxiset constitue une alternative particulièrement intéressante à l'approche minimax. Elle consiste à déterminer l'ensemble maximal sur lequel une procédure d'estimation donnée atteint une vitesse de convergence donnée. Cette approche est à ce jour étudiée dans les seuls problèmes d'estimation non-paramétriques. Elle a été initiée par Kerkyacharian et Picard (2000, 2002). Signalons que l'approche maxiset pour l'estimation n'est pas directement transposable aux problèmes de test d'hypothèses non-paramétriques. Ce projet est actuellement en cours de réalisation en collaboration avec Christophe Pouet.
- Contexte dépendant. Une extension possible de nos articles [4] et [7], est de considérer le cas de variables dépendantes. Deux formes distinctes de dépendance sont envisagées. Elles sont caractérisées par :
  -un processus mélangeant,
  -un coefficient de mélange d'un champ aléatoire i.e. une dépendance spatiale.
  Pour chacun des deux cas, notre objectif est de déterminer la vitesse minimax de test et de fournir un test statistique atteignant cette vitesse. Ce projet est actuellement en cours de réalisation en collaboration avec Sophie Niang et Christophe Pouet.
- Copules. Je mène actuellement un travail en collaboration avec Karine Tribouley sur un problème de test minimax sur la fonction copule ou plus précisément sur la densité associée à la fonction de copule. Nous nous plaçons dans un cadre bidimensionnel. Nous nous intéressons à déterminer la vitesse mimimax de test de H_0

contre H_1. L'hypothèse nulle est caractérisée par la donnée d'une famille paramétrique $\mathcal{F}_\Theta$. L'hypothèse alternative est définie par une boule de Besov de densités distantes de $\mathcal{F}_\Theta$ par la norme L_2.

Chapitre 3

Statistique Bayésienne non-paramétrique.

Je me suis intéressée aux propriétés asymptotiques (consistance et vitesse de convergence) d'estimation Bayésienne d'ensembles à niveau. Ce travail a été mené en collaboration avec Judith Rousseau. Il a fait l'objet de deux articles : étude de la consistance [9] et vitesse de convergence [8].

La première partie de ce chapitre constitue une introduction générale sur la statistique Bayésienne non-paramétrique. Puis, suivent une présentation de nos résultats sur l'estimation d'ensembles à niveau, et une conclusion comprenant des commentaires sur les résultats obtenus et les perspectives de recherche.

3.1 Une introduction.

L'approche Bayésienne non-paramétrique s'est développée depuis la fin des années 90 grâce à l'émergence de méthodes de simulation performantes et notamment des méthodes de simulation par chaînes de Markov (MCMC). Celles-ci ont l'avantage de pouvoir implémenter relativement facilement les estimateurs Bayésiens non-paramétriques. Sur le plan théorique, il existe depuis les années 50-60 des résultats théoriques sur le comportement asymptotique des procédures Bayésiennes quand la dimension du paramètre est très grande, voire non-finie. Ces résultats ont permis d'établir la convergence de la loi a posteriori à condition que la loi a priori vérifie certaines conditions. Les travaux pionniers sur ce sujet sont dus, entre autres à Doob (1949), Le Cam (1958), Freedman (1963) et Schwartz (1965). Doob (1949) a montré que sous des conditions relativement peu restrictives, la consistance est obtenue pour Π-presque toutes valeurs du paramètre (Π désigne la loi a priori). Son étude ne permet pas de savoir pour quelles valeurs du paramètre la consistance est effective. Le Cam (1958), sous des conditions générales mais restrictives, a établi la consistance de l'estimateur Bayésien pour toute valeur du paramètre. Freedman (1963) a étudié la consistance et la normalité asymptotique des estimateurs Bayésiens dans le cas particulier d'observations de loi discrète. Schwartz (1965) a prouvé la consistance de la loi a posteriori, à condition que la loi a priori charge positivement, au sens strict, tous les voisinages de Kullback de la vraie loi et qu'il existe un test uniformément convergeant du vrai paramètre contre le complémentaire de tout voisinage du vrai paramètre. Une unification et une extension de ces résultats ont été réalisées par Diaconis et Freedman (1986). Ils ont notamment étudié la validité de la consistance dans différents modèles et pour différentes lois a priori. Il en résulte que dans le cadre non-paramétrique, l'étude de la convergence de la loi a posteriori est une étape indispensable permettant de s'assurer que le choix de la loi a priori est raisonnable. En particulier, un choix de loi a priori subjective n'est pas envisageable dans un contexte non-paramétrique.

Depuis la fin des années 90, l'étude de la consistance de loi a posteriori a connu un regain d'intérêt. C'est notamment grâce à l'article de Barron, Schervish et Wasserman (1999), dans lequel les conditions, que doit vérifier la loi a priori chargeant une densité afin de garantir la consistance de la loi a posteriori, sont données sous une forme explicite. L'extension naturelle de l'étude de cette consistance est la détermination de la vitesse de convergence de la loi a posteriori. Cela permet notamment de construire des lois a priori plus *objectives*. Se fixant une loi a priori, il est également possible de comparer les vitesses de convergence des procédures Bayésiennes aux vitesses de convergence minimax, pour un modèle donné. Lorsque le paramètre d'intérêt est une densité de probabilité, Ghosal, Ghosh et van der Vaart (2000), Shen et Wasserman (2001) et Ghosal et van der Vaart (2007a) ont obtenu les vitesses de convergence de la loi a posteriori sur le complémentaire de voisinages de la vraie densité exprimés à partir de la distance de Hellinger ou de distances équivalentes à la distance de Hellinger.

Il faut souligner que les résultats des articles de Barron *et al.* (1999), de Ghosal *et al.* (2000), de Shen *et al.* (2001) et de Ghosal et van der Vaart (2007a) sont généraux : ils établissent les conditions que doit vérifier la loi a priori, sous lesquelles la consistance de la loi a posteriori est garantie, respectivement sous lesquelles une vitesse de convergence de la loi a posteriori est fournie. Plus précisément, et en notant $\mathcal{P}$ l'ensemble des distributions, ces conditions sont :

- la loi a priori charge suffisamment les voisinages de Kullback de la vraie distribution,
- il existe une suite $(\mathcal{P}_n)_{n\in\mathbb{N}}$ de sous-ensembles de $\mathcal{P}$ telle que :
  - la suite $(\mathcal{P}_n)_n$ approche $\mathcal{P}$ et cette approximation est mesurée via la loi a priori,
  - l'entropie métrique de $\mathcal{P}_n$ est contrôlée.

Ces résultats généraux ont engendré un nombre important de résultats de consistance de la loi a posteriori ou de vitesses de convergence pour un choix particulier de loi a priori comme les mélanges de Dirichlet (Ghosal, Ghosh et Ramamoothy, (1999b) et Ghosal et van der Vaart (2007b)), les processus Gaussiens ou encore les arbres de Polya (Ghosal, Ghosh et Ramamoothy, 1999a). Un résumé de tous ces résultats est disponible dans Ghosh et Ramamoorthy (2003).

3.2 Ensembles à niveau définis à partir d'une densité de probabilité.

En collaboration avec Judith Rousseau, nous nous sommes intéréssées à l'estimation Bayésienne non-paramétrique d'ensembles à niveau à partir de $X^{(n)} = (X_1, \ldots, X_n)$, un échantillon i.i.d. d'observations d dimensionnelles ($d \geq 2$). L'ensemble à niveau est construit à partir de f (la densité commune des observations). Il est défini par l'ensemble des points pour lesquels la densité est supérieure à λ (un niveau fixé), c'est à dire par

$$G_{f,\lambda} = \{y \in E; f(y) \geq \lambda\}, \tag{3.1}$$

où E est le support des observations, et f est inconnue.

Le problème d'estimation d'ensembles à niveau a été largement étudié, ces ensembles apparaissant naturellement dans de nombreuses applications, comme en contrôle de qualité (en considérant que le processus est hors contrôle si une nouvelle observation n'appartient pas à $G_{f,\lambda}$) ou encore pour déterminer le nombre de modes et leur localisation (en effectuant des tests de multimodalités pour différentes valeurs de λ).

En fréquentiste, un certain nombre d'estimateurs d'ensembles à niveau sont définis comme des arguments maximisant un critère sur une classe d'ensembles (le plus répandu des critères est celui que fournit l'approche "excès de masse"). Leur implémentation est d'autant plus problématique que la classe des ensembles est complexe. Dans le cas de classes simples telles que la classe des ensembles convexes (Hartigan, 1987) ou encore la classe des ellipsoïdes (Nolan, 1991), des algorithmes pour implémenter les estimateurs existent. Il existe également des estimateurs fréquentistes pour lesquels des vitesses de convergence ont été établies : Polonik (1995) (estimateur défini par l'approche excès de masse), Tsybakov (1997) (estimateur défini par une approche excès de masse locale sur une classe d'ensembles étoilés et réguliers), Walther (1997) (estimateur construit sur des arguments géométriques qui s'avère être moins performant que l'estimateur excès de masse dès lors que la densité n'est pas suffisamment régulière) et plus récemment Rigollet et Vert (2007) (estimateur plug-in).

