Estimation dans des modèles à variables cachées

Catherine Matias

ORSAY
N^o D'ORDRE :

UNIVERSITÉ DE PARIS SUD
U.F.R SCIENTIFIQUE D'ORSAY

THÈSE

présentée

pour obtenir

Le TITRE de DOCTEUR EN SCIENCES
DE L'UNIVERSITÉ PARIS XI ORSAY

SPÉCIALITÉ : MATHÉMATIQUES

par

Catherine MATIAS

Sujet : **ESTIMATION DANS DES MODÈLES À VARIABLES CACHÉES**

Rapporteurs : M. Peter BICKEL
M. Aad van der VAART

Soutenue le 21 décembre 2001 devant le jury composé de :

M.	Jean BRETAGNOLLE	Président
Mme	Élisabeth GASSIAT	Directrice de Thèse
M.	Éric MOULINES	Examinateur
M.	Alexandre TSYBAKOV	Examinateur
M.	Aad van der VAART	Rapporteur

Table des matières

Introduction

Contents

Cette thèse est consacrée à l'estimation dans des modèles à variables cachées. Les observations sont des variables aléatoires issues d'une suite de variables aléatoires non observée (le signal) et d'une autre suite de variables aléatoires, de loi connue ou non (le bruit). Le cadre de ces estimations pourra être aussi bien paramétrique que semi-paramétrique ou encore non-paramétrique.

Le premier chapitre de cette thèse concerne l'étude d'une chaîne de Markov cachée. Le signal est une chaîne de Markov et la forme du bruit, indépendant du signal, n'est pas spécifiée (ce bruit peut être additif ou encore multiplicatif). La transition de la chaîne de Markov ainsi que la fonctionnelle qui définit les observations sont connues au paramètre θ près et nous nous intéressons à l'estimation de ce paramètre θ. Nous prouvons que l'estimateur du maximum de vraisemblance pour ce paramètre θ est consistant, asymptotiquement normal et efficace.

Le second chapitre concerne le modèle de convolution où le signal est de loi inconnue et le bruit de loi gaussienne centrée et de variance inconnue σ^2. Nous nous intéressons à l'estimation semi-paramétrique de la variance σ^2 du bruit et de la densité g de la loi du signal. Nous exhibons différents estimateurs de la variance et calculons leurs vitesses de convergence. Nous donnons également une borne inférieure du risque minimax pour l'estimation de cette variance σ^2 ainsi que pour la densité g du signal.

Le troisième et dernier chapitre concerne l'estimation non-paramétrique de certaines fonctionnelles dans le modèle de convolution où le signal est de loi inconnue et le bruit est supposé de loi entièrement connue gaussienne centrée réduite. Nous nous intéressons à la vitesse d'estimation de ces fonctionnelles. Nous donnons des bornes inférieures et des bornes supérieures pour ces vitesses d'estimation pour la norme de $\mathbb{L}_p(\mathbb{R})$, ainsi que pour le risque quadratique ponctuel.

Nous allons maintenant décrire plus précisément les modèles mentionnés ci-dessus, énoncer les résultats obtenus dans chacun de ces modèles, expliquer les liens qui les unissent, et évoquer les divers prolongements possibles à ce travail.

Estimation paramétrique dans le modèle de Markov caché

État de l'art

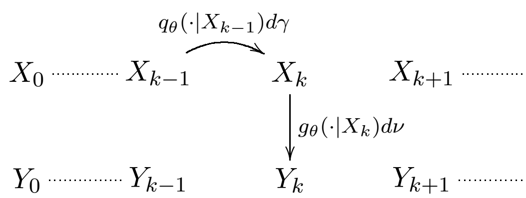
Les modèles de Markov cachés ont fait l'objet d'un très grand développement ces dernières années, ceci étant dû à de nombreuses applications pratiques dans des domaines aussi variés que la reconnaissance de la parole (Juang and Rabiner 1991), la neurophysiologie (Fredkin and Rice 1987), la biologie (Churchill 1989), l'économétrie (Kim, Shephard, and Chib 1998), ou encore l'analyse des séries temporelles (DeJong and Shephard 1995; Chan and Ledolter 1995). On trouvera également dans MacDonald et Zucchini (1997) des références plus complètes.

Le premier chapitre de cette thèse présente un travail réalisé en collaboration avec Randal Douc et portant sur l'estimation des paramètres d'un modèle de Markov caché. Commençons tout d'abord par décrire le modèle étudié. Nous considérons une chaîne de Markov homogène $\{X_n\}_{n\geq 0}$ sur un espace métrique $\mathcal{X}$ muni de la tribu borélienne $\mathcal{B}or(\mathcal{X})$ et d'une mesure positive γ. La transition $q_\theta d\gamma$ de la chaîne est connue au paramètre θ près, dont la vraie valeur est notée

θ^* et sa loi initiale est notée $\pi^* d\gamma$. Cette chaîne n'est pas observée et sera donc dite cachée. Les observations sont constituées par une suite de variables aléatoires $\{Y_n\}_{n\geq 0}$ définies sur un espace métrique séparable et complet $\mathcal{Y}$ muni de la tribu borélienne $\mathcal{B}or(\mathcal{Y})$ et d'une mesure positive ν. Ces observations vérifient l'hypothèse suivante : la loi du n-uplet $(Y_0, \dots, Y_n)$, conditionnelle à la donnée des variables cachées $(X_0, \dots, X_n)$ est égale au produit tensoriel des lois de chaque Y_i, conditionnelles à la seule donnée de X_i, pour i variant de 0 à n :

$$\forall\, n \geq 0, \quad \mathcal{L}(Y_0, \dots, Y_n | X_0, \dots, X_n) = \bigotimes_{i=0}^{n} \mathcal{L}(Y_i | X_i).$$

La densité de cette loi conditionnelle de Y_i sachant X_i (par hypothèse, indépendante du numéro i de l'observation) est notée $g_\theta d\nu$ et supposée connue au paramètre θ près dont la vraie valeur est θ^*. Nous pouvons schématiser ce modèle de la façon suivante :



Le but de l'étude réside dans l'estimation basée sur l'observation de la suite $\{Y_n\}_{n\geq 0}$, du vrai paramètre θ^*, appartenant à l'intérieur d'un compact Θ de $\mathbb{R}^p$.

Ce travail s'inscrit dans le prolongement des travaux antérieurs portant sur l'estimation des paramètres d'un modèle de Markov caché. Les premiers résultats datent de 1966 avec les travaux de Baum et Petrie dans le cadre très particulier où les variables aléatoires cachées ainsi que les variables observées ne prennent qu'un nombre fini de valeurs (i.e. les espaces d'états $\mathcal{X}$ et $\mathcal{Y}$ sont finis). Les auteurs établissent la consistance et la normalité asymptotique de l'estimateur du maximum de vraisemblance pour le paramètre θ, sous l'hypothèse d'un régime stationnaire pour la chaîne de Markov $\{X_n\}_{n\geq 0}$ (autrement dit, la loi initiale $\pi^* d\gamma$ de la chaîne est connue et égale à sa loi stationnaire). Le travail de Bakry et al. (1997) propose d'étendre les résultats précédents au cas non stationnaire par une technique de couplage.

Récemment, ces résultats ont été étendus au cas d'un espace d'états plus général pour les observations $\{Y_n\}_{n\geq 0}$, comprenant le cas de $\mathbb{R}^d$. Cependant, la chaîne de Markov est encore supposée stationnaire à espace d'états $\mathcal{X}$ fini. Les travaux de Leroux (1992) établissent la consistance de l'estimateur du maximum de vraisemblance des paramètres. Bickel et Ritov (1996) étudient la normalité asymptotique locale (LAN) de cet estimateur. Par la suite, Bickel, Ritov et Rydèn (1998) prouvent la normalité asymptotique de l'estimateur du maximum de vraisemblance pour le paramètre θ, sous l'hypothèse de sa consistance.

Parallèlement à ces travaux, Legland et Mevel (2000b, 2000a) ont développé une technique différente pour établir la consistance et la normalité asymptotique de l'estimateur du maximum de vraisemblance dans le cadre d'un espace d'états $\mathcal{Y}$ séparable complet pour la suite des observations et d'un espace d'états $\mathcal{X}$ fini pour la chaîne de Markov. Un des avantages de cette technique est qu'elle ne nécessite aucune hypothèse de stationnarité sur la chaîne de Markov (ce qui était le cas dans les travaux précédemment cités, sauf pour Bakry et al.). C'est sur ces techniques que s'appuie notre étude des modèles de Markov cachés, généralisant les travaux de Mevel (1997) au cadre d'un espace d'états pour la chaîne de Markov non plus fini mais métrique compact. Dans ce cadre, nous établissons la consistance, la normalité asymptotique et l'efficacité de l'estimateur du maximum de vraisemblance des paramètres.

L'article de Jensen et Petersen (1999) est un travail parallèle au notre. Par des techniques analogues à celles initiées par Baum et Petrie (1966) et par Bickel, Ritov et Rydèn (1998), les auteurs prouvent la normalité asymptotique de l'estimateur du maximum de vraisemblance dans le cas d'un espace d'états $\mathcal{Y}$ topologique général pour les observations, et d'un espace d'états $\mathcal{X}$ compact pour la chaîne cachée. Ce résultat ne s'applique que dans le cadre d'une chaîne stationnaire. Les auteurs établissent la normalité asymptotique de l'estimateur du maximum de vraisemblance sous l'hypothèse de sa consistance.

Une des différences essentielles entre ce travail et notre étude réside dans le fait que nous établissons un résultat de consistance (qui est une extension du résultat obtenu par Leroux (1992)) sous l'hypothèse usuelle d'identifiabilité du modèle, point qui n'est abordé ni dans le travail de Mevel (1997), ni dans celui de Jensen et Petersen (1999). Les techniques utilisées ici sont bien adaptées pour l'étude des estimateurs récursifs (où l'estimateur est réévalué à chaque nouvelle observation) mais elles nécessitent cependant des hypothèses plus fortes que celles utilisées par Bickel et al. (1998) et par Jensen et Petersen (1999).

Enfin, dans un travail postérieur au notre, Douc, Moulines et Rydèn (2001) établissent la consistance et la normalité asymptotique de l'estimateur du maximum de vraisemblance dans un modèle auto-régressif à régime markovien (la variable observée Y_n ne dépend plus seulement de la variable cachée X_n mais également de $Y_{n-1}, \dots Y_{n-d}$ où d est un entier fixé). L'approche s'inspire des travaux de Bickel et al. (1998) et de ceux de Jensen et Petersen (1999), ce qui permet d'affaiblir certaines hypothèses. Une technique de couplage permet d'étendre ces résultats au cas non-stationnaire.

Techniques utilisées

Précisons en quelques mots les principales étapes de nos preuves ainsi que les techniques utilisées. Nous supposerons à présent que la chaîne de Markov $\{X_n\}_{n\geq 0}$ est à valeurs dans un sous-ensemble compact $\mathbf{K}$ de $\mathcal{X}$. En effet, nous utilisons l'hypothèse suivante sur la transition de la chaîne de Markov :

$$\forall \theta \in \Theta, \quad \inf_{x,x'} q_\theta(x, x') > 0. \tag{1}$$

Or, q_θ étant le noyau de transition d'une loi de probabilité, il vérifie l'égalité $\int q_\theta(x, y)dy = 1$, ce qui oblige, avec l'hypothèse (1), l'espace d'intégration à être de mesure finie. Aucune hypothèse n'est faite sur la loi initiale $\pi^* d\gamma$ de cette chaîne et les paramètres du modèle sont donc $(\theta; \zeta)$ où θ appartient à Θ et ζ est une densité sur $(\mathcal{X}, \mathcal{B}or(\mathcal{X}), \gamma)$.

La méthode classique de Wald (1949) pour prouver la consistance de l'estimateur du maximum de vraisemblance comporte trois points : prouver la convergence de la log-vraisemblance renormalisée vers une fonction de contraste limite, montrer que ce contraste limite est maximum uniquement pour la valeur du vrai paramètre et enfin établir un critère d'uniformité par rapport au paramètre dans la convergence de ce contraste limite.

Le premier point (à savoir la convergence de la log-vraisemblance renormalisée vers un contraste limite) se base le plus souvent sur le théorème ergodique (voir par exemple le livre de Billingsley (1965)). Dans notre cas, ce théorème n'est pas directement applicable car la log-vraisemblance fait intervenir une somme de quantités aléatoires dont le comportement n'est pas stationnaire (voir ci-dessous l'expression de cette log-vraisemblance).

En utilisant des conditionnements successifs, cette log-vraisemblance s'écrit sous la forme de la somme pour k variant de 0 à $n-1$ de la loi de Y_k conditionnelle au passé $Y_{k-1}, \dots Y_0$. Or en introduisant la variable X_k et en conditionnant par rapport à cette variable, nous sommes amenés à considérer la loi de X_k conditionnelle au passé des observations $Y_{k-1}, \dots Y_0$. Cette quantité est appelée le filtre de prédiction du modèle et c'est sur son étude que se base notre approche.

Formalisons cette écriture de la log-vraisemblance. Le filtre de prédiction $f^{\zeta}_{\theta,n}$ du modèle est défini de la façon suivante :

$$\forall v \in \mathbf{K}, \begin{cases} f^{\zeta}_{\theta,0}(v) & = & \zeta(v) \\ f^{\zeta}_{\theta,n+1}(v) & \triangleq & \Phi_1(Y_n, f^{\zeta}_{\theta,n}; \theta) \triangleq \frac{\int_u q_\theta(u,v) g_\theta(Y_n|u) f^{\zeta}_{\theta,n}(u) d\gamma(u)}{\int_u g_\theta(Y_n|u) f^{\zeta}_{\theta,n}(u) d\gamma(u)} \end{cases} .$$

Lorsque θ est égal à la vraie valeur θ^* du paramètre et lorsque ζ vaut la densité initiale π^* de la loi de la chaîne $\{X_n\}_{n\geq 0}$, le filtre $f^{\pi^*}_{\theta^*,n}$ représente la densité par rapport à la mesure dominante γ, de la loi de la variable cachée X_n, conditionnelle au passé des observations $Y_{n-1}, \dots, Y_0$. La log-vraisemblance des paramètres (θ, ζ) pour les observations $Y_0, \dots, Y_n$ s'écrit alors de la façon suivante :

$$\ell_n(\theta, \zeta) = \sum_{k=0}^{n-1} \log \left(\int_{\mathbb{K}} g_\theta(Y_k|x) f^{\zeta}_{\theta,k}(x) d\gamma(x) \right) .$$

Il est à noter que $f^{\zeta}_{\theta,n+1}$ est une fonction déterministe de $(Y_n, f^{\zeta}_{\theta,n})$, ce qui permet d'obtenir, en définissant $Z_n = (X_n, Y_n, f^{\zeta}_{\theta,n})$ pour tout entier n positif, une nouvelle chaîne de Markov. Il se trouve que cette nouvelle chaîne (dite chaîne étendue) est d'un intérêt tout particulier puisque la log-vraisemblance renormalisée $n^{-1}\ell_n(\theta, \zeta)$ des paramètres s'écrit comme moyenne empirique d'une certaine fonction de cette chaîne $\{Z_n\}_{n\geq 0}$.

L'étude de cette chaîne étendue (et en particulier la preuve de son ergodicité, voir Section 1.2.2) nous permet alors d'établir la convergence presque sure de la log-vraisemblance renormalisée $n^{-1}\ell_n(\theta, \zeta)$ vers un contraste limite $\ell(\theta)$ qui ne dépend pas de la loi initiale ζ de la chaîne de Markov (Proposition 1.2.6). L'étape suivante consiste à établir que ce contraste limite est maximum uniquement pour la vraie valeur du paramètre, à savoir θ^*.

Pour cela, l'existence de ce contraste en tant que limite et sans caractérisation supplémentaire pose problème. Il se trouve que dans le cas d'un espace d'états fini pour la chaîne de Markov, l'espace d'états de la chaîne étendue est localement compact, et le Théorème de représentation de Riesz s'applique et permet d'écrire le contraste limite comme l'intégrale d'une certaine fonction sous la mesure de probabilité invariante de la chaîne étendue. Lorsque l'espace d'états de la chaîne de Markov n'est plus compact, l'espace dans lequel varie le filtre n'est plus de dimension finie (auparavant il s'agissait de $\mathbb{R}^M$ où M est le nombre fini de valeurs prises par la chaîne de Markov) et le Théorème de Riesz n'est plus valable.

Pour contourner ce problème, nous utilisons une suite de mesures appropriée qui est asymptotiquement tendue, donc admet une sous-suite convergente en loi, d'après le Théorème de Prohorov (1956) (voir par exemple le livre de van der Vaart and Wellner (1996)). Ceci nous permet d'établir l'existence d'une mesure invariante pour la chaîne étendue (Sunyach 1975) et de représenter le contraste limite $\ell(\theta)$ comme l'intégrale d'une certaine fonction sous cette mesure λ_θ (Proposition 1.2.6).

D'autre part, nous utilisons le fait que la chaîne étendue $\{Z_n\}_{n\geq 0}$ qui est réellement utilisée n'est pas exactement celle que nous avons décrite ci-dessus, mais plutôt une variante de celle-ci qui prend en compte le vrai filtre de prédiction du modèle : $f^{\pi^*}_{\theta^*,n}$, où θ^* et π^* sont les vraies valeurs des paramètres θ et ζ (voir les définitions (1.3) et (1.4) de la transition et de la loi initiale de cette chaîne). Cette astuce nous permet alors de représenter les contrastes limites $\ell(\theta)$ et $\ell(\theta^*)$ sous la même mesure de probabilité λ_θ, et donc de pouvoir ainsi les comparer (voir Théorème 1.2.7). Nous pouvons alors en déduire la consistance de l'estimateur du maximum de vraisemblance (Théorème 1.2.9), via la méthode initiée par Wald (1949), en assurant l'équicontinuité de la log-vraisemblance.

Enfin nous montrons la normalité asymptotique et l'efficacité de cet estimateur avec les mêmes outils que ceux utilisés dans la preuve de sa consistance. Nous introduisons de nouvelles chaînes étendues dont les premières coordonnées sont celles de la chaîne étendue $\{Z_n\}_{n\geq 0}$ et qui intègrent les dérivées premières et secondes par rapport à θ du filtre de prédiction (Section 1.3). La preuve est classiquement basée sur trois points : prouver la convergence en loi du gradient de la log-vraisemblance renormalisée par $n^{-1/2}$ vers une loi normale centrée de variance l'inverse de l'information de Fisher (Proposition 1.3.1), établir la convergence presque sure de la matrice hessienne des dérivées secondes de la log-vraisemblance renormalisée par n^{-1} vers l'opposé de cette information de Fisher (Proposition 1.3.2), et enfin prouver l'uniformité par rapport au paramètre θ de la convergence en probabilité de cette matrice hessienne (Proposition 1.3.3).

Passage à un cadre semi-paramétrique : motivations

La question qui se pose assez naturellement est la suivante : que deviennent ces résultats si la transition de la chaîne est maintenant supposée entièrement inconnue, la fonctionnelle qui définit les observations étant encore connue à paramètre près ? Autrement dit, si nous observons une chaîne de Markov homogène de transition inconnue à travers un bruit (additif ou multiplicatif) : que pouvons nous dire de l'estimation semi-paramétrique des lois et des paramètres qui entrent en jeu ?

Précisons tout d'abord notre modèle. Nous observons les variables aléatoires $Y_n = X_n + \varepsilon_n$, pour $n \geq 0$, où la loi de la chaîne de Markov $\{X_n\}_{n\geq 0}$ est entièrement inconnue, la loi du bruit blanc $\{\varepsilon_n\}_{n\geq 0}$ est connue au paramètre de variance σ^2 près, et les deux suites $\{X_n\}_{n\geq 0}$ et $\{\varepsilon_n\}_{n\geq 0}$ sont indépendantes. Nous nous intéressons alors à l'estimation de la variance σ^2.

Le problème de l'estimation de la variance du bruit dans un modèle où le signal est observé à un bruit additif près semble assez naturel. Les travaux portant sur l'estimation de la loi du signal supposent généralement la loi du bruit totalement connue. Dans de nombreuses applications, il semble réaliste de supposer que cette variance est inconnue et de chercher à l'estimer.

La première question qui se pose est celle de l'identifiabilité de ce paramètre. Lorsque le signal est composé de variables aléatoires indépendantes et identiquement distribuées, Kiefer et Wolfowitz (1956) ont établi que l'identifiabilité du modèle est équivalente à l'hypothèse que la loi du signal ne possède pas de composante gaussienne. Plus précisément, notons Φ_σ la densité de la loi gaussienne centrée de variance σ^2. La densité g de la loi du signal ne possède pas de composante gaussienne si l'écriture $g = \Phi_\sigma * g'$ où g' est une densité et $*$ désigne le produit de convolution, implique que $\sigma = 0$ et $g = g'$ presque sûrement.

Nous avons établi un critère permettant d'assurer l'identifiabilité de la variance du bruit lorsque le signal n'est plus une suite de variables aléatoires indépendantes et identiquement distribuées mais une chaîne de Markov. Dans la suite, toutes les chaînes de Markov considérées sont supposées homogènes. Nous noterons $\mathcal{X}$ l'espace d'états (inclus dans $\mathbb{R}$) de la chaîne de Markov $\{X_n\}_{n\geq 0}$.

Hypothèse I.1. *Le noyau de transition q n'est pas trivial :*

$$\exists x_1 \neq x_2 \in \mathcal{X} \text{ tels que } q(x_1,\cdot) \neq q(x_2,\cdot)\ .$$

Sous l'Hypothèse I.1, nous dirons que la chaîne de Markov $\{X_n\}_{n\geq 0}$ de transition q est une “vraie” chaîne de Markov i.e. n'est pas une suite de variables aléatoires indépendantes.

Hypothèse I.2. *Soit q un noyau de transition sur $(\mathcal{X}, \mathcal{B}or(\mathcal{X}), \nu)$ vérifiant : pour toute fonction h intégrable sur $(\mathcal{X}, \mathcal{B}or(\mathcal{X}), \nu)$,*

$$\int_{\mathcal{X}} h(x) q(x;\cdot) d\nu(x) = 0 \quad \text{si et seulement si} \quad h = 0.$$

Cette hypothèse assure que si les chaînes $\{X_n\}_{n\geq 0}$ et $\{X'_n\}_{n\geq 0}$ ont pour transition q et pour lois initiales respectives μ_1 et μ_2 avec $\mu_1 \neq \mu_2$, alors $\{X_n\}_{n\geq 0}$ et $\{X'_n\}_{n\geq 0}$ n'ont pas la même loi.

Proposition I.1. *Soient $\{X_n\}_{n\geq 0}$ une chaîne de Markov homogène dont le noyau de transition q est inconnu et vérifie les Hypothèses I.1 et I.2, et $\{\varepsilon_n\}_{n\geq 0}$ un bruit blanc indépendant de $\{X_n\}_{n\geq 0}$, de loi normale centrée et de variance σ^2 inconnue, alors dans le modèle où l'on observe la suite de variables aléatoires $\{Y_n\}_{n\geq 0}$ données par*

$$\forall n \geq 0\ ,\quad Y_n = X_n + \varepsilon_n,$$

le paramètre σ^2 est identifiable.

Ce résultat n'est pas repris dans le corps de la thèse. Sa preuve est différée en Annexe A de l'Introduction.

L'estimation de σ^2 devient alors théoriquement possible, mais reste un problème difficile à résoudre. Nous avons donc commencé par étudier le cas où les variables aléatoires $\{X_n\}_{n\geq 0}$ sont indépendantes et identiquement distribuées (et non plus de transition markovienne). Ceci fait l'objet du second chapitre de cette thèse.

Estimation semi-paramétrique dans le modèle de convolution

État de l'art

Nous considérons donc dans ce second chapitre le modèle de convolution à variance du bruit inconnue, et nous cherchons à estimer cette variance σ_0^2 ainsi que la densité g_0 de la loi du signal non observé. Plus précisément, nous considérons la suite $Y_n = X_n + \varepsilon_n$ pour $n \geq 0$, où les suites $\{X_n\}_{n\geq 0}$ et $\{\varepsilon_n\}_{n\geq 0}$ sont indépendantes et constituées de variables aléatoires indépendantes et identiquement distribuées. Les variables X_n ont une loi de densité g_0 inconnue et les variable ε_n sont gaussiennes centrées de variance σ_0^2 inconnue.

L'hypothèse classiquement utilisée dans ce modèle est celle de la loi du bruit entièrement connue et il s'agit alors d'inférer la densité g_0 de la loi du signal, qui est supposée dans une certaine classe de régularité. Ceci fait l'objet de très nombreux travaux (voir par exemple les articles de Carroll et Hall (1988), Devroye (1989), Fan (1991c, 1991a, 1991b), Liu et Taylor (1989), Stefanski et Carroll (1990), Stefanski (1990), Zhang (1990)).

En particulier, Fan (1991c) a prouvé que l'estimation de g_0 est d'autant plus lente que la densité de la loi du bruit est régulière, un bruit Gaussien étant en ce sens très régulier. Il a considéré les espaces de régularité suivants pour la densité g_0 de la loi du signal :

$$\mathcal{C}_{m,\alpha,\beta} = \left\{ g \in \mathbb{L}_1(\mathbb{R}), g \geq 0, \int g(x)dx = 1; \forall x \in \mathbb{R}, \forall \delta > 0, |g(x+\delta) - g(x)| \leq \beta\delta^{\alpha} \right\}.$$

Lorsque le bruit est de loi gaussienne (entièrement connue), la vitesse d'estimation minimax de la densité g_0 pour le risque quadratique ponctuel et pour le risque pour la norme de $\mathbb{L}_p(\mathbb{R})$, dans ces espaces de régularité $\mathcal{C}_{m,\alpha,\beta}$, est égale à $(\log n)^{-(m+\alpha)}$, et est donc très lente.

Pensky et Vidakovic ont sensiblement amélioré cette vitesse d'estimation sous l'hypothèse supplémentaire que la densité g de la loi du signal est “super régulière”. En notant g^* la transformée de Fourier de g, nous dirons que la densité g est super-régulière lorsqu'elle appartient à l'espace suivant :

$$\mathcal{SS}_{\alpha,\nu,\rho}(A) = \left\{ g \in \mathbb{L}_1(\mathbb{R}); \int |g^*(t)|^2 (t^2+1)^{\alpha} \exp(2\rho|t|^{\nu}) dt \leq A_{\alpha} \right\},$$

où α, ν, ρ et A sont des constantes positives. Lorsque la densité g du signal appartient à cet espace de régularité $\mathcal{SS}_{\alpha,\nu,\rho}(A)$, les auteurs construisent un estimateur qui converge, pour $\nu < 2$, à une vitesse qui est du type $n^{-\eta}(\log n)^{\xi}$ où les constantes η et ξ sont strictement positives. Plus précisément, en définissant le risque quadratique intégré (MISE) de l'estimateur $\hat{g}_n$ par :

$$\text{MISE}(\hat{g}_n) = \mathbb{E} \int (\hat{g}_n(x) - g(x))^2 dx,$$

Pensky et Vidakovic (1999) établissent que

$$\sup_{g \in \mathcal{SS}_{\alpha,\nu,\rho}(A_\alpha)} \text{MISE}(\hat{g}_n) = \begin{cases} O(n^{-\eta}(\log n)^{\xi}), & \text{si} \quad \nu < 2, \\ O((\log n)^{-\alpha} \exp(-\zeta(\log n)^{\nu/2})) & \text{si} \quad \nu \geq 2, \end{cases}$$

où η, ξ et ζ ont des formes explicites (nous renvoyons au Théorème 4 de l'article de Pensky et Vidakovic (1999) pour davantage de détails).

Il est naturel de s'interroger alors sur la persistance de ces résultats lorsque la variance du bruit n'est plus connue. Nous prouvons que lorsque la variance du bruit est inconnue, la vitesse d'estimation de g_0 est alors fortement dégradée. C'est le résultat principal du Chapitre 2.

Résultats et techniques utilisées

Trois types d'hypothèses sont imposées à la densité inconnue g_0. Sous chacune de ces hypothèses, nous établissons des bornes inférieures ou des bornes supérieures pour le risque quadratique d'estimation ponctuelle de la densité g_0 ou le risque quadratique d'estimation de la variance σ_0^2. Enfin, sans aucune hypothèse sur g_0 (autre que celle assurant l'identifiabilité du modèle) nous construisons un estimateur de la variance σ_0^2 et établissons sa consistance.

Nous nous sommes tout d'abord placés dans un cadre où la densité g_0 appartient à une classe de fonctions suffisamment régulières (qui englobe les deux exemples susmentionnés $\mathcal{C}_{m,\alpha,\beta}$ et $\mathcal{SS}_{\alpha,\nu,\rho}(A)$) et ne contient pas de fonctions ayant une composante gaussienne (afin d'assurer l'identifiabilité du modèle). Notons $\mathcal{G}$ l'ensemble des densités sans composante gaussienne et considérons des espaces de fonctions $\mathcal{R}$ qui contiennent une fonction $\tilde{g}_0$ satisfaisant les hypothèses suivantes :

Hypothèse *Pour tous $\sigma > 0$, $\tau > 0$ suffisamment petit, et tout nombre réel $0 < t \leq \tau$, le produit de convolution $\tilde{g}_0 * (\Phi_{\sqrt{t}\,\sigma} 1\!\!1_{\tau|\cdot| \leq \sqrt{t}\,\sigma})$ n'a pas de composante gaussienne et appartient à l'espace $\mathcal{R}$.*

Cette condition est assez faible si on garde à l'esprit le fait que la convolution régularise les fonctions. Si l'espace $\mathcal{R}$ est un espace de fonctions régulières, il semble raisonnable d'imposer sa stabilité sous le produit de convolution.

Hypothèse *La fonction $\tilde{g}_0$ est de classe $\mathcal{C}^3$, avec $\sup_{x \in \mathbb{R}^*} |\tilde{g}_0'(x)/x| < +\infty$, $\tilde{g}_0''(0) \neq 0$ et $\|\tilde{g}_0^{(3)}\|_\infty < +\infty$.*

Sous ces hypothèses, le risque quadratique minimax pour l'estimation ponctuelle de g_0 est minoré par une constante que divise $\sqrt{\log n}$. Autrement dit, si l'ensemble de densités $\mathcal{R}$ contient une fonction $\tilde{g}_0$ qui satisfait les hypothèses précédentes, nous prouvons la minoration suivante, valable pour tout voisinage $\mathcal{V}(\sigma_0)$ de σ_0 :

$$\liminf_{n \to \infty} \inf_{\hat{T}_n} \sup_{\sigma \in \mathcal{V}(\sigma_0)} \sup_{g \in \mathcal{G} \cap \mathcal{R}} \sqrt{\log n}\, \mathbb{E}^{1/2}(\hat{T}_n - g(x_0))^2 > 0,$$

où l'infimum est pris sur tous les estimateurs $\hat{T}_n$ basés sur les observations $Y_1, \dots Y_n$. Ceci signifie en d'autres termes que la déconvolution (c'est à dire l'inférence du signal) ne peut se faire

à vitesse plus rapide que $(\log n)^{-1/2}$, autrement dit une vitesse beaucoup plus lente que celles obtenues à variance du bruit connue.
Notons que ce résultat est en fait une conséquence d'un résultat analogue sur la variance du bruit (Proposition 2.2.1). Plus précisément, sous ces hypothèses imposées sur la densité inconnue g_0, le risque minimax pour l'estimation de la variance σ_0^2 est également minoré par une constante que divise $\sqrt{\log n}$.