Dérouler l'approche Bayésienne pour estimer des ensembles à niveau a été motivé par les points suivants :

- à notre connaissance il n'existe pas d'autre résultat théorique en estimation Bayésienne non-paramétrique d'ensembles à niveau,

- l'approche Bayésienne ne nécessite aucune spécification a priori de la classe des ensembles à niveau à laquelle le vrai ensemble appartient.

Et les points positifs résultant de notre étude sont :

- l'obtention de résultats généraux sur les vitesses de convergence du type de ceux de Ghosal *et al.* (2000) et Shen *et al.* (2001). Mais la spécificité de notre objet d'intérêt rend les résultats de ces derniers inadéquats dans ce contexte (sous peine de sous-optimalité),
- la validation de la consistance de loi a posteriori pour une famille de lois a priori générale,
- un estimateur explicite fourni par l'approche Bayésienne,
- la facile implémentation de l'estimateur Bayésien dès lors que l'on sait simuler dans la loi a posteriori, ou à défaut, en utilisant des méthodes de simulation reposant sur des algorithmes MCMC.
- D'un point de vue théorique, il est toujours intéressant de comparer les méthodes d'estimation, et dans ce cadre, de comparer les approches Bayésienne et fréquentiste.

3.2.1 Estimateur Bayésien d'ensembles à niveau.

L'approche Bayésienne requière :

- une loi a priori Π chargeant $\mathcal{F} = \{f\ :\ \text{densités définies sur } E\}$,
- une fonction de perte $L(G_1, G_2) = \mu(G_1 \Delta G_2)$, qui correspond à la mesure de Lebesgue (μ : mesure de Lebesgue sur $\mathbb{R}^d$) de la différence symétrique entre les ensembles G_1 et G_2. Cette fonction de perte est usuelle dans les problèmes d'inférence sur des ensembles.

L'estimateur Bayésien $\hat{G}_\lambda^\Pi$ de $G_{f,\lambda}$ défini par (3.1), est l'argument minimisant le risque de Bayes associé à Π et pour la perte L, c'est à dire

$$\begin{aligned}\hat{G}_\lambda^\Pi &= \arg\min_G \mathbb{E}^\Pi[L(G, G_{f,\lambda})|X^{(n)}] \\ &= \{y \in E; \Pi[y \in G_{f,\lambda}|X^{(n)}] \geq \Pi[y \in G_{f,\lambda}^c|X^{(n)}]\},\end{aligned} \tag{3.2}$$

où $\mathbb{E}^\Pi[\cdot|X^{(n)}]$ désigne l'espérance a posteriori quand la loi a priori est Π. Soulignons à nouveau, que dès lors que l'on est capable de simuler dans la loi a posteriori qui charge ici $G_{f,\lambda}$ (y étant non-aléatoire), l'estimateur défini par la relation (3.2), est facilement implémentable : il présente l'avantage d'avoir une forme explicite.

Nous utiliserons les notations $\mathbb{P}_0$, $\mathbb{E}_0$, f_0 et $G_{0,\lambda}$ pour désigner la vraie distribution des observations, l'espérance et la densité relativement à $\mathbb{P}_0$, et le vrai ensemble à niveau.

3.2.2 Consistance de la loi a posteriori.

La consistance de la loi a posteriori a été établie dans le cas bidimensionnel i.e. $E \subset \mathbb{R}^d$, avec $d = 2$. La partie théorique de l'article [9] pourrait être généralisée au cas $d > 2$. En revanche, la partie implémentation serait probablement non transposable au cas $d > 2$, de part la complexité de l'espace nécessaire à explorer.

Etant en dimension 2, il nous a semblé approprié de construire une famille de lois a priori engendrées par un processus ponctuel spatial ξ. Plus précisément, la loi a priori Π charge les densités du type

$$f(y) = \sum_{k=1}^{K} \frac{\exp(\theta_k)}{\sum_{k=1}^{K} \exp(\theta_k)\mu(V_k(\xi))} \mathbb{1}_{y \in V_k(\xi)}, \tag{3.3}$$

au travers des lois

- de $(\xi_k, k)_{1 \leq k \leq K}$
- de $\theta = (\theta_k)_{1 \leq k \leq K}$ sachant ξ.

Les cellules $(V_k)_{1 \leq k \leq K}$ forment une partition de E et sont les éléments d'une tesselation de Voronoi engendrés par les $(\xi_k)_{1 \leq k \leq K}$. Pour cette famille de lois a priori, nous avons étudié la consistance de la loi a posteriori en termes de distance de Hellinger. Avant de préciser ce résultat, définissons pour tout $\epsilon > 0$,

$$A_\epsilon = \{f \in \mathcal{F} \ : \ h^2(f, f_0) < \epsilon\},$$

où h^2 désigne le carré de la distance de Hellinger entre f et f_0. Remarquons que A_ϵ est un ϵ-voisinage de f_0 exprimé à partir de la distance de Hellinger.

Nous avons établi *la consistance de loi a posteriori* i.e. pour tout $\epsilon > 0$

$$\lim_{n \to \infty} \mathbb{P}^{\Pi}\left(A_\epsilon | X^{(n)}\right) = 1 \ \text{ p.s. } [\mathbb{P}_0], \tag{3.4}$$

à condition que Π vérifie les conditions suivantes :

- [C1] la probabilité que K charge les entiers supérieurs à 3 et cette probabilité devient exponentiellement petite dès lors que K est supérieur à $O(n/\log(n))$,
- [C2] sachant K, la probabilité de ξ est strictement positive sur E^K,
- [C3] sachant ξ, la probabilité jointe du vecteur θ est strictement positive sur $\mathbb{R}^{K-1}$ (pour des raisons d'identifiabilité la dimension du paramètre θ est réduite à $K-1$),
- [C4] sachant ξ, et pour deux éléments *voisins* V_j et V_l, la probabilité que θ_j et θ_l soient éloignés d'une distance supérieure à une puissance de n renormalisée par la distance entre ξ_j et ξ_l, est exponentiellement petite.

Les éléments V_j et V_l sont *voisins* si ils ont une frontière commune.

Le résultat de consistance de la loi a posteriori exprimé en termes de voisinages de Hellinger (3.4) implique la convergence p.s. de l'estimateur Bayésien $\hat{G}_{\lambda}^{\Pi}$ vers $G_{0,\lambda}$ dès lors que la mesure de Lebesgue de la frontière de $G_{0,\lambda}$ est nulle. Autrement dit, sous les conditions [C1],[C2], [C3], [C4] et [C0] ([C0] : $\mu(\partial G_{0,\lambda}) = 0$, où $\partial G_{0,\lambda}$ désigne la frontière de $G_{0,\lambda}$), alors

$$\lim_{n\to\infty} L(\hat{G}_{\lambda}^{\Pi}, G_{0,\lambda}) = 0 \text{ p.s. } [\mathbb{P}_0].$$

Soulignons que l'hypothèse [C0] n'autorise pas f_0 à être égale à λ sur un ensemble de mesure non nulle et que c'est une hypothèse classique pour l'estimation d'ensembles à niveau.

Le prolongement naturel de ces résultats, établis sous des conditions relativement générales, est leur application pour un choix particulier de loi a priori chargeant les densités du type (3.3). Nous avons considéré :
- K de loi de Poisson tronquée,
- sachant K, les $(\xi_k)_{1\leq k\leq K}$ sont des réalisations i.i.d. de Processus de Poisson,
- sachant ξ, les $(\theta_k)_{1\leq k\leq K}$ sont des réalisations d'un champ aléatoire de Markov telles que la loi jointe de $(\theta_k)_{1\leq k\leq K-1}$ est une loi propre.

Pour cette loi a priori particulière, nous avons montré la consistance de l'estimateur en vérifiant les hypothèses [C1]-[C4] et nous avons implémenté l'estimateur Bayésien en adaptant un algorithme du type sauts réversibles MCMC proposé par Heikkinen (1998).

Remarque 1 Nous nous sommes également intéressées à l'estimation Bayésienne d'ensembles à niveau quand ces ensembles sont définis à partir d'une fonction de régression i.e. lorsque f dans l'expression (3.1), désigne la fonction de régression. Nous avons considéré le modèle de régression à entrées aléatoires indépendantes $X_i,\ i = 1,\ldots,n$ (leur densité p est connue) et avec des erreurs Gaussiennes. Dans ce cadre, la loi a priori charge la classe des fonctions de régression qui vérifient $\mathbb{E}_p[f^2(X)] < \infty$. Comme pour les ensembles à niveau définis à partir de la densité, nous avons établi la consistance de la loi a posteriori sous des conditions similaires aux hypothèses [C1]-[C4], puis nous en avons déduit la convergence p.s. de l'estimateur Bayésien à la condition que la frontière du vrai ensemble à niveau soit de μ-mesure nulle.