Nos minorations du risque minimax sont prouvées à l'aide de l'inégalité de van Trees (Gill et Levit 1995). La preuve se fait en deux étapes que nous allons décrire de façon assez informelle. La première étape consiste à exhiber une sous-famille paramétrique unidimensionnelle de lois $\{P_t\}_{|t|\leq t_n}$ pour laquelle l'estimation du paramètre $\psi(P_t)$ basée sur les observations $Y_1, \dots Y_n$ de loi P_0, se fait le plus difficilement possible. La suite $\{t_n\}_{n\geq 0}$ est une suite de nombres réels qui décroît vers 0 et qui sera choisie ultérieurement. La famille $\{P_t\}_{|t|\leq t_n}$ est construite de sorte que l'information de Fisher

$$I(t) = \mathbb{E}_t \left[\frac{\partial \log P_t(Y)}{\partial t} \right]^2 = \int \left(\frac{\partial P_t(y)}{\partial t} \right)^2 P_t(dy)$$

dans le modèle P_t est très petite pour des valeurs de t non nulles (P_t et P_0 sont des modèles très proches mais les paramètres $t \neq 0$ et 0 sont distincts).
Nous pouvons alors minorer le risque sur toute la classe par le risque sur la sous-famille :

$$\inf_{\widehat{T}_n} \sup_P \mathbb{E}_P [\widehat{T}_n - \psi(P)]^2 \geq \inf_{\widehat{T}_n} \sup_{|t|\leq t_n} \mathbb{E}_{P_t} [\widehat{T}_n - \psi(t)]^2,$$

où l'infimum est pris sur tous les estimateurs $\hat{T}_n$ basés sur les observations $Y_1, \dots Y_n$.
Nous appliquons alors l'inégalité de van Trees qui est une forme bayésienne de la borne de Cramer-Rao. En notant λ_0 une densité de probabilité sur l'intervalle $[-1;1]$ vérifiant certaines hypothèses de régularité, nous construisons $\lambda_n = t_n^{-1}\lambda_0(\cdot/t_n)$ densité de probabilité sur l'intervalle $[-t_n;t_n]$. L'information de Fisher J_n pour cette mesure est définie par :

$$J_n = \mathbb{E}_{\lambda_n} \left[\frac{\partial \log \lambda_n(t)}{\partial t} \right]^2 = \int \left(\frac{\partial \lambda_n(t)}{\partial t} \right)^2 \lambda_n(t) dt = \frac{J_0}{t_n^2}.$$

L'inégalité de van Trees donne alors la minoration suivante :

$$\inf_{\widehat{T}_n} \sup_{|t|\leq t_n} \mathbb{E}_{P_t} [\widehat{T}_n - \psi(P_t)]^2 \geq \frac{[\int (\partial\psi(P_t))/(\partial t) \lambda_n(t) dt]^2}{n \int I(t)\lambda_n(t)dt + J_0 t_n^{-2}}.$$

Le choix de la suite $\{t_n\}_{n\geq 0}$ qui se fait de façon à réaliser un compromis entre les deux termes apparaissant au dénominateur fournit alors une borne inférieure pour le risque minimax.

Nous proposons alors trois estimateurs la variance σ_0^2 suivant les hypothèses faites sur la classe de densités g_0.
Le premier de ces estimateurs est construit pour des densités qui appartiennent à un ensemble de fonctions dont la transformée de Laplace ne croît pas trop vite à l'infini. Plus précisément, définissons :

$$\mathcal{L}_{v,M} = \left\{ g \in \mathbb{L}_1(\mathbb{R}) \ / \ g \geq 0 \ , \ \int g(x)dx = 1 \ , \ \left| \log \left(\int e^{tx} g(x)dx \right) \right| \leq M t^2 v(t) \right\}.$$

Lorsque la densité du signal appartient à cet espace, le comportement de sa transformée de Laplace à l'infini permet d'identifier le paramètre de variance σ_0^2. En effet, σ_0^2 vérifie :

$$\sigma_0^2 = \lim_{t\to\infty} \frac{2}{t^2} \log \mathbb{E}(e^{tY}),$$

ce qui conduit à définir notre estimateur de la façon suivante :

$$\hat{\sigma}_{L,n}^2(t_n) = -\frac{2}{t_n^2} \log \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n} e^{it_nY_j}\right),$$

pour une suite croissante vers l'infini de points $(t_n)_{n\geq 0}$ bien choisie. Nous établissons la consistance de cet estimateur et calculons sa vitesse de convergence (Lemme 2.3.1), qui dépend explicitement du type de croissance imposé sur la transformée de Laplace de g_0 (autrement dit du choix de la fonction v). Pour tout $\Sigma > 0$, choisissons $t_n = \sqrt{\alpha \log n}/(2\Sigma)$ où $0 < \alpha < 1$, alors nous prouvons la majoration :

$$\sup_{\sigma\in[0;\Sigma]} \sup_{g\in\mathcal{L}_{v,M}} \left[\mathbb{E}_{\sigma,g}(\hat{\sigma}_{L,n}^2(t_n) - \sigma^2)\right]^{1/2} \leq \frac{8\Sigma^2}{\alpha(\log n)n^{(1-\alpha)/2}} + 2Mv(t_n).$$

Dans le cas particulier où g_0 est à support inclus dans $[-M; M]$ (avec M fixé), le résultat précédent s'applique avec $v : t \mapsto 1/t$ et nous obtenons un estimateur convergeant à la vitesse $(\log n)^{-1/2}$ (Proposition 2.3.2). Cette vitesse n'est pourtant pas minimax, car les hypothèses de régularité sous le produit de convolution qui avaient été faites ci-dessus ne sont pas valables dans ce cadre. Pourtant, nous montrons que cet estimateur est presque minimax. En effet, nous prouvons que le risque quadratique minimax est minoré par une constante que divise $\log(n)$ (autrement dit l'estimation ne peut se faire à vitesse plus rapide que $(\log n)^{-1}$). Notons $\mathcal{G}_M$ l'espace des densités à support inclus dans le compact $[-M; M]$, alors pour tout voisinage $\mathcal{V}(\sigma_0)$ de σ_0, il existe des constantes positives C_1 et C_2 telles que :

$$\frac{C_1}{\log n} \leq \liminf_{n\to\infty} \inf_{\hat{g}_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{g\in\mathcal{G}_M} \mathbb{E}^{1/2}(g(x_0) - \hat{g}_n)^2 \leq \frac{C_2}{\sqrt{\log n}},$$

où l'infimum est pris sur tous les estimateurs $\hat{g}_n$ basés sur les observations $Y_1, \ldots Y_n$.

En utilisant une méthode analogue, nous proposons un second estimateur de la variance σ_0^2 convergeant lorsque la transformée de Fourier est à décroissance contrôlée à l'infini (voir la définition (2.12) de l'ensemble $\mathcal{F}_{v,M}$), dont nous établissons la consistance. Cependant, bien que sa construction soit assez similaire à celle du précédent estimateur, le calcul de sa vitesse de convergence ne peut pas s'obtenir par les mêmes méthodes.

Enfin, sans aucune hypothèse sur g_0 autre que celle permettant d'assurer l'identifiabilité du modèle, nous proposons un estimateur convergent de σ_0 (Section 2.4). L'idée de cette construction repose sur le fait que l'identifiabilité même du modèle impose à la densité g_0 du signal de ne pas avoir de composante gaussienne. Considérons alors la fonction α qui est le produit de la transformée de Fourier de h_0 (la densité de la loi des observations) avec la fonction $(\Phi_\sigma^*)^{-1}$:

$$\forall\zeta \in \mathbb{R},\ \forall\sigma \geq 0, \quad \alpha(\zeta;\sigma) = g_0^*(\zeta)e^{-\zeta^2(\sigma_0^2-\sigma^2)/2} = h_0^*(\zeta)e^{\zeta^2\sigma^2/2}.$$

Lorsque σ est égal à la vraie valeur du paramètre σ_0, la fonction $\alpha(\cdot;\sigma_0)$ est la fonction caractéristique de la loi des variables cachées g_0^*. Remarquons alors les propriétés suivantes de α :

- lorsque $\sigma \leq \sigma_0$, la fonction $\alpha(\cdot;\sigma)$ est la transformée de Fourier de la mesure positive $g_0 * \Phi_{\sqrt{\sigma_0^2-\sigma^2}}$.

- lorsque $\sigma > \sigma_0$, la fonction $\alpha(\cdot;\sigma)$ n'est plus une transformée de Fourier car g_0 ne contient pas de composante gaussienne. Donc d'après le Théorème de Bochner (voir par exemple le livre de Feller (1971)), la fonction $\alpha(\cdot;\sigma)$ n'est plus définie positive (i.e. il existe un entier $n \geq 1$ et une famille de réels $\{t_k\}_{1\leq k\leq n}$ tels que la matrice symétrique $(\alpha(t_k - t_l;\sigma))_{1\leq k,l\leq n}$ n'est pas définie positive).

Il s'ensuit que le vrai paramètre σ_0 est la plus grande valeur de σ pour laquelle la fonction $\alpha(\cdot;\sigma)$ reste définie positive. En d'autres termes, σ_0 est la plus grande valeur de σ telle que pour tout entier n, et toute famille de nombres réels $\{t_k; 1 \leq k \leq n\}$, la plus petite valeur propre de la matrice $(\alpha(t_k - t_l;\sigma))_{1\leq k,l\leq n}$ est positive. Nous estimons la fonction α par son estimateur empirique $\hat{\alpha}_n$ et nous utilisons une famille de points $\{t_{k,n}\}_{n\geq 0}$ dense dans $\mathbb{R}$, telle que pour un entier n assez grand, les matrices $(\hat{\alpha}_n(t_{k,n}-t_{l,n};\sigma))_{k,l}$ possèdent également une plus petite valeur propre positive (à une constante ϵ_n près) tant que σ reste inférieur à σ_0.

Plus précisément, la fonction α est donc estimée par son estimateur empirique :

$$\hat{\alpha}_n(\zeta;\sigma) = \left(\frac{1}{n}\sum_{p=1}^{n} e^{i\zeta Y_p}\right) e^{\zeta^2\sigma^2/2}.$$

Considérons alors deux suites $(k_n)_{n\geq 0}$ et $(\ell_n)_{n\geq 0}$ de réels croissantes vers l'infini, de sorte que le rapport ℓ_n/k_n tende aussi vers l'infini (afin d'obtenir une famille dense dans $\mathbb{R}$). Les points $t_{k,n} = k/k_n$ forment une partition de l'intervalle $[-\ell_n/k_n; \ell_n/k_n]$ pour k variant entre $-\ell_n$ et ℓ_n. Considérons de plus une suite $\{\epsilon_n\}_{n\geq 0}$ de réels décroissante vers zéro. Nous définissons notre estimateur de σ_0 de la façon suivante :

$$\hat{\sigma}_n = \sup\left\{ s \geq 0 \ : \ \forall \sigma \leq s, \inf_{u;\ \|u\|=1} \sum_{-\ell_n\leq k,l\leq \ell_n} u_k \hat{\alpha}_n(t_{k,n} - t_{l,n};\sigma)\bar{u}_l \geq -\epsilon_n \right\}.$$

Nous prouvons alors la convergence en probabilité de cet estimateur vers la vraie valeur σ_0, ceci pour un choix adéquat des paramètres ℓ_n, k_n, ϵ_n et v_n (voir Hypothèse 2.4 et Théorème 2.4.1). Cette preuve est basée sur l'étude du processus empirique

$$\mathbb{G}_n(f) = n^{-1/2}\sum_{j=1}^{n}\left(f(Y_j) - \int f(y)h_0(y)dy\right),$$

lorsque f prend ses valeurs dans l'ensemble de fonctions $\{x \mapsto e^{itx} \ ; \ |t| \leq \ell_n/k_n\}$.

Décrivons la procédure qui permet d'obtenir un calcul effectif de cet estimateur. Considérons les matrices $T_n(\sigma) = \{\hat{\alpha}_n(t_{k,n}-t_{l,n};\sigma)\}_{-\ell_n\leq k,l\leq \ell_n}$. Le graphe de la fonction $\sigma \mapsto \lambda_{min}(T_n(\sigma))$, où $\lambda_{min}(T)$ désigne la plus petite valeur propre de la matrice T, nous permet d'obtenir la valeur de

$\hat{\sigma}_n$ en considérant la première valeur de σ pour laquelle $\lambda_{min}(T_n(\sigma)) \leq -\epsilon_n$. Les conditions que doivent satisfaire les paramètres ℓ_n, k_n, ϵ_n et v_n ne fournissent qu'un ordre de grandeur asymptotique de ces quantités. Le calibrage des constantes peut se faire par exemple par une étude de simulations, mais ceci est en dehors du cadre de cette thèse.

En conclusion, nous avons vu que l'hypothèse de connaissance de la variance du bruit, bien que peu réaliste, n'est pas facile à lever. En effet, dans les cas que nous avons considérés, l'estimation de cette variance se fait à des vitesses de convergence très lentes. De plus, la non connaissance de cette variance influe fortement sur la vitesse d'estimation de g_0. Notamment, la vitesse d'estimation de la densité g_0 est fortement dégradée lorsque la variance du bruit σ_0^2 est inconnue.

Une question qui se pose est de savoir si cette vitesse pourrait être améliorée si l'on se place dans le cadre d'un signal $\{X_n\}_{n\geq 0}$ markovien et non plus de variables indépendantes. Avec des hypothèses idoines sur le noyau de transition de la chaîne $\{X_n\}_{n\geq 0}$, il se peut que la suite d'observations $\{Y_n\}_{n\geq 0}$ comporte davantage d'information sur le modèle, et donc que l'estimation de la variance du bruit σ_0^2 se fasse à une meilleure vitesse. Cette étude fera l'objet d'un travail ultérieur.

Estimation minimax non-paramétrique de fonctionnelles dans le modèle de convolution

Motivations

Le troisième chapitre de cette thèse présente un travail réalisé en collaboration avec Marie-Luce Taupin et qui porte sur l'estimation de fonctionnelles linéaires particulières de la densité du signal, dans le modèle de convolution. Ce travail complète un certain nombre de résultats obtenus dans la thèse de Taupin (1998). En particulier, nous montrons que pour certaines de ces fonctionnelles (fonctionnelles issues d'un polynôme ou bien d'une fonction trigonométrique), l'estimateur proposé par Taupin (2001) atteint la vitesse de convergence du risque minimax pour la norme de $\mathbb{L}_p(\mathbb{R})$.

Précisons tout d'abord le modèle, les hypothèses et les notations, ainsi que la forme de ces fonctionnelles. Dans le modèle de convolution, nous disposons de n observations indépendantes satisfaisant :

$$Y_k = X_k + \varepsilon_k \ , \quad \forall\, 0 \leq k \leq n.$$

La suite $\{X_n\}_{n\geq 0}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de loi de densité g inconnue par rapport à la mesure de Lebesgue sur $\mathbb{R}$. La suite $\{\varepsilon_n\}_{n\geq 0}$ est un bruit blanc de loi normale centrée réduite dont la densité est notée Φ_1. Les suites $\{X_n\}_{n\geq 0}$ et $\{\varepsilon_n\}_{n\geq 0}$ sont indépendantes. La densité de la loi de l'observation Y_n par rapport à la mesure de Lebesgue sur $\mathbb{R}$ est le produit de convolution $g * \Phi_1$ et est notée h. On notera

$$\mathcal{H} = \left\{ g * \Phi_1 \ ; \ g \geq 0, \int g(x)dx = 1 \right\},$$

l'ensemble dans lequel varie cette densité.

Nous nous intéressons alors à l'estimation de fonctionnelles linéaires de la densité inconnue g de la forme suivante :

$$\Gamma_f(y) = \int f(x)g(x)\Phi_1(x-y)dx \quad ; \quad \forall y \in \mathbb{R},$$

en concentrant notre étude au cadre d'une fonction f polynomiale ou bien d'une fonction f trigonométrique.

La motivation pour l'étude de ces fonctionnelles découle du problème de l'estimation semi-paramétrique dans le modèle de régression avec erreurs sur les variables suivant. Considérons le modèle où les observations sont des variables aléatoires $\{(Y_n, Z_n)\}_{n\geq 0}$ satisfaisant les relations suivantes :

$$\begin{cases} Z_n &= f_\beta(X_n) + \eta_n \\ Y_n &= X_n + \varepsilon_n \end{cases} , \qquad \forall n \geq 0,$$

où la fonction f_β est connue au paramètre fini-dimensionnel β près, les erreurs $\{\eta_n\}_{n\geq 0}$ et $\{\varepsilon_n\}_{n\geq 0}$ sont des suites de variables aléatoires indépendantes, identiquement distribuées, centrées et de variances respectives σ_η^2 et $\sigma_\epsilon^2 = 1$ (pour simplifier les notations), la suite $\{\varepsilon_n\}_{n\geq 0}$ étant de plus de loi gaussienne, et la suite $\{X_n\}_{n\geq 0}$ non observée est une suite de variables aléatoires indépendantes de même loi de densité g inconnue. Enfin, les suites $\{X_n\}_{n\geq 0}$, $\{\eta_n\}_{n\geq 0}$ et $\{\varepsilon_n\}_{n\geq 0}$ sont supposées indépendantes.

Dans ce modèle de régression avec erreurs sur les variables, l'objectif est d'estimer le paramètre β en présence d'une nuisance, à savoir la densité g inconnue des variables cachées. Une méthode proposée par Taupin (2001) consiste à utiliser le critère suivant :

$$\frac{1}{n}\sum_{i=1}^{n}(Z_i - \mathbb{E}(f_\beta(X_i)|Y_i))^2,$$

dans lequel l'espérance conditionnelle $\mathbb{E}(f_\beta(X_i)|Y_i)$ est remplacée par un estimateur basé sur les observations $Y_1, \ldots Y_n$. La structure de convolution permet d'écrire cette espérance conditionnelle sous la forme :

$$\mathbb{E}(f_\beta(X_n)|Y_n = y) = \frac{\int f_\beta(x)g(x)\Phi_1(x-y)dx}{\int g(x)\Phi_1(x-y)dx} = \frac{\Gamma_{f_\beta}(y)}{h(y)}.$$

L'estimation de cette espérance conditionnelle se fait alors en estimant séparément le numérateur Γ_f et le dénominateur h, correspondant au cas particulier $f = 1$ pour Γ_f.

Il se trouve que les vitesses de convergence obtenues pour l'estimation de β dépendent souvent des vitesses de convergence ponctuelle et uniforme de l'estimateur de Γ_{f_β}. Il est donc naturel de s'intéresser au problème de l'estimation ponctuelle et de l'estimation pour la norme de $\mathbb{L}_p(\mathbb{R})$, où $1 \leq p \leq \infty$, de ces fonctionnelles. Taupin (2001) fournit un estimateur $\widehat{\Gamma}_{f,n}$ de ces fonctionnelles et calcule des majorations du risque quadratique ponctuel et du risque en norme uniforme de cet estimateur, pour certaines classes de fonctions f (espaces de Sobolev, fonctions analytiques, polynomiales, ...).

Dans le Chapitre 3, nous étendons les résultats obtenus par Taupin (2001) dans le cas de fonctionnelles issues d'un polynôme ou d'une fonction trigonométrique. Le choix particulier de ces fonctionnelles repose sur les remarques suivantes (Lemmes 3.2.1 et 3.2.2). Lorsque f est une fonction polynomiale $f : x \mapsto \sum_{j=0}^{\ell} \beta_j x^j$, où ℓ est un entier positif fixé et les points $\{\beta_j\}_{0\leq j\leq \ell}$ sont des réels, il existe une famille de polynômes $\{f_j\}_{1\leq j\leq \ell-1}$ de degré $deg(f_j) = j$ telle que

$$\Gamma_f = \beta_\ell h^{(\ell)} + \sum_{j=0}^{\ell-1} f_{\ell-j} h^{(j)},$$

où $h^{(j)}$ est la dérivée j-ème de la densité h. De plus, lorsque f est une fonction trigonométrique du type $\mathcal{C}_{\ell,\beta} : x \mapsto \sum_{j=0}^{\ell} \beta_j \cos(jx)$ où ℓ est un entier positif fixé et les points $\{\beta_j\}_{0\leq j\leq \ell}$ sont des réels, alors

$$\Gamma_f(y) = \sum_{j=0}^{\ell} \beta_j \left(e^{ijy} h(y+ij) + e^{-ijy} h(y-ij)\right)$$

(une formule analogue existe pour $\mathcal{S}_{\ell,\beta} : x \mapsto \sum_{j=0}^{\ell} \beta_j \sin(jx)$). L'estimation de la fonctionnelle Γ_f est donc, dans ces deux cas, liée au problème d'estimation de la densité h et de ses dérivées.

Contrairement au problème de déconvolution à variance du bruit inconnue abordé dans le Chapitre 2, il n'est pas nécessaire de faire des hypothèses de régularité sur la densité g inconnue dans le problème d'estimation de ces fonctionnelles. Plus précisément, la construction de l'estimateur $\widehat{\Gamma}_{f,n}$ et les calculs de vitesse ne nécessitent aucune hypothèse sur la densité g. En effet, la régularité des fonctionnelles découle de l'hypothèse d'un bruit de loi normale centrée réduite. Ainsi, le calcul des vitesses de convergence (bornes inférieures et bornes supérieures) repose sur la connaissance de la fonction f qui définit la fonctionnelle Γ_f et de la densité de la loi du bruit. L'estimateur proposé pour ces fonctionnelles est automatiquement adaptatif. De plus, tous les problèmes classiques de la déconvolution inhérents au choix de la fenêtre et à celui du noyau ne se posent pas dans notre cadre.

Résultats antérieurs

Citons plus précisément les résultats obtenus par Taupin (2001). Nous noterons $\mathcal{G}_f$ l'ensemble des fonctions, vues comme fonctionnelles linéaires de la densité g au point f, autrement dit :

$$\mathcal{G}_f = \left\{ \Gamma_f(\cdot) = \int g(x) f(x) \Phi_1(x - \cdot) dx \,;\, g \text{ densité de probabilité} \right\}.$$

Résultats de Taupin (2001)

Rappelons tout d'abord les résultats de convergence du risque quadratique ponctuel. Lorsque f est une fonction polynomiale de degré inférieur où égal à l'entier ℓ positif, Taupin établit

$$\limsup_{n\to\infty} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}} \mathbb{E}^{1/2}(\Gamma_f(y_0) - \widehat{\Gamma}_{f,n}(y_0))^2 < \infty,$$

où $\widehat{\Gamma}_{f,n}$ est l'estimateur qu'elle propose.

Dans le cas où f est une fonction de la forme $\mathcal{C}_{\ell,\beta}$ ou bien $\mathcal{S}_{\ell,\beta}$ (définies plus haut) pour un entier $\ell \geq 1$ et des réels $\{\beta_j\}_{1\leq j\leq \ell}$ fixés,

$$\limsup_{n\to\infty} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{\exp(\ell\sqrt{\log n})} \mathbb{E}^{1/2}(\Gamma_f(y_0) - \widehat{\Gamma}_{f,n}(y_0))^2 < \infty.$$

Dans le cas particulier de la densité (f=1), Taupin (1998) prouve la minoration suivante :

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{h\in\mathcal{H}} \frac{\sqrt{n}}{(\log n)^{1/4}} \mathbb{E}^{1/2}(h(y_0) - T_n)^2 > 0,$$

où l'infimum est pris sur tous les estimateurs T_n basés sur les observations $Y_1, \ldots, Y_n$.
Dans le cas de l'estimation de la densité, la vitesse de convergence en norme uniforme obtenue par Taupin est la suivante :

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}} \frac{\sqrt{n}}{(\log n)^{1/4}\sqrt{\log\log n}} \mathbb{E} \left\| h(y) - \widehat{h}_n(y) \right\|_\infty < \infty.$$

Les vitesses de convergence en norme uniforme pour les fonctionnelles polynomiales ou trigonométriques sont obtenues pour tout compact Z_C de $\mathbb{R}$. Lorsque f est une fonction polynomiale de degré inférieur où égal à l'entier $\ell \geq 1$,

$$\limsup_{n\to\infty} \sup_{\Gamma_f\in\mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{(2\ell+3)/4}} \mathbb{E} \sup_{y\in Z_C} |\Gamma_f(y) - \widehat{\Gamma}_{f,n}(y)| < \infty.$$

Dans le cas où f est une fonction de la forme $\mathcal{C}_{\ell,\beta}$ ou bien $\mathcal{S}_{\ell,\beta}$ pour un entier $\ell \geq 1$ et des réels $\{\beta_j\}_{1\leq j\leq \ell}$ fixés, la majoration suivante est valide pour tout compact Z_C de $\mathbb{R}$:

$$\limsup_{n\to\infty} \sup_{\Gamma_f\in\mathcal{G}_f} \frac{\sqrt{n}}{\exp(\ell\sqrt{\log n})(\sqrt{\log n})} \mathbb{E} \sup_{y\in Z_C} |\Gamma_f(y) - \widehat{\Gamma}_{f,n}(y)| < \infty.$$

Nouveaux résultats

Nous étendons tout d'abord les résultats concernant le risque quadratique ponctuel. Nous établissons que l'estimateur $\widehat{\Gamma}_{f,n}$ est minimax dans le cas polynômial et presque minimax (la perte est en $(\log n)^{1/4}$) dans le cas trigonométrique. Plus précisément, nous établissons que lorsque f est une fonction polynomiale de degré inférieur où égal à l'entier ℓ positif,

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Gamma_f\in\mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{\frac{2\ell+1}{4}}} \mathbb{E}^{1/2}(\Gamma_f(y_0) - T_n)^2 > 0,$$

et lorsque f est une fonction de la forme $\mathcal{C}_{\ell,\beta}$ ou de la forme $\mathcal{S}_{\ell,\beta}$ pour un entier $\ell \geq 1$ et des réels $\{\beta_j\}_{1\leq j\leq \ell}$ fixés :

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Gamma_f\in\mathcal{G}_f} \frac{\sqrt{n}(\log n)^{1/4}}{\exp(\ell\sqrt{\log n})} \mathbb{E}^{1/2}(\Gamma_f(y_0) - T_n)^2 > 0,$$

(les infima sont pris sur tous les estimateurs T_n basés sur les observations $Y_1, \ldots, Y_n$).
Ces résultats sont obtenus en appliquant l'inégalité de van Trees (Gill and Levit 1995), via une adaptation de la preuve de Taupin (1998) dans le cas de la densité. Ils sont présentés dans la Section 3.3.

Notre apport le plus conséquent concerne l'étude du risque minimax pour la norme de $\mathbb{L}_p(\mathbb{R})$ lorsque $1 \leq p \leq \infty$. Dans le cas de la densité ainsi que des fonctionnelles polynomiales, nous calculons la vitesse de convergence pour la norme de $\mathbb{L}_p(\mathbb{R})$ de l'estimateur $\widehat{\Gamma}_{f,n}$ (Section 3.4) et nous établissons que ces vitesses sont en fait optimales au sens du risque minimax (Section 3.5).

Considérons tout d'abord le cadre de la norme uniforme. Les vitesses de convergence que nous obtenons sont améliorées. Taupin obtenait des résultats en norme uniforme sur un compact fixé grâce à une méthode chaînage qui se traduisait par une perte de vitesse en $\sqrt{\log n}$ entre le risque quadratique ponctuel et le risque en norme uniforme. Notre approche (basée sur l'inégalité de Rosenthal (1970)) réduit cette perte à un facteur $\sqrt{\log\log n}$ qui est en fait la vraie perte puisque nous prouvons également que cette vitesse est minimax dans le cas polynômial. Notons également que cette approche permet de s'affranchir du cadre de la norme uniforme sur un compact fixé Z_C de $\mathbb{R}$.

Plus précisément, lorsque f est une fonction polynomiale de degré $\ell \geq 0$, nous prouvons les inégalités suivantes :

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{\frac{2\ell+1}{4}} \sqrt{\log\log n}} \mathbb{E}\|\Gamma_f - T_n\|_\infty > 0$$

et

$$\limsup_{n\to\infty} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{\frac{2\ell+1}{4}} \sqrt{\log\log n}} \mathbb{E}\|\Gamma_f - \widehat{\Gamma}_{f,n}\|_\infty < \infty.$$

(Corollaire 3.5.7 pour la borne inférieure et Théorèmes 3.4.1 et 3.4.5 pour la borne supérieure). Lorsque $\ell = 0$, on retrouve la borne supérieure déjà connue.
Lorsque f est une fonction du type $\mathcal{C}_{\ell,\beta}$ ou $\mathcal{S}_{\ell,\beta}$ pour un entier $\ell \geq 1$ et des réels $\{\beta_j\}_{1\leq j\leq \ell}$ fixés, l'estimateur $\widehat{\Gamma}_{f,n}$ est de vitesse sensiblement égale à la vitesse minimax puisque nous avons :

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}(\log n)^{3/4}}{\exp(\ell\sqrt{\log n})} \mathbb{E}\|\Gamma_f - T_n\|_\infty > 0$$

et

$$\limsup_{n\to\infty} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{\exp(\ell\sqrt{\log n})\sqrt{\log\log n}} \mathbb{E}\|\Gamma_f - \widehat{\Gamma}_{f,n}\|_\infty < \infty.$$

(Théorème 3.5.9 pour la borne inférieure et Théorème 3.4.8 pour la borne supérieure).

En ce qui concerne le risque pour la norme de $\mathbb{L}_p(\mathbb{R})$, aucun résultat n'avait été obtenu jusqu'ici. Nous calculons les vitesses de convergence de notre estimateur pour les fonctionnelles de type polynômial. Nous utilisons pour cela deux méthodes différentes. La première est basée sur les travaux de Ibragimov et Hasminskii (1983) portant sur l'estimation de densité. La seconde repose sur l'inégalité de Rosenthal (1970) avec les constantes exactes données par Pinelis (1995). Ces calculs de la vitesse sont valables pour la norme de $\mathbb{L}_p(\mathbb{R})$ où $p \geq 2$. Nous prouvons ensuite que ces vitesses sont minimax en utilisant le lemme de Fano (1952).

Plus précisément, lorsque f est une fonction polynomiale de degré $\ell \geq 0$ et pour $p \geq 2$ (Théorèmes 3.4.1 et 3.4.5)

$$\limsup_{n\to\infty} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{\frac{2\ell+1}{4}}} \mathbb{E}\|\Gamma_f - \widehat{\Gamma}_{f,n}\|_p < \infty,$$

et lorsque $p > 1$ dans le cas $\ell = 0$ et $p \geq 1$ dans le cas $\ell \geq 1$, nous prouvons (Théorèmes 3.5.1 et 3.5.7)

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{\frac{2\ell+1}{4}}} \mathbb{E}\|\Gamma_f - T_n\|_p > 0.$$

Outils utilisés

Les vitesses de convergence de l'estimateur proposé pour les fonctionnelles de la densité Γ_f, pour la norme de $\mathbb{L}_p(\mathbb{R})$, sont calculées par deux méthodes différentes. Plus précisément, le contrôle de la différence pour la norme de $\mathbb{L}_p(\mathbb{R})$ entre l'estimateur $\widehat{\Gamma}_{f,n}$ et la valeur du paramètre Γ_f se décompose, via l'inégalité triangulaire, en un contrôle d'un terme dit de biais et d'un terme correspondant au terme de variance pour l'estimation ponctuelle.

$$\mathbb{E}\|\widehat{\Gamma}_{f,n} - \Gamma_f\|_p \leq \|\Gamma_f - \mathbb{E}\widehat{\Gamma}_{f,n}\|_p + \mathbb{E}\|\widehat{\Gamma}_{f,n} - \mathbb{E}\widehat{\Gamma}_{f,n}\|_p.$$

Le contrôle du biais est traité de façon directe. Le contrôle de l'analogue du terme de variance pour la norme de $\mathbb{L}_p(\mathbb{R})$ (i.e le second terme dans l'inégalité précédente) est quant à lui traité par deux méthodes différentes.

La première utilise les résultats de Ibragimov et Hasminskii (1983) concernant l'estimation de densité dans certaines classes de régularité. Nous utilisons en particulier une généralisation du Lemme 4 de Ibragimov et Hasminskii (1983) qui permet de majorer l'analogue du terme de variance pour la norme de $\mathbb{L}_p(\mathbb{R})$ (plus précisément le terme $\mathbb{E}\|\widehat{\Gamma}_{f,n} - \mathbb{E}\widehat{\Gamma}_{f,n}\|_p$) d'un estimateur à noyau de la densité. La seconde se base sur l'inégalité de Rosenthal (1970) avec des constantes exactes données par Pinelis (1995).

Les constantes exactes sont très importantes pour traiter le cas de la norme uniforme. En effet, le contrôle du terme analogue au terme de variance se fait, pour $p = \infty$, via une inégalité énoncée dans Nikol′skii (1969) pour des fonctions entières, et qui permet de contrôler le terme $\mathbb{E}\|\widehat{\Gamma}_{f,n} - \mathbb{E}\widehat{\Gamma}_{f,n}\|_\infty$ en fonction de la norme dans $\mathbb{L}_p(\mathbb{R})$ de ce même terme. La majoration de ce terme pour les valeurs de p finies fait apparaître un facteur $\sqrt{p}$ qui, lorsque l'on s'autorise à prendre $p = \log\log n$, permet d'obtenir le contrôle souhaité sur $\mathbb{E}\|\widehat{\Gamma}_{f,n} - \mathbb{E}\widehat{\Gamma}_{f,n}\|_\infty$.

Le lemme de Fano (1952) (voir par exemple Cover et Thomas (1991)) dans sa version appliquée à la statistique par Ibragimov and Has′minskiĭ (1981) –voir aussi les articles de Birgé (1983, 2001)– permet de donner une borne inférieure du risque minimax pour la norme de $\mathbb{L}_p(\mathbb{R})$. L'application de ce lemme passe par la construction d'une famille finie de lois de probabilité dans le modèle qui sont proches (au sens de la distance de Kullback) alors que les paramètres à estimer correspondants sont à distance strictement positive.

Rappelons en quelques lignes l'utilisation de ce lemme. Nous considérons un espace métrique (Θ, d) et un ensemble de mesures de probabilité $\mathcal{P}$ indexé par Θ : $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$. À partir de l'observation de la variable aléatoire X de loi P_θ, nous cherchons à obtenir une borne inférieure pour le risque d'estimation du paramètre θ :

$$\mathrm{R}(\Theta) = \inf_{\hat{\theta}(X)} \sup_{\theta\in\Theta} \mathbb{E}_\theta(d(\theta; \hat{\theta}(X)))$$

où l'infimum est pris sur tous les estimateurs $\hat{\theta}(X)$ à valeurs dans Θ. Considérons un sous-ensemble fini Θ' de Θ de cardinal $|\Theta'| \geq 3$ et tel que pour tous paramètres (θ, θ') distincts dans Θ', la distance $d(\theta, \theta')$ est minorée par $\delta > 0$. Des calculs classiques permettent alors d'établir la minoration suivante :

$$\mathrm{R}(\Theta) \geq \frac{\delta}{2} \inf_{\hat{T}(X)} \left(1 - \inf_{\theta \in \Theta'} \mathrm{P}_\theta \left(\hat{T}(X) = \theta\right)\right)$$

où l'infimum concerne à présent les estimateurs $\hat{T}(X)$ prenant des valeurs uniquement dans le sous-ensemble Θ'. Fixons un point θ_0 dans Θ'. Nous notons $\mathcal{K}(P_\theta; P_{\theta_0})$ la distance de Kullback-Leibler entre P_θ et P_{θ_0} et nous définissons

$$\overline{\mathcal{K}} = \frac{1}{|\Theta'|} \sum_{\theta \in \Theta'} \mathcal{K}(P_\theta; P_{\theta_0}).$$

L'inégalité de Fano (1952) permet d'obtenir l'existence d'une constante universelle α telle que :

$$\mathrm{R}(\Theta) \geq \frac{\delta}{2} \left(1 - \alpha \vee \frac{\overline{\mathcal{K}}}{\log(|\Theta'| + 1)}\right).$$

L'application de ce lemme passe donc par la construction d'une famille finie Θ' de paramètres qui sont à distance minorée par un certain $\delta > 0$. Si les modèles induits $\{P_\theta\}_{\theta \in \Theta'}$ sont suffisamment proches au sens de la distance de Kullback, le terme $\overline{\mathcal{K}}/(\log(|\Theta'|+1))$ dans l'inégalité précédente peut être contrôlé, et le risque minimax d'estimation sur Θ est alors minoré par une constante que multiplie δ.

L'inégalité de van Trees (Gill et Levit 1995) permet d'établir une borne inférieure du risque minimax quadratique ponctuel. Nous avons déjà abordé lors de la présentation du second chapitre de cette thèse le principe de fonctionnement de cette inégalité.

Dans notre problème d'estimation, les bornes inférieures pour le risque minimax sont établies via le choix d'une famille finie de points (pour le lemme de Fano) ou bien via celui d'un chemin (pour l'inégalité de van Trees) qui appartiennent à l'ensemble de paramètres à estimer et qui réalisent le pire cas pour l'estimation de ces paramètres. Il faut ensuite pouvoir relier cette famille ou ce chemin aux points correspondants dans l'espace des densités. La construction des sous-familles finies ou des chemins est plus aisée dans le cas des fonctionnelles polynomiales ou trigonométriques du fait de l'écriture explicite de Γ_f en fonction de la densité h et de ses dérivées. Nous avons donc commencé par traiter ces cas. Dans un cas plus général, la relation entre Γ_f et la densité h est implicite. La construction des sous-familles finies ou des chemins pour lesquels l'estimation de Γ_f est la plus difficile est donc plus compliquée.

L'étape suivante sera de construire une borne inférieure pour l'estimation de Γ_f lorsque la fonction f appartient à une classe de régularité plus générale que celles que nous avons abordées.

Cas d'une variance de bruit inconnue pour l'estimation des fonctionnelles

Terminons enfin par quelques mots sur le devenir de ces résultats lorsque la variance du bruit σ^2 n'est plus connue égale à 1 mais inconnue. Le premier résultat notable est que la variance du bruit n'entre pas en jeu dans l'estimation de la densité h des observations. En effet, si nous observons $Y_n = X_n + \varepsilon_n$ pour tout n positif, où $\{X_n\}_{n\geq 0}$ et $\{\varepsilon_n\}_{n\geq 0}$ sont deux suites indépendantes de variables aléatoires indépendantes et identiquement distribuées, la première de loi de densité g inconnue, la seconde de loi normale centrée de variance σ^2 inconnue, alors la densité h est estimée par un estimateur à noyau qui ne dépend pas de la variance σ^2 inconnue, convergeant à la vitesse minimax $n^{-1/2}(\log n)^{1/4}$.

Cependant, l'estimation des fonctionnelles polynomiales de degré supérieur ou égal à 1 fait à présent entrer en jeu non seulement les dérivées de la densité h, mais également la variance σ^2 du bruit. En effet nous avons, par exemple pour $f : x \mapsto x$, l'égalité suivante :

$$\forall y \in \mathbb{R} \ , \quad \Gamma_f(y) = yh(y) + \sigma^2 h'(y).$$

Puisque l'estimation de cette variance σ^2 se fait à vitesse très lente (inférieure à $(\log n)^{-1/2}$ dans les cas réguliers d'après les résultats obtenus dans le Chapitre 2), l'estimation de ces fonctionnelles est donc sensiblement dégradée. Nous allons énoncer un résultat précis permettant de quantifier cette dégradation. Notons $\mathcal{G}_P(\mathcal{R})$ l'ensemble des fonctionnelles issues du polynôme P de degré inférieur ou égal à ℓ, lorsque la densité g varie dans un espace "régulier" noté $\mathcal{R}$:

$$\mathcal{G}_P(\mathcal{R}) = \left\{ \Gamma = \int P(x)\Phi_\sigma(\cdot - x)g(x)dx \ ; \ g \in \mathcal{R}, \ \sigma > 0 \right\}.$$

L'inégalité de van Trees couplée au chemin utilisé dans la preuve de la Proposition 2.2.1 permet de montrer le résultat suivant (extrait du Chapitre 2) :

Soit $\mathcal{R}$ un sous-ensemble de $\mathbb{L}_1(\mathbb{R})$ inclus dans l'ensemble des densités sans composantes gaussiennes, contenant une fonction bornée $\tilde{g}_0$ dérivable ℓ fois qui vérifie les Hypothèses 2.1 et 2.3 avec de plus $\|\tilde{g}_0^{(\ell)}\|_\infty < \infty$. Alors, pour tout y_0 fixé dans $\mathbb{R}$ pour tout entier $\ell \geq 1$ et pour toute fonction polynomiale P de degré supérieur ou égal à ℓ, nous avons :

$$\liminf_{n\to\infty} \inf_{\widehat{\Gamma}_n} \sup_{\Gamma \in \mathcal{G}_P(\mathcal{R})} \sqrt{\log n}\,\mathbb{E}^{1/2}(\widehat{\Gamma}_n - \Gamma(y_0))^2 > 0,$$

l'infimum étant pris sur tous les estimateurs $\widehat{\Gamma}_n$ basés sur les observations $Y_1, \dots Y_n$.

Nous retrouvons ici le fait que l'hypothèse de connaissance de la variance du bruit est une hypothèse difficile à lever et que sans cette hypothèse, la qualité d'estimation des fonctionnelles polynomiales est nettement dégradée.

Chapitre 1

Asymptotics of the maximum likelihood estimator for general hidden Markov models *

Résumé

Nous étudions la consistance, la normalité asymptotique et l'efficacité de l'estimateur du maximum de vraisemblance dans un modèle de Markov caché, éventuellement non stationnaire, pour lequel l'espace d'états de la chaîne cachée est séparable et compact (non nécessairement fini) et où nous supposons que le noyau de transition de la chaîne de Markov ainsi que la loi conditionnelle à la chaîne cachée des observations sont connues à un paramètre θ près appartenant à l'intérieur d'un compact Θ de $\mathbb{R}^k$. Sous l'hypothèse d'identifiabilité du modèle, nous établissons la consistance et la normalité asymptotique de l'estimateur du maximum de vraisemblance en utilisant l'oubli exponentiel du filtre de prédiction et l'ergodicité géométrique de nouvelles chaînes bien choisies.

Contents

*joint work with Randal DOUC : *Dpt. TSI / CNRS URA 820, cole Nationale Suprieure des Tlcommunications, Paris, France. E-mail :* `douc@tsi.enst.fr`

1.1 Introduction

Hidden Markov Models (HMMs) form a wide class of discrete-time stochastic processes, used in different areas such as speech recognition (Juang and Rabiner 1991), neurophysiology (Fredkin and Rice 1987), biology (Churchill 1989), econometrics (Kim, Shephard, and Chib 1998) and time series analysis (DeJong and Shephard (1995), Chan and Ledolter (1995), and see also MacDonald and Zucchini (1997) and the references therein).

Most works on maximum likelihood estimation in such models have focused on iterative numerical methods, suitable for approximating the maximum likelihood estimator. For finite hidden state space models, the pioneering contribution is due to Baum et al. (1970), who presented an early and non trivial application of the expectation-maximization principle (Dempster et al. 1977), known as the "forward-backward" procedure. The more challenging issue of hidden Markov models with continuous state space has been much studied throughout the 1990s, mostly using simulation-based approaches allowed by the recent advances of Markov chain Monte Carlo methods (Kim et al. (1998), Durbin and Koopman (1997), DeJong and Shephard (1995), Chan and Ledolter (1995)).

By contrast, the statistical issues regarding the asymptotic properties of the maximum likelihood estimator for hidden Markov models have been largely ignored until recently. Baum and Petrie (1966) have shown the consistency and asymptotic normality of the maximum likelihood estimator in the particular case where both the observed and the latent variables take only finitely many values. These results have been extended recently in series of papers by Leroux (1992), Bickel and Ritov (1996), Bickel et al. (1998)(referred to as BRR), and Bakry et al. (1997). BRR generalize the method followed by Baum and Petrie (1966) to the case where the hidden Markov process X_n takes a finite number of values, the observations belonging to a general topological space. Around the same time, LeGland and Mevel (2000b, 2000a), independently developed a different technique to prove the consistency and asymptotic normality of the maximum likelihood estimator (Mevel 1997) for hidden Markov models with finite hidden state space. The work of LeGland and Mevel (2000b, 2000a), is based on the remark that the likelihood can be expressed as an additive function of an extended Markov chain. The key of the proof by LeGland and Mevel (2000b) consists in showing that under appropriate conditions, this extended chain is geometrically ergodic, in a sense given below. This proof by LeGland and Mevel (2000b), applies even when the Hidden Markov Chain is non stationary, i.e the initial distribution of the chain does not correspond to its invariant distribution while Baum and Petrie (1966), Leroux (1992) and BRR (1998) assume stationarity (see however Bakry et al. (1997), for an extension to the non-stationary case, based on a coupling technique).

In this paper, we study the maximum likelihood estimation when the Hidden Markov Chain takes value in a topological space, assumed to be compact. The main result is the consistency of the Maximum Likelihood Estimator (MLE), under the usual condition of identifiability of the parameters. We also identify the limit of the likelihood process as the Kullback-Leibler divergence. In addition, we prove the asymptotic normality of the MLE under both the stationary and the non stationary case. Up to our best knowledge, the only contribution dealing with the maximum likelihood for Hidden Markov Models taking values in a general topological space (see Meyn and Tweedie (1993) for the definition of Markov chains in such spaces) is Jensen and Petersen (1999). The authors show the asymptotic normality (but not the consistency) of the maximum likelihood estimator when the hidden Markov chain is stationary, the technique being based on an extension of BRR (1998). The technique used in our contribution to prove the consistency and the asymptotic normality of the maximum likelihood estimator for general hidden Markov models is completely different and is essentially based on the technique developed by LeGland and Mevel (2000b, 2000a) in a finite state space. A distinctive advantage of this technique is that it makes it possible to study in a common framework the stationary and the non-stationary case. The main difficulty encountered when trying to use the "extended chain method" for hidden Markov models with a continuous state space is that some components of the extended chain lie in an infinite dimensional functional space. This is in sharp contrast with the case of a finite hidden state space in which the predictive density can always be identified with an element of $\mathbb{R}^M$, where M is the number of different possible values for X_n. Instead of the Riesz representation theorem (Mevel 1997), we investigate another approach based on the asymptotic tightness of a sequence of measures to prove the existence of an invariant measure for the extended chain (Sunyach 1975). It is then possible to extend the result of Leroux (1992) under the basic assumption that the model is identifiable (in the sense that different parameter values lead to different distributions for the stationarized version of the observed process; see the definition below), a point which was not covered in the work of LeGland and Mevel (1997) nor by Jensen and Petersen (1999). We stress that the technique developped in Leroux (1992) to prove consistency, which is based on the Kingman sub-additive theorem cannot be immediately adapted to continuous state space. For technical reasons detailed in the sequel, we restrain our attention to cases where the hidden state space can be assumed to be compact.

The rest of the paper is organized as follows: in Section 2, we first prove the geometric ergodicity of the extended chain. As a corollary, pointwise convergence of the normalized log-likelihood is obtained and its limit is identified with the generalized Kullback-Leibler divergence. The final item of Section 2 deals with the consistency of the maximum likelihood estimator. Section 3 is devoted to the asymptotic normality of the maximum likelihood estimator, using the technique of Section 2 applied to a further extended Markov chain, encompassing the gradient and the Hessian of the prediction filter.

1.2 Consistency of the maximum likelihood estimator

Let $\{X_n\}_{n=0}^{\infty}$ be a Markov Chain on a compact set $\mathbf{K} \subset \mathcal{X}$, where $\mathcal{X}$ is a separable state space, equipped with a metrizable topology, and $\mathcal{B}(\mathbf{K})$ is the associated Borel σ-field. We denote $\{Q_\theta(x, \mathcal{A}), x \in \mathbf{K}, \mathcal{A} \in \mathcal{B}(\mathbf{K})\}$, the Markov transition kernel (see Meyn and Tweedie 1993) of the chain. We also let $\{Y_n\}_{n=0}^{\infty}$ be a sequence of random variables in a topological, separable and complete set $\mathcal{Y}$, such that, conditional on $\{X_n\}_{n=0}^{\infty}$, $\{Y_n\}_{n=0}^{\infty}$ is a sequence of conditionally independent random variables, Y_n with conditional density $g_\theta(y|X_n)$ with respect to some σ-finite measure ν on the Borel σ-field $\mathcal{B}(\mathcal{Y})$. Usually, $\mathcal{X}$ and $\mathcal{Y}$ are subsets of $\mathbb{R}^s$ and $\mathbb{R}^t$ respectively, but they may also be higher dimensional spaces. Moreover, both Q_θ and g_θ depend on a parameter θ in Θ, where Θ is a compact subset of $\mathbb{R}^p$. Θ is equipped with a norm denoted by $\|.\|$. The true parameter value will be denoted $\theta^*$and is assumed to be in the interior of Θ.

Assume that each transition kernel Q_θ has a density q_θ with respect to the same σ-finite dominating measure γ on $\mathcal{X}$. For notational simplicity, it is assumed that the initial distribution of $\{X_k\}$ under θ^* has a density with respect to γ denoted by π^*. We stress that π^* is not necessarily the invariant distribution of the chain. Denote P^* the probability distribution of $\{Y_n\}_{n\geq 0}$ induced by the parameter (θ^*, π^*) and $\mathbb{E}^*$ the associated expectation. The convergence results that will be proved throughout the paper hold under P^*. In the following, for $m \leq n$, denote Y_n^m the family of random variables $(Y_n, \ldots, Y_m)$. The $\mathbf{L}_p(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$-norm will be denoted $\|.\|_p$, and C, C',... will denote unspecified finite constants which may take different values upon each appearance. Moreover, for any measurable function f on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$, denote ess sup$(f) = \inf\{M \geq 0, \gamma(\{M < |f|\}) = 0\}$ and if f is non-negative, ess inf$(f) = \sup\{M \geq 0, \gamma(\{M > f\}) = 0\}$ (with obvious conventions if those sets are empty).

As outlined in the introduction, the central idea consists in writing the log-likelihood of the observations as an additive function of a Markov Chain, comprising the observations $\{Y_n\}$, the current state X_n of the initial Markov chain and the predictive density of the state, often referred to as the filter in the HMM literature.

We define a family of probability density functions (pdf) $\{f_{\theta,n}^{\zeta}\}_{n \geq 0}$ over $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$, by the following recurrence equations

$$\forall v \in \mathbf{K}, \begin{cases} f_{\theta,0}^{\zeta}(v) &= \zeta(v) \\ f_{\theta,n+1}^{\zeta}(v) &\triangleq \Phi_1(Y_n, f_{\theta,n}^{\zeta}; \theta) \triangleq \frac{\int_u q_\theta(u,v) g_\theta(Y_n|u) f_{\theta,n}^{\zeta}(u) d\gamma(u)}{\int_u g_\theta(Y_n|u) f_{\theta,n}^{\zeta}(u) d\gamma(u)} \end{cases} \tag{1.1}$$

where ζ is an initial pdf. Note that $f_{\theta,n+1}^{\zeta}$ is a deterministic function of $(Y_n, f_{\theta,n}^{\zeta})$. If $\zeta = \pi^*$, it is easily checked that $f_{\theta^*,n}^{\pi^*}$ is the predictive density distribution, i.e. the density of the conditional distribution of X_n given Y_{n-1}^0. Otherwise, $f_{\theta,n}^{\zeta}$ is usually named the prediction filter.

We consider the following family of contrast functions

$$\ell_n(\theta, \zeta) = \sum_{m=0}^{n-1} \log \left(\int_{\mathbf{K}} g_\theta(Y_m|x) f_{\theta,m}^{\zeta}(x) d\gamma(x) \right) \tag{1.2}$$

When ζ is fixed and equal to π^*, then $\ell_n(\theta, \pi^*)$ is the log-likelihood function for the parameter θ. We show below that the choice of the initial predictive density function ζ does not affect the limiting value of $\frac{1}{n}\ell_n(\theta, \zeta)$. This parameter can thus be chosen by the user. Moreover, the parameter θ^* is the only one that can be consistently estimated. As usual, for ergodic non stationary chains, the initial distribution can not be consistently estimated.

We define the state space $\mathbf{E} \triangleq \mathbf{K} \times \mathcal{Y} \times {\mathbf{S}^+}^2$ where $\mathbf{S}^+ \triangleq \{f \in \mathbf{L}_1(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma) \ ; \ f \geq 0, \|f\|_1 = 1\}$. In the sequel, $\mathbf{S}^+$ is equipped with the topology induced by the $\mathbf{L}_1$-norm, $\mathcal{B}(\mathbf{S}^+)$ denotes the corresponding Borel σ-field. We denote $\mathcal{B}(\mathbf{E})$ the Borel σ-field on the space $\mathbf{E}$ itself, induced by the product topology. On the state space $\mathbf{E}$, define the extended Markov chain $Z_n \triangleq (X_n, Y_n, F_n, F_n^*)$ by an initial law λ, and the following transition kernel: For all (x, y, f, f^*) in $\mathbf{E}$ and $\mathcal{A}_\mathcal{K} \times \mathcal{A}_\mathcal{Y} \times \mathcal{A}_\mathcal{S} \times \mathcal{A}_{\mathcal{S}^*}$ in $\mathcal{B}(\mathbf{K}) \times \mathcal{B}(\mathcal{Y}) \times \mathcal{B}(\mathbf{S}^+) \times \mathcal{B}(\mathbf{S}^+)$,

$$\begin{aligned} &\Pi_\theta((x,y,f,f^*); \mathcal{A}_\mathcal{K} \times \mathcal{A}_\mathcal{Y} \times \mathcal{A}_\mathcal{S} \times \mathcal{A}_{\mathcal{S}^*}) \\ &\quad = P[(X_{n+1}, Y_{n+1}, F_{n+1}, F_{n+1}^*) \in \mathcal{A}_\mathcal{K} \times \mathcal{A}_\mathcal{Y} \times \mathcal{A}_\mathcal{S} \times \mathcal{A}_{\mathcal{S}^*} \mid (X_n, Y_n, F_n, F_n^*) = (x,y,f,f^*)] \\ &\qquad = \int_{\mathcal{A}_\mathcal{K} \times \mathcal{A}_\mathcal{Y}} q_{\theta^*}(x,x')\, g_{\theta^*}(y'|x')\, d\gamma(x') d\nu(y') \mathbb{1}_{\mathcal{A}_\mathcal{S}}(\Phi_1(y,f;\theta)) \mathbb{1}_{\mathcal{A}_{\mathcal{S}^*}}(\Phi_1(y,f^*;\theta^*)) \end{aligned} \tag{1.3}$$

(where $\Phi_1(y, f; \theta)$ is defined in (1.1)).
In the sequel, we will often consider the particular initial distribution $\lambda(\zeta)$ defined for each pdf ζ on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$ by

$$\lambda(\zeta)(\mathcal{A}_\mathcal{K} \times \mathcal{A}_\mathcal{Y} \times \mathcal{A}_\mathcal{S} \times \mathcal{A}_{\mathcal{S}^*}) = \int_{\mathcal{A}_\mathcal{K} \times \mathcal{A}_\mathcal{Y}} \pi^*(x) g_{\theta^*}(y|x) d\gamma(x) d\nu(y) \mathbb{1}_{\mathcal{A}_\mathcal{S}}(\zeta) \mathbb{1}_{\mathcal{A}_{\mathcal{S}^*}}(\pi^*) \tag{1.4}$$

This initial distribution generates a Markov chain $\{Z_n\}$ where the first coordinate corresponds to the Markov chain $\{X_n\}$, the second to the observations $\{Y_n\}$, the third to the prediction filter $\{f^{\zeta}_{\theta,n}\}$ and the fourth coordinate to the predictive density of X_n given Y^0_{n-1}, that is $\{f^{\pi^*}_{\theta^*,n}\}$.

Finally, we denote $P_{\theta,\lambda}$ the probability distribution of the chain $\{Z_n\}$ induced by the initial distribution λ and the transition kernel Π_θ ($P_{\theta,\lambda}$ is classically defined on $\bigvee_{n=1}^{\infty} \mathcal{B}(\mathbf{E}_n)$, where $(\mathbf{E}_n)_{n\geq 0}$ is a sequence of copies of $\mathbf{E}$).

Using this definition, and assuming that $\{Z_n\}$ has the initial distribution $\lambda(\zeta)$, the normalized log-likelihood may be expressed as

$$\frac{1}{n}\ell_n(\theta,\zeta) = \frac{1}{n}\sum_{m=0}^{n-1} h_\theta(Z_m) \quad P_{\theta,\lambda(\zeta)} - a.s. \tag{1.5}$$

$$\text{where } h_\theta(x,y,f,f^*) \triangleq \log\left(\int_u g_\theta(y|u) f(u) d\gamma(u)\right) \tag{1.6}$$

Equation (1.5) shows that the normalized log-likelihood of the observations for the parameter θ can be written as an additive functional of Z_n. Thus, following ideas of LeGland and Mevel (2000b, 2000a), geometric ergodicity of the extended Markov chain will be the key ingredient to prove the consistency of the maximum likelihood estimator.

It is clear from (1.2) that the log-likelihood is indeed a function of the first three components of Z_n which themselves form a Markovian process. Extending the chain Z_n so as to include the prediction density corresponding to the actual value θ^* of the parameter is however needed to obtain the consistency result of Section 1.2.4 (see Leroux 1992, for a closely related idea).

The following definitions and assumptions are stated for future reference. For y in $\mathcal{Y}$, define

$$\delta(y) \triangleq \sup_{\theta\in\Theta} \delta_\theta(y) \triangleq \sup_{\theta\in\Theta}\left[\frac{\operatorname{esssup}_x g_\theta(y|x)}{\operatorname{essinf}_x g_\theta(y|x)}\right] \tag{1.7}$$

$$\epsilon \triangleq \inf_{\theta\in\Theta} \epsilon_\theta \triangleq \inf_{\theta\in\Theta} \frac{\operatorname{essinf}_{x,x'} q_\theta(x,x')}{\operatorname{esssup}_{x,x'} q_\theta(x,x')} \tag{1.8}$$

Assumption 1.1. $0 < \epsilon < 1$.

For $s > 0$, define $\Delta_s \triangleq \operatorname{esssup}_x \int \sup_\theta [\delta_\theta(y)]^s g_{\theta^*}(y|x) d\nu(y)$.

Assumption 1.2. Δ_1 *is finite.*

For $s > 0$, define $\Gamma_s \triangleq \operatorname{esssup}_x \int \sup_\theta \left[k_\theta(y)\right]^s g_{\theta^*}(y|x) d\nu(y)$,
where $k_\theta(y) \triangleq \operatorname{esssup}_x |\log g_\theta(y|x)|$.

Assumption 1.3. *There exists some* $s > 1$ *such that* Γ_s *is finite.*

If assumption 1.1 holds, then $\inf_\theta \operatorname{essinf}_{x,x'} q_\theta(x,x') > 0$ (note that $\operatorname{esssup}_{x,x'} q_\theta(x,x') \geq \gamma(\mathbf{K})^{-1}$). Doeblin's condition is then satisfied for the Markov chain associated with Q_θ (see Meyn and Tweedie 1993, p.391). Moreover, the kernel Q_θ is strongly aperiodic (see Meyn and Tweedie 1993, p.118). Thus, by Theorems 16.0.2 and 16.2.4 from Meyn and Tweedie (p.384), there exists some real number $0 < \rho_0 < 1$ (uniform in θ), such that, for all θ in Θ, there exists a probability measure on $\mathcal{B}(\mathbf{K})$, having density α_θ with respect to γ, such that, for all x in $\mathbf{K}$,

$$\|Q^n_\theta(x,.) - \alpha_\theta(.)d\gamma(.)\|_{TV} \leq \rho_0^n, \tag{1.9}$$

where $\|.\|_{TV}$ is the total variation norm, and Q_θ^n is the kernel Q_θ iterated n times. Then $\alpha_\theta(.)d\gamma(.)$ is the unique invariant probability measure of the chain. Thus, for all θ in Θ, the Markov chain associated with the kernel Q_θ is uniformly ergodic.

Remark 1.1. Note that in assumption 1.1, we may relax $\epsilon \in]0,1[$ and derive the same results under the assumption

Assumption 1.1(bis) *There exists m in $\mathbb{N}^*$ such that for all θ in Θ, Q_θ is aperiodic and*

$$0 < \inf_\theta \frac{\operatorname{essinf}_{x,x'} q_\theta^m(x,x')}{\operatorname{esssup}_{x,x'} q_\theta^m(x,x')} < 1$$

where q_θ^m denotes the density of the the kernel Q_θ iterated m times.

In that case, all the properties concerning the Markov Chain associated with Q_θ are derived from properties concerning the m-skeleton (see definition in Meyn and Tweedie 1993, p.68).