3.2.3 Vitesses de convergence.

Il me faut dans un premier temps signaler que la méthodologie utilisée dans l'article [8] (basée sur l'étude de rapport de vraisemblances) permet uniquement d'obtenir une majoration de la vitesse de convergence de l'estimateur Bayésien d'ensembles à niveau. Cela signifie qu'il est possible que la vraie vitesse de convergence de notre estimateur soit meilleure que la majoration que nous obtenons. Dans la suite de cette partie, j'utiliserai abusivement le terme de "vitesse de convergence" pour désigner "la majoration de la vitesse de convergence" de l'estimateur Bayésien.

Le résultat fondamental de l'article [8] est l'obtention de la vitesse de convergence de l'estimateur Bayésien sous des hypothèses générales. Nous avons ensuite appliqué ce résultat en prenant une loi a priori particulière, ce qui nous a permis de comparer notre vitesse à celle de l'estimateur fréquentiste proposé par Polonik (1995).

Nous considérons le cas de $E \subset \mathbb{R}^d$, avec $d \geq 2$. Trouver une majoration de la vitesse de convergence de $\hat{G}_\lambda^\Pi$ consiste à déterminer τ_n tel que

$$\mathbb{E}_0^n[L(\hat{G}_\lambda^\Pi, G_{0,\lambda})] \leq \kappa\tau_n, \tag{3.5}$$

où κ est une constante positive, $\tau_n \searrow 0, n \longrightarrow \infty$ et $\mathbb{E}_0^n$ désigne l'espérance sous la vraie loi quand on dispose de n variables. Cette majoration s'effectue au travers de l'étude de $\Pi(B_n^c|X^{(n)})$ (qui doit être rendue exponentiellement petite) :

$$\begin{aligned} \Pi(B_n^c|X^{(n)}) &= \frac{\int_{B_n^c} \prod_{i=1}^n f(X_i) d\Pi(f)}{\int_{\mathcal{F}} \prod_{i=1}^n f(X_i) d\Pi(f)} \\ &= \frac{\int_{B_n^c} \prod_{i=1}^n \frac{f}{f_0}(X_i) d\Pi(f)}{\int_{\mathcal{F}} \prod_{i=1}^n \frac{f}{f_0}(X_i) d\Pi(f)}, \end{aligned} \tag{3.6}$$

où $B_n^c = \{f \in \mathcal{F} : L(G_{0,\lambda}, G_{f,\lambda}) > M\tau_n\}$, $M = cste > 0$ et $\Pi(\cdot|X^{(n)})$ désigne la probabilité a posteriori. Soulignons que

- les B_n correspondent à des L-voisinages qui se concentrent fortement en $G_{0,\lambda}$ quand n grandit,
- τ_n décrit la manière dont s'effectue cette concentration en $G_{0,\lambda}$.

Notre approche pour la détermination de τ_n est similaire à celle utilisée dans Ghosal *et al.* (2000) et Shen *et al.* (2001). Ils ont obtenu un contrôle de $\Pi(B_n^c|X^{(n)})$ dans le cas où B_n est un voisinage de Hellinger autour de la vraie densité. Dans leur problème, les distances intervenant dans l'équation (3.6), sont des distances sur les densités qui sont comparables. Pour les ensembles, il n'en est plus de même puisque les distances intervenant dans l'équation (3.6) sont une distance entre ensembles et, une distance entre densités. En conséquence, il serait contre-indiqué d'utiliser leur résultat en considérant le fait que deux ensembles à niveau distants de $\tau > 0$ implique une distance de $\eta(\tau) > 0$ entre les densités associées, à défaut la vitesse engendrée par une telle procédure serait *sous-optimale*.

Nous considérons les hypothèses suivantes :

- [A0] "Un contrôle du volume autour de la frontière de $G_{0,\lambda}$" i.e. $\mu\{x \in E : |f_0(x) - \lambda| \leq \eta\} \leq c_3\eta^\alpha$, $\forall \eta > 0$ tel que $\eta < \eta_0$, où η_0 est fixé et $c_3 > 0$.

$\alpha \to 0^+$ correspond à des parties "plates du graphe de f_0" autour de $\partial G_{0,\lambda}$,
$\alpha \to +\infty$ correspond à une discontinuité de f_0 en $\partial G_{0,\lambda}$.

- [A1] "Π charge suffisamment les ρ_n-voisinages de Kullback de f_0" i.e. $\Pi(S_n(\rho_n)) \geq exp(-c_3 n\rho_n),\ c_3 > 0,\ \rho_n > 0,\ n\rho_n \to +\infty,\ \rho_n \to 0$, où $S_n(t) = \{f \in \mathcal{F}; \mathcal{K}(f_0, f) \leq t/2; M_0(f_0, f) \leq t; \|f_0/f\|_\infty \leq T\}$, où $\mathcal{K}(\cdot,\cdot)$ est la divergence de Kullback, $M_0(f_0, f) = \mathbb{E}_{f_0}[(\log f_0(X) - \log f(X))^2]$ et $T > 0$.
- [A2] "$\mathcal{F}$ est approchée par $\mathcal{F}_n$" i.e. $\mathcal{F}_n \subseteq \mathcal{F} : \Pi(\mathcal{F}_n^c) \leq \exp(-(c_3+1)n\rho_n)$
- [A3] "La complexité de $\mathcal{F}_n$" est mesurée par l'entropie $\mathcal{D}_n \leq c_5 \tau_n^{-r},\ r > 0,\ c_5 > 0$, où $\mathcal{D}_n$ est le logarithme du cardinal du plus petit ensemble contenant les paires (G_L, G_U) dans E telles que $\forall f \in \mathcal{F}_n,\ \exists (G_L, G_U)$ telle que $G_L \subset G_{f,\lambda} \subset G_U$ et $\int_{G_U \Delta G_L} |f(x) - \lambda| d\mu(x) \leq \tau_n^{(\alpha+1)/\alpha}$.
  En réalité, [A3] concerne chaque sous-ensemble $\mathcal{F}_{n,j},\ j \in \mathbb{N}$,

  $$\mathcal{F}_{n,j} = \{f \in \mathcal{F}_n \ :\ j\tau_n^{(\alpha+1)/\alpha} \leq \int_{G_{0,\lambda}\Delta G_{f,\lambda}} |f_0(x) - \lambda| d\mu(x) \leq (j+1)\tau_n^{(\alpha+1)/\alpha}\},$$

  de la suite $(\mathcal{F}_{n,j})_{j\in\mathbb{N}}$ d'ensembles emboités.

Sous les hypothèses [A0]-[A3], la relation (3.5) est obtenue avec τ_n défini par :

$$\tau_n = \max(n^{-\frac{\alpha}{\alpha+2+r\alpha}}, \rho_n^{\frac{\alpha}{\alpha+2}}).$$

Notons que l'hypothèse [A0] est plus précise que l'hypothèse [C0] puisqu'elle fournit un contrôle explicite du comportement le f_0 autour de la frontière de $G_{0,\lambda}$. Signalons que les hypothèses [A1]-[A3] sont des hypothèses qui apparaissent usuellement lorsqu'on étudie la distribution a posteriori dans un contexte non-paramétrique (voir Ghosal *et al* (2000) et Shen *et al.* (2001)). Et remarquons que l'hypothèse [A1] garantit à la loi a priori de placer suffisamment de masse dans les voisinages de Kullback de f_0.

Le prolongement naturel de ces résultats, établis sous des conditions générales, est leur application pour un choix particulier de loi a priori vérifiant [A1]-[A3]. Il faut souligner que la vérification des hypothèses [A1]-[A3] pour une loi a priori donnée est loin d'être évidente. C'est la raison pour laquelle, nous n'avons pas appliqué notre résultat à la famille de lois a priori du type log-spline avec noeuds libres chargeant les densités du type (3.3).