1.2.1 Exponential forgetting of the prediction filter

Ergodic properties for $\{Y_n\}$ are inherited from those of $\{X_n\}$ (see for instance Leroux 1992, Celeux et al. (1993)). The point of this section is to obtain those of the extended chain $Z_n = (X_n, Y_n, F_n, F_n^*)$ from the exponential forgetting for the prediction filter. Iterating the recurrence relation (1.1) $n-m+1$ times backwards ($m \leq n$) yields $f^\zeta_{\theta,n+1} = \Phi_{n-m+1}(Y_n^m, f^\zeta_{\theta,m};\theta)$, where

$$\Phi_{n-m+1}(Y_n^m, f;\theta)(v) \triangleq \frac{\int_{u_n} q_\theta(u_n,v)g_\theta(Y_n|u_n)\ldots\int_{u_m} q_\theta(u_m,u_{m+1})g_\theta(Y_m|u_m)f(u_m)d\gamma(u_n)\ldots d\gamma(u_m)}{\int_{u_n} g_\theta(Y_n|u_n)\ldots\int_{u_m} q_\theta(u_m,u_{m+1})g_\theta(Y_m|u_m)f(u_m)d\gamma(u_n)\ldots d\gamma(u_m)}$$

Using this expression, we may show the following exponential forgetting inequality (which is a non-trivial adaptation from Mevel (1997, Proposition 2.2, part A)).

Proposition 1.2.1. *Under assumption 1.1, for all f, f' in $\mathbf{S}^+$,*

$$\|\Phi_{n-m+1}(Y_n^m, f;\theta) - \Phi_{n-m+1}(Y_n^m, f';\theta)\|_1 \leq 2{\epsilon_\theta}^{-1}\,\delta_\theta(Y_m)(1-\epsilon_\theta)^{n-m+1}\|f-f'\|_1 \qquad (1.10)$$

The proof is given in Appendix B.1. The upper-bound of the right hand side only depends on the initial conditions (Y_m, f, f'). Note that assumption 1.1 implies that ϵ_θ is lower bounded by a strictly positive ϵ. Hence, $(1-\epsilon_\theta)$ is upper-bounded by $(1-\epsilon)$ and the geometrical rate of the right-hand side of equation (1.10) can be bounded uniformly for θ in Θ, as soon as $\delta(Y_m)$ is finite.

1.2.2 Geometric ergodicity of the extended Markov Chain

In this section, it is shown that the exponential forgetting of the prediction filter implies the geometric ergodicity (in a sense to be defined) of the extended chain $\{Z_n\}$.

Definition 1.2.1. $\mathrm{Lip}(\mathbf{E})$ *is the set of real valued measurable functions h on $(\mathbf{E}, \mathcal{B}(\mathbf{E}))$ such that, for all (x,y) in $\mathbf{K}\times\mathcal{Y}$, there exists* $\mathrm{lip}(h,x,y)$ *and* $\mathrm{k}(h,x,y)$ *such that for all f_1, f_1^*, f_2, f_2^* in $\mathbf{S}^+$,*

$$\begin{aligned} |h(x,y,f_1,f_1^*) - h(x,y,f_2,f_2^*)| &\leq \mathrm{lip}(h,x,y)\left(\|f_1 - f_2\|_1 + \|f_1^* - f_2^*\|_1\right) \\ |h(x,y,f_1,f_1^*)| &\leq \mathrm{k}(h,x,y) \end{aligned}$$

and

$$\begin{cases} \mathrm{lip}(h) \triangleq \mathrm{esssup}_x \int \mathrm{lip}(h,x,y)\, g_{\theta^*}(y|x) d\nu(y) < \infty \\ \mathrm{k}(h) \triangleq \mathrm{esssup}_x \int \mathrm{k}(h,x,y)\, g_{\theta^*}(y|x) d\nu(y) < \infty \end{cases}$$

These functions are of Lipschitz type with respect to the two last components and are bounded independently from the two last components by a function which satisfies an integrability condition. We will prove in the sequel that, for all h in $\mathrm{Lip}(\mathbf{E})$ verifying some moment condition (see (1.15)), a law of large numbers holds for $\{h(Z_n)\}_{n\geq 0}$.

Proposition 1.2.2. *Under assumptions 1.1,1.2, define*

$$\rho \triangleq \max(\rho_0^{\frac{1}{2}}, (1-\epsilon)^{\frac{1}{2}}) \tag{1.11}$$

(where ϵ and ρ_0 are defined in (1.8) and (1.9)).
There exists a constant $C > 0$ such that, for all θ in Θ, for all h in $\mathrm{Lip}(\mathbf{E})$*, for all z, z' in* $\mathbf{E}$*, and for all $n \geq 1$*

$$|\Pi_\theta^n h(z) - \Pi_\theta^n h(z')| \leq C[\mathrm{lip}(h) + \mathrm{k}(h)]\, \rho^n$$

where Π_θ^n denotes the nth iterate of Π_θ and $\Pi_\theta^n h(z) = \int \Pi_\theta^n(z, dw) h(w)$.

The proof of this proposition stands in Appendix B.2. As a consequence, the extended Markov chain satisfies a geometric ergodicity property, stated in the following corollaries.

Corollary 1.2.3. *Under assumptions 1.1 and 1.2, there exists a constant $C > 0$ such that, for all θ in Θ and for all h in* $\mathrm{Lip}(\mathbf{E})$*, there exists $\Lambda_\theta(h) < \infty$ such that for all z in* $\mathbf{E}$ *and for all $n \geq 1$,*

$$|\Pi_\theta^n h(z) - \Lambda_\theta(h)| \leq C\,[\mathrm{lip}(h) + \mathrm{k}(h)]\, \frac{\rho^n}{1-\rho}, \tag{1.12}$$

This inequality may be used in the following way : for any initial probability measure λ on $(\mathbf{E}, \mathcal{B}(\mathbf{E}))$*, and for any function h in* $\mathrm{Lip}(\mathbf{E})$

$$\Lambda_\theta(h) = \lim_{n\to\infty} \mathbb{E}\mathbb{E}_{\theta,\lambda}(h(Z_n)). \tag{1.13}$$

In addition, for all h in Lip$(\mathbf{E})$*, there exists a unique solution V of the Poisson equation*

$$[I - \Pi_\theta]\, V(z) = h(z) - \Lambda_\theta(h).$$

Corollary 1.2.4. *Under assumptions 1.1 and 1.2, the kernel Π_θ admits an unique invariant probability measure λ_θ on* $(\mathbf{E}, \mathcal{B}(\mathbf{E}))$ *and for any function h in* $\mathrm{Lip}(\mathbf{E})$

$$\Lambda_\theta(h) = \int h(z) \lambda_\theta(dz). \tag{1.14}$$

The type of convergence in Corollary 1.2.3 does not coincide with the usual definition of uniform geometric ergodicity (see Meyn and Tweedie 1993, Theorem 16.0.2 p.384) since our result applies here only to some kind of Lipschitz functions on $\mathbf{E}$, whereas in the standard literature about Markov chains, the ergodicity is expressed in terms of exponential decrease to zero of the total variation norm, i.e implying that (1.12) applies to all measurable bounded functions.
The proofs are given in Appendix B.2. Note that it is easy to show that $\Lambda_\theta(h)$ is a linear form. Nevertheless, it is not straightforward to prove that there exists a measure λ_θ on $\mathbf{E}$ representing $\Lambda_\theta(h)$. This is because the Riesz representation theorem does not apply. To prove the existence of an invariant probability measure, we use a technique based on asymptotic tightness of an appropriately defined sequence of probability measures, exploiting the basic ideas of Sunyach (1975).

Remark 1.2. We will often use in the paper the following property : Let $\mathbf{E}_n$ be a sequence of copies of the space $\mathbf{E}$, $\mathcal{A}$ in $\bigvee_{n\geq 0}\mathcal{B}(\mathbf{E}_n)$ and Υ a real valued function on $\prod_{n\geq 0}\mathbf{E}_n$, measurable with respect to $\bigvee_{n\geq 0}\mathcal{B}(\mathbf{E}_n)$. Note that, due to the initial distribution $\lambda(\zeta)$, (F_n, F_n^*) is $P_{\theta,\lambda(\zeta)}$-a.s equal to $(f^{\zeta}_{\theta,n}, f^{\pi^*}_{\theta^*,n})$ for all n in $\mathbb{N}$. It implies that

$$\Upsilon\left[\{Z_n\}_{n\geq 0}\right] = \Upsilon\left[\left\{(X_n, Y_n, f^{\zeta}_{\theta,n}, f^{\pi^*}_{\theta^*,n})\right\}_{n\geq 0}\right]\ P_{\theta,\lambda(\zeta)} - a.s$$

Thus,

$$P_{\theta,\lambda(\zeta)}(\Upsilon(\{Z_n\}_{n\geq 0}) \in \mathcal{A}) = P^*\left[\Upsilon(\{(X_n, Y_n, f^{\zeta}_{\theta,n}, f^{\pi^*}_{\theta^*,n})\}_{n\geq 0}) \in \mathcal{A}\right]$$

since the marginal distribution of $P_{\theta,\lambda(\zeta)}$ on events only depending on $\{(X_n, Y_n)\}_{n\geq 0}$ is equal to P^*.

Proposition 1.2.5. *Under assumptions 1.1 and 1.2, for any θ in Θ, for any pdf ζ on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$, and for any function h in* $\mathrm{Lip}(\mathbf{E})$ *such that there exists $s > 1$ satisfying*

$$\operatorname*{esssup}_x \int \mathrm{k}^s(h, x, y) g_{\theta^*}(y|x) d\nu(y) < \infty \tag{1.15}$$

the following convergence results hold

i) $\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} h(Z_m) = \Lambda_\theta(h) \quad P_{\theta,\lambda(\zeta)} - a.s$

ii) $\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} h(X_m, Y_m, f^{\zeta}_{\theta,m}, f^{\pi^*}_{\theta^*,m}) = \Lambda_\theta(h) \quad P^* - a.s$

The proof of this proposition (see Appendix B.3) is based on the ergodicity of $\{Z_n\}$ and more precisely, on the existence of a solution to the Poisson equation stated in Corollary 1.2.3 using a classical martingale decomposition method (the proof is analogous to Meyn and Tweedie 1993, Theorem 17.4.3). Note that the statement ii) is a direct consequence of i) using the remark 2.

1.2.3 Pointwise convergence

The following proposition ensures that, under some regularity conditions, the normalized log-likelihood of the observations converges almost everywhere to a finite function of θ, and identifies the limit function. It is a straightforward application of Proposition 1.2.5.

Proposition 1.2.6. *Under assumptions 1.1 to 1.3, for any θ in Θ, there exists a finite value $\ell(\theta)$ such that for any initial pdf ζ on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$,*

$$\lim_{n\to\infty} \frac{1}{n}\ell_n(\theta,\zeta) = \ell(\theta) \ \ P^* - a.s$$

Moreover,

$$\ell(\theta) = \int \log\left[\int g_\theta(y|u)f(u)d\gamma(u)\right]\lambda_\theta(dx, dy, df, df^*) \tag{1.16}$$

Note that the limiting value $\ell(\theta)$ does not depend on the initial distribution $\lambda(\zeta)$. This means that the limiting value of $\ell_n(\theta,\zeta)/n$ does not depend on the initial distribution of the chain $\{X_n\}$ (which does not necessarily coincide with the stationary distribution of the chain, in contrast with BRR and Jensen and Petersen (1999)) nor on the initial value of the prediction filter ζ, which may be chosen arbitrarily. The limiting value $\ell(\theta)$ is the integral of h_θ with respect to the invariant measure of $\{Z_n\}$.

1.2.4 Consistency of the maximum likelihood estimator

In order to prove the consistency of the maximum likelihood estimator, we need to check that the limit of the normalized log-likelihood is maximized only at θ^*, the true value of the parameter, i.e $\ell(\theta) \leq \ell(\theta^*)$ with equality if and only if $\theta = \theta^*$. Define

$$g_{\theta,n+1}(y_n^0|X_0) \triangleq g_\theta(y_0|X_0)\int q_\theta(X_0,u_1)g_\theta(y_1|u_1)\ldots \\ \ldots q_\theta(u_{n-1},u_n)g_\theta(y_n|u_n)d\gamma(u_1)\ldots d\gamma(u_n) \tag{1.17}$$

the density under the parameter θ of $\{Y_n^0\}$ conditionally to X_0 with respect to the measure $\nu^{\otimes(n+1)}$.

We assume that

Assumption 1.4.

$$\int_{\mathbf{K}\times\mathcal{A}_0\times\ldots\times\mathcal{A}_p} g_{\theta,p+1}(y_0,\ldots,y_p|u_0)d\alpha_\theta(u_0)d\nu(y_0)\ldots d\nu(y_p) = \\ \int_{\mathbf{K}\times\mathcal{A}_0\times\ldots\times\mathcal{A}_p} g_{\theta',p+1}(y_0,\ldots,y_p|u_0)d\alpha_{\theta'}(u_0)d\nu(y_0)\ldots d\nu(y_p) \tag{1.18}$$

holds for all non negative p and all Borel sets $\mathcal{A}_i$ in $\mathcal{B}(\mathcal{Y})$, $0 \leq i \leq p$ if and only if $\theta = \theta'$.

(Recall that α_β is the invariant measure for the kernel q_β).
This condition is equivalent to :
Assumption 1.4(bis)

$$\int q_\theta(u,u_1)...q_\theta(u_{p-1},u_p)g_\theta(y_0|u)...g_\theta(y_p|u_p)d\alpha_\theta(u)d\gamma(u_1)...d\gamma(u_p) = \\ \int q_{\theta'}(u,u_1)...q_{\theta'}(u_{p-1},u_p)g_{\theta'}(y_0|u)...g_{\theta'}(y_p|u_p)d\alpha_{\theta'}(u)d\gamma(u_1)...d\gamma(u_p)$$

almost everywhere with respect to the measure $\nu^{\otimes(p+1)}$*, for all non negative* p*, if and only if* $\theta = \theta'$*.*

Note that when the state space $\mathbf{K}$ is finite, assumption 1.4 is implied by condition 2 in Leroux (1992). In fact, assumption 1.4 is the minimal assumption guaranteeing identifiability for the model.

Remark 1.3. In many models of interest, the parameter itself is identifiable, only up to a permutation of states. It is most often possible to reparametrize the model to avoid this type of degeneracy. When it is not possible, one should define a slightly extended notion of identifiability, based on an equivalence relation e.g. $\theta \equiv \theta'$ iff θ is equal to θ' up to a permutation of its components. The results presented below can be straightforwardly adapted in such situation.

Theorem 1.2.7. *Under assumptions 1.1 to 1.4,* $\ell(\theta) \leq \ell(\theta^*)$ *and* $\ell(\theta) = \ell(\theta^*)$ *iff* $\theta = \theta^*$*.*

The key point in the proof is that the difference $\ell(\theta^*) - \ell(\theta)$ can be written as the expected difference between two functions under the same measure. This is possible, since the fourth coordinate of Z_n is the prediction filter associated with the true value of the parameter θ^*. This is the main motivation for extending (X_n, Y_n, F_n) to (X_n, Y_n, F_n, F_n^*).

Proof. (Theorem 1.2.7) The first part of the proof follows an idea from Leroux (1992, Lemma 6), but the conclusion has to be adapted to cope with general state space. We first express $\ell(\theta)$ in a more tractable way.

Using Corollary 1.2.3,

$$2\ell(\theta) = \lim_n \mathbb{E}_{\theta,\lambda(\zeta)} \left(h_\theta(Z_n) + h_\theta(Z_{n+1})\right)$$

By direct application of the recurrence relation between $f^\zeta_{\theta,n}$ and $f^\zeta_{\theta,n+1}$, we have

$$\begin{aligned} h_\theta(Z_n) + h_\theta(Z_{n+1}) &= \log\left[\int g_{\theta,2}(Y_n, Y_{n+1}|u) f^\zeta_{\theta,n}(u) d\gamma(u)\right] \quad P_{\theta,\lambda(\zeta)} - a.s \\ &\triangleq H_\theta(Y_n, Y_{n+1}, f^\zeta_{\theta,n}) \end{aligned}$$

(where $g_{\theta,2}$ is defined by (1.17)).
Then, using the remark 2,

$$2\ell(\theta) = \lim_n \mathbb{E}^*\left(\mathbb{E}^*\left(H_\theta(Y_n, Y_{n+1}, f^\zeta_{\theta,n}) \middle| Y^0_{n-1}\right)\right) \tag{1.19}$$

Denote $G_\theta(x, y, f, f^*) \triangleq \int H_\theta(y_0, y_1, f) g_{\theta^*,2}(y_0, y_1|u) f^*(u) d\nu(y_0) d\nu(y_1) d\gamma(u)$. It is easy to check that G_θ is in Lip($\mathbf{E}$). Thus, by Corollary 1.2.3,

$$\Lambda_\theta(G_\theta) = \lim_n \mathbb{E}_{\theta,\lambda(\zeta)}(G_\theta(Z_n)) \tag{1.20}$$

Using the equality

$$\begin{aligned} \mathbb{E}^*[H_\theta(Y_n, Y_{n+1}, f^\zeta_{\theta,n})|Y^0_{n-1}] &= \mathbb{E}^*\left(\mathbb{E}^*\left[H_\theta(Y_n, Y_{n+1}, f^\zeta_{\theta,n}) \middle| X_n, Y^0_{n-1}\right] \middle| Y^0_{n-1}\right) \\ &= G_\theta(X_n, Y_n, f^\zeta_{\theta,n}, f^{\pi^*}_{\theta^*,n}) \end{aligned}$$

combined with (1.19) and (1.20), it yields that

$$\begin{aligned} 2\ell(\theta) &= \Lambda_\theta(G_\theta) \\ &= \int \log\left[\int g_{\theta,2}(y_0,y_1|u_0)f(u_0)d\gamma(u_0)\right] \\ &\quad \times g_{\theta^*,2}(y_0,y_1|u)f^*(u)d\nu(y_0)d\nu(y_1)d\gamma(u)\lambda_\theta(\mathbf{K},\mathcal{Y},df,df^*) \end{aligned}$$

Define the function $h^*(x,y,f,f^*) \triangleq \log(\int g_{\theta^*}(y|u)f^*(u)d\gamma(u))$. Replacing the function h_θ by h^*, it is easily checked in the same way that

$$2\ell(\theta^*) = \int \log\left[\int g_{\theta^*,2}(y_0,y_1|u_0)f^*(u_0)d\gamma(u_0)\right] \\ \times g_{\theta^*,2}(y_0,y_1|u)f^*(u)d\nu(y_0)d\nu(y_1)d\gamma(u)\lambda_\theta(\mathbf{K},\mathcal{Y},\mathbf{S}^+,df^*)$$

Hence,

$$\begin{aligned} 2(\ell(\theta^*)-\ell(\theta)) &= \int \left[\int g_{\theta^*,2}(y_0,y_1|u)f^*(u)d\gamma(u)\right] \times \log\left(\frac{\int g_{\theta^*,2}(y_0,y_1|u)f^*(u)d\gamma(u)}{\int g_{\theta,2}(y_0,y_1|u)f(u)d\gamma(u)}\right) \\ &\quad d\nu(y_0)d\nu(y_1)\lambda_\theta(\mathbf{K},\mathcal{Y},df,df^*) \end{aligned}$$

which is the integral of a Kullback-Leibler divergence and is thus non-negative. We now show that the equality $\ell(\theta)=\ell(\theta^*)$ implies that $\theta=\theta^*$ under the assumption of identifiability 1.4. If $\ell(\theta)=\ell(\theta^*)$, the Kullback-Leibler divergence is null and thus,

$$\int g_{\theta^*,2}(y_0,y_1|u)f^*(u)d\gamma(u) = \int g_{\theta,2}(y_0,y_1|u)f(u)d\gamma(u) \tag{1.21}$$

almost everywhere with respect to the measure $d\nu(y_0)d\nu(y_1)\lambda_\theta(\mathbf{K},\mathcal{Y},df,df^*)$. This technique can be easily generalized computing the quantities $(m+1)\ell_n(\theta,\zeta)$ for each integer m, such that $\ell(\theta)=\ell(\theta^*)$ implies almost everywhere with respect to the measure $\lambda_\theta(\mathbf{K},\mathcal{Y},df,df^*)$,

$$\forall m \geq 1,\ \int g_{\theta^*,m+1}(y_0,\ldots,y_m|u)f^*(u)d\gamma(u) = \int g_{\theta,m+1}(y_0,\ldots,y_m|u)f(u)d\gamma(u)\ , \\ \nu(dy_0)\otimes\ldots\otimes\nu(dy_m) - a.e \tag{1.22}$$

Now, let us prove that the condition of assumption 1.4 holds. Let (f_0,f_0^*) be in $\mathbf{S}^+\times\mathbf{S}^+$ such that the property (1.22) is satisfied. Let $p\leq m$ and h be any bounded, continuous function on $\mathcal{Y}^{p+1}$. Multiplying both sides of the equation by $h(y_{m-p},\ldots y_m)$, and then taking the integral with respect to $d\nu(y_0)\ldots d\nu(y_m)$, yields

$$\int F_\theta(u_{m-p})q_\theta^{m-p}(u_0,u_{m-p})f_0(u_0)d\gamma(u_0)d\gamma(u_{m-p}) = \\ \int F_{\theta^*}(u_{m-p})q_{\theta^*}^{m-p}(u_0,u_{m-p})f_0^*(u_0)d\gamma(u_0)d\gamma(u_{m-p}) \tag{1.23}$$

where for any β,

$$F_\beta(u) \triangleq \int h(y_{m-p},...,y_m)g_{\beta,p+1}(y_{m-p},...,y_m|u)d\nu(y_{m-p})\ldots d\nu(y_m) \tag{1.24}$$

which is continuous and bounded by $\operatorname{esssup}_{u\in\mathcal{Y}^{p+1}} |h(u)|$ on $\mathbf{K}$. Now, using Equation (1.9), we have, for any u in $\mathbf{K}$,

$$\left|\int F_\theta(u')q_\theta^{m-p}(u,u')d\gamma(u') - \int F_\theta(u')d\alpha_\theta(u')\right| \leq \rho_0^{m-p} \times \operatorname*{esssup}_{u\in\mathcal{Y}^{p+1}} |h(u)|$$

This implies, by dominated convergence

$$\lim_{m\to\infty} \int F_\theta(u')q_\theta^{m-p}(u,u')f_0(u)d\gamma(u)d\gamma(u') = \int F_\theta(u')d\alpha_\theta(u')$$

This result also holds with the parameter θ^* instead of θ. Now, combining with (1.23) yields

$$\int F_\theta(u')d\alpha_\theta(u') = \int F_{\theta^*}(u')d\alpha_{\theta^*}(u')$$

Replacing F_θ and F_{θ^*} by their definition in (1.24) and knowing that h is any bounded continuous function in $\mathcal{Y}^{p+1}$,

$$\int g_{\theta,p+1}(y_0,\dots,y_p|u_0)d\alpha_\theta(u_0) = \int g_{\theta',p+1}(y_0,\dots,y_p|u_0)d\alpha_{\theta'}(u_0)$$

almost everywhere with respect to the measure $d\nu(y_0)...d\nu(y_p)$, for any p in $\mathbb{N}$. Thus $\theta = \theta^*$ by the identifiability assumption 1.4. This completes the proof. □

Let $\hat{\theta}_n(\zeta) \triangleq \operatorname{argmax}_\theta \ell_n(\theta,\zeta)$ be the maximum likelihood estimator for a fixed initial pdf ζ (following assumptions will assure its existence). It has been proved in Proposition 1.2.6 and Theorem 1.2.7 that the normalized log-likelihood for a fixed θ converges almost surely under P^* to $\ell(\theta)$ which is maximum at the point θ^*. Under additional assumptions on the regularity of some functions with respect to θ, uniformity (in a sense to be defined) in the convergence of the normalized log-likelihood may be shown, and thus, consistency can be stated, using the basic strategy invented by Wald (1949).

For all y in $\mathcal{Y}$ and $\eta > 0$, denote $\omega^g(y,\eta) \triangleq \sup_{\|\theta-\theta'\|\leq\eta} \operatorname{esssup}_x |g_\theta(y|x) - g_{\theta'}(y|x)|$. Let $\eta > 0$ and θ, θ' in Θ such that $\|\theta-\theta'\| < \eta$. Then, for all pdf ζ on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$,

$$\begin{aligned}
\ell_n(\theta,\zeta) - \ell_n(\theta',\zeta) &= \sum_{m=0}^{n-1} \log\left(1 + \frac{\int \left[g_\theta(Y_m|u)f^\zeta_{\theta,m}(u) - g_{\theta'}(Y_m|u)f^\zeta_{\theta',m}(u)\right] d\gamma(u)}{\int g_{\theta'}(Y_m|u)f^\zeta_{\theta',m}(u)d\gamma(u)}\right) \\
&\leq \sum_{m=0}^{n-1} \left(\frac{\int |g_\theta(Y_m|u) - g_{\theta'}(Y_m|u)| f^\zeta_{\theta,m}(u)d\gamma(u)}{\int g_{\theta'}(Y_m|u)f^\zeta_{\theta',m}(u)d\gamma(u)}\right. \\
&\qquad\qquad \left. + \frac{\int g_{\theta'}(Y_m|u)|f^\zeta_{\theta',m}(u) - f^\zeta_{\theta,m}(u)|d\gamma(u)}{\int g_{\theta'}(Y_m|u)f^\zeta_{\theta',m}(u)d\gamma(u)}\right) \\
|\ell_n(\theta,\zeta) - \ell_n(\theta',\zeta)| &\leq 2\sum_{m=0}^{n-1} \left(\frac{\omega^g(Y_m,\eta)}{\inf_\theta \operatorname{essinf}_x g_\theta(Y_m|x)} + \delta(Y_m)\|f^\zeta_{\theta,m} - f^\zeta_{\theta',m}\|_1\right) \qquad (1.25)
\end{aligned}$$

Define for all y in $\mathcal{Y}$, for all s,t in $\mathbb{N}$ and for all $\eta > 0$,

$$\begin{aligned}
\delta'(y,\eta) &\triangleq \left[\frac{\omega^g(y,\eta)}{\inf_\theta \operatorname{essinf}_x g_\theta(y|x)}\right] && (1.26)\\
\Delta'_s(\eta) &\triangleq \operatorname*{esssup}_x \int [\delta'(y,\eta)]^s g_{\theta^*}(y|x) d\nu(y) && (1.27)\\
(\Delta\Delta')_{s,t}(\eta) &\triangleq \operatorname*{esssup}_x \int [\delta(y)]^s [\delta'(y,\eta)]^t g_{\theta^*}(y|x) d\nu(y) && (1.28)\\
\omega^q(\eta) &\triangleq \sup_{\|\theta-\theta'\|\le\eta} \int \operatorname*{esssup}_u |q_\theta(u,v) - q_{\theta'}(u,v)| d\gamma(v) && (1.29)
\end{aligned}$$

Assumption 1.5. *For all y in $\mathcal{Y}$,*

$$\begin{aligned}
\lim_{\eta\to 0} \omega^q(\eta) &= 0\\
\lim_{\eta\to 0} \omega^g(y,\eta) &= 0
\end{aligned}$$

Assumption 1.6. *Δ_2 is finite and for some positive η_0, $\Delta'_1(\eta_0)$ and $(\Delta\Delta')_{1,1}(\eta_0)$ are finite.*

We prove in Appendix B.3 the following lemma

Lemma 1.2.8. *Under assumption 1.1,*

$$\sup_{\zeta\in\mathbf{S}^+} \|f^\zeta_{\theta,m} - f^\zeta_{\theta',m}\|_1 \le 2\epsilon^{-1} \sum_{k=0}^{m-1} \delta(Y_{k+1})\rho^{m-k}[\omega^q(\eta) + 2\delta'(Y_k,\eta)] + [\omega^q(\eta) + 2\delta'(Y_m,\eta)] \quad (1.30)$$

Combining (1.25) with this lemma, we prove under assumption 1.5 that $\theta \mapsto \ell_n(\theta,\zeta)$ is continuous. The continuity of $\theta \mapsto \ell(\theta)$ is then straightforward : the family of functions $(\Pi^n_\theta h_\theta(z))_{n\ge 0}$ converges uniformly in θ to $\ell(\theta)$ as we have

$$|\Pi^n_\theta h_\theta(z) - \Lambda_\theta(h_\theta)| \le C\,[\Delta_1 + \Gamma_1]\,\frac{\rho^n}{1-\rho}.$$

Moreover :

$$\Pi^n_\theta h_\theta(z) = \int \log\left(\int g_\theta(y_n|u) f^\zeta_{\theta,n}(u) d\gamma(u)\right) q_{\theta^*}(x,x_1)\cdots q_{\theta^*}(x_{n-1},x_n)$$
$$\times\, g_{\theta^*}(y_1|x_1)\cdots g_{\theta^*}(y_n|x_n) d\gamma(x_1)\cdots d\gamma(x_n) d\nu(y_1)\cdots d\nu(y_n)$$

The functions $\theta \mapsto \log(\int g_\theta(y_n|u) f^\zeta_{\theta,n}(u)d\gamma(u))$ are continuous (same argument as for $\ell_n(\theta,\zeta)$) and uniformly integrable as Γ_s is finite for some $s > 1$. Then we obtain that $\theta \mapsto \Pi^n_\theta h_\theta(z)$ is continuous, and so is $\theta \mapsto \ell(\theta)$.

Now, denote $\omega^\ell_n(\eta) \triangleq \sup\{\frac{1}{n}|\ell_n(\theta,\zeta) - \ell_n(\theta',\zeta)|, \|\theta-\theta'\| < \eta\}$. Combining (1.25), (1.30) and assumptions 1.5 and 1.6, we obtain that, P^*-a.s,

$$\lim_k \limsup_n \left[\omega^\ell_n\left(\frac{1}{k}\right)\right] = 0 \quad (1.31)$$

Theorem 1.2.9. *Under assumptions 1.1 to 1.6, for all initial pdf ζ,*

$$\hat{\theta}_n(\zeta) \underset{n\to\infty}{\longrightarrow} \theta^* \quad \textit{in } P^* - \textit{probability.}$$

Proof. Let $\mathcal{C}$ a compact set not including θ^*. Since $\theta \mapsto \ell(\theta)$ is a continuous function, there exists $\epsilon > 0$ such that, for all $\theta \in \mathcal{C}$, $\ell(\theta) < \ell(\theta^*) - 2\epsilon$. By (1.31), there exists P^*-a.s finite random integers k_1 and N_1 such that for all $n \geq N_1$, $\omega_n^\ell\left(\frac{1}{k_1}\right) \leq \frac{\epsilon}{2}$. Let $\{\theta_i\}_{1\leq i\leq M}$ an $\frac{1}{k_1}$-net, covering $\mathcal{C}$. Then, there exists a P^*-a.s finite random integer N_2 such that for all $n \geq N_2$, and all $1 \leq i \leq N_2$, $\frac{1}{n}\ell(\theta_i, \zeta) \leq \ell(\theta_i) + \frac{\epsilon}{2}$.
Now, let $\theta \in \mathcal{C}$, there exists a point θ_i such that $\|\theta - \theta_i\| \leq \frac{1}{k_1}$. Then, for all $n \geq \max(N_1, N_2)$,

$$\begin{aligned}
\frac{1}{n}\ell_n(\theta,\zeta) &\leq \frac{1}{n}\ell_n(\theta_i,\zeta) + \omega_n^\ell\left(\frac{1}{k_1}\right) \\
&\leq \ell(\theta_i) + \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&\leq \ell(\theta^*) - \epsilon
\end{aligned}$$

Thus, there exists a P^*-a.s finite integer N such that for all $n \geq N$, $\sup_{\theta\in\mathcal{C}} \ell_n(\theta,\zeta) < \ell_n(\theta^*, \zeta)$. The proof is completed. ∎

1.2.5 Examples

Let us give some illustrations.
Example 1 : Markov chain in additive noise.
Define: $Y_t = h(X_t) + B_t$ where $\{B_t\}$ is a sequence of independent and identically distributed centered Gaussian random variables, with variance σ^2, $\{X_t\}$ is a Markov chain on a compact set of $\mathbb{R}^d$, independent of $\{B_t\}$, and with transition Q_α. The parameter here is $\theta = (\alpha, \sigma^2)$.
Let ϕ_t be the Log-Laplace function of the marginal distributions of $\{Y_t\}$.

$$\phi_t(u) = \log\left(\mathbb{E}(e^{uY_t})\right) = \log\mathbb{E}(e^{uh(X_t)}) + \log\left(e^{\frac{u^2\sigma^2}{2}}\right)$$

$$\text{so that} \quad \frac{2}{u^2}\phi_t(u) = \sigma^2 + \frac{2}{u^2}\log\left(\mathbb{E}(e^{uh(X_t)})\right) \underset{u\to\infty}{\longrightarrow} \sigma^2$$

σ^2 is thus identifiable. If different parameters α lead to different stationary distributions for the process $\{h(X_t)\}$, then assumption 1.4 holds.
Such models are used in the following applications:

- Underwater acoustics (Quinn 1996), where $h(x) = A\cos(2\pi x)$, and A is a constant.
- Speech processing, count data analysis... (Pagano 1974), where $h(x) = x$ and $\{X_t\}$ is an autoregressive process of first order, with bounded innovations, or more generally an ARMA process. In this last case, just notice that Theorems 1 and 2 may be generalized when assumption 1.1 holds only for a power of the transition Q_α, that is when assumption 1.1(bis) holds.

Example 2 : Markov chain in multiplicative noise.
Define $Y_t = B_t.h(X_t)$ where $\{B_t\}$ is a process of independent and identically distributed, centered Gaussian random variables, with variance σ^2, $\{X_t\}$ is a Markov chain on a compact subset of $\mathbb{R}$, independent of $\{B_t\}$, and h is any measurable function. Those models are used in different areas (see Kim et al. (1998), about econometrics).

1.3 Asymptotic normality of the maximum likelihood estimator

The asymptotic normality of the maximum likelihood estimator involves a second order Taylor expansion of the log-likelihood. We shall use similar arguments to what was done with the extended chain Z_n, but we will now consider Markov chains involving both the gradient and the Hessian of the prediction filter. Those chains have an exponential forgetting of the initial condition and then we prove that they are uniformly ergodic. Denote $\partial_k()$ the first order partial derivative with respect to the component θ_k of θ, and ∇, ∇^2 the Jacobian and the Hessian operator with respect to the parameter θ.

Assumption 1.7. *(a) For any (x, x', y) in $\mathbf{K} \times \mathbf{K} \times \mathcal{Y}$, the functions $\theta \mapsto g_\theta(y|x)$ and $\theta \mapsto q_\theta(x, x')$ are differentiable.*

Define for all y in $\mathcal{Y}$,

$$\delta'(y) \triangleq \sup_{\theta,\theta'} \max_{1\le k\le p} \left[\frac{\operatorname{esssup}_x |\partial_k g_\theta(y|x)|}{\operatorname{essinf}_x g_{\theta'}(y|x)} \right] \tag{1.32}$$

and, for all s, t in $\mathbb{N}$,

$$\Delta'_s \triangleq \operatorname*{esssup}_x \int [\delta'(y)]^s g_{\theta^*}(y|x) d\nu(y) \tag{1.33}$$

$$(\Delta\Delta')_{s,t} \triangleq \operatorname*{esssup}_x \int [\delta(y)]^s [\delta'(y)]^t g_{\theta^*}(y|x) d\nu(y) \tag{1.34}$$

Assumption 1.7. *(b) Moreover,*

$$\sup_\theta \operatorname*{esssup}_{u,v} q_\theta(u, v) < \infty \,, \qquad \max_{1\le k\le p} \sup_\theta \operatorname*{esssup}_{u,v} |\partial_k q_\theta(u, v)| < \infty.$$

For some $s > 1$, Δ'_{2s}, Δ_4, $(\Delta\Delta')_{4,1}$ and $(\Delta\Delta')_{3,2}$ are finite.

The first step consists in differentiating the recurrence relation between $f^\zeta_{\theta,n}$ and $f^\zeta_{\theta,n+1}$ (which is possible under assumption 1.7). Then, $\{\nabla f^\zeta_{\theta,n}\}$ appears as a functional auto-regressive chain. For all $1 \le k \le p$,

$$\partial_k f^\zeta_{\theta,n+1}(v) = \int_u a_\theta(Y_n, f^\zeta_{\theta,n})(u, v)\, \partial_k f^\zeta_{\theta,n}(u) d\gamma(u) + U_{\theta,k}(Y_n, f^\zeta_{\theta,n})(v) \tag{1.35}$$

$$\text{where } a_\theta(y, f)(u, v) \triangleq \left(q_\theta(u, v) - \frac{\int q_\theta(s, v) g_\theta(y|s) f(s) d\gamma(s)}{\int g_\theta(y|s) f(s) d\gamma(s)} \right) \frac{g_\theta(y|u)}{\int g_\theta(y|s) f(s) d\gamma(s)}$$

$$U_{\theta,k}(y, f)(v) \triangleq \frac{\int \partial_k q_\theta(u, v) g_\theta(y|u) f(u) d\gamma(u)}{\int g_\theta(y|u) f(u) d\gamma(u)}$$

$$+ \int \left(q_\theta(u, v) - \frac{\int q_\theta(s, v) g_\theta(y|s) f(s) d\gamma(s)}{\int g_\theta(y|s) f(s) d\gamma(s)} \right) \times \frac{\partial_k g_\theta(y|u) f(u)}{\int g_\theta(y|s) f(s) d\gamma(s)} d\gamma(u)$$

Denote U_θ the vector function $(U_{\theta,k})_{1\le k\le p}$.

We consider the space $\mathbf{E'} = \mathbf{K} \times \mathcal{Y} \times \mathbf{S}^+ \times \mathbf{\Sigma}$, where

$$\mathbf{\Sigma} \triangleq \{\sigma = (\sigma_k)_{1\le k\le p}; \sigma_k \in \mathbf{L}_1(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma); \int_u \sigma_k(u) d\gamma(u) = 0\}.$$

The space $\mathbf{\Sigma}$ is equipped with the topology induced by the L_1-norm, and denote $\mathcal{B}(\mathbf{E'})$ the Borel σ-field induced by the product topology on $\mathbf{E'}$. Denote for all (y, f, σ) in $\mathcal{Y} \times \mathbf{S}^+ \times \mathbf{\Sigma}$, θ in Θ, and v in $\mathbf{K}$

$$\Psi_1(y, f, \sigma; \theta)(v) \triangleq \int a_\theta(y, f)(u, v)\sigma(u)d\gamma(u) + U_\theta(y, f)(v)$$

(Note that $\Psi_1 = (\Psi_1^k)_{1\leq k\leq p}$ is a vector in $\mathbb{R}^p$).
We define an extended Markov chain $\tilde{Z}_n = (X_n, Y_n, F_n, G_n)$ on $(\mathbf{E'}, \mathcal{B}(\mathbf{E'}))$ by its initial law $\tilde{\lambda}$ and the transition kernel $\tilde{\Pi}_\theta$ given by, for all (x, y, f, σ) in $\mathbf{E'}$ and $\mathcal{A}_\mathcal{K} \times \mathcal{A}_\mathcal{Y} \times \mathcal{A}_\mathcal{S} \times \mathcal{A}_\pm$ in $\mathcal{B}(\mathbf{K}) \times \mathcal{B}(\mathcal{Y}) \times \mathcal{B}(\mathbf{S}^+) \times \mathcal{B}(\mathbf{\Sigma})$,

$$\begin{aligned}
&\tilde{\Pi}_\theta((x, y, f, \sigma); \mathcal{A}_\mathcal{K} \times \mathcal{A}_\mathcal{Y} \times \mathcal{A}_\mathcal{S} \times \mathcal{A}_\pm)\\
&\qquad = P(X_{n+1} \in \mathcal{A}_\mathcal{K}, Y_{n+1} \in \mathcal{A}_\mathcal{Y}, F_{n+1} \in \mathcal{A}_\mathcal{S}, G_{n+1} \in \mathcal{A}_\pm \\
&\qquad\qquad \mid X_n = x, Y_n = y, F_n = f, G_n = \sigma)\\
&\qquad = \int_{\mathcal{A}_\mathcal{K}\times\mathcal{A}_\mathcal{Y}} q_{\theta^*}(x, x')\, g_{\theta^*}(y'|x')\, d\gamma(x')d\nu(y')\mathbb{1}_{\mathcal{A}_\mathcal{S}}(\Phi_1(y, f; \theta))\mathbb{1}_{\mathcal{A}_\pm}(\Psi_1(y, f, \sigma; \theta))
\end{aligned}$$

In the sequel, we will often consider the particular initial distribution $\tilde{\lambda}(\zeta)$ defined for each pdf ζ on $(\mathbf{K}, \mathcal{B}(\mathbf{K}), \gamma)$ by

$$\tilde{\lambda}(\zeta)(\mathcal{A}_\mathcal{K} \times \mathcal{A}_\mathcal{Y} \times \mathcal{A}_\mathcal{S} \times \mathcal{A}_\pm) = \int_{\mathcal{A}_\mathcal{K}\times\mathcal{A}_\mathcal{Y}} \pi^*(x)g_{\theta^*}(y|x)d\gamma(x)d\nu(y)\mathbb{1}_{\mathcal{A}_\mathcal{S}}(\zeta)\mathbb{1}_{\mathcal{A}_\pm}(0) \tag{1.36}$$

We denote $\tilde{P}_{\theta,\tilde{\lambda}(\zeta)}$ the law of $\{\tilde{Z}_n\}$ induced by $\tilde{\lambda}(\zeta)$ and $\tilde{\Pi}_\theta$ and $\tilde{\mathbb{E}}_{\theta,\tilde{\lambda}(\zeta)}$ the corresponding expectation. The exponential ergodicity of the extended chain $\{\tilde{Z}_n\}$ (similar to the result obtained in Proposition 1.2.1) and convergence of normalized sums of some functions of $\{\tilde{Z}_n\}$, are stated in Appendix D.
These properties, combined with the results of Propositions 1.3.2 and 1.3.3, ensure the convergence of the gradient of the normalized log-likelihood.

Assumption 1.8. *For any (x, x', y) in $\mathbf{K} \times \mathbf{K} \times \mathcal{Y}$, the functions $\theta \mapsto g_\theta(y|x)$ and $\theta \mapsto q_\theta(x, x')$ are twice differentiable. We also assume that:*
$\sup_\theta \max_{1\leq k,l\leq p} \operatorname{esssup}_{u,v} |\partial^2_{k,l} q_\theta(u, v)| < \infty$.

Define for all y in $\mathcal{Y}$, and for all s, t, r in $\mathbb{N}$,

$$\begin{aligned}
\delta''(y) &\triangleq \sup_{\theta,\theta'} \max_{1\leq k,l\leq p} \left[\frac{\operatorname{esssup}_x |\partial^2_{k,l} g_\theta(y|x)|}{\operatorname{essinf}_x g_{\theta'}(y|x)}\right]\\
\Delta''_s &\triangleq \operatorname*{esssup}_x \int [\delta''(y)]^s g_{\theta^*}(y|x)d\nu(y)\\
(\Delta\Delta'')_{s,t} &\triangleq \operatorname*{esssup}_x \int [\delta(y)]^s [\delta''(y)]^t g_{\theta^*}(y|x)d\nu(y)\\
(\Delta\Delta'\Delta'')_{s,t,r} &\triangleq \operatorname*{esssup}_x \int [\delta(y)]^s [\delta'(y)]^t [\delta''(y)]^r g_{\theta^*}(y|x)d\nu(y)
\end{aligned}$$

Assumption 1.9. *There exists some $s > 1$ such that Δ''_s is finite. Moreover, the following quantities Δ_{11}, $(\Delta\Delta')_{11,1}$, $(\Delta\Delta')_{10,2}$, $(\Delta\Delta')_{9,3}$, $(\Delta\Delta')_{8,4}$, $(\Delta\Delta'')_{7,1}$, $(\Delta\Delta'\Delta'')_{7,1,1}$ and $(\Delta\Delta'\Delta'')_{6,2,1}$ are supposed to be finite.*

Proposition 1.3.1. *Under assumptions 1.1, 1.7 to 1.9 and the hypothesis that the maximum likelihood estimator is consistent (see Theorem 1.2.9), there exists a matrix J_{θ^*} such that*

$$n^{-\frac{1}{2}}\nabla\ell_n(\theta^*,\zeta) \longrightarrow \mathcal{N}(0,J_{\theta^*}) \quad \textit{weakly under } P^*.$$

Moreover, $J_{\theta^*} = \lim_{n\to\infty}(1/n)\mathbb{E}^*(\nabla\ell_n(\theta^*,\zeta)^t.\nabla\ell_n(\theta^*,\zeta))$.

This result extends Theorem 3.1 in Jensen and Petersen (1999) to the non-stationary case, and its proof stands in Appendix D.1. The second order Taylor expansion of the log-likelihood involves also its Hessian matrix. Differentiating equation (1.35), we obtain $\{\nabla^2 f^{\zeta}_{\theta,n}\}$ as a functional autoregressive chain.

$$\nabla^2 f^{\zeta}_{\theta,n+1} \triangleq \alpha_1(Y_n, f^{\zeta}_{\theta,n}, \nabla f^{\zeta}_{\theta,n}, \nabla^2 f^{\zeta}_{\theta,n};\theta)$$

As we have done with Z_n and $\tilde{Z}_n$, we consider now $\check{Z}_n = (X_n, Y_n, F_n, G_n, H_n)$ an extended Markov chain on a space **E"**defined by $\mathbf{E"} = \mathbf{K}\times\mathcal{Y}\times\mathbf{S}^+\times\mathbf{\Sigma}\times\mathbf{\Sigma_2}$, where

$$\mathbf{\Sigma_2} = \{\tau = (\tau_{k,l}); \tau_{k,l} \in \mathbf{L}_1(\mathbf{K},\mathcal{B}(\mathbf{K}),\gamma) \text{ and } \forall v\in\mathbf{K}, \tau(v) \text{ symetric and positive definite}\}.$$

(The matrix $M = (M_{i,j})_{1\le i,j\le n}$ is said to be positive definite if for all $(a_i)_{1\le i\le n}$ in $\mathbb{R}$, we have $\sum_{1\le i,j\le n} a_i M_{i,j} a_j \ge 0$, with equality if and only if $a_i = 0$ for all $1\le i\le n$).
The space $\mathbf{\Sigma_2}$ is equipped with the topology induced by the L_1-norm and $\mathcal{B}(\mathbf{E"})$ will denote the Borel σ-field on **E"** induced by the product topology. The Markov chain $\check{Z}_n$ is defined by its initial law $\check{\lambda}$ and its transition kernel $\check{\Pi}_\theta$: for all (x,y,f,σ,τ) in **E"** and $\mathcal{A}_{\mathcal{K}}\times\mathcal{A}_{\mathcal{Y}}\times\mathcal{A}_{\mathcal{S}}\times\mathcal{A}_{\pm}\times\mathcal{A}_{\pm'}$ in $\mathcal{B}(\mathbf{K})\times\mathcal{B}(\mathcal{Y})\times\mathcal{B}(\mathbf{S}^+)\times\mathcal{B}(\mathbf{\Sigma})\times\mathcal{B}(\mathbf{\Sigma_2})$,

$$\begin{aligned}
&\check{\Pi}_\theta[(x,y,f,\sigma,\tau);\mathcal{A}_{\mathcal{K}}\times\mathcal{A}_{\mathcal{Y}}\times\mathcal{A}_{\mathcal{S}}\times\mathcal{A}_{\pm}\times\mathcal{A}_{\pm'}] \\
&\quad = P[(X_{n+1},Y_{n+1},F_{n+1},G_{n+1},H_{n+1}) \in \mathcal{A}_{\mathcal{K}}\times\mathcal{A}_{\mathcal{Y}}\times\mathcal{A}_{\mathcal{S}}\times\mathcal{A}_{\pm}\times\mathcal{A}_{\pm'} \\
&\qquad\qquad \mid (X_n,Y_n,F_n,G_n,H_n) = (x,y,f,\sigma,\tau)] \\
&\quad = \int_{\mathcal{A}_{\mathcal{K}}\times\mathcal{A}_{\mathcal{Y}}} q_{\theta^*}(x,x')\, g_{\theta^*}(y'|x')\, d\gamma(x')d\nu(y') \\
&\qquad\qquad \times \mathbb{1}_{\mathcal{A}_{\mathcal{S}}}(\Phi_1(y,f;\theta))\mathbb{1}_{\mathcal{A}_{\pm}}(\Psi_1(y,f,\sigma;\theta))\mathbb{1}_{\mathcal{A}_{\pm'}}(\alpha_1(y,f,\sigma,\tau;\theta)).
\end{aligned}$$

We will consider the particular initial distribution $\check{\lambda}(\zeta)$ defined for each pdf ζ on the space $(\mathbf{K},\mathcal{B}(\mathbf{K}),\gamma)$ by

$$\begin{aligned}
\check{\lambda}(\zeta)(\mathcal{A}_{\mathcal{K}}\times\mathcal{A}_{\mathcal{Y}}\times\mathcal{A}_{\mathcal{S}}\times\mathcal{A}_{\pm}\times\mathcal{A}_{\pm'}) = \int_{\mathcal{A}_{\mathcal{K}}\times\mathcal{A}_{\mathcal{Y}}} \pi^*(x) g_{\theta^*}(y|x) d\gamma(x) d\nu(y) \\
\times \mathbb{1}_{\mathcal{A}_{\mathcal{S}}}(\zeta)\mathbb{1}_{\mathcal{A}_{\pm}}(0)\mathbb{1}_{\mathcal{A}_{\pm'}}(0).
\end{aligned}$$

We denote $\check{P}_{\theta,\check{\lambda}(\zeta)}$ the law of $\{\check{Z}_n\}_n$ induced by $\check{\lambda}(\zeta)$ and $\check{\Pi}_\theta$, and $\check{\mathbb{E}}_{\theta,\check{\lambda}(\zeta)}$ the corresponding expectation. In Appendix B.4.3, we prove the exponential forgetting of the initial condition for this chain, its ergodicity, and the convergence of normalized sums of some functions of this chain. Then we are able to prove the convergence of the Hessian of the normalized log-likelihood.

Proposition 1.3.2. *Under assumptions 1.1 and 1.7 to 1.9, there exists a matrix I_θ such that*

$$\frac{1}{n}\nabla^2\ell_n(\theta,\zeta) \longrightarrow -I_\theta \quad P^* - a.s$$

Moreover, I_{θ^*} is the Fisher information matrix, that is

$$I_{\theta^*} = -\lim_{n\to\infty}\frac{1}{n}\mathbb{E}^*(\nabla^2\ell_n(\theta^*,\zeta)) = \lim_{n\to\infty}\frac{1}{n}\mathbb{E}^*(\nabla\ell_n(\theta^*,\zeta)^t\nabla\ell_n(\theta^*,\zeta)) = J_{\theta^*}$$

Denote for all $\eta > 0$ and y in $\mathcal{Y}$,

$$\begin{aligned}
\omega^{q'}(\eta) &\triangleq \sup_{\|\theta-\theta^*\|<\eta}\max_{1\le k\le p}\int \operatorname{esssup}_u |\partial_k q_\theta(u,v) - \partial_k q_{\theta^*}(u,v)|d\gamma(v)\\
\omega^{q''}(\eta) &\triangleq \sup_{\|\theta-\theta^*\|<\eta}\max_{1\le k,l\le p}\int \operatorname{esssup}_u |\partial^2_{k,l} q_\theta(u,v) - \partial^2_{k,l} q_{\theta^*}(u,v)|d\gamma(v)\\
\omega^{g''}(y,\eta) &\triangleq \sup_{\|\theta-\theta^*\|<\eta}\max_{1\le k,l\le p}\operatorname{esssup}_u |\partial^2_{k,l} g_\theta(y|u) - \partial^2_{k,l} g_{\theta^*}(y|u)|\\
\Delta'''(\eta) &\triangleq \int \frac{\omega^{g''}(y,\eta)}{\inf_\theta \operatorname{esssup}_x g_\theta(y|x)} g_{\theta^*}(y|x)d\nu(y)
\end{aligned}$$

Assumption 1.10. *Assume that*

$$\lim_{\eta\to 0}\omega^{q'}(\eta) = \lim_{\eta\to 0}\omega^{q''}(\eta) = 0$$

Moreover, there exists some $\eta_0 > 0$ such that $\Delta'''(\eta_0)$ is finite.

We finally prove

Proposition 1.3.3. *Under assumptions 1.1, 1.5, 1.6, 1.9 and 1.10, we have for any k,l in $\{1\cdots p\}$,*

$$\lim_{\eta\to 0}\mathbb{E}^*\left(\sup_{\|\theta-\theta^*\|\le\eta}\frac{1}{n}|\partial^2_{k,l}\ell_n(\theta,\zeta) - \partial^2_{k,l}\ell_n(\theta^*,\zeta)|\right) = 0$$

Remark 1.4. This proposition ensures that $\nabla\left[\lim_{n\to\infty}\frac{1}{n}\nabla\ell_n(\theta,\zeta)\right]_{\theta=\theta^*} = -I_{\theta^*}$ and that $\nabla\left[\Lambda_\theta(h_\theta)\right]_{\theta=\theta^*} = 0$.

These properties are sufficient to get the asymptotic normality of the maximum likelihood estimator, using the following assumption.

Assumption 1.11. *I_{θ^*} is an invertible matrix.*

Theorem 1.3.4. *Under assumptions 1.7 to 1.11, and the hypothesis that $\hat\theta_n(\zeta)$ is consistent,*

$$n^{\frac{1}{2}}(\hat\theta_n - \theta^*) \longrightarrow \mathcal{N}(0, I^{-1}_{\theta^*}) \quad \textit{weakly under } P^*.$$

Remark 1.5. Combining Propositions 1.3.2 and 1.3.3 with Theorem 1.3.4, we obtain a confidence interval for θ^*. Moreover, we also have the local asymptotic normality (LAN) conditions for our model, and the maximum likelihood estimator is locally asymptotically minimax (LAM) (see Bickel et al. (1998)).

Chapitre 2

Semiparametric deconvolution with unknown variance

Résumé

Ce travail porte sur l'estimation semi-paramétrique dans un modèle de convolution, où le bruit suit une loi Gaussienne centrée de variance σ^2 inconnue. Les modèles non-paramétriques de convolution, où la loi du bruit est entièrement connue, ont été largement étudiés, et il s'avère que la vitesse d'estimation de la densité g du signal est d'autant plus lente que la loi du bruit est "régulière" (Fan 1991c). Cependant, la régularité imposée sur g permet d'améliorer ces vitesses d'estimation. Dans cet article, nous montrons que lorsque la loi du bruit (qui est supposée Gaussienne centrée) possède une variance σ^2 inconnue (ce qui en pratique est toujours le cas), les vitesses d'estimation de la densité g du signal sont dégradées par rapport au cas où la loi du bruit est entièrement connue. Plus précisemment, se restreindre à des densités pour le signal régulières n'améliore pas la vitesse d'estimation de g qui est toujours plus lente que $(\log n)^{-1}$. Une conséquence en est la détérioration des vitesses d'estimation des paramètres dans un modèle de régression non-linéaire avec erreurs sur les variables. Nous construisons alors divers estimateurs de la variance, dont un consistant dès que le signal a un moment d'ordre 1 fini.

Contents

2.1 Introduction and main result

Consider a deconvolution model $Y = X + \varepsilon$, where X has an unknown density g with respect to the Lebesgue measure on $\mathbb{R}$, and the error measurement ε is supposed to be Gaussian, centered, with variance σ^2, and independent of X. Deconvolution density estimation has been studied in depth by several authors. Recent related work include Carroll and Hall (1988), Devroye (1989), Fan (1991a, 1991b, 1991c), Liu and Taylor (1989), Stefanski and Carroll (1990), Stefanski (1990), Zhang (1990). When σ^2 is known, Fan (1991c) proved that for all fixed point x_0 in $\mathbb{R}$, $g(x_0)$ can be approximated at the optimal rate of convergence $(\log n)^{-(m+\alpha)/2}$, when g is supposed to belong to the set

$$\mathcal{C}_{m,\alpha,\beta} = \{g \in \mathbb{L}_1(\mathbb{R}) : g \geq 0; \text{ and } \forall x \in \mathbb{R},\ \forall \delta > 0,\ |g^{(m)}(x) - g^{(m)}(x+\delta)| \leq \beta\delta^\alpha\} \quad (2.1)$$

where m in $\mathbb{N}$, $\beta > 0$ and $0 < \alpha \leq 1$ are known constants. More precisely,

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{g\in\mathcal{C}_{m,\alpha,\beta}} (\log n)^{(m+\alpha)} \mathbb{E}[\hat{T}_n - g(x_0)]^2 > 0,$$

and this rate of convergence is attained. Results about convergence in $\mathbb{L}_p(\mathbb{R})$-norm were also obtained in (Fan 1991b) and (Fan 1993), where it is proved:

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{g\in\mathcal{C}_{m,\alpha,\beta}} (\log n)^{(m+\alpha)/2} \mathbb{E}\|\hat{T}_n - g\|_p > 0,$$

and this rate is attained. This result is consequently improved by Pensky and Vidakovic (1999), when the density g is super-smooth. Denote by g^* the Fourier transform of g, and assume that the density g belongs to the set

$$\mathcal{SS}_{\alpha,\nu,\rho}(A) = \left\{ g \in \mathbb{L}_1(\mathbb{R}) : \int |g^*(t)|^2 (t^2+1)^\alpha \exp(2\rho|t|^\nu) dt \leq A \right\}, \quad (2.2)$$

for some positive constants α, ν, ρ and A. Pensky and Vidakovic (1999) constructed an estimator $\hat{g}_n$ of g, whose mean-square integrated error

$$\mathrm{MISE}(\hat{g}_n) = \mathbb{E} \int (\hat{g}_n(x) - g(x))^2 dx \quad (2.3)$$

satisfies

$$\sup_{g\in\mathcal{SS}_{\alpha,\nu,\rho}(A)} \mathrm{MISE}(\hat{g}_n) = \begin{cases} O(n^{-\eta}(\log n)^\xi), & \text{if } \nu < 2, \\ O((\log n)^{-\alpha} \exp(-\zeta(\log n)^{\nu/2})) & \text{if } \nu \geq 2, \end{cases}$$

where η, ξ and ζ have explicit forms. So that assuming the density of the signal super-smooth ensures faster rates of convergence, in the case of an entirely known noise.

The question that naturally arises is what happens when σ^2 is unknown? This problem becomes a particular case of a semiparametric model (van der Vaart 1998), and more precisely, of mixture models (Lindsay 1986), known as the normal mean mixture model. This problem of measurements being contaminated with errors is used in many different areas such as physics or biology (practical problems of deconvolution can be found in Medgyessy (1973)).

Semiparametric mixture models are studied by Ishwaran (1999). The author shows a loss of information for the finite-dimensional parameter when the model is constrained to allow for only discrete mixtures. In the normal mean mixture model, allowing discrete mixtures to have limit points leads to a breakdown of the classical $\sqrt{n}$ inference for the finite-dimensional parameter. Can we be more precise in quantifying this breakdown? The answer is yes, and we will see that for regular mixtures, the estimation of the finite-dimensional parameter σ happens at a slower rate than $(\log n)^{-1}$.

Assuming that the error density is perfectly known seems to be unrealistic in many practical applications. Neumann (1997) gives a scheme to estimate σ^2 when observations of the noise sequence are available. Our estimator of σ^2 is based only on the observations of the convolution model. More precisely, we assume that we observe the random variables:

$$Y_n = X_n + \varepsilon_n \; ; \;\; n \in \mathbb{N},$$

where $\{X_n\}_{n\geq 0}$ is a sequence of independent and identically distributed random variables on $\mathbb{R}$, and $\{\varepsilon_n\}_{n\geq 0}$ is a centered and normally distributed white noise. The sequences $\{X_n\}_{n\geq 0}$ and $\{\varepsilon_n\}_{n\geq 0}$ are supposed to be independent. We denote by σ^2 the variance of the ε-sequence, by g (resp. h) the density of the distribution of X (resp. Y) with respect to the Lebesgue measure on $\mathbb{R}$. In this paper, Φ_σ denotes the density of a Gaussian centered random variable, with variance equal to σ^2, and the notation $g * \Phi_\sigma$ stands for the convolution product between g and Φ_σ. A density function g (with respect to the Lebesgue measure on $\mathbb{R}$) is said to have "no Gaussian component" if the equality $g = g' * \Phi_\sigma$ where g' is a density function on $\mathbb{R}$ implies $\sigma = 0$ (and then $g = g'$). We consider:

$$\mathcal{G} = \{g \in \mathbb{L}_1(\mathbb{R}) \;:\; g \text{ density without Gaussian component}\},$$

the set of densities for the signal, with respect to the Lebesgue measure on $\mathbb{R}$. Note that the restriction on g enables us to identify the parameter σ^2. Denote also

$$\mathcal{H} = \{g * \Phi_\sigma \;:\; g \in \mathcal{G} \text{ and } \sigma > 0\},$$

the set of densities for the observed sequence, with respect to the Lebesgue measure on $\mathbb{R}$. In this paper, the notation $\mathbb{E}_{\sigma,g}$ or the expression *under* (σ, g) stand when the distribution of the observations is supposed to have density $g * \Phi_\sigma$.

When σ^2 is unknown, we prove that pointwise estimation of g is deteriorated in many cases. In fact, the minimax quadratic risk over a set of regular densities, for the estimation of σ^2 is lower bounded by a constant divided by $\log n$. Then, the estimation of $g(x_0)$ (for some fixed point x_0 in $\mathbb{R}$) will not be possible, in regular cases, at a faster than $(\log n)^{-1}$ rate. This main result is stated in the following theorem. We consider a set of functions $\mathcal{R}$ containing a bounded function g_0 which satisfies the following assumptions.

Assumption 2.1. *For all $\sigma > 0$, for all $\tau > 0$ small enough, and all real numbers $0 < t \leq \tau^2$, the convolution product $g_0 * (\Phi_{\sqrt{t}\sigma} 1\!\!1_{\tau|\cdot|\leq\sqrt{t}\sigma})$ has no Gaussian component and belongs to $\mathcal{R}$.*

This condition is rather weak as convolution regularizes the functions so that if $\mathcal{R}$ is some "regular" space it is most likely stable under the convolution product. Moreover, g_0 can always be chosen such that this convolution has no Gaussian component.