Afin de définir la loi a priori, nous nous restreignons au cas $E = [0,1]^2$ et nous supposons que f_0 vérifie [A0] et [H0] ([H0] : f_0 appartient à la classe de Hölder de régularité $\gamma = m + \beta,\ m \in \mathbb{N},\ \gamma - 1 \leq m < \gamma,\ \beta \in]0,1]$). La classe des densités log-splines approximant naturellement les densités régulières au sens de Hölder, nous considérons la loi a priori Π chargeant les densités log-splines définies sur $[0,1]^2$ i.e. chargeant les densités f_θ définies pour tout $x = (x_1, x_2) \in [0,1]^2$ par :

$$f_\theta(x) = \exp(\theta_{(1)}^t B(x_1) + \theta_{(2)}^t B(x_2) + \theta_{(3)}^t \overline{B}(x_1, x_2) - c(\theta)),$$

où $B = (B_1, \ldots, B_J)$ est la base naturelle de splines univariés, $J = K + m$ avec K le nombre d'éléments d'une partition régulière fixe de $[0, 1]$, $\forall i = 1, 2,\ \theta_{(i)} \in \mathbb{R}^J$ et $\theta_{(3)} \in \mathbb{R}^{J^2}$, $\overline{B}(x_1, x_2) = (B_1(x_1)B_1(x_2), B_1(x_1)B_2(x_2), \ldots, B_J(x_1)B_J(x_2))^t$, $\forall i = 1, 2, 3,\ \theta^t_{(i)}$ est le transposé du vecteur $\theta_{(i)}$, $c(\theta)$ est la constante de renormalisation qui assure à f_θ d'être une densité de probabilité. Nous renvoyons à Stone (1994) pour une description détaillée des splines dans $\mathbb{R}^2$.
La loi a priori Π charge les densités du type f_θ au travers d'une loi $\mathbb{P}_1$ sur K et d'une loi π^K sur $\theta = (\theta_{(1)}, \theta_{(2)}, \theta_{(3)})$ sachant K (θ est de dimension $2J + J^2$). L'ensemble $\mathcal{F}_n$ intervenant dans l'hypothèse [A2] est alors $\mathcal{F}_n = \{f_\theta\ :\ \theta \in \mathbb{R}^{2J+J^2},\ J = K + m,\ K \leq \rho_n^{1/(2\gamma)}\}$, où ρ_n apparaît dans l'hypothèse [A2].
Si f_0 vérifie [H0] et [A0] et si $\Pi(f) = \mathbb{P}_1(K)\pi^K(\theta)$, avec quelques conditions sur les lois $\mathbb{P}_1$ et π^K, alors

$$\tau_n \leq r_n = \begin{cases} \left(\frac{n}{\log(n)}\right)^{-\frac{\gamma\alpha}{\alpha+2\gamma+2}} & \text{quand } \gamma \leq 1, \\ \left(\frac{n}{\log(n)}\right)^{-\frac{\gamma\alpha}{(\alpha+2)(\gamma+1)}} & \text{quand } \gamma > 1. \end{cases} \tag{3.7}$$

Afin de mesurer la performance de l'estimateur Bayésien, nous avons comparé la vitesse de convergence r_n définie par (3.7) à la vitesse de convergence obtenue par l'estimateur fréquentiste proposé par Polonik (1995). Cette comparaison a nécessité au préalable de déterminer la vitesse v_n^P de Polonik pour la classe des densités vérifiant [H0] et [A0].
Il s'avère qu'aucun des deux estimateurs n'est plus performant que l'autre uniformément en (α, γ) autrement dit, certains choix de (γ, α) conduisent à "la vitesse r_n est meilleure que la vitesse v_n^P" et d'autres conduisent à "la vitesse v_n^P est meilleure que la vitesse r_n".

3.2.4 Commentaires et perspectives.

- Les articles [9] et [8] fournissent, sous des conditions générales, un résultat de consistance et des vitesses de convergence de l'estimateur Bayésien d'ensembles à niveau. Ces résultats sont établis à partir de l'étude de la loi a posteriori et en ce sens, ils sont semblables aux résultats publiés par Barron *et al.* (1999), Ghosal *et al.* (2000) et Shen *et al.* (2001). Pour autant, l'étude Bayésienne des ensembles à niveau ne peut s'effectuer au travers de l'étude Bayésienne d'une distribution de probabilité, puisqu'à défaut, les résultats y seraient sous-optimaux.
- Pour la classe d'ensembles $\mathcal{G}(\gamma, \alpha)$, associée à la classe des densités vérifiant [H0] et [C0], Rigollet et Vert (2007) ont établi la vitesse minimax d'estimation d'ensembles à niveau. Elle est de l'ordre de $v_{n,M} = n^{-\frac{\alpha\gamma}{2\gamma+2}}$. Cette vitesse correspond à la vitesse de convergence $v_n^P = n^{-\alpha/(\alpha+2+\rho\alpha)}$ établie par Polonik (1995) lorsque $\rho < 1$ et pour ρ désignant l'exposant dans le contrôle supérieur de l'entropie à crochet de la classe $\mathcal{G}(\gamma, \alpha)$.

Il faut souligner que les paires (γ, α) pour lesquelles r_n, défini par (3.7), est meilleure que v_n^p sont celles qui correspondent à $\rho > 1$. Autrement dit, l'estimateur Bayésien associé à la loi a priori log-splines, est plus performant que celui de Polonik (1995), dans le cas où ce dernier n'atteint pas la vitesse minimax.

- Les méthodes Bayésiennes ou plus généralement les méthodes basées sur la vraisemblance nécessitent d'imposer des hypothèses globales sur la densité ; dans le cadre fréquentiste (Polonik (1995), Tsybakov (1997), Rigollet et Vert (2007)) seules des hypothèses locales au voisinage de la frontière du vrai ensemble à niveau sont nécessaires. Ce type de contraintes peut impliquer qu'aucun estimateur Bayésien d'un ensemble à niveau ne soit asymptotiquement minimax pour la classe $\mathcal{G}(\gamma, \alpha)$. Se pose alors la question d'existence d'une loi a priori permettant d'atteindre la vitesse minimax pour la classe d'ensembles $\mathcal{G}(\gamma, \alpha)$. La loi a priori log-splines manque de flexibilité, notamment dans l'exploration de l'espace : nous avons considéré une partition à pas fixe de l'espace $[0, 1]^2$. La loi a priori du type log-spline avec noeuds libres chargeant les densités du type (3.3) semble plus adéquate mais reste alors le problème non-trivial de vérification des hypothèses [A1]-[A3].
  Soulignons que lorsque l'objet d'intérêt est la distribution de probabilité, un choix convenable de lois a priori implique que l'estimateur Bayésien de la densité atteint la vitesse minimax pour une classe de distributions donnée (Ghosal *et al.* (2000), Shen *et al.* (2001)).
- Nous envisageons d'étudier l'estimation adaptative des ensembles de niveau par une approche Bayésienne. En effet, l'approche Bayésienne se prête naturellement à ce genre de problème, en mettant une loi a priori sur la régularité γ de la densité. Soulignons que l'estimateur Bayésien associé à la loi a priori des log-splines, est construit indépendamment de α (α correspond à l'indicateur du comportement de la densité au voisinage de la frontière de son ensemble à niveau).
- Tous les points précédents soulèvent des questions relatives aux interactions entre les approches Bayésiennes et fréquentistes. D'un point de vue pratique, l'estimateur Bayésien a l'avantage d'être explicite et facile à implémenter (l'implémentation de l'estimateur Bayésien associé à la loi a priori log-splines à pas fixes est relativement aisée). D'un point de vue théorique, l'estimateur Bayésien associé à la loi a priori log-spline à pas fixes s'avère être compétitif en termes de vitesse de convergence (cf la comparaison avec les vitesses obtenues pas Polonik, 1995) même si il ne permet pas d'atteindre la vitesse minimax sur la classe $\mathcal{G}(\gamma, \alpha)$.

Chapitre 4

Inférence paramétrique dans un modèle de déconvolution aveugle bruitée.

Je me suis intéressée à estimer le niveau de bruit, le filtre inverse ainsi que la distribution du signal d'entrée dans un modèle de déconvolution aveugle bruitée. Cette étude menée en collaboration avec Emmanuelle Gautherat, est actuellement sous la forme d'un article (article [10]), en cours de soumission.

Ce chapitre se décompose en une partie détaillant le modèle et les hypothèses, une partie dédiée à la description des procédures d'estimation et des résultats obtenus et une partie dans laquelle figurent mes commentaires.

4.1 Modèle.

Le processus observé $(Y_t)_{t\in\mathbb{Z}}$ est généré par un modèle de déconvolution aveugle bruitée :

$$Y_t \;=\; (u\star X)_t+\sigma_0 W_t=\sum_{k\in\mathbb{Z}} u_k X_{t-k}+\sigma_0 W_t, \tag{4.1}$$

où $u=(u_t)_{t\in\mathbb{Z}}$ est un filtre inconnu et déterministe, $X=(X_t)_{t\in\mathbb{Z}}$ est le signal d'entrée qui est un processus inobservé à valeurs dans $\mathbb{C}$, dont la loi est supposée discrète à p points du support, σ_0 est le niveau de bruit inconnu, et $(W_t)_{t\in\mathbb{Z}}$ est une suite de variables complexes Gaussiennes qui sont indépendantes de $(X_t)_{t\in\mathbb{Z}}$.
Notre objectif est de restituer la loi du processus $X=(X_t)_{t\in\mathbb{Z}}$, et d'estimer les quantités σ_0 et $(u_t)_{t\in\mathbb{Z}}$. Le filtre u est supposé sommable ($u\in l_1(\mathbb{Z})$) et inversible ($\exists\theta=(\theta_t)_{t\in\mathbb{Z}}\in l_1(\mathbb{Z})$ tel que $\sum_{t\in\mathbb{Z}} u_t\theta_{k-t}=\delta_0(k)$, où $\delta_0(\cdot)$ désigne le symbole de Kronecker).