Assumption 2.2. *The function* $\alpha_0(y) = \int_{-1}^{1} g_0(y+u)du$ *satisfies* $\int [\alpha_0(y)]^{-1} e^{-y^2/2} dy < \infty$.

This assumption is satisfied for example with $g_0(y) = \pi^{-1}/(1+y^2)$.

Assumption 2.3. *The function* g_0 *is three times continuously differentiable with* $\sup_{x\in\mathbb{R}^*} \left|\frac{g_0'(x)}{x}\right| < +\infty$, $g_0''(0) \neq 0$ *and* $\|g_0^{(3)}\|_\infty < +\infty$.

In Section 2.2, we give examples of such sets of functions. This Section also contains our main theorem:

Theorem 2.2 *Assume that* $\mathcal{R}$ *is a subset of* $\mathbb{L}_1(\mathbb{R})$ *containing a bounded function* g_0 *which satisfies Assumptions 2.1, 2.2 and 2.3, then for all real number* x_0*, for all* $\sigma_0 > 0$*, and every neighborhood* $\mathcal{V}(\sigma_0)$ *of* σ_0*, we have*

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{g\in\mathcal{G}\cap\mathcal{R}} (\log n) \left[\mathbb{E}_{\sigma,g}(\hat{T}_n - g(x_0))^2\right]^{1/2} > 0,$$

where the infimum is taken over all estimators $\hat{T}_n$ *based on the observations* $Y_1, \ldots, Y_n$.

A consequence of this result concerns the non-linear errors-in-variables model. Consider the model where the observations $\{(Y_n; Z_n)\}_{n\geq 0}$ satisfy the following relations:

$$\left\{ \begin{array}{lcl} Z_n & = & f_\beta(X_n) + \eta_n \\ Y_n & = & X_n + \varepsilon_n \end{array} \right. , \qquad \forall\, n \geq 0, \tag{2.4}$$

where the function f_β is known up to the finite dimensional parameter β, the errors $\{(\eta_n; \varepsilon_n)\}_{n\geq 0}$ are independent, identically distributed and centered with respective variances σ_η^2 and $\sigma_\epsilon^2 = 1$, the variables ε_n being Gaussian and the sequence $\{X_n\}_{n\geq 0}$ is not observed and is a sequence of independent and identically distributed random variables with distribution admitting a density g with respect to the Lebesgue measure on $\mathbb{R}$. The purpose is to estimate the parameter β in this model where g is considered as a nuisance. Taupin (2001) constructed an estimator of β based on the estimation of the conditional expectation $\mathbb{E}(f_\beta(X_n)|Y_n)$. The fact is that this conditional expectation writes in the following form:

$$\mathbb{E}(f_\beta(X_n)|Y_n = y) = \frac{\int f_\beta(x)g(x)\Phi_1(x-y)dx}{\int g(x)\Phi_1(x-y)dx}.$$

Taupin (2001) constructed an estimator based on the observations $\{Y_n\}_{n\geq 0}$, of the linear functional Γ_f of the density g, defined by the formula:

$$\Gamma_f(y) = \int f(x)g(x)\Phi_1(x-y)dx \ ; \quad \forall y \in \mathbb{R}.$$

When f is identically equal to one, the functional Γ_f equals to the density h of the observations Y_n. Rates of convergence for this estimator of the functionals when f is either a polynomial function or a trigonometric function of the form $x \mapsto \sum_{j=0}^{\ell} \beta_j \cos(jx)$ or of the form $x \mapsto \sum_{j=0}^{\ell} \beta_j \sin(jx)$ for some integer ℓ and real fixed parameters $(\beta_j)_{0\leq j\leq \ell}$ are given in Taupin (2001) and are shown to be minimax in Matias and Taupin (2001). Typically, the minimax rate of convergence in $\mathbb{L}_p(\mathbb{R})$-norm or in pointwise quadratic risk for a functional Γ_f where f is a polynomial function of degree less or equal to ℓ is equal to $(\log n)^{(2\ell+1)/4}/\sqrt{n}$ (Taupin 2001). In the case of the estimation of h, those rates of convergence are not deteriorated when σ^2 becomes unknown. Now consider the case where f is a polynomial function with degree ℓ greater or equal to one. We prove that the estimation of Γ_f is seriously deteriorated in this case when σ^2 is unknown as the minimax quadratic risk becomes lower bounded by a constant divided by $\log n$ (Theorem 2.2.4).

This paper is organized as follows. Section 2.2 gives lower bounds for the estimations of σ^2 and $g(x_0)$ in the convolution model with unknown variance, using the van Trees inequality (Gill and Levit 1995). It also gives a result about the degradation of the estimation of linear functionals coming from polynomial functions in the errors-in-variables model. Section 2.3 gives some estimators of σ^2 consistent when the Laplace or the Fourier transform of g has some decrease at infinity. Moreover, in Section 2.4, we construct an estimator of σ that is consistent, assuming nothing but a first order moment on g. Technical proofs are given in Section 2.5.

2.2 Lower bounds

2.2.1 Results in the convolution model with unknown variance

In this part, we will give lower bounds for the minimax quadratic risks for the estimation of σ and $g(x_0)$, when x_0 is a fixed point in $\mathbb{R}$:

$$\inf_{\hat{T}_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{g\in\mathcal{G}\cap\mathcal{R}} \left[\mathbb{E}_{\sigma,g}(\hat{T}_n-\sigma^2)^2\right]^{1/2} \quad \text{and} \quad \inf_{\hat{T}_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{g\in\mathcal{G}\cap\mathcal{R}} \left[\mathbb{E}_{\sigma,g}(\hat{T}_n-g(x_0))^2\right]^{1/2},$$

where $\hat{T}_n$ ranges over all estimators based on the observations $Y_1,\dots Y_n$, $\mathcal{R}$ is a set of densities, and $\mathcal{V}(\sigma_0)$ is any small enough neighborhood of a fixed point $\sigma_0>0$. We use the van Trees inequality (Gill and Levit 1995) on a suited one-dimensional sub-model. In fact, the difficulty of the model is contained in a "worst-case" sub-model, that is to say a worst one-dimensional sub-model. Fix the points x_0 in $\mathbb{R}$ and σ_0 in $\mathbb{R}^+$, a real positive parameter τ, a bounded function g_0 in $\mathcal{R}$ satisfying Assumptions 2.1 and 2.2 and consider the following notations: for all $0<t\leq\tau^2$, we denote

$$p_t(u)=\mathcal{C}_\tau\Phi_{\sqrt{t}\,\sigma_0}(u)1\!\!1_{\tau|u|\leq\sqrt{t}\,\sigma_0},$$

where the normalizing constant $\mathcal{C}_\tau$ is equal to $\left(\int\Phi_1(z)1\!\!1_{\tau|z|\leq 1}dz\right)^{-1}$. Now, let us construct a path in $\mathcal{G}\cap\mathcal{R}$:

$$\forall\, 0<t\leq\tau^2,\;\; h_t=g_0(\cdot-x_0)*p_t*\Phi_{\sqrt{1-t}\,\sigma_0} \qquad \text{and} \qquad h_0=g_0(\cdot-x_0)*\Phi_{\sigma_0}. \tag{2.5}$$

The function p_t is the truncated density of a centered Gaussian random variable, with variance equal to $t\sigma_0^2$. But the convolution of $\Phi_{\sqrt{t}\,\sigma_0}$ with $\Phi_{\sqrt{1-t}\,\sigma_0}$ is equal to Φ_{σ_0}, so that we can expect that the densities h_t (for $t>0$) and h_0 are close to each other (in a sense to be precised). For all $t>0$, we denote by σ_t^2 the variance of the noise involved in the definition of h_t, that is to say $\sigma_t^2=(1-t)\sigma_0^2$.
Our aim is to give lower bounds for the estimation of the variance of the noise and for the estimation of g in some classes.

Proposition 2.2.1. *Assume that $\mathcal{R}$ is a subset of $\mathbb{L}_1(\mathbb{R})$ containing a bounded function g_0 which satisfies Assumptions 2.1 and 2.2, then for all $\sigma_0 > 0$ and every small enough neighborhood $\mathcal{V}(\sigma_0)$ of σ_0, we have*

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{g\in\mathcal{G}\cap\mathcal{R}} (\log n) \left[\mathbb{E}_{\sigma,g}(\hat{T}_n - \sigma^2)^2\right]^{1/2} > 0, \tag{2.6}$$

where the infimum is taken over all estimators $\hat{T}_n$ based on the observations $Y_1, \dots Y_n$.

The key of the proof is the use of the van Trees inequality (Gill and Levit 1995). As always in this kind of proof, the purpose is to exhibit densities h_t and h_0 which are close to each other, in the sense that the Fisher information $I(t)$ of the model is small for fixed $t > 0$, whereas the parameters $\sigma_t^2 = (1-t)\sigma_0^2$ and σ_0^2 are well-separated.
A slight adaptation of this path gives the corresponding result on the pointwise estimation of g at x_0.

Theorem 2.2.2. *Assume that $\mathcal{R}$ is a subset of $\mathbb{L}_1(\mathbb{R})$ containing a bounded function g_0 which satisfies Assumptions 2.1, 2.2 and 2.3, then for all real number x_0, for all $\sigma_0 > 0$, and every small enough neighborhood $\mathcal{V}(\sigma_0)$ of σ_0, we have*

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{g\in\mathcal{G}\cap\mathcal{R}} (\log n) \left[\mathbb{E}_{\sigma,g}(\hat{T}_n - g(x_0))^2\right]^{1/2} > 0,$$

where the infimum is taken over all estimators $\hat{T}_n$ based on the observations $Y_1, \dots, Y_n$.

Its proof stands in Section 2.5. We now give some examples of sets $\mathcal{R}$ containing bounded functions that satisfy Assumptions 2.1, 2.2 and 2.3, and compare our result with existing ones.

Example 2.2.1. *Consider the set of functions $\mathcal{C}_{m,\alpha,\beta}$ defined by (2.1). The regularity assumption on the function g has been used by Fan (1991c) when the distribution of the noise sequence is known. Fan proved that the minimax rate of convergence when the density of the noise sequence is known and super smooth, is $(\log n)^{-(m+\alpha)}$. Adding the condition on Gaussian components, in order to have an identifiable model, we prove that when σ is unknown, the rate of convergence in the pointwise estimation of g is seriously deteriorated, as it becomes slower than $(\log n)^{-1}$.*

Example 2.2.2. *Consider now the set of functions $\mathcal{SS}_{\alpha,\nu,\rho}(A)$ defined by (2.2). Pensky and Vidakovic (1999) gave the following improvement on the results of Fan: when the density g of the distribution of the signal is known to be super-smooth, they constructed an estimator of g, whose MISE error (2.3) is upper bounded by a constant times some power of $\log n$, divided by some power of n (see Pensky and Vidakovic (1999) for more details). But when the variance of the noise sequence is unknown, the fact that g is super-smooth does not improve the rate of convergence of its pointwise estimation. It seems to be also the case for the MISE error in the estimation of g.*

We now give the proof of Proposition 2.2.1.

Proof. Without loss of generality, we assume that $x_0 = 0$ and $\sigma_0 = 1$. The first thing to check is that the parametric path belongs to the model. But g_0 is chosen so that for all $0 \le t \le \tau^2$, the density $g_t = g_0 * p_t$ has no Gaussian component and belongs to $\mathcal{R}$ (see Assumption 2.1). By applying definition (2.5) of the density h_t, we can write, for all $0 < t \leq \tau^2$:

$$\begin{aligned} h_t(y) &= \mathcal{C}_\tau h_0(y) - (\mathcal{C}_\tau h_0 - h_t)(y) \\ &= \mathcal{C}_\tau h_0(y) - g_0 * (\mathcal{C}_\tau \Phi_1 - p_t * \Phi_{\sqrt{1-t}})(y). \end{aligned} \tag{2.7}$$

Note that we have the following identity: $p_t = \mathcal{C}_\tau \Phi_{\sqrt{t}}(1 - 1\!\!1_{\tau|\cdot|>\sqrt{t}})$, so that

$$(\mathcal{C}_\tau \Phi_1 - p_t * \Phi_{\sqrt{1-t}})(y) = \mathcal{C}_\tau \left[\Phi_1 - \left(\Phi_{\sqrt{t}} - \Phi_{\sqrt{t}} 1\!\!1_{\{\tau|.|>\sqrt{t}\}} \right) * \Phi_{\sqrt{1-t}} \right](y).$$

The convolution between the normal densities $\Phi_{\sqrt{t}}$ and $\Phi_{\sqrt{1-t}}$ is equal to Φ_1, and then

$$(\mathcal{C}_\tau \Phi_1 - p_t * \Phi_{\sqrt{1-t}})(y) = \mathcal{C}_\tau (\Phi_{\sqrt{t}} 1\!\!1_{\{\tau|\cdot|>\sqrt{t}\}}) * \Phi_{\sqrt{1-t}}(y).$$

Combining with (2.7), we get

$$\begin{aligned} h_t(y) &= \mathcal{C}_\tau h_0(y) - \mathcal{C}_\tau g_0 * (\Phi_{\sqrt{t}} 1\!\!1_{\{\tau|\cdot|>\sqrt{t}\}}) * \Phi_{\sqrt{1-t}}(y) \\ &= \mathcal{C}_\tau h_0(y) - \mathcal{C}_\tau \iint \Phi_{\sqrt{1-t}}(y-u)\, \Phi_{\sqrt{t}}(u-v)\, 1\!\!1_{\tau|u-v|>\sqrt{t}} g_0(v) dv du. \end{aligned} \tag{2.8}$$

We have :

$$\begin{aligned} &\Phi_{\sqrt{t}}(u-v)\, \Phi_{\sqrt{1-t}}(y-u) = \\ &\qquad \frac{1}{2\pi\sqrt{t(1-t)}} \exp\left[-\frac{(u-(1-t)v-ty)^2}{2t(1-t)} - \frac{v^2}{2t} - \frac{y^2}{2(1-t)} + \frac{[(1-t)v+ty]^2}{2t(1-t)} \right] \\ &\qquad\qquad = \Phi_{\sqrt{t(1-t)}}\ (u-(1-t)v-ty)\ \Phi_1(v-y), \end{aligned}$$

and then returning to (2.8),

$$\begin{aligned} h_t(y) &= \mathcal{C}_\tau h_0(y) \\ &\quad - \frac{\mathcal{C}_\tau}{\sqrt{2\pi}} \int \left[\int \frac{1}{\sqrt{t(1-t)}} \Phi_1 \left(\frac{u-(1-t)v-ty}{\sqrt{t(1-t)}} \right) 1\!\!1_{\tau|u-v|>\sqrt{t}} du \right] e^{-\frac{(v-y)^2}{2}} g_0(v) dv \\ &= \mathcal{C}_\tau h_0(y) - \frac{\mathcal{C}_\tau}{\sqrt{2\pi}} \int \left[\int_{\frac{(1/\tau)+\sqrt{t}(v-y)}{\sqrt{1-t}}}^{+\infty} \Phi_1(z) dz + \int_{-\infty}^{\frac{-(1/\tau)+\sqrt{t}(v-y)}{\sqrt{1-t}}} \Phi_1(z) dz \right] e^{-\frac{(y-v)^2}{2}} g_0(v) dv. \end{aligned} \tag{2.9}$$

This expression of h_t is useful in the computation of the corresponding Fisher information. In the rest of this section, the notation $\mathbb{E}_t$ stands as an abbreviation for $\mathbb{E}_{\sigma_t, g_t}$. The Fisher information associated to our path is defined by:

$$I(t) = \mathbb{E}_t \left(\frac{\partial \log h_t}{\partial t}(Y) \right)^2 = \int \left(\frac{\partial h_t}{\partial t} \right)^2 (y) h_t^{-1}(y) dy\ , \qquad \text{for all } 0 \leq t \leq \tau^2. \tag{2.10}$$

Lemma 2.2.3. *The Fisher information satisfies:*

$$\forall\, 0 \leq t \leq \tau^2\ ,\ I(t) = I(0)(1+o(1)) = \frac{\mathcal{C}_\tau^2}{4\pi^2} \tau^{-2} e^{-1/\tau^2} \left(\int f_0(y)^2 h_0(y)^{-1} dy \right) (1+o(1)),$$

where f_0 is defined by

$$f_0(y) = \int [1-(v-y)^2] e^{-\frac{(y-v)^2}{2}} g_0(v) dv.$$

This Lemma is proved in Section 2.5.
Now, let us consider $\lambda_0(t)dt$ a probability measure on $[0;1]$ satisfying the following conditions

- $\lambda_0(0) = \lambda_0(1) = 0$.
- $t \mapsto \lambda_0(t)$ is continuously differentiable in $]0;1[$.
- $\lambda_0(t)dt$ has finite Fisher information:

$$J_0 = \int_0^1 \frac{\lambda_0'^2(t)}{\lambda_0(t)} dt,$$

where the prime stands for derivation with respect to the parameter t.
Rescaling this measure on the interval $[0;\tau^2]$ we define:

$$\lambda(t)dt = \frac{1}{\tau^2}\lambda_0\left(\frac{t}{\tau^2}\right) dt$$

which has the Fisher information J_0/τ^4. Denote by $\mathbb{E}_\lambda(I)$ the quantity $\int I(t)\lambda(t)dt$ (t is seen as a random variable with values in $[0;\tau^2]$, distributed according to the probability measure $\lambda(t)dt$). The van Trees inequality (Gill and Levit 1995) for the estimation of the variance of the noise in the convolution model gives us for small enough τ:

$$\begin{aligned}\inf_{\hat{T}_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{g\in\mathcal{G}\cap\mathcal{R}} \mathbb{E}_{\sigma,g}(\hat{T}_n - \sigma^2)^2 &\geq \inf_{\hat{T}_n} \int_0^{\tau^2} \mathbb{E}_t(\hat{T}_n - \sigma_t^2)^2 \lambda(t)dt \\ &\geq \left(\int \frac{\partial \sigma_t^2}{\partial t}(t)\lambda(t)dt\right)^2 \left(n\mathbb{E}_\lambda(I) + J_0/\tau^4\right)^{-1}.\end{aligned}$$

And then

$$\inf_{\hat{T}_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{g\in\mathcal{G}\cap\mathcal{R}} \mathbb{E}_{\sigma,g}(\hat{T}_n - \sigma^2)^2 \geq \left(n\mathbb{E}_\lambda(I) + J_0/\tau^4\right)^{-1}.$$

Using Lemma 2.2.3,

$$\mathbb{E}_\lambda(I) = I(0)(1+o(1)) = \frac{\mathcal{C}_\tau^2}{4\pi^2}\tau^{-2}e^{-1/\tau^2}\left(\int f_0(y)^2 h_0(y)^{-1}dy\right)(1+o(1)),$$

and then, returning to the van Trees inequality

$$\inf_{\hat{T}_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{g\in\mathcal{G}\cap\mathcal{R}} \mathbb{E}_{\sigma,g}(\hat{T}_n - \sigma^2)^2 \geq \left[n\frac{\mathcal{C}_\tau^2}{4\pi^2}\left(\int f_0(y)^2 h_0(y)^{-1}dy\right)\tau^{-2}e^{-1/\tau^2}(1+o(1)) + J_0/\tau^4\right]^{-1}.$$

Choosing $\tau^{-1} = \sqrt{\log n}$, we get:

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{\sigma\in\mathcal{V}(\sigma_0)} \sup_{g\in\mathcal{G}\cap\mathcal{R}} (\log n)^2 \mathbb{E}_{\sigma,g}(\hat{T}_n - \sigma^2)^2 > 0,$$

which achieves the proof of Proposition 2.2.1. □

2.2.2 Consequence in the errors-in-variables model

Consider in our model the estimation of the linear functional $\Gamma_{f,\sigma}$ defined by:

$$\Gamma_{f,\sigma}(y) = \int f(x)g(x)\Phi_\sigma(x-y)dx \ ; \quad \forall y \in \mathbb{R},$$

motivated by the the errors-in-variables model described in the introduction (formulae (2.4)), when the parameter σ^2 is unknown.

More precisely, we are interested in the functionals given by a polynomial function. Let P be a polynomial function of degree ℓ greater or equal to one, and define the set

$$\mathcal{G}_P(\mathcal{R}) = \left\{ \Gamma = \int P(x)\Phi_\sigma(\cdot - x)g(x)dx \ ; \ g \in \mathcal{R} \cap \mathcal{G}, \ \sigma > 0 \right\},$$

of those functionals constructed with a density g lying in some "regular" space $\mathcal{R}$ and without Gaussian component. We have the following theorem concerning the estimation of a functional Γ in $\mathcal{G}_P(\mathcal{R})$.

Theorem 2.2.4. *Let $\mathcal{R}$ be a subset of $\mathbb{L}_1(\mathbb{R})$ containing a bounded function g_0 that is supposed to be ℓ times continuously differentiable, satisfying Assumptions 2.1, 2.2 and 2.3 and such that $\|g_0^{(\ell)}\|_\infty < \infty$. Then, for all fixed real number y_0 and every polynomial function P of degree ℓ greater or equal to one, we have:*

$$\liminf_{n\to\infty} \inf_{\widehat{\Gamma}_n} \sup_{\Gamma \in \mathcal{G}_P(\mathcal{R})} (\log n) \left[\mathbb{E}(\widehat{\Gamma}_n - \Gamma(y_0))^2 \right]^{1/2} > 0,$$

where the infimum is taken over all the estimators $\widehat{\Gamma}_n$ based on the observations $Y_1, \dots Y_n$.

The proof of this result is given in Section 2.5.

2.3 Some upper bounds

In this part, we propose an estimator of σ, assuming that the distribution of the signal possesses a Laplace transform not rapidly increasing at infinity. This framework contains for example the case of g having a support included in some fixed compact set. We start by the general construction of this estimator, and then give the result on the fixed compact support case.
The construction of this estimator of σ is based on the behaviour of the Laplace transform of g at infinity: as the distribution of the signal has no Gaussian component, its Laplace transform should increase lower than e^{Ct^2} at infinity. Assume that the density of the hidden variable X_n belongs to the following set of functions:

$$\mathcal{L}_{v,M} = \left\{ g \in \mathbb{L}_1(\mathbb{R}) \ / \ g \geq 0 \ , \ \int g(x)dx = 1 \ , \ \left| \log\left(\int e^{tx} g(x)dx \right) \right| \leq Mt^2 v(t) \right\}, \tag{2.11}$$

for some positive function v vanishing at infinity and a real positive number M. Note for example that when $v(t) = 1/|t|$, the set of functions $\mathcal{L}_{v,M}$ contains the set of functions with support included in $[-M; M]$. In the rest of the paper, we will use the abbreviated notation:

$$\mathcal{L}(M) = \left\{ g \in \mathbb{L}_1(\mathbb{R}) \ / \ g \geq 0 \ , \ \int g(x)dx = 1 \ , \ \left| \log\left(\int e^{tx} g(x)dx \right) \right| \leq M|t| \right\}.$$

When the density g of the signal sequence belongs to $\mathcal{L}_{v,M}$ for some function v and some real positive number M, the variance of the noise is identifiable, as can be seen by computing the Laplace transform of the observations. In fact, we have the equality

$$\forall g \in \mathcal{L}_{v,M} \ , \qquad \sigma^2 = \lim_{t\to\infty} \frac{2}{t^2} \log \mathbb{E}_{\sigma,g}\left(e^{tY_1}\right).$$

It is then natural to define the following empirical estimator of σ

$$\hat{\sigma}^2_{L,n}(t_n) = \frac{2}{t_n^2} \log\left(\frac{1}{n}\sum_{j=1}^n e^{t_n Y_j}\right),$$

where $(t_n)_{n\geq 0}$ is some sequence of positive numbers increasing to infinity (the subscript L stands for Laplace).

Proposition 2.3.1. *For any $\Sigma > 0$, define $t_n = \sqrt{\alpha \log n}/(2\Sigma)$ for some $0 < \alpha < 1$. Then the following estimator converges to σ^2 in probability:*

$$\forall g \in \mathcal{L}_{v,M},\ \forall \sigma \in [0;\Sigma], \qquad \hat{\sigma}^2_{L,n}(t_n) \underset{n\to\infty}{\longrightarrow} \sigma^2 \qquad \textit{in probability under } (\sigma, g).$$

Moreover, for all real positive number M and all positive function v vanishing at infinity:

$$\sup_{\sigma\in[0;\Sigma]} \sup_{g\in\mathcal{L}_{v,M}} \left[E_{\sigma,g}(\hat{\sigma}^2_{L,n}(t_n) - \sigma^2)\right]^{1/2} \leq \frac{8\Sigma^2}{\alpha(\log n)n^{(1-\alpha)/2}} + 2Mv(t_n).$$

The proof of this proposition stands in Section 2.5.
Now, we consider the special case $v(t) = 1/|t|$, that is to say the set $\mathcal{L}(M)$ containing the functions with a support included in $[-M;M]$.

Corollary 2.3.2. *For all $M > 0$ and all $\Sigma > 0$ there exists a positive constant C such that:*

$$\inf_{\hat{T}_n} \sup_{\sigma\in[0;\Sigma]} \sup_{g\in\mathcal{L}(M)} \left[E_{\sigma,g}(\hat{T}_n - \sigma^2)\right]^{1/2} \leq \frac{C}{\sqrt{\log n}},$$

where the infimum is taken over all estimators $\hat{T}_n$ based on the observations $Y_1,\dots Y_n$.

This inequality is a straightforward application of Proposition 2.3.1.

The same construction can be made using the Fourier transform instead of the Laplace transform. Consider the following set for the density g of the signal:

$$\mathcal{F}_v = \left\{ g \in \mathbb{L}_1(\mathbb{R}) \ / \ g \geq 0 \ , \ \int g(x)dx = 1 \ , \ \lim_{|t|\to\infty} \frac{\left|\log\left(\int e^{itx}g(x)dx\right)\right|}{t^2 v(t)} < +\infty \right\} \qquad (2.12)$$

for some positive function v vanishing at infinity. This is the case for example when g has the form $Cp(x)e^{-\lambda x}\mathbb{1}_{x>0}$, where λ is a positive constant, p a polynomial function and C a normalizing constant. With this condition on the density of X_n, the parameter σ stays identifiable. In fact, using the Fourier transform of the observations, we get:

$$\sigma^2 = \lim_{t\to+\infty} -\frac{2}{t^2} \log \mathbb{E}\left(e^{itY}\right).$$

It is then natural to consider the following empirical estimator of σ:

$$\hat{\sigma}^2_{F,n}(t_n) = -\frac{2}{t_n^2}\log\left(\frac{1}{n}\sum_{j=1}^{n} e^{it_nY_j}\right)$$

where $(t_n)_{n\geq 0}$ is some sequence of positive numbers increasing to infinity (the subscript F stands for Fourier).

Proposition 2.3.3. *For any $\Sigma > 0$, define $t_n = \sqrt{\alpha \log n/(2\Sigma^2)}$ for some $0 < \alpha < 1$. Then the following estimator converges to σ^2 in probability:*

$$\forall g \in \mathcal{F}_v \text{ , } \forall \sigma \in [0;\Sigma], \qquad \hat{\sigma}^2_{F,n}(t_n) \underset{n\to\infty}{\longrightarrow} \sigma^2 \qquad \textit{in probability under } (\sigma, g).$$

The proof of this proposition stands in Section 2.5.

2.4 Universal estimation of the noise variance

The preceding estimator $\hat{\sigma}^2_{L,n}$ of σ is interesting as its rate of convergence is entirely known, but it would be nice to have an estimator of σ converging whatever the form of g is, without Gaussian component. This is the purpose of this section, as we only assume here a first order moment on the signal g.
We denote by g^* the Fourier transform of g and for all ζ in $\mathbb{R}$, for all fixed $\sigma > 0$ and all $\tau > 0$,

$$\alpha(\zeta;\tau) = g^*(\zeta)e^{-\zeta^2(\sigma^2-\tau^2)/2} = h^*(\zeta)e^{\zeta^2\tau^2/2}. \tag{2.13}$$

The function α is the product of the Fourier transform of h (the density of the distribution of the observations) and the function $(\Phi^*_\tau)^{-1}$. When τ is equal to the true value of the parameter σ, the function $\alpha(\cdot;\sigma)$ is equal to the characteristic function of the distribution of the hidden sequence. We observe the following properties of α:

- when $\tau \leq \sigma$, the function $\alpha(\cdot;\tau)$ is the Fourier transform of the positive measure $g*\Phi_{\sqrt{\sigma^2-\tau^2}}$.
- when $\tau > \sigma$, the function $\alpha(\cdot;\tau)$ is no longer a Fourier transform of a probability measure on the real line (as g is supposed to belong to the set $\mathcal{G}$ and then has no Gaussian component). By Bochner's theorem (see for example Feller (1971)), the function $\alpha(\cdot;\tau)$ will not be positive definite.

We define an empirical estimator of α by:

$$\hat{\alpha}_n(\zeta;\tau) = \left(\frac{1}{n}\sum_{p=1}^{n} e^{i\zeta Y_p}\right) e^{\zeta^2\tau^2/2} \tag{2.14}$$

Now, we use the remarks on the properties of the function α. The idea is that the real parameter σ is the largest value of τ for which the function $\alpha(\cdot;\tau)$ stays positive definite. It means that σ is the largest value of τ satisfying that for all integer n, and all n-tuples of real numbers $\{t_k; 1 \leq k \leq n\}$, the smallest eigenvalue of the matrix $(\alpha(t_k - t_l;\tau))_{1\leq k,l\leq n}$ is positive. We approximate the function α by its empirical estimator $\hat{\alpha}_n$ and use a dense family $\{t_{k,n}\}_{n\geq 0}$ of points in $\mathbb{R}$, so that for n large enough, the matrices $(\hat{\alpha}_n(t_{k,n}-t_{l,n};\tau))_{k,l}$ will have their smallest eigenvalue positive (to within about some ϵ_n constant) until τ reaches σ. In the rest of the section, u is used to denote a point in $\mathbb{C}^p$, where p may change along the lines, and $\|u\|$ denotes the norm $(\sum_{j=1}^p |u_j|^2)^{1/2}$. Let us define

$$\hat{\sigma}_n = \sup\left\{ s \geq 0 \ : \ \forall \tau \leq s, \inf_{u;\ \|u\|=1} \sum_{-\ell_n \leq k,l \leq \ell_n} u_k \hat{\alpha}_n(t_{k,n}-t_{l,n};\tau)\bar{u}_l \geq -\epsilon_n \right\} \tag{2.15}$$

where k_n and ℓ_n are two sequences of numbers increasing to infinity, in such a way that ℓ_n/k_n also increases to infinity. The points $t_{k,n} = k/k_n$ form a partition of the interval $[-\ell_n/k_n;\ell_n/k_n]$ when the integer k ranges from $-\ell_n$ to ℓ_n; and ϵ_n is a sequence of numbers decreasing to zero.

Assumption 2.4. *Fix $\Sigma > 0$ and choose the parameters $\ell_n; k_n$ and ϵ_n in the following way:*

- $\frac{\ell_n}{k_n} = \frac{1}{\Sigma}\sqrt{\frac{a \log n}{2}}$ *for some $0 < a < 1/2$.*
- $k_n = n^{1/2-a-b}$ *for some $b > 0$ and $2a + b < 1/2$.*
- $\epsilon_n = \frac{\sqrt{2a}}{\Sigma}\frac{\sqrt{\log n}}{n^b} v_n$ *where v_n is an increasing sequence of numbers converging to infinity in such a way that ϵ_n converges to zero, and $v_n^{-1} = o((\log\log n)^{-1/2})$.*

Theorem 2.4.1. *For all $\Sigma > 0$, under Assumption 2.4, and for all $\sigma \in]0;\Sigma]$, all g in $\mathcal{G}$ such that $\int |x|g(x)dx < \infty$, we have:*

$$\hat{\sigma}_n \underset{n\to\infty}{\longrightarrow} \sigma \text{ in } \textit{probability under } (\sigma, g).$$

This Theorem is proved in Section 2.5.
Note that the estimator $\hat{\sigma}_n$ can be computed considering the matrices $T_n(\tau) = \{\hat{\alpha}_n(t_{k,n} - t_{l,n};\tau)\}_{-\ell_n \leq k,l \leq \ell_n}$. The graph of the function $\tau \mapsto \lambda_{min}(T_n(\tau))$, where $\lambda_{min}(T)$ denotes the smallest eigenvalue of the matrix T, gives the value of $\hat{\sigma}_n$ by considering the first value of τ such that $\lambda_{min}(T_n(\tau)) \leq -\epsilon_n$. Note also that we only give the asymptotic behaviour of the parameters ℓ_n, k_n, ϵ_n and v_n. The choice of values would be justified by practical applications, but this is beyond the scope of this paper.

2.5 Proofs

Proof. Proof of Lemma 2.2.3.
We calculate the derivative of h_t with respect to the parameter t, using equation (2.9):

$$\frac{\partial h_t}{\partial t}(y) = -\frac{\mathcal{C}_\tau}{\sqrt{2\pi}}\int\left[-\Phi_1\left(\frac{(1/\tau)+\sqrt{t}(v-y)}{\sqrt{1-t}}\right)\left(\frac{(v-y)}{2\sqrt{t(1-t)}}+\frac{(1/\tau)+\sqrt{t}(v-y)}{2(1-t)^{3/2}}\right)\right.$$
$$\left.+\Phi_1\left(\frac{-(1/\tau)+\sqrt{t}(v-y)}{\sqrt{1-t}}\right)\left(\frac{(v-y)}{2\sqrt{t(1-t)}}+\frac{-(1/\tau)+\sqrt{t}(v-y)}{2(1-t)^{3/2}}\right)\right]e^{-\frac{(y-v)^2}{2}}g_0(v)dv$$

But the equalities

$$\begin{aligned}
\Phi_1\left(\frac{(1/\tau)+\sqrt{t}(v-y)}{\sqrt{1-t}}\right) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\tau^{-2}+t(v-y)^2}{2(1-t)}\right)\exp\left(-\frac{\sqrt{t}(v-y)}{\tau(1-t)}\right)\\
\Phi_1\left(\frac{-(1/\tau)+\sqrt{t}(v-y)}{\sqrt{1-t}}\right) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\tau^{-2}+t(v-y)^2}{2(1-t)}\right)\exp\left(+\frac{\sqrt{t}(v-y)}{\tau(1-t)}\right)
\end{aligned}$$

lead to:

$$\begin{aligned}
\frac{\partial h_t}{\partial t}(y) &= \frac{-\mathcal{C}_\tau}{2\pi}\int e^{-\frac{\tau^{-2}+t(v-y)^2}{2(1-t)}}\left[\left(\frac{(v-y)}{2\sqrt{t(1-t)}}+\frac{-(1/\tau)+\sqrt{t}(v-y)}{2(1-t)^{3/2}}\right)e^{+\frac{\sqrt{t}(v-y)}{\tau(1-t)}}\right.\\
&\left.\quad -\left(\frac{(v-y)}{2\sqrt{t(1-t)}}+\frac{(1/\tau)+\sqrt{t}(v-y)}{2(1-t)^{3/2}}\right)e^{-\frac{\sqrt{t}(v-y)}{\tau(1-t)}}\right]e^{-\frac{(y-v)^2}{2}}g_0(v)dv.
\end{aligned}$$

Simple calculations will give:

$$\begin{aligned}
\frac{\partial h_t}{\partial t}(y) &= \frac{-\mathcal{C}_\tau}{2\pi(1-t)^{3/2}}\int(v-y)e^{-\frac{\tau^{-2}+t(v-y)^2}{2(1-t)}}\frac{1}{\sqrt{t}}\mathrm{sh}\left(\frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}}\right)e^{-\frac{(y-v)^2}{2}}g_0(v)dv\\
&\quad+\frac{\mathcal{C}_\tau}{2\pi(1-t)^{3/2}\tau}\int e^{-\frac{\tau^{-2}+t(v-y)^2}{2(1-t)}}\mathrm{ch}\left(\frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}}\right)e^{-\frac{(y-v)^2}{2}}g_0(v)dv \qquad (2.16)
\end{aligned}$$

The second term in the right hand side of the last equality is well-defined when t is equal to zero. The first one is continuous at $t=0$, observing that:

- $\frac{1}{\sqrt{t}}\mathrm{sh}\left(\frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}}\right)\longrightarrow_{t\to 0}\frac{v-y}{\tau}$
- Using a Taylor formula:

$$\frac{1}{\sqrt{t}}\mathrm{sh}\left(\frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}}\right)=\frac{(v-y)}{2\sqrt{1-t}}\int_{-1/\tau}^{1/\tau}e^{\frac{x\sqrt{t}(v-y)}{\sqrt{1-t}}}dx,$$

  we get an upper-bound, valid for all $0<t\le\min(1/2;\tau^2)$,

$$\left|\frac{1}{\sqrt{t}}\mathrm{sh}\left(\frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}}\right)\right|\le\frac{\sqrt{2}}{\tau}|v-y|e^{|v-y|}. \qquad (2.17)$$

  This leads to the domination, valid for all $0<t\le\min(1/2;\tau^2)$,

$$\begin{aligned}
&\left|(v-y)e^{-\frac{\tau^{-2}+t(v-y)^2}{2(1-t)}}\frac{1}{\sqrt{t}}\mathrm{sh}\left(\frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}}\right)e^{-\frac{(y-v)^2}{2}}g_0(v)\right|\\
&\qquad\le\frac{\sqrt{2}}{\tau}(v-y)^2e^{-\tau^{-2}/2}e^{|v-y|}e^{-\frac{(y-v)^2}{2}}g_0(v)
\end{aligned}$$

  and the dominating function, as a function of the variable v, belongs to $\mathbb{L}_1(\mathbb{R})$.

Then, dominated convergence combined with the expression (2.16) gives the continuity of $[(\partial h_t)/(\partial t)](y)$ at $t=0$:

$$\begin{aligned}\left(\frac{\partial h_t}{\partial t}\right)_{|t=0}(y) &= \frac{\mathcal{C}_\tau e^{-1/2\tau^2}}{2\pi\ \tau}\int [1-(v-y)^2]e^{-\frac{(y-v)^2}{2}}g_0(v)dv = \frac{\mathcal{C}_\tau e^{-1/2\tau^2}}{2\pi\ \tau}f_0(y)\\ \text{where}\quad & f_0(y) = \int [1-(v-y)^2]e^{-\frac{(y-v)^2}{2}}g_0(v)dv\end{aligned}$$

belongs to $\mathbb{L}_1(\mathbb{R})$. Then, we have by definition

$$I(t) = E_t\left(\frac{\partial \log h_t}{\partial t}(Y)\right)^2 = \int \left(\frac{\partial h_t}{\partial t}\right)^2 (y)h_t^{-1}(y)dy.$$

And now, we claim the continuity of $t\mapsto I(t)$ at $t=0$, using that:

- $t\mapsto \left(\frac{\partial h_t}{\partial t}\right)^2(y)h_t^{-1}(y)$ is continuous at $t=0$ for all $y\in\mathbb{R}$.
- Consider the expression (2.16), and use the inequality $|a+b|^2\le 2(a^2+b^2)$, combined with the Cauchy-Schwarz inequality (recall that g_0 is a probability density). We get the following domination:

$$\begin{aligned}\left|\frac{\partial h_t}{\partial t}(y)\right|^2 &\le \frac{\mathcal{C}_\tau^2}{2\pi^2(1-t)^3}\int |v-y|^2 e^{-\frac{\tau^{-2}+t(v-y)^2}{(1-t)}}\frac{1}{t}\left|\text{sh}\left(\frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}}\right)\right|^2 e^{-(v-y)^2}g_0(v)dv\\ &\quad+\frac{\mathcal{C}_\tau^2\tau^{-2}}{2\pi^2(1-t)^3}\int e^{-\frac{\tau^{-2}+t(v-y)^2}{(1-t)}}\left|\text{ch}\left(\frac{\sqrt{t}(v-y)}{\tau\sqrt{1-t}}\right)\right|^2 e^{-(v-y)^2}g_0(v)dv.\end{aligned}$$

  Now, assume that $0<t\le\min(1/2;\tau^2)$, use the upper-bound (2.17) and the inequality $|\text{ch}(x)|\le e^{|x|}$ to obtain:

$$\begin{aligned}\left|\frac{\partial h_t}{\partial t}(y)\right|^2 &\le \frac{8\mathcal{C}_\tau^2 e^{-1/\tau^2}}{\pi^2\ \tau^2}\int |v-y|^4 e^{2|v-y|}e^{-(v-y)^2}g_0(v)dv\\ &\quad+\frac{4\mathcal{C}_\tau^2 e^{-1/\tau^2}}{\pi^2\ \tau^2}\int e^{2|v-y|}e^{-(v-y)^2}g_0(v)dv\\ &\le \frac{8\mathcal{C}_\tau^2 e^{-1/\tau^2}}{\pi^2\ \tau^2}\int (1+|v-y|^4)e^{2|v-y|}e^{-(v-y)^2}g_0(v)dv. \quad (2.18)\end{aligned}$$

- Remember equality (2.9)

$$\begin{aligned}h_t(y) &= \mathcal{C}_\tau h_0(y) - \frac{\mathcal{C}_\tau}{\sqrt{2\pi}}\int\left[\int_{\frac{\tau^{-1}+\sqrt{t}(v-y)}{\sqrt{1-t}}}^{+\infty}\Phi_1(z)dz + \int_{-\infty}^{\frac{-\tau^{-1}+\sqrt{t}(v-y)}{\sqrt{1-t}}}\Phi_1(z)dz\right]\\ &\qquad\times e^{-(v-y)^2/2}g_0(v)dv\\ &= \frac{\mathcal{C}_\tau}{\sqrt{2\pi}}\int\left[\int_{\frac{-\tau^{-1}+\sqrt{t}(v-y)}{\sqrt{1-t}}}^{\frac{\tau^{-1}+\sqrt{t}(v-y)}{\sqrt{1-t}}}\Phi_1(z)dz\right]e^{-(v-y)^2/2}g_0(v)dv.\end{aligned}$$

We lower bound this integral by its restriction to the set $\{|v-y| \leq 1\}$ so that for small enough τ, it can also be lower bounded by

$$h_t(y) \geq \frac{1}{2\sqrt{2\pi}} \int 1\!\!1_{|v-y|\leq 1} e^{-(v-y)^2/2} g_0(v)dv,$$

and we obtain

$$h_t(y) \geq \frac{e^{-1/2}}{2\sqrt{2\pi}} \int_{-1}^{1} g_0(y+u)du.$$

Now, by assumption, the function $\alpha_0(y) = \int_{-1}^{1} g_0(y+u)du$ satisfies

$$\int [\alpha_0(y)]^{-1} e^{-y^2/2} dy < \infty.$$

Combining with (2.18), we obtain a domination on the quantity

$$\left|\frac{\partial h_t}{\partial t}(y)\right|^2 h_t(y)^{-1}$$

by an integrable function of the variable y.

This achieves the proof of the continuity of the function $t \to I(t)$ at the point $t = 0$. In conclusion, we have

$$I(t) = I(0)(1+o(1)) = \frac{\mathcal{C}_\tau^2}{4\pi^2} \tau^{-2} e^{-1/\tau^2} \left(\int f_0(y)^2 h_0(y)^{-1} dy \right) (1+o(1)).$$

□

Proof. Proof of Theorem 2.2.2.
Without loss of generality, we will assume $x_0 = 0$ and $\sigma_0 = 1$. We use the same path as in the proof of Proposition 2.2.1, assuming moreover that g_0 satisfies Assumption 2.3. The only thing to care about is the functional to be estimated. Here we have :

$$\begin{aligned} g_t(0) &= \int g_0(u)p_t(u)du = \mathcal{C}_\tau \int_{|v|\leq(1/\tau)} g_0(\sqrt{t}v) e^{-v^2/2} \frac{dv}{\sqrt{2\pi}} \\ \frac{\partial g_t(0)}{\partial t} &= \frac{\mathcal{C}_\tau}{2\sqrt{t}} \int_{|v|\leq(1/\tau)} v g_0'(\sqrt{t}v)\Phi_1(v)dv. \end{aligned}$$

Expanding g_0' in the neighborhood of 0, there exists a point $\bar{v}$ between 0 and $\sqrt{t}v$ such that

$$g_0'(\sqrt{t}v) = g_0'(0) + \sqrt{t}v g_0''(0) + tv^2 g_0^{(3)}(\bar{v})$$

Then, using the same notations as in the proof of Proposition 2.2.1,

$$\begin{aligned} \frac{\partial g_t(0)}{\partial t} &= g_0''(0) \frac{\mathcal{C}_\tau}{2} \left(\int_{|v|\leq(1/\tau)} v^2 \Phi_1(v)dv \right) (1+O(\sqrt{t})) \\ \left(\int \frac{\partial g_t(0)}{\partial t}(t)\lambda(t)dt \right)^2 &= [g_0''(0)]^2 \frac{\mathcal{C}_\tau^2}{4} \left(\int_{|v|\leq(1/\tau)} v^2 \Phi_1(v)dv \right)^2 (1+O(\tau)). \end{aligned}$$

Now, applying the van Trees inequality, we obtain that for small enough τ:

$$\begin{aligned}\inf_{\hat{T}_n}\sup_{\sigma\in\mathcal{V}(\sigma_0)}\sup_{g\in\mathcal{G}\cap\mathcal{R}}\mathbb{E}_{\sigma,g}(\hat{T}_n-g(0))^2 &\geq \inf_{\hat{T}_n}\int_0^{\tau}\mathbb{E}_t(\hat{T}_n-g_t(0))^2\lambda(t)dt\\ &\geq \left(\int\frac{\partial g_t(0)}{\partial t}(t)\lambda(t)dt\right)^2\left(n\mathbb{E}_\lambda(I)+J_0/\tau^4\right)^{-1}\\ &= \left[g_0''(0)\right]^2\frac{\mathcal{C}_\tau^2}{4}\left(\int_{|v|\leq(1/\tau)}v^2\Phi_1(v)dv\right)^2(1+o(1))\left(n\mathbb{E}_\lambda(I)+J_0/\tau^4\right)^{-1}.\end{aligned}$$

Using Lemma 2.2.3, we have the equality

$$\mathbb{E}_\lambda(I)=I(0)(1+o(1))=\frac{\mathcal{C}_\tau^2}{4\pi^2}\frac{e^{-1/\tau^2}}{\tau^2}\left(\int f_0(y)^2h_0(y)^{-1}dy\right)(1+o(1)).$$

And then, returning to the van Trees inequality

$$\inf_{\hat{T}_n}\sup_{\sigma\in\mathcal{V}(\sigma_0)}\sup_{g\in\mathcal{G}\cap\mathcal{R}}\mathbb{E}_{\sigma,g}(\hat{T}_n-g(0))^2\geq\frac{[g_0''(0)]^2\frac{\mathcal{C}_\tau^2}{4}\left(\int_{|v|\leq(1/\tau)}v^2\Phi_1(v)dv\right)^2(1+o(1))}{n\frac{\mathcal{C}_\tau^2}{4\pi^2}\left(\int f_0(y)^2h_0(y)^{-1}dy\right)\tau^{-2}e^{-1/\tau^2}(1+o(1))+J_0/\tau^4}.$$

Choosing $\tau^{-1}=\sqrt{\log n}$, we get

$$\liminf_{n\to\infty}\inf_{\hat{T}_n}\sup_{\sigma\in\mathcal{V}(\sigma_0)}\sup_{g\in\mathcal{G}\cap\mathcal{R}}(\log n)^2\mathbb{E}_{\sigma,g}(\hat{T}_n-g(0))^2>0.$$

□

Proof. Proof of Theorem 2.2.4.
In order to simplify notations, we only give the proof for $\ell=1$ and $P:x\mapsto x$, the same arguments apply for greater values of ℓ. Note first that the density h of the distribution of the observations Y_n is the convolution product between g and Φ_σ. Differentiating this expression, we get that

$$\Gamma_x(y_0)\triangleq\int xg(x)\Phi_\sigma(x-y_0)dx=y_0h(y_0)+\sigma^2h'(y_0).$$

Now use the same path as in the proof of Proposition 2.2.1 (with the conventions $x_0=0$ and $\sigma_0=1$), assuming moreover that g_0 is continuously differentiable with $\|g_0'\|_\infty<\infty$. The linear functional of interest writes:

$$\Gamma_t(y_0)\triangleq\int xg_t(x)\Phi_{\sigma_t}(x-y_0)dx=y_0h_t(y_0)+\sigma_t^2h_t'(y_0)=y_0h_t(y_0)+(1-t)h_t'(y_0)\ ,\quad 0\leq t\leq\tau^2.$$

The van Trees inequality gives us:

$$\begin{aligned}\inf_{\widehat{\Gamma}_n}\sup_{\Gamma\in\mathcal{G}_P(\mathcal{R})}\mathbb{E}(\widehat{\Gamma}_n-\Gamma(y_0))^2 &\geq \inf_{\widehat{\Gamma}_n}\int_0^{\tau^2}\mathbb{E}_t(\widehat{\Gamma}_n-\Gamma_t(y_0))^2\lambda(t)dt\\ &\geq \left[\int\frac{\partial}{\partial t}\Gamma_t(y_0)\lambda(t)dt\right]^2\left(n\mathbb{E}_\lambda(I)+\frac{J_0}{\tau^4}\right)^{-1}.\end{aligned}$$

We already computed the denominator of this expression in the proof of Proposition 2.2.1 and the only thing to do is to compute the numerator. We have:

$$\frac{\partial}{\partial t}\Gamma_t(y_0) = y_0\frac{\partial}{\partial t}h_t(y_0) - h_t'(y_0) + (1-t)\frac{\partial}{\partial t}h_t'(y_0).$$

By the continuity of $t \mapsto (\partial h_t)/(\partial t)(y_0)$ at the point $t = 0$ (see the proof of Lemma 2.2.3), we get

$$\frac{\partial}{\partial t}h_t(y_0) = \frac{\mathcal{C}_\tau e^{-1/2\tau^2}}{2\pi\ \tau} f_0(y_0)(1+o(1)) = O(\tau).$$

The same argument applies to prove that

$$\frac{\partial}{\partial t}h_t'(y_0) = \frac{\mathcal{C}_\tau e^{-1/2\tau^2}}{2\pi\ \tau} \tilde{f}_0(y_0)(1+o(1)) = O(\tau),$$

where $\tilde{f}_0(y_0) = \int [1-(v-y_0)^2]e^{-(y-v)^2/2}g_0'(v)dv$. Using the continuity of $t \mapsto h_t(y_0)$ at $t = 0$, we have:

$$h_t'(y_0) = -g_0' * \Phi_1(y_0) + O(\tau).$$

This enables us to write:

$$\int \frac{\partial}{\partial t}\Gamma_t(y_0)\lambda(t)dt = -g_0' * \Phi_1(y_0) + O(\tau).$$

Return to the van Trees inequality to get that:

$$\begin{aligned}\inf_{\widehat{\Gamma}_n}\sup_{\Gamma\in\mathcal{G}_P(\mathcal{R})} \mathbb{E}(\widehat{\Gamma}_n - \Gamma(y_0))^2 &\geq \left(-g_0' * \Phi_1(y_0) + O(\tau)\right)^2 \left(n\mathbb{E}_\lambda(I) + \frac{J_0}{\tau^4}\right)^{-1} \\ &\geq \left(-g_0' * \Phi_1(y_0) + O(\tau)\right)^2 \left[n\frac{\mathcal{C}_\tau^2}{4\pi^2}\left(\int f_0(y)^2 h_0(y)dy\right)\tau^{-2}e^{-1/\tau^2}(1+o(1)) + \frac{J_0}{\tau^4}\right]^{-1}.\end{aligned}$$

Choose $\tau^{-1} = \sqrt{\log n}$ to obtain the desired result.

□

Proof. Proof of Proposition 2.3.1.
We denote by Ψ_g the Laplace transform of g: $\Psi_g(t) = \int e^{tx}g(x)dx$. The difference between σ and its estimator writes

$$\begin{aligned}\hat{\sigma}_{L,n}^2 - \sigma^2 &= \frac{2}{t_n^2}\log\left(\frac{1}{n}\sum_{j=1}^n e^{t_nY_j - \sigma^2t_n^2/2}\right) \\ &= \frac{2}{t_n^2}\log\left[1 + \frac{1}{n}\sum_{j=1}^n\left(\frac{e^{t_nX_j}}{\Psi_g(t_n)}e^{t_n\varepsilon_j - \sigma^2t_n^2/2} - 1\right)\right] + \frac{2}{t_n^2}\log\Psi_g(t_n) \\ &= \frac{2}{t_n^2}\log(1+S_n) + \frac{2}{t_n^2}\log\Psi_g(t_n) \qquad (2.19)\end{aligned}$$

where S_n is the empirical mean of independent, identically distributed centered random variables:

$$S_n \equiv \frac{1}{n}\sum_{j=1}^n\left(e^{t_nX_j}\Psi_g(t_n)^{-1}e^{t_n\varepsilon_j - \sigma^2t_n^2/2} - 1\right) \equiv \frac{1}{n}\sum_{j=1}^n U_{j,n}. \qquad (2.20)$$

In (2.19), the second term in the right hand side is deterministic and converges to zero as n tends to infinity (see the definition (2.11) of the set $\mathcal{L}_{v,M}$). Moreover

$$\mathrm{Var}(S_n) = \frac{1}{n}\left[\mathrm{E}\left(\frac{e^{2t_nX_1}}{\Psi_g(t_n)^2}e^{2t_n\varepsilon_1-\sigma^2t_n^2}\right)-1\right] \leq \frac{1}{n}\times\frac{\Psi_g(2t_n)}{\Psi_g(t_n)^2}e^{\sigma^2t_n^2}.$$

Using the definition of $\mathcal{L}_v$ (see (2.11)), we obtain that for all integer n:

$$e^{-Mt_n^2v(t_n)} \leq \Psi_g(t_n) \leq e^{Mt_n^2v(t_n)}$$

and then

$$\mathrm{Var}(S_n) \leq \frac{1}{n}e^{(\sigma^2+2Mv(t_n)+4Mv(2t_n))t_n^2} \leq \frac{e^{2t_n^2\Sigma^2}}{n}$$

for sufficiently large n (depending on M), and all $\sigma \leq \Sigma$.
With our choice of the parameters $(t_n)_{n\geq 0}$, we get that S_n converges in the $\mathbb{L}_2(\mathbb{R})$-norm, and also in probability (uniformly with respect to $\sigma \in [0;\Sigma]$ and to g in $\mathcal{L}_{v,M}$). Then equality (2.19) leads to the convergence of $\hat{\sigma}^2_{L,n}$ to σ^2 in probability under (σ, g), uniformly in σ in $[0;\Sigma]$ and in g in $\mathcal{L}_{v,M}$.
Now we compute its rate of convergence in the $\mathbb{L}_2(\mathbb{R})$-norm. We first restrict our attention to the behaviour of the term $\log(1+S_n)$ appearing in (2.19). Note that

$$\mathbb{E}[\log(1+S_n)]^2 = \mathbb{E}\left[\int_0^1 \frac{S_n}{1+tS_n}dt\right]^2 \leq \int_0^1 \mathbb{E}\left(\frac{S_n^2}{(1+tS_n)^2}\right)dt$$

Using the convexity of the function $x \mapsto 1/x^2$ on $\mathbb{R}^{+*}$, we get:

$$\frac{1}{(1+tS_n)^2} = \left[\frac{1}{n}\sum_{j=1}^n(1+tU_{j,n})\right]^{-2} \leq \frac{1}{n}\sum_{j=1}^n\frac{1}{(1+tU_{j,n})^2},$$

so that:

$$\mathbb{E}\left(\frac{S_n^2}{(1+tS_n)^2}\right) \leq \mathbb{E}\left(\frac{S_n^2}{(1+tU_{1,n})^2}\right) = \frac{1}{n^2}\sum_{1\leq k,l\leq n}\mathbb{E}\left(\frac{U_{k,n}U_{l,n}}{(1+tU_{1,n})^2}\right).$$

Now, if k differs from l and is greater or equal to 2, we have:

$$\mathbb{E}\left(\frac{U_{k,n}U_{l,n}}{(1+tU_{1,n})^2}\right) = \mathbb{E}(U_{k,n})\mathbb{E}\left(\frac{U_{l,n}}{(1+tU_{1,n})^2}\right) = 0.$$

So that the sum reduces to:

$$\sum_{1\leq k,l\leq n}\mathbb{E}\left(\frac{U_{k,n}U_{l,n}}{(1+tU_{1,n})^2}\right) = \mathbb{E}\left(\frac{U_{1,n}^2}{(1+tU_{1,n})^2}\right) + (n-1)\mathbb{E}(U_{1,n}^2)\mathbb{E}\left(\frac{1}{(1+tU_{1,n})^2}\right).$$

Returning to the upper bound on $\mathbb{E}[\log(1+S_n)]^2$, we get:

$$\begin{aligned}\mathbb{E}[\log(1+S_n)]^2 &\leq \int_0^1\left[\frac{1}{n^2}\mathbb{E}\left(\frac{U_{1,n}^2}{(1+tU_{1,n})^2}\right) + \frac{1}{n}\left(1-\frac{1}{n}\right)\mathbb{E}(U_{1,n}^2)\mathbb{E}\left(\frac{1}{(1+tU_{1,n})^2}\right)\right]dt \\ &\leq \frac{1}{n^2}\int_0^1\mathbb{E}\left(\frac{U_{1,n}^2}{(1+tU_{1,n})^2}\right)dt + \frac{1}{n}\left(1-\frac{1}{n}\right)\mathbb{E}(U_{1,n}^2)\int_0^1\mathbb{E}\left(\frac{1}{(1+tU_{1,n})^2}\right)dt. \quad (2.21)\end{aligned}$$

Remember the definition of the random variable $U_{1,n}$ (see (2.20)) to get that:

$$\mathbb{E}(U_{1,n}^2) \leq \Psi_g(2t_n)\Psi_g(t_n)^{-2} \leq e^{2t_n^2\Sigma^2}. \tag{2.22}$$

Moreover

$$\begin{aligned}\int_0^1 \mathbb{E}\left(\frac{1}{(1+tU_{1,n})^2}\right)dt = \mathbb{E}\int_0^1 \frac{dt}{1+tU_{1,n}}^2 = \mathbb{E}\left[\frac{-1}{U_{1,n}(1+tU_{1,n})}\right]_{t=0}^1 \\ = \mathbb{E}\left(\frac{1}{U_{1,n}}\right) - \mathbb{E}\left(\frac{1}{U_{1,n}(1+U_{1,n})}\right) = \mathbb{E}\left(\frac{1}{1+U_{1,n}}\right) \\ = \mathbb{E}\left(\Psi_g(t_n)e^{-t_n(X_1+\varepsilon_1)+\sigma^2t_n^2/2}\right) = \Psi_g(t_n)\Psi_g(-t_n)e^{\sigma^2t_n^2} \leq e^{2t_n^2\Sigma^2}, \quad (2.23)\end{aligned}$$

for sufficiently large n (depending on M), and all $\sigma \leq \Sigma$. And at least:

$$\begin{aligned}\int_0^1 \mathbb{E}\left(\frac{U_{1,n}^2}{(1+tU_{1,n})^2}\right)dt = \mathbb{E}\left[\frac{-U_{1,n}}{(1+tU_{1,n})}\right]_{t=0}^1 = \mathbb{E}\left(U_{1,n}\right) - \mathbb{E}\left(\frac{U_{1,n}}{(1+U_{1,n})}\right) \\ = -\mathbb{E}\left(\frac{U_{1,n}}{1+U_{1,n}}\right) = \mathbb{E}\left[\Psi_g(t_n)e^{-t_n(X_1+\varepsilon_1)+\sigma^2t_n^2/2}\left(1-e^{t_nX_1+t_n\varepsilon_1+\sigma^2t_n^2/2}\Psi_g(t_n)^{-1}\right)\right] \\ = \Psi_g(t_n)\Psi_g(-t_n)e^{\sigma^2t_n^2} - e^{\sigma^2t_n^2} \leq e^{2t_n^2\Sigma^2}, \quad (2.24)\end{aligned}$$

for sufficiently large n (depending on M), and all $\sigma \leq \Sigma$. Combining the upper bounds (2.22), (2.23) and (2.24) with the inequality (2.21), we get that for sufficiently large n (depending on M), and all $\sigma \leq \Sigma$:

$$\mathbb{E}[\log(1+S_n)]^2 \leq \frac{e^{2t_n^2\Sigma^2}}{n^2} + \frac{1}{n}\left(1-\frac{1}{n}\right)e^{4t_n^2\Sigma^2} \leq \frac{e^{4t_n^2\Sigma^2}}{n} \leq \frac{1}{n^{1-\alpha}}.$$

Returning to the equality (2.19) and using a Cauchy-Schwarz inequality combined with the preceding result, we get:

$$\begin{aligned}\sup_{\sigma\in[0;\Sigma]}\sup_{g\in\mathcal{L}_{v,M}}\left[\mathbb{E}(\hat{\sigma}_{L,n}-\sigma^2)^2\right]^{1/2} &\leq \frac{2}{t_n^2}\left[\mathbb{E}(\log(1+S_n))^2\right]^{1/2} + 2Mv(t_n) \\ &\leq \frac{8\Sigma^2}{\alpha(\log n)n^{(1-\alpha)/2}} + 2Mv(t_n),\end{aligned}$$

which achieves the proof.