Pour notre étude, nous adoptons l'approche qui consiste à estimer le filtre inverse θ plutôt que le filtre lui-même. Le problème d'estimation de θ et/ou de σ_0 et/ou de la loi de X lorsque la loi des entrées est discrète avec un nombre fini de valeurs possibles, a engendré un certain nombre d'articles parmi lesquels Li (1995), Gamboa et Gassiat (1996), Gassiat et Gautherat (1998, 1999) et Gautherat (2002). Li (1995) a étudié, dans le cas réel et complexe, pour des données non-bruitées indépendantes et en supposant connus les points du support de X, l'existence d'un filtre inverse pour u. Gamboa et Gassiat (1996) ont proposé, dans un contexte général (points du support, nombre de points du support et distribution de X inconnus, variables $(X_t)_{t\in\mathbb{Z}}$ dépendantes) pour des données non-bruitées réelles, un estimateur consistant du filtre inverse. Gassiat et Gautherat (1998) ont généralisé l'article de Gamboa et Gassiat (1996) en considérant des données bruitées : elles ont établi la convergence en probabilité et presque sûre des estimateurs de θ et de σ_0, construits à partir d'une fonction de contraste empirique pénalisée. Dans un cadre similaire à celui de leur article de 1998, Gassiat et Gautherat (1999) ont établi la distribution limite de l'estimateur du filtre inverse et de celui du niveau de bruit lorsque θ est supposé appartenir à une famille paramétrique. Pour des observations bruitées, Gautherat (1997, 2002) a obtenu la distribution asymptotique des estimateurs des points du support et de ceux des probabilités du signal d'entrée.

La seule connaissance a priori du caractère discret et fini des valeurs possibles du processus d'entrée permet d'estimer avec consistance une très large classe de filtres inverses (Gamboa et Gassiat (1996) pour des données non-bruitées, Gassiat et Gautherat (1998, 1999) pour des données bruitées) et pour des processus $(X_t)_{t\in\mathbb{Z}}$ non nécessairement indépendants (Gamboa et Gassiat (1996), Gassiat et Gautherat (1998, 1999)), ou bien non nécessairement stationnaires (Li, 1995), ou bien non nécessairement causaux (Li (1995), Gamboa et Gassiat (1996), Gassiat et Gautherat (1998, 1999)), ou bien pour lesquels les points du support et leur nombre sont inconnus (Gamboa et Gassiat (1996), Gassiat et Gautherat (1998, 1999)). Gamboa et Gassiat (1996) ont exploité la structure des modèles linéaires discrets (relation (4.1), avec $\sigma_0 = 0$) à partir de laquelle ils ont proposé une nouvelle méthode d'estimation du filtre inverse qui s'avère être super-efficace (en termes de vitesse). Leurs procédures d'estimation reposent sur la minimisation d'une fonction de contraste empirique. Dans le modèle (4.1), la présence du bruit rend plus complexe la déconvolution aveugle des données, qui permet de caractériser le filtre inverse et le niveau de bruit. Dans ce cadre généralisant celui de Gamboa et Gassiat (1996), Gassiat et Gautherat (1998, 1999) ont proposé une méthode d'estimation reposant sur la minimisation d'une fonction de contraste empirique pénalisée.

Pour ce modèle et dans le cas complexe, nous avons considéré et étudié une nouvelle procédure d'estimation du filtre inverse et du niveau de bruit en exploitant les propriétés spécifiques des formes de Hankel. Cette procédure d'estimation est construite à partir d'un critère empirique qui est entièrement explicite, ce qui n'était pas le cas dans les travaux de Gassiat et Gautherat (1998, 1999) et de Gautherat (1997). Notre méthode d'estimation s'avère être performante tant sur le plan théorique (résultat de consistance et distribution limite) que d'un point de vue pratique. Elle présente l'avantage d'être facilement implémentable comparée à la méthode de Gassiat et Gautherat (1998). Leur méthode consiste à rechercher les minima d'une fonction de contraste empirique pénalisée : elle est très sensible au point de départ de l'algorithme (minima locaux) et elle requière le calibrage du terme de pénalité.

4.1.1 Hypothèses.

- [M1] $(X_k)_{k\in\mathbb{Z}}$ est une suite de variables aléatoires complexes, discrètes, de loi commune $\Pi = (\pi_j)_{\{j=1,\ldots,p\}}$ associée au vecteur $a = (a_1, \ldots, a_p) \in \mathbb{C}^p$, dont les composantes sont distinctes deux à deux ($p \geq 2$ est connu).
- [M2] $(X_k)_{k\in\mathbb{Z}}$ est stationnaire et ergodique.
- [M3] $\forall n \in \mathbb{N}^\star$, $\forall (j_1, \ldots, j_n) \in \{1, \ldots, p\}^n$, $\mathbb{P}(X_1 = a_{j_1}, \ldots, X_n = a_{j_n}) > 0$.
- [M4] $\forall k \in \mathbb{Z}$, $W_k = W_k^R + iW_k^I$, où $W^R = (W_k^R)_{k\in\mathbb{Z}}$ et $W^I = (W_k^I)_{k\in\mathbb{Z}}$ sont indépendants. $(W_k^R)_{k\in\mathbb{Z}}$ et $(W_k^I)_{k\in\mathbb{Z}}$ sont des variables aléatoires réelles i.i.d., de loi normale $\mathcal{N}(0, 1/2)$ et elles sont indépendantes de $(X_k)_{k\in\mathbb{Z}}$.
- [M5] $U(x) = \sum_k u_k e^{ikx}$ est continue et ne s'annule pas sur $[0, 2\pi]$.
- [M6] Le filtre inverse de u est noté θ. Tous les filtres appartiennent à $\Theta = \{s(\xi) \in l_1(\mathbb{Z}), \xi \in \mathcal{K}\}$, où la fonction $s \in \mathcal{C}^1$ est connue et injective. Le paramètre ξ est inconnu. Le vrai filtre inverse θ est supposé appartenir à Θ. $\mathcal{K} \subset \mathbb{R}^d$ est compact tel que si $\xi, \tilde{\xi} \in \mathcal{K}$ vérifient $s_k(\xi) = rs_{k-l}(\tilde{\xi})$, $\forall k \in \mathbb{Z}$, alors $r = 1$, $l = 0$, et $\xi = \tilde{\xi}$.

Soulignons que notre problème est paramétrique au travers de l'argument ξ apparaissant dans Θ (hypothèse [M6]). L'hypothèse [M6] permet de contourner les problèmes d'identifiabilité liés à un possible changement d'échelle ou de décalage des filtres de Θ. Notons également que s injective implique l'existence de $\xi_0 \in int(\mathcal{K})$ tel que $s(\xi_0) = \theta$. Estimer θ revient alors à estimer ξ_0.

Le filtre $(u_t)_{t\in\mathbb{Z}}$ étant inversible, le processus initial $(Y_t)_{t\in\mathbb{Z}}$ peut être transformé en un processus $(Z_t)_{t\in\mathbb{Z}}$ en appliquant n'importe quel filtre s à $(Y_t)_{t\in\mathbb{Z}}$, autrement dit, le processus résultant $(Z_t)_{t\in\mathbb{Z}}$ est défini par

$$\forall t \in \mathbb{Z}, Z_t(s) = (s \star Y)_t. \tag{4.2}$$

Notons que dans le cas non-bruité i.e. $\sigma_0 = 0$ dans l'équation (4.1), $Z_t(\theta) = X_t, \forall t \in \mathbb{Z}$. C'est à partir du processus $Z_t(s)$ défini par (4.2) avec $s \in \Theta$, que nous allons développer notre procédure d'estimation.

4.1.2 Caractérisations.

Dans le cas non bruité ($\sigma_0 = 0$), Gamboa et Gassiat (1996) ont montré que sous [M1], [M3] et [M5], $Z_t(s)$ a au plus p valeurs distinctes ssi $s = \theta$ à une échelle et à un décalage près. La définition de Θ comme sous-ensemble de $l_1(\mathbb{Z})$ permet de lever l'ambiguïté sur l'échelle et le décalage. La caractérisation de σ_0 et θ étant valide pour tout filtre inverse sommable, s désigne dans la suite de ce paragraphe, n'importe quel filtre sommable. La variable aléatoire $Z_t(s)$ avec $s \neq \theta$ admettant au moins p points du support, la caractérisation de σ_0 et θ s'effectue au travers d'une fonction de contraste H, capable de distinguer les variables aléatoires discrètes dont le support est de cardinal p, de celles dont le support est de cardinal supérieur à p.

Soit $d(s)$, le $(p+1)^2$-vecteur des moments conjugués de $Z_0(s)$ défini par $d_{j(p+1)+(k+1)}(s) = \mathbb{E}\left((Z_0(s))^k\overline{(Z_0(s))^j}\right), \forall (j,k) \in \{0,\ldots,p\}^2$. Le vecteur $d_{j(p+1)+(k+1)}(s)$ réarrangé en une matrice carrée $(p+1)\times(p+1)$ forme ainsi une matrice de Hankel. Pour $\sigma \geq 0$, les pseudo-moments conjugués $\tilde{d}(\sigma,s)$ sont solutions du système inversible suivant :

$$d(s) = A(\sigma\|s\|_2)\tilde{d}(\sigma,s),$$

où $\|s\|_2$ est la norme l_2 de s, $A(\sigma\|s\|_2)$ est une $(p+1)^2\times(p+1)^2$-matrice inversible telle que pour toute paire $(j,k) \in \{0,\ldots,p\}\times\{0,\ldots,p\}$ et $\forall 0 \leq m \leq j,\ 0 \leq l \leq k,\ \forall \beta \in \mathbb{R},\ A_{j(p+1)+k+1,m(p+1)+l+1}(\beta) = C_k^l C_j^m \gamma_{j-m,k-l}\beta^{k-l+j-m}$, avec $\gamma_{m,l} = \mathbb{E}((W_0)^l(\overline{W_0})^m)$. Sous [M4], notons que $\gamma_{m,l} = \mathbb{1}_{m=l}m!$, ce qui permet d'obtenir une expression explicite de la matrice inverse A^{-1}. Définissons la fonction H à valeurs réelles définie par :

$$H(\sigma,s) = \det(\tilde{D}(\sigma,s)),$$

où $\tilde{D}$ est le réarrangement du vecteur $\tilde{d}$ en une $((p+1)\times(p+1))$–matrice. Sous ce formalisme, Gassiat et Gautherat (1999) pour σ_0 et θ et Gautherat (1997) pour la distribution du processus $(X_t)_{t\in\mathbb{Z}}$ ont établi les résultats suivants :
Sous [M1], [M3] et [M5], σ_0 et θ vérifient

$$\begin{array}{ccc} \forall \sigma < \sigma_0, & H(\sigma,s) > 0 & \forall s \in l_1(\mathbb{Z}), \\ H(\sigma_0,s) = 0 & \text{ssi} & s = \theta \text{ à une échelle et un décalage près.} \end{array}$$

Sous [M1], les composantes du vecteur $a = (a_i)_{\{i=1,\ldots,p\}}$ sont les zéros dans $\mathbb{C}[X]$ du polynôme de degré p dont les coefficients sont les composantes du vecteur $v^\star$, qui est le vecteur propre associé à la plus petite valeur propre de la matrice $\tilde{D}(\sigma_0,\theta)$.
Sous [M1], le vecteur des probabilités $\Pi = (\pi_1,\ldots,\pi_p)$ est l'unique solution du système linéaire : $\forall k = 0,\ldots,p-1,\ \ \mathbb{E}(X_0^k) = \sum_{i=1}^p q_i a_i^k$.

4.2 Procédures d'estimation et résultats.

Pour tout $\xi \in \mathcal{K}$, notons $\bar{s}(\xi)$ dans Θ, les filtres tronqués définis pour tout $k \in \mathbb{Z}$ par $\bar{s}_k(\xi) = s_k(\xi)\ \mathbb{1}_{\{|k|\leq k(n)\}}$, où $k(n)$ est une suite de nombres strictement positifs, décroissante vers 0, quand n grandit. L'estimation est basée sur n observations

$Z_t(\bar{s}(\xi))$, $t = 1 + k(n), \ldots, n - k(n)$, définies comme la convolution discrète tronquée entre $s(\xi)$ et $(Y_t)_{t=1+k(n),\ldots,n-k(n)}$. Nous considérons $d_n(\bar{s}(\xi))$ la version empirique du vecteur des moments conjugués de $(Z_t(\bar{s}(\xi)))_{t=1+k(n),\ldots,n-k(n)}$, qui combinée avec A^{-1}, nous permet d'en déduire $\tilde{d}_n(\sigma, \bar{s}(\xi))$, la version empirique des pseudo-moments conjugués. Finalement, la fonction critère H_n est définie comme étant le determinant de la matrice carrée $\tilde{D}_n(\sigma, \bar{s}(\xi))$, dont les éléments sont les composantes du vecteur des pseudo-moments empiriques $\tilde{d}_n(\sigma, \bar{s}(\xi))$.

$$\begin{aligned}
d_{j(p+1)+k+1,n}(\bar{s}(\xi)) &= \frac{1}{n-2k(n)} \sum_{t=1+k(n)}^{n-k(n)} Z_t^k(\bar{s}(\xi)) \overline{Z_t^j(\bar{s}(\xi))}, \\
\tilde{d}_n(\sigma, \bar{s}(\xi)) &= A^{-1}(\sigma \|s(\xi)\|_{n,2}) d_n(\bar{s}(\xi)), \\
\text{où } \|s(\cdot)\|_{n,2} &= \left(\sum_{k=-k(n)}^{k(n)} s_k(\cdot)^2 \right)^{1/2}, \\
H_n(\sigma, \bar{s}(\xi)) &= \det(\tilde{D}_n(\sigma, \bar{s}(\xi))).
\end{aligned}$$

Les estimateurs $\widehat{\sigma}_{0,n}, \widehat{\xi}_n$ sont solutions du système suivant :

$$\begin{cases} H_n(\widehat{\sigma}_{0,n}, \bar{s}(\widehat{\xi}_n)) = 0 \\ \widehat{\sigma}_{0,n} = \min \left\{ \sigma \in \mathbb{R}_+^\star \text{ pour lesquels } \exists \xi \in \mathcal{K} :\ H_n(\sigma, \bar{s}(\xi)) = 0 \right\} \end{cases}$$

L'estimateur du filtre inverse est défini par $\widehat{\theta}_n = \bar{s}(\widehat{\xi}_n)$.
Les estimateurs des points du support $(\widehat{a}_1, \ldots, \widehat{a}_p) \in \mathbb{C}^p$ sont les racines dans $\mathbb{C}[X]$ de $p_{\widehat{v}^\star}(x) = \sum_{j=0}^{p} \widehat{v}_j^\star x^j$, où $\widehat{v}^\star = (\widehat{v}_0^\star, \ldots, \widehat{v}_p^\star)$ est le vecteur propre associé à la plus petite valeur propre de $\tilde{D}_n(\widehat{\sigma}_{0,n}, \bar{s}(\widehat{\xi}_n))$.
L'estimateur des probabilités des entrées $\widehat{\Pi} = (\widehat{\pi}_1, \ldots, \widehat{\pi}_p)$ est solution unique du système linéaire : $\forall j = 0, \ldots, p-1$, $\tilde{d}_{j(p+1)+1,n}(\widehat{\sigma}_{0,n}, \bar{s}(\widehat{\xi}_n)) = \sum_{i=1}^p q_i \widehat{a}_i^j$. Pour ces estimateurs, nous obtenons les résultats suivants :

Sous [M1]-[M6], si $k(n) = o(1/\sqrt{n})$, si $\sum_{|k|>k(n)} |s_k(\xi_0)| = o((1/\sqrt{n}))$, alors quand n tend vers l'infini, $\widehat{\sigma}_{0,n}$ converge p.s. vers σ_0, $\|\widehat{\xi}_n - \xi_0\|$ converge p.s. vers 0, $\|\widehat{a} - a\|$ et $\|\widehat{\Pi} - \Pi\|$ convergent p.s. vers 0, où $\|\cdot\|$ désigne la norme Euclidienne dans $\mathbb{R}^d$ ou bien dans $\mathbb{R}^p$.

Sous [M1]-[M6], si $k(n) = o(1/\sqrt{n})$ et si $\sum_{|k|>k(n)} |s_k(\xi_0)| = o((1/\sqrt{n}))$, sous des hypothèses plus fortes concernant la fonction s et une hypothèse de normalité asymptotique du vecteur $d_n(s(\xi_0))$ et de certaines de ses dérivées, quand n tend vers l'infini, les normalités asymptotiques des quantités centrées suivantes $\sqrt{n}\begin{pmatrix} \widehat{\xi}_n - \xi \\ \widehat{\sigma}_{0,n} - \sigma_0 \end{pmatrix}$, $\sqrt{n}(\hat{a}-a)$ et $\sqrt{n}(\hat{\Pi}-\Pi)$ sont alors établies avec des variances asymptotiques explicites.