□

Proof. Proof of Proposition 2.3.3.
The proof is essentially the same as in the proof of Proposition 2.3.1. We denote by g^* the Fourier transform of g: $g^*(t) = \int e^{itx}g(x)dx$. The difference between σ and its estimator writes

$$\begin{aligned}\hat{\sigma}_{F,n}^2 - \sigma^2 &= -\frac{2}{t_n^2}\log\left(\frac{1}{n}\sum_{j=1}^n e^{it_nY_j+\sigma^2t_n^2/2}\right) \\ &= -\frac{2}{t_n^2}\log\left[1+\frac{1}{n}\sum_{j=1}^n\left(\frac{e^{it_nX_j}}{g^*(t_n)}e^{it_n\varepsilon_j+\sigma^2t_n^2/2}-1\right)\right] - \frac{2}{t_n^2}\log g^*(t_n) \\ &= -\frac{2}{t_n^2}\log(1+\bar{S}_n) - \frac{2}{t_n^2}\log g^*(t_n) \quad (2.25)\end{aligned}$$

where $\bar{S}_n$ is the empirical mean of independent, identically distributed centered random variables:

$$\bar{S}_n \equiv \frac{1}{n}\sum_{j=1}^{n}\left(\frac{e^{it_nX_j}}{g^*(t_n)}e^{it_n\varepsilon_j+\sigma^2t_n^2/2}-1\right) \equiv \frac{1}{n}\sum_{j=1}^{n}\bar{U}_{j,n}. \tag{2.26}$$

In (2.25), the second term in the right hand side is deterministic and converges to zero as n tends to infinity (see the definition (2.12) of the set $\mathcal{F}_v$), and we will focus on the first one. We have

$$\mathrm{Var}(\bar{S}_n) = \frac{1}{n}\left(\frac{e^{\sigma^2t_n^2}}{|g^*(t_n)|^2}-1\right) \leq \frac{e^{\sigma^2t_n^2}}{n|g^*(t_n)|^2}.$$

Using the definition of $\mathcal{F}_v$ (see (2.12)), we obtain the existence of some constant M depending on g such that for large enough n:

$$e^{-Mt_n^2v(t_n)} \leq |g^*(t_n)| \leq e^{Mt_n^2v(t_n)}.$$

It leads to an upper-bound on the quantity $\mathrm{Var}(\bar{S}_n)$:

$$\mathrm{Var}(\bar{S}_n) \leq \frac{1}{n}e^{(\sigma^2+2Mv(t_n))t_n^2} \leq \frac{e^{2t_n^2\Sigma^2}}{n}$$

for sufficiently large n depending on M, and all $\sigma \leq \Sigma$.
We choose the sequence $(t_n)_{n\geq 0}$ in the following way

$$t_n = \sqrt{\frac{\alpha}{2\Sigma^2}\log(n)}$$

for some $0 < \alpha < 1$, so that we get, for all g in $\mathcal{F}_v$

$$\sup_{\sigma\in[0;\Sigma]} \mathrm{Var}(\bar{S}_n) \underset{n\to\infty}{\longrightarrow} 0.$$

In particular, we get that $\bar{S}_n$ converges to zero in $\mathbb{L}_2(\mathbb{R})$ and in probability. Combining with (2.25), we get that for all g in $\mathcal{F}_v$ and uniformly with respect to σ in $[0;\Sigma]$, the estimator $\hat{\sigma}^2_{F,n}$ converges to σ^2 in probability under (σ, g).

□

Proof. Proof of Theorem 2.4.1.
We have to choose the parameters to ensure that $\hat{\sigma}_n$ will converge (in a sense to be specified) to the true value of the parameter σ. We will first study the quantity

$$\hat{\alpha}_n(t;\tau) - \alpha(t;\tau)$$

that represents the difference between the value of α and its estimator. We have:

$$\hat{\alpha}_n(t;\tau) - \alpha(t;\tau) = e^{t^2\tau^2/2}\frac{1}{\sqrt{n}}\mathbb{G}_n(f_t)$$

where $\mathbb{G}_n$ is the empirical process associated to the observations (that is to say $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P_{\sigma,g})$ where $\mathbb{P}_n$ is the empirical probability measure for the observations: $\mathbb{P}_n(f) = (1/n)\sum_{i=1}^n f(Y_i)$), and $f_t(u) = e^{itu}$. This leads to the upper-bound:

$$\sup_{u:\|u\|=1}\sup_{\tau\in[0;\Sigma]}\left|\sum_{-\ell_n\leq k,l\leq\ell_n} u_k\left[\hat{\alpha}_n(t_{k,n}-t_{l,n};\tau) - \alpha(t_{k,n}-t_{l,n};\tau)\right]\bar{u}_l\right| \leq 2\ell_n\Big(\sup_{|t|\leq\frac{\ell_n}{k_n}}|\mathbb{G}_nf_t|\Big)\frac{e^{2\ell_n^2\Sigma^2/k_n^2}}{\sqrt{n}}$$

We impose the following condition on the parameters

$$\frac{\ell_n}{k_n} = \frac{1}{\Sigma}\sqrt{\frac{a \log n}{2}} \quad \text{where } 0 < a < 1/2, \tag{2.27}$$

so that we get: $e^{2\ell_n^2\Sigma^2/k_n^2} = n^a$. It leads to

$$\sup_{u,\tau}\left|\sum_{-\ell_n\le k,l\le \ell_n} u_k[\hat{\alpha}_n(t_{k,n}-t_{l,n};\tau)-\alpha(t_{k,n}-t_{l,n};\tau)]\bar{u}_l\right| \le \frac{\sqrt{2a}}{\Sigma}\frac{\sqrt{\log n}k_n}{n^{1/2-a}}\Big(\sup_{|t|\le\frac{\ell_n}{k_n}}|\mathbb{G}_n f_t|\Big), \tag{2.28}$$

and we impose

$$k_n \le n^{1/2-a-b} \text{ where } b > 0$$

so that the quantity

$$\frac{\sqrt{2a}}{\Sigma}\frac{\sqrt{\log n}k_n}{n^{1/2-a}}$$

converges to zero. Recall that the parameter ϵ_n is chosen as

$$\epsilon_n = \frac{\sqrt{2a}}{\Sigma}\frac{\sqrt{\log n}k_n}{n^{1/2-a}} \times v_n,$$

where v_n is an increasing sequence of numbers converging to infinity such that ϵ_n tends to zero:

$$v_n \underset{n\to\infty}{\longrightarrow} \infty \quad , \quad \epsilon_n \underset{n\to\infty}{\longrightarrow} 0.$$

In fact, we choose the sequence $(v_n)_{n\ge 0}$ increasing to infinity, under the constraint:

$$v_n = o\left(\frac{n^{1/2-a}}{\sqrt{\log n}k_n}\right).$$

This choice of the parameters is made in order to ensure the convergence to zero of the term:

$$\Delta_n = P_{\sigma,g}\left(\sup_{u:\|u\|=1}\sup_{\tau\in[0;\Sigma]}\left|\sum_{-\ell_n\le k,l\le\ell_n} u_k[\alpha(t_{k,n}-t_{l,n};\tau)-\hat{\alpha}(t_{k,n}-t_{l,n};\tau)]\bar{u}_l\right| \ge \epsilon_n\right). \tag{2.29}$$

Let us prove this convergence. Using the domination (2.28), we have:

$$\Delta_n \le P_{\sigma,g}\left(\frac{\sqrt{2a}}{\Sigma}\frac{\sqrt{\log n}k_n}{n^{1/2-a}}\Big(\sup_{|t|\le\frac{\ell_n}{k_n}}|\mathbb{G}_n f_t|\Big) \ge \epsilon_n\right).$$

So that,

$$\Delta_n \le P_{\sigma,g}\left(\sup_{|t|\le\frac{\ell_n}{k_n}}|\mathbb{G}_n f_t| \ge v_n\right) \le \frac{1}{v_n}\mathbb{E}_{\sigma,g}\left(\sup_{|t|\le\frac{\ell_n}{k_n}}|\mathbb{G}_n f_t|\right). \tag{2.30}$$

Now we use a maximal inequality to control the mean of the empirical process. The following notations can be found in more details in van der Vaart and Wellner (1996). We consider the class of functions $\mathcal{F}_n$ defined by $\{f_t; |t| \le \ell_n/k_n\}$ (note that this class has an envelope function equal to one). The complexity of this family stands in its entropy defined through the bracketing numbers for this class. Theorem 2.7.11 in van der Vaart and Wellner (1996) applies in our context: denote by $F(x) = 2|x|$ the function such that for all s, t in $T_n = [-\ell_n/k_n; \ell_n/k_n]$

$$|f_t(x) - f_s(x)| = |e^{itx} - e^{isx}| \leq |s-t|F(x).$$

This theorem asserts that the bracketing numbers for the class $\mathcal{F}_n$ (that means the minimal number of brackets of size ϵ needed to cover $\mathcal{F}_n$) are controlled by the covering numbers of T_n (i.e the minimal number of balls of radius ϵ needed to cover T_n):

$$N_{[\,]}(\epsilon; \mathcal{F}_n; \|\cdot\|_{P_{\sigma,g},2}) \leq N\left(\frac{\epsilon}{2\|F\|_{P_{\sigma,g},2}}; T_n; |\cdot|\right)$$

(here, $\|\cdot\|_{P_{\sigma,g},2}$ denotes the $\mathbb{L}_2(\mathbb{R})$-norm under the measure $P_{\sigma,g}$). But it is easy to bound the covering numbers for T_n

$$N\left(\frac{\epsilon}{2\|F\|_{P_{\sigma,g},2}}; T_n; |\cdot|\right) \leq \frac{4\ell_n\|F\|_{P_{\sigma,g},2}}{k_n\epsilon}.$$

So that we obtain the following control on the bracketing numbers for the class $\mathcal{F}_n$

$$N_{[\,]}(\epsilon; \mathcal{F}_n; \|\cdot\|_{P_{\sigma,g},2}) \leq \frac{4\ell_n\|F\|_{P_{\sigma,g},2}}{k_n\epsilon}. \tag{2.31}$$

Let us define the entropy of this class by the formula:

$$J_{[\,]}(1; \mathcal{F}_n; \|\cdot\|_{P_{\sigma,g},2}) = \int_0^1 \sqrt{1 + \log N_{[\,]}(\epsilon; \mathcal{F}_n; \|\cdot\|_{P_{\sigma,g},2})}d\epsilon. \tag{2.32}$$

Now we apply Theorem 2.14.2 in van der Vaart and Wellner (1996): there exist a universal constant C such that

$$\mathbb{E}_{\sigma,g}\left(\sup_{|t|\leq\frac{\ell_n}{k_n}} |\mathbb{G}_n f_t|\right) \leq CJ_{[\,]}(1; \mathcal{F}_n; \|\cdot\|_{P_{\sigma,g},2}).$$

Combining with the definition of the entropy (2.32), with inequality (2.31) and using (2.27), we obtain that there exist some constant κ such that

$$\mathbb{E}_{\sigma,g}\left(\sup_{|t|\leq\frac{\ell_n}{k_n}} |\mathbb{G}_n f_t|\right) \leq \kappa(\log\log n)^{1/2}.$$

Now return to the quantity Δ_n defined in (2.29) and to its upper-bound (2.30):

$$\Delta_n \leq \frac{1}{v_n}\mathbb{E}_0\left(\sup_{|t|\leq\frac{\ell_n}{k_n}} |\mathbb{G}_n f_t|\right) \leq \frac{\kappa}{v_n}(\log\log n)^{1/2}. \tag{2.33}$$

Recall that the parameter v_n satisfies the constraint

$$v_n^{-1} = o((\log\log n)^{-1/2}),$$

in order to get the convergence of Δ_n to zero:

$$\Delta_n \underset{n\to\infty}{\longrightarrow} 0. \tag{2.34}$$

We will now prove the consistency of the estimator $\hat{\sigma}_n$. We start by studying what happens if $\hat{\sigma}_n$ is greater than or equal to $\sigma+\epsilon$ for some arbitrary $\epsilon>0$. Using the definition of $\hat{\sigma}_n$ (see (2.15))

$$P_{\sigma,g}(\hat{\sigma}_n \geq \sigma+\epsilon) \quad \leq \quad P_{\sigma,g}\left(\inf_{u:\|u\|=1} \sum_{-\ell_n\leq k,l\leq \ell_n} u_k \hat{\alpha}_n(t_{k,n}-t_{l,n};\sigma+\epsilon)\bar{u}_l \geq -\epsilon_n\right).$$

But this quantity involves the estimator $\hat{\alpha}$ whereas the real parameter α is more tractable. Using the definition of Δ_n (see(2.29)) we write:

$$P_{\sigma,g}(\hat{\sigma}_n \geq \sigma+\epsilon) \leq P_{\sigma,g}\left(\inf_{u:\|u\|=1} \sum_{-\ell_n\leq k,l\leq \ell_n} u_k \alpha(t_{k,n}-t_{l,n};\sigma+\epsilon)\bar{u}_l \geq -2\epsilon_n\right) + \Delta_n. \tag{2.35}$$

But Δ_n tends to zero as n tends to infinity. We will show that the first term in the right hand side is null for n large enough. At the point $\sigma+\epsilon$, the function $\alpha(\cdot;\sigma+\epsilon)$ is not positive definite, and then there exists n_0 in $\mathbb{N}$, $(\zeta_1,\dots\zeta_{n_0})$ in $\mathbb{R}^{n_0}$, and $u^0=(u_1^0,\dots u_{n_0}^0)$ in $\mathbb{C}^{n_0}$ with $\|u^0\|=1$ such that

$$\sum_{1\leq k,j\leq n_0} u_k^0\alpha(\zeta_k-\zeta_j;\sigma+\epsilon)\bar{u}_j^0 < 0.$$

But with the choice (2.27) for the parameters, for n large enough, we obtain that for all k in $\{1;\dots n_0\}$, the point ζ_k satisfies $|\zeta_k|\leq \frac{\ell_n}{k_n}$, and there exists some $\phi_n(k)$ with

$$\left|\zeta_k - t_{\phi_n(k),n}\right| \leq \frac{1}{k_n}.$$

Moreover,

$$\begin{aligned}
\frac{\partial}{\partial\zeta}\alpha(\zeta;\tau) &= e^{\zeta^2\tau^2/2}\left(\zeta\tau^2\int e^{i\zeta u}h(u)du + \int iue^{i\zeta u}h(u)du\right)\\
\left|\frac{\partial}{\partial\zeta}\alpha(\zeta;\tau)\right| &\leq e^{\zeta^2\Sigma^2/2}\left(|\zeta|\Sigma^2+\mathbb{E}_{\sigma,g}|Y_1|\right)
\end{aligned}$$

and then for all ζ,ζ' belonging to the interval $[-\ell_n/k_n;\ell_n/k_n]$, we get

$$|\alpha(\zeta;\tau)-\alpha(\zeta';\tau)| \leq |\zeta-\zeta'|e^{2\ell_n^2\Sigma^2/k_n^2}\left(\frac{\ell_n}{k_n}\Sigma^2+\mathbb{E}_{\sigma,g}|Y_1|\right)$$

and then

$$\begin{aligned}
|\alpha(\zeta_k-\zeta_j;\tau)-\alpha(t_{\phi_n(k),n}-t_{\phi_n(j),n};\tau)| &\leq \frac{2}{k_n}e^{2\ell_n^2\Sigma^2/k_n^2}\left(\frac{\ell_n}{k_n}\Sigma^2+\mathbb{E}_{\sigma,g}|Y_1|\right)\\
\left|\sum_{1\leq k,j\leq n_0} u_k^0\left[\alpha(\zeta_k-\zeta_j;\tau)-\alpha(t_{\phi_n(k),n}-t_{\phi_n(j),n};\tau)\right]\bar{u}_j^0\right| &\leq Cn_0\frac{n^a\sqrt{\log n}}{k_n}
\end{aligned}$$

where C is a constant. We choose

$$k_n = n^{1/2-a-b} \text{ where } \left(\frac{1}{2} - a - b\right) > a \text{ ; i.e } 2a + b < \frac{1}{2}$$

(note that this implies in fact that $a < 1/4$). Then we get

$$\sup_{\tau \in]0;\Sigma]} \left| \sum_{1 \leq k,j \leq n_0} u_k^0 \left[\alpha(\zeta_k - \zeta_j; \tau) - \alpha(t_{\phi_n(k),n} - t_{\phi_n(j),n}; \tau) \right] \bar{u}_j^0 \right| \underset{n\to\infty}{\longrightarrow} 0.$$

So we can conclude that the first term in (2.35) is bounded in the following way, for large enough n

$$P_{\sigma,g}\left(\inf_{u:\|u\|=1} \sum_{-\ell_n \leq k,l \leq \ell_n} u_k \alpha(t_{k,n} - t_{l,n}; \sigma + \epsilon) \bar{u}_l \geq -2\epsilon_n \right) \\ \leq P_{\sigma,g}\left(\sum_{-\ell_n \leq k,l \leq \ell_n} v_k \alpha(t_{k,n} - t_{l,n}; \sigma + \epsilon) \bar{v}_l \geq -2\epsilon_n \right)$$

where

$$v_k = \begin{cases} u_j^0 \text{ when } k = \phi_n(j) \text{ and } 1 \leq j \leq n_0, \\ 0 \text{ otherwise.} \end{cases}$$

and $\|v\|^2 = \sum_{k=-\ell_n}^{\ell_n} v_k^2 = \sum_{j=0}^{n_0} u_j^2 = 1$. So we have

$$\begin{aligned} & P_{\sigma,g}\left(\inf_{u:\|u\|=1}; \sum_{-\ell_n \leq k,l \leq \ell_n} u_k \alpha(t_{k,n} - t_{l,n}; \sigma + \epsilon) \bar{u}_l \geq -2\epsilon_n \right) \\ \leq \quad & P_{\sigma,g}\left(\sum_{1 \leq k,l \leq n_0} u_k^0 \alpha(\zeta_k - \zeta_l; \sigma + \epsilon) \bar{u}_l^0 \right. \\ & \left. \geq -2\epsilon_n + \sum_{1 \leq k,l \leq n_0} u_k^0 \left[\alpha(\zeta_k - \zeta_l; \sigma + \epsilon) - \alpha(t_{\phi_n(k),n} - t_{\phi_n(l),n}; \sigma + \epsilon) \right] \bar{u}_l^0 \right). \end{aligned}$$

We are interested in what occurs in the probability appearing in the right hand side of this inequality. All the quantities are deterministic. The first one is negative and the second one converges to zero: for n large enough, the probability of this event is null. So, for n large enough (depending on ϵ), we have

$$P_{\sigma,g}(\hat{\sigma}_n - \sigma \geq \epsilon) \leq \Delta_n \underset{n\to\infty}{\longrightarrow} 0.$$

We are studying now the probability that $\hat{\sigma}_n$ would be less than or equal to $\sigma - \epsilon$ for some arbitrary $\epsilon > 0$.

$$\begin{aligned}
& P_{\sigma,g}(\hat{\sigma}_n - \sigma \leq -\epsilon) \\
\leq \quad & P_{\sigma,g}\left(\exists \sigma_n \leq \sigma - \epsilon, \exists u \in \mathbb{C}^{2\ell_n}, \|u\| = 1, \sum_{-\ell_n \leq k,l \leq \ell_n} u_k \hat{\alpha}_n(t_{k,n} - t_{l,n}; \sigma_n)\bar{u}_l < -\epsilon_n\right) \\
\leq \quad & P_{\sigma,g}\left(\exists \sigma_n \leq \sigma - \epsilon, \exists u \in \mathbb{C}^{2\ell_n}, \|u\| = 1, \sum_{-\ell_n \leq k,l \leq \ell_n} u_k \alpha(t_{k,n} - t_{l,n}; \sigma_n)\bar{u}_l < 0\right) \\
& + P_{\sigma,g}\left(\sup_{\tau \in [0;\Sigma]} \sup_{u; \|u\|=1} \left| \sum_{-\ell_n \leq k,l \leq \ell_n} u_k(\hat{\alpha}_n(t_{k,n} - t_{l,n}; \tau) - \alpha(t_{k,n} - t_{l,n}; \tau))\bar{u}_l \right| > \epsilon_n \right).
\end{aligned}$$

The first term in the right hand side is equal to zero as $\alpha(\cdot, \sigma_n)$ is positive definite for σ_n less or equal to $\sigma - \epsilon$. The second one, Δ_n, as we have already seen, tends to zero as n tends to infinity. In conclusion

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall n \geq N, \quad P_{\sigma,g}(|\hat{\sigma}_n - \sigma| \geq \epsilon) \leq 2\Delta_n \underset{n \to \infty}{\longrightarrow} 0. \tag{2.36}$$

We proved the convergence in probability of our estimator to the true value of the parameter σ. Note that the convergence in the $\mathbb{L}_2(\mathbb{R})$-norm is a consequence of the fact that $\hat{\sigma}_n$ is almost surely bounded.

□

Chapitre 3

Minimax estimation of some linear functionals in the convolution model *

Résumé

Dans le modèle de convolution $Y = X + \varepsilon$, où le signal X est de densité g inconnue et le bruit ε est de loi normale centrée réduite, nous nous intéressons aux propriétés d'estimation, à partir d'un échantillon $Y_1, ..., Y_n$, de fonctionnelles linéaires intégrales de g de la forme $\Gamma_f(y) = \int f(x)\Phi_1(y-x)g(x)dx$ où Φ_1 désigne la densité du bruit ε et f est une fonction connue. Nous étendons les résultats de Taupin (2001, 1998), dans le cas où la fonction f est soit un polynôme soit une fonction trigonométrique, en établissant des minorations du risque quadratique ponctuel, du risque par rapport à la norme de $\mathbb{L}_\infty(\mathbb{R})$, ainsi que des majorations et minorations du risque par rapport à la norme de $\mathbb{L}_p(\mathbb{R})$ lorsque $1 \leq p < \infty$. Nous montrons que l'estimateur proposé par Taupin (2001) atteint les vitesses optimales dans le cas où f est un polynôme et est presque minimax dans le cas où f est une fonction trigonométrique, avec une perte en $(\log n)^{1/4}$ pour le risque quadratique et $(\log n)^{3/4}/\sqrt{\log\log n}$ pour le risque en norme $\mathbb{L}_\infty(\mathbb{R})$.

Contents

*joint work with Marie-Luce TAUPIN : *CNRS UMR C 8628. Équipe de probabilités, statistique et modélisation. Bât 425. Université Paris Sud. 91405 Orsay Cedex, France. E-mail :* Marie-Luce.Taupin@math.u-psud.fr

3.1 Introduction

This part is devoted to the estimation of some linear integral functionals of a density in the convolution model. Consider two sequences of independent and identically distributed real random variables $\{X_k\}_{k\geq 1}$ and $\{\varepsilon_k\}_{k\geq 1}$, the first one with unknown density g with respect to the Lebesgue measure on $\mathbb{R}$, the second one being Gaussian distributed, with zero mean and variance equal to one. Assume moreover that these two sequences are independent. Our statistical model is based on the observation of the sequence $\{Y_k\}_{k\geq 1}$ given by the relation

$$Y_k = X_k + \varepsilon_k \ , \quad k \geq 1.$$

Due to the independence between X_n and ε_n, the density h of the observation Y_n is simply the convolution product $h = g * \Phi_1$, where Φ_1 stands for the standard Gaussian density.

In this model we aim at estimating linear integral functionals of the unknown density g of the following form:

$$\Gamma_f(y) = \int f(x)g(x)\Phi_1(x-y)dx \ , \ \forall y \in \mathbb{R}. \tag{3.1}$$

We focus here on the cases of f being either polynomial or a trigonometric function with a particular interest on the special case $f \equiv 1$ corresponding to the density h of the observations also denoted by:

$$\Gamma_0(y) = \int g(x)\Phi_1(x-y)dx = g * \Phi_1(x) = h(y) \ , \ \forall y \in \mathbb{R}.$$

Let us first motivate the interest in those functionals. Consider a nonlinear errors-in-variables regression model described by the observation of the random variables $\{(Y_k, Z_k)\}_{1\leq k\leq n}$ satisfying the relations:

$$\begin{cases} Z_k &= f_\beta(X_k) + \eta_k \\ Y_k &= X_k + \varepsilon_k \end{cases} , \qquad \forall k \geq 1,$$

where the function f is known up to the finite dimensional parameter β, the errors $\{(\eta_n, \varepsilon_n)\}_{n\geq 0}$ are centered, independent and identically distributed with respective variances σ_η^2 and $\sigma_\varepsilon^2 = 1$ for sake of simplicity. We assume furthermore that the errors ε_n are normally distributed. The sequence $\{X_n\}_{n\geq 0}$ is not observed and is a sequence of independent identically distributed random variables, with unknown density g. Moreover, the sequences $\{X_n\}_{n\geq 0}$, $\{\eta_n\}_{n\geq 0}$ and $\{\varepsilon_n\}_{n\geq 0}$ are independent. In this errors-in-variables regression model, the purpose is to estimate the parameter β in the presence of the unknown density g of the unobserved variables considered as a nuisance parameter. In this context, Taupin (2001) proposes an estimator of this parameter β based on the criterion

$$n^{-1}\sum_{i=1}^{n} W(Y_i)[Z_i - \mathbb{E}(f_\beta(X_i)|Y_i)]^2,$$

W being a compactly supported weight function, and where the conditional expectation $\mathbb{E}(f_\beta(X_i)|Y_i)$ is replaced by a non-parametric estimation based on the sample $Y_1, .., Y_n$. Denoting by X, Y and ε the generic variables, and using the independence between X and ε, the conditional expectation $\mathbb{E}(f_\beta(X)|Y=y)$ can be written in the following form:

$$\mathbb{E}(f_\beta(X)|Y=y) = \frac{\int f_\beta(x)g(x)\Phi_1(x-y)dx}{\int g(x)\Phi_1(x-y)dx} = \frac{\Gamma_{f_\beta}(y)}{h(y)}.$$

This conditional expectation is estimated by estimating separately the numerator and the denominator. Taupin (2001) constructed an estimator of this functional for general functions f and gave the corresponding upper bounds on the pointwise quadratic risk and on the risk with respect to uniform norm for various classes of functions f such as polynomial functions, exponential functions: $f(x) = \exp\{\beta x\}$ or $f(x) = \exp\{\beta x^2\}$ where β is a fixed real number, sums of cosines: $f(x) = \sum_{j=0}^{\ell} \beta_j \cos(jx)$ for fixed parameters $(\beta_j)_{0\le j\le \ell}$, and more generally functions f admitting an analytic continuation into the strip $\mathcal{B}_\gamma = \{x + iy \; ; \; (x, y) \in \mathbb{R}^2 \; , \; |y| \le \gamma\}$ and functions f belonging to Sobolev classes.

The upper bound for the rate of convergence of the estimator of β often depends on the pointwise asymptotic properties of the estimator of Γ_f and also on those with respect to the $\mathbb{L}_\infty(\mathbb{R})$-norm. Consequently it is interesting to extend the study about linear functionals, and especially to provide upper and lower bounds for the risk with respect to different loss functions.

Let us be more precise. Consider the following sets: $\mathcal{H}$ is the set of possible densities h with respect to the Lebesgue measure on $\mathbb{R}$ in this model:

$$\mathcal{H} = \left\{ g * \Phi_1 \; ; \; g \ge 0, \int_{\mathbb{R}} g(x)dx = 1 \right\},$$

$\mathcal{G}$ is the set of probability densities g with respect to the Lebesgue measure on $\mathbb{R}$:

$$\mathcal{G} = \left\{ g \; ; \; g \ge 0, \int_{\mathbb{R}} g(x)dx = 1 \right\},$$

and when f is a fixed function, $\mathcal{G}_f$ is the set of linear functionals of the unknown density g:

$$\mathcal{G}_f = \left\{ \Gamma_f = \int f(x)g(x)\Phi_1(x - \cdot)dx \; ; \; g \in \mathcal{G} \text{ such that } \forall y \in \mathbb{R}, fg\Phi_1(\cdot - y) \in \mathbb{L}_1(\mathbb{R}) \right\}. \quad (3.2)$$

Note that when f is a polynomial function of degree equal to ℓ, the functionals Γ_f belonging to $\mathcal{G}_f$ are defined for densities g having at least ℓ finite moments. This condition relies on the fact that we use the functional Γ_f to estimate $\mathbb{E}(f(X)|Y)$ and then when f is a polynomial function of degree ℓ, we need that $\mathbb{E}(|X|^\ell) < \infty$. When f is a trigonometric function, the functional Γ_f exists whatever the density g is.

Our aim is to obtain upper and lower bounds on the minimax risk for the estimation of Γ_f belonging to the class of functionals $\mathcal{G}_f$. We consider different minimax risks: the pointwise minimax quadratic risk for the estimation of $\Gamma_f(y_0)$ when y_0 is fixed and the minimax risk in $\mathbb{L}_p(\mathbb{R})$-norm $(1 \le p \le \infty)$ for the global estimation of Γ_f.

The pointwise minimax quadratic risk over the set $\mathcal{G}_f$ is defined by:

$$\forall y_0 \in \mathbb{R} \; , \quad R_n(f) = \inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \mathbb{E}^{1/2}[\Gamma_f(y_0) - T_n]^2, \quad (3.3)$$

where the infimum is taken over all the estimators T based on the observations $Y_1, \ldots Y_n$. Moreover, the minimax risk in $\mathbb{L}_p(\mathbb{R})$-norm is defined by

$$\forall \; 1 \le p \le \infty, \quad R_{n,p}(f) = \inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \mathbb{E}\|\Gamma_f - T_n\|_p, \quad (3.4)$$

where the infimum is taken over all the estimators T_n based on the observations $Y_1, \dots Y_n$.

Following Taupin (2001), the functional Γ_f is estimated by

$$\widehat{\Gamma}_{f,n}(y_0) = \frac{1}{2\pi n} \sum_{j=1}^{n} \mathrm{Re} \int [f\Phi_1(y-\cdot)]^\star(t) e^{t^2/2} e^{-itY_j} K_n^*(t) dt,$$

where $K_n = C_n K(C_n \cdot)$ is a kernel to be chosen and $\{C_n\}_{n\geq 0}$ is a sequence converging to infinity. In the same way, h is estimated by the kernel estimator

$$\widehat{h}_n(y_0) = \frac{C_n}{n} \sum_{j=1}^{n} K[C_n(y_0 - Y_j)].$$

For instance, k could be equal to the analogue of the de La Vallée-Poussin kernel V (see Nikol'skii (1969)) with $C_n = \sqrt{\log n}$.

We start with the problem of estimating the density $h = \Phi_1 * g$. Classical calculations on kernel estimation provide the following upper bound for the pointwise estimation of the density h. Applying Ibragimov and Hasminskii's result (see Ibragimov and Hasminskii (1983) and Taupin (2001)) the following result holds

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{(\log n)^{1/4}} \sup_{h\in\mathcal{H}} \mathbb{E}^{1/2}[\widehat{h}_n(y_0) - h(y_0)]^2 \leq \|V\|_2. \tag{3.5}$$

Moreover,

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{(\log n)^{1/4}\sqrt{\log\log n}} \sup_{h\in\mathcal{H}} \mathbb{E} \parallel \widehat{h}_n(y_0) - h \parallel_\infty < +\infty. \tag{3.6}$$

Note that, due to the structure of convolution model, the regularity of the density h is strongly related to the ε's density, which is known and equals the standard Gaussian density. Therefore some drawbacks arising from the use of such a kernel estimator