4.3 Commentaires et remarques.

- La contribution principale de cette étude a été d'exploiter au mieux les caractérisations détaillées dans la partie 4.1.2 de ce chapitre, en particulier en exhibant explicitement la matrice inverse A^{-1}. A partir de la forme explicite de A^{-1}, nous avons démontré qu'en tout n, il existe un couple de niveau de bruit et de filtre inverse qui annule H_n. Les estimateurs sont alors définis comme les zéros de la fonction empirique H_n.
- L'hypothèse de normalité des erreurs n'est pas nécessaire à la résolution de ce problème : seule l'infiniment divisibilité de la loi des erreurs est requise. Dans le cas d'une loi des erreurs différente de la loi normale, les moments conjugués de cette loi devront être calculés de manière à rendre A et A^{-1} explicites : ce qui n'est pas toujours réalisable théoriquement.
- Nous avons testé numériquement notre méthode d'estimation sur des modèles simples (cas particuliers de l'équation (4.1)) : un mélange bruité (correspond à un filtre $u_t = 1\!\!\mathrm{I}_{t=0} = \theta_t$) et un modèle autorégressif d'ordre 2 bruité (correspond à un filtre inverse θ ayant 3 composantes). Les résultats sont relativement performants : lorsque le vrai niveau de bruit n'est pas très élevé ($\sigma_0 = 0.05$), les valeurs des estimés sont bonnes même avec un petit nombre d'observations ($n = 100$) ; lorsque le niveau de bruit est élevé ($\sigma_0 = 1$), un grand nombre d'observations ($n = 15000$) est nécessaire pour obtenir des valeurs des estimés proches des vraies quantités.
- Nous envisageons de résoudre ce problème d'estimation dans un cadre non-paramétrique. Il s'agira alors de supposer que les filtres et les filtres inverses appartiennent à une boule fonctionnelle (la plus naturelle ici étant un ellipsoïde de $l_2(\mathbb{Z})$) et de dérouler l'approche minimax afin de déterminer les vitesses minimax et d'exhiber les estimateurs asymptotiquement optimaux.

Références

1. Baraud, Y., Huet, S. et , B. (2003), Adaptive tests of linear hypothesis by model selection, *Ann. Statist.*, **31**, 225–251.
2. Barron, A., Birgé, L. et Massart, P. (1999), Risk bound for model selection via penalization, *Probab. Theory and Relat. Fields*, **113**, 301–413.
3. Barron, A., Schervish, M.J. et Wasserman, L. (1999), The consistency of posterior distribution in nonparametric problems, *Ann. Statist.*, **27**, 536–561.
4. Bickel, P.J. et Ritov, Y. (1988), Estimating integrated squared density derivatives : sharp best order of convergence estimates, *Sankhya*, **50**, 381–393.
5. Birgé, L. (1983), Approximation dans les espaces métriques et théorie de l'estimation, *Z. Wahrsch. Ver. Geb.*, **65**, 181–237.
6. Birgé L. et Massart, P. (1995), Estimation of integral functionals of a density, *Ann. Statist.*, **23**, 11–29.
7. Birgé, L. et Massart, P. (1997), From model selection to adaptive estimation, *Festschrift for Lucien Le Cam : Research Papers in Probability and Statistics (D. Pollard, E. Torgersen, and G. Yang, eds)*, 55–87, Springer-Verlag, NY.
8. Cavalier, L. et Tsybakov, A.B. (2001), Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation, *Mathematical Methods of Statistics*, **10**, 247–282.
9. Chu, C.K. et Cheng, P.E. (1996), Estimation of jump points and jump values of a density function, *Statistica Sinica*, **6**, 79–95.
10. Diaconis, P. et Freedman, D.A. (1986), On the consistency of Bayes estimates , *Ann. Statist.*, **14**, 1–26.
11. Donoho, D.L. et Johnstone, I.M. (1994a), Minimax risk for l_q losses over l_p-balls, *Probability Theory and Related Fields*, **99**, 277–303.
12. Donoho, D.L. et Johnstone, I.M. (1994b), Ideal spatial adaptation via wavelet shrinkage, *Biometrika*, **81**, 425–455.
13. Donoho, D.L. et Johnstone, I.M. (1995), Adpating to unknown smoothness via wavelet shrinkage, *J. Amer. Statist. Assos.*, **90**, 1200–1224.
14. Donoho, D.L. et Johnstone, I.M. (1996), Neo-classical minimax problems, thresholding and adaptive function estimation, *Bernoulli*, **2**, 39–62.
15. Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. et Picard, D. (1995), Wavelet Shrinkage : Asymptotia ?, *J.R. Statist. Soc. B*, **57**, 301–369.
16. Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. et Picard, D. (1996), Density estimation by wavelet thresholding, *Ann. Statist.*, **24**, 508–539.

17. Donoho, D.L. et Nussbaum, M. (1990), Minimax quadratic estimation of a quadratic functional, *J. Complexity*, **6**, 290–323.

18. Doob, J.L. (1949), Application of the theory of martingales, *Le Calcul des Probabilités et ses applications*, Colloques Internationaux du C.N.R.S. (Paris), 23–27.

19. Efromovich, S.Yu. et Pinsker, M.S. (1984), A learning algorithm for nonparametric filtering, *Autom. Telem.*, **11**, 1434–1440.

20. Efromovich, S.Yu. et Low, M. (1996), On optimal adaptive estimation of a quadratic functional, *Ann. Statist.*, **24**, 1106–1125.

21. Ermakov, M.S. (1990), Minimax detection of a signal in a white Gaussian noise, *Theory Probab. Appl.*, **35**, 667–679.

22. Freidlin, M.I. et Korostelev, A.P. (1995), Image processing for plane domains : change-point problems for the domain's area, *Problems of Information Transmission*, **31**, 27–45.

23. Fromont, M. et Lévy-Leduc, C. (2003), Adaptive tests for periodic signals detection with applications to laser vibrometry, *ESAIM Probability and Statistics*, **10**, 46–75.

24. Farrel, R. (1967), On the lack of a uniformly consistent sequence of estimate of a density function in certain cases, *Ann. Math. Statist.*, **38**, 471–474.

25. Freedman, D.A. (1963), On the asymptotic behavior of Bayes estimates in the discrete case, *Ann. Math. Statist.*, **34**, 1386–1403.

26. Gamboa, F. et Gassiat, E. (1996), Blind Deconvolution of discrete linear systems, *Ann. Statist.*, **24**, 1964–1981.

27. Gassiat, E. et Gautherat, E. (1998), Identification of noisy linear systems with discrete random input, *IEEE Transaction on Information Theory*, **44**, 1941–1952.

28. Gassiat, E. et Gautherat, E. (1999), Speed of convergence for the blind deconvolution of a linear systems with discrete random input, *Ann. Statist.*, **27**, 1684–1705.

29. Gautherat, E. (1997), Déconvolution aveugle des systèmes linéaires aléatoires discrets bruités ou non, Thèse, Université Evry-Val d'Essonne.

30. Gautherat, E. (2002), Déconvolution aveugle bruitée : estimation de la distribution du processus source, *Document de travail, LS-CREST*, **22**.

31. Ghosal, S., Ghosh, J.K. et Ramamoorthy, R.V. (1999a), Consistency issues in Bayesian non- parametrics, In *Asymptotics Nonparametrics and Times Series : A tribute to Madan Lal Puri*, eds : S. Ghosh, 639–668.

32. Ghosal, S., Ghosh, J.K. et Ramamoorthy, R.V. (1999b), Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.*, **27**, 143–158.

33. Ghosal, S., Ghosh, J.K. et van der Vaart, A. (2000), Convergence rates of posterior distributions, *Ann. Statist.*, **28**, 500–531.

34. Ghosal, S. et van der Vaart, A. (2007a), Convergence rates of posterior distributions for noniid observations, *Ann. Statist.*, **35**,192–223.

35. Ghosal, S. et van der Vaart, A. (2007b), Posterior convergence rates of dirichlet mixtures at smooth densities, *Ann. Statist.*, **35**, 697–723.

36. Ghosh, J.K. et Ramamoorthy, R.V. (2003), *Bayesian Nonparametrics*, Springer Series in Statistics, Springer–Verlag, NY.

37. Goldenshluger, A., Tsybakov, A.B. et Zeevi, A. (2006), Optimal change-point estimation from indirect observations, *Ann. Statist.*, **34**, 350–372.

38. Golubev, G.K. (1987), Adaptive asymptotically minimax estimates for smooth signals. *Problems of Information Transmission*, **23**, 57–67.

39. Golubev, G.K. (1992), Nonparametric estimation of smooth densities of a distribution in L_2, *Problems Information Transmission*, **28**, 44–54.

40. Golubev G., Levit B., et Tsybakov A.B. (1996), Asymptotically efficient estimation of analytic functions in Gaussian noise, *Bernoulli*, **2**, 167–181.

41. Golubev, G.K. et Nussbaum, M. (1992), Adaptive spline estimates in a nonparametric regression model, *Theory Probab. Appl.*, **37**, 521–529.

42. Guerre, E. et Lavergne, P. (2002), Optimal minimax rates for nonparametric specification testing in regression models, *Econom. Theory*, **18**, 1139–1171.

43. Härdle, W., Kerkyacharian, G., Picard, D. et Tsybakov, A.B. (1998), *Wavelets, Approximation and Statistical Applications*, Lecture Notes in Statistics, **129**, Springer, NY.

44. Hartigan, J.A. (1987), Estimation of a convex density contour in two dimensions, *J. Amer. Statist. Assoc.*, **82**, 267–270.

45. Heikkinen, J. (1998), Curve and surface estimation using dynamic step functions. In *Practical Nonparametric and semiparametric Bayesian statistics*, D. Dey, P. Müller and D. Sinha, ed, N.Y : Springer–Verlag, 255–272.

46. Horowitz, J.L. et Spokoiny, V.G. (2001), An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative, *Econometrica*, **69**, 599–631.

47. Ibragimov, I. et Khas'minskii, R.Z. (1981), *Statistical Estimation*, Springer, NY.

48. Ibragimov, I.A., Khas'minskii, R.Z. et Nemirovskii, A.S. (1986), Some problems on nonparametric estimation in Gaussian white noise, *Theory Probab. Appl.*, **31**, 391–406.

49. Ingster, Yu.I. (1982), On minimax nonparametric detection of a signal in Gaussian white noise, *Probl. Inf. Trans.*, **18**, 61–73.

50. Ingster, Yu.I. (1986), Minimax testing of nonparametric hypotheses on a distribution density in L_p-metrics, *Theory Probab. Appl.*, **31**, 333–337.

51. Ingster, Yu.I. (1990), Minimax detection of signals in L_p-metrics, *Zap. Nauchn. Sem. Leningr. Otdel. Mat. Inst. Steklov.*, **184**, 152–168.

52. Ingster, Yu.I. (1993), Asymptotically minimax hypothesis testing for nonparametric alternatives, I, II, III., *Math. Methods Stat.*, **2**, 85–114 ; 171–189 ; 249–268.

53. Ingster, Y. et Suslina, I.A. (2002), *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*, **Vol. 169**, Lecture Note in Statistics.

54. Kerkyacharian, G. et Picard, D. (2000), Thresholding algorithms, maxisets and well-concentrated bases, *Test*, **9**, 283–344.

55. Kerkyacharian, G. et Picard, D. (2002), Minimax or maxisets ?, *Bernoulli*, **8**, 219–253.

56. Khas'minskii, R.Z. et Lebedev, V.S. (1990), On the properties of parametric estimators for areas of a discontinuous image, *Problems of Control and Information Theory*, **19**, 375–385.

57. Korostelev, A.P. (1987), On minimax estimation of a discontinuous signal, *Theory Probab. Appl.*, **32**, 727–730.

58. Korostelev, A.P. et Tsybakov, A.B., (1993), *Minimax theory of image reconstruction*, Lecture Notes in Statistics, Springer–Verlag, NY.

59. Laurent, B. (1996), Efficient estimation of integral functionals of a density, *Ann. Statist.*, **24**, 659–681.

60. Laurent, B. (2005), Adaptive estimation of a quadratic functional of a density by model selection, *ESAIM P & S*, **9**, 1–18.

61. Laurent, B. et Massart, P. (2000), Adaptive estimation of a quadratic functional by model selection, *Ann. Statist.*, **28**, 1302–1338.

62. Le Cam, L.M. (1958), Les Propriétés Asymptotiques des Solutions de Bayes, *Inst. Statist. Univ. Paris*, **7**, 17–35.

63. Lepskii, O.V. (1990), On a problem of adaptive estimation in Gaussian white noise, *Theory Probab. Appl.*, **35**, 454–466.

64. Lepskii, O.V. (1991), Asymptotically minimax adaptive estimation. I. Upper bounds, *Theory Probab. Appl.*, **36**, 682–697.

65. Lepskii, O.V. (1993), On asymptotically exact testing of nonparametric hypotheses, *Discussion paper*, Université de Louvain.

66. Lepskii, O.V., Nemirovski, A. et Spokoiny, V.G. (1999), On estimation of the L_r norm of a regression function, *Probab. Theory Relat. Fields*, **113**, 221–253.

67. Lepskii, O.V. et Spokoiny, V.G. (1999), Minimax nonparametric hypothesis testing : the case of inhomogeneous alternative, *Bernoulli*, **5**, 333–358.

68. Lepskii, O.V. et Tsybakov, A.B. (2000), Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point, *Probab. Theory Relat. Fields*, **117**, 17–48.

69. Li, T.H. (1995), Blind deconvolution of linear systems with multilevel nonstationary inputs, *Ann. Statist.*, **23**, 690–704.

70. Loubes, J-M. et van de Geer, S.A. (2002), Adaptive estimation in regression, using soft thresholding type penalties, *Statistica Neerlandica*, **56**, 453–478.

71. Neumann, M.H. (1997), Optimal change-point in inverse problems, *Scand. J. Statist.*, **24**, 503–521.

72. Nolan, D. (1991), The excess mass ellipsoid, *J. Multivariate Anal.*, **39**, 348–371.

73. Nussbaum M. (1985), Spline smoothing in regression Models and Asymptotic Efficiency in L_2, *Ann. Statist.*, **13**, 984–997.

74. Picard, D. et Tribouley K. (2000), Adaptive confidence interval for pointwise curve estimation, *Ann. Statist.*, **28**, 298–335.

75. Pinsker, M.S. (1980), Optimal filtering of square integrable signals in Gaussian white noise, *Problems of Information Transmission*, **16**, 120–133.

76. Polonik, W. (1995), Measuring mass concentrations and estimating density contours clusters–an excess mass approach, *Ann. Statist.*, **23**, 3, 855–881.

77. Raimondo, M. (1998), Minimax estimation of sharp change points, *Ann. Statist.*, **26**, 1379–1397.

78. Rigollet, P. et Vert, R. (2007), Fast rates for plug-in estimators of density level sets, *Preprint*, http ://hal.archives-ouvertes.fr/hal-00114180.

79. Shen, X. et Wasserman, L. (2001), Rates of convergence of posterior distributions, *Ann. Statist.*, **29**, 687–714.

80. Schwartz, L. (1965), On Bayes Procedures, *Z. Wahrsch. Ver. Geb.*, **4**, 10–26.

81. Spokoiny, V.G. (1996), Adaptive hypothesis testing using wavelets, *Ann. Statist.*, **24**, 2477–2498.

82. Spokoiny, V.G. (1998), Adaptive and spatially adaptive testing of a nonparametric hypothesis, *Math. Methods Stat.*, **7**, 245–273.

83. Stone, C. (1994), The use of polynomial splines and their tensor products in multivariate function estimation, *Ann. Statist.*, **22**, 118–184.

84. Suslina, I.A. (1993), Minimax detection of a signal for l_p–ball, *Zap. Nauchn. Semin. POMI*, **207**, 127–137.

85. Tsybakov, A.B. (1994), Multidimensional change-point problems and boundary esti mation. In : *Change-Point Problems* (E. Carlstein, H.G. Müller and D. Siegmund, eds.), IMS Lecture Notes Monograph Series, vol.23, Hayward, California, 317–329. IMS, Hayward, CA.

86. Tsybakov, A.B. (1997), On nonparametric estimation of density level sets, *Ann. Statist.*, **25**, 948–969.

87. Tsybakov, A.B. (2004), *Introduction à l'estimation non-paramétrique*, Springer.

88. Walther, G. (1997), Granulometric smoothing, *Ann. Statist.*, **25**, 6, 2273–2299.

Bibliographie

[1] Gayraud, G. (1997), Estimation of functionals of density support, *Mathematical Methods of Statistics*, **6**, 1, 26-47.

[2] Gayraud, G., Tribouley, K. (1999), Adaptive Estimation and Confidence Interval for a Quadratic Functional by Wavelet, *Statistics and probability letters*, **44**, 2, 109–122.

[3] Gayraud, G. (2001), Minimax hypotheses testing about the density support, *Bernoulli*, **7**, 3, 507–526.

[4] Gayraud, G., Pouet, Ch. (2001), Minimax Testing composite null hypotheses in the discrete regression scheme, *Math. Methods Stat.*, **4** , 375–394.

[5] Gayraud, G., (2002), Minimax estimation of a discontinuity for the density, *Journal of Nonparametric Statistics*, **14**, 1–2, 59–66, 2002.

[6] Gayraud, G., Tsybakov, A.B., (2002), Testing hypotheses about contours in images, *Journal of Nonparametric Statistics*, **14**, 1–2, 67–85.

[7] Gayraud, G., Pouet, Ch. (2005), Adaptive minimax testing in the discrete regression scheme, *Probability Theory and Related Fields*, **4**, 531–558.

[8] Gayraud, G., Rousseau, J. (2005), Rates of Convergence for a Bayesian Level set estimation, *Scandinavian Journal of Statistics*, **32**, 639-660.

[9] Gayraud, G., Rousseau, J. (2007), Consistency results on nonparametric Bayesian estimation of level sets using spatial priors, *Test*, **16**, 90-108.

[10] Gautherat, E et Gayraud, G, Parametric Estimation in Noisy Blind Deconvolution Model : a New Estimation Procedure, soumis à EJS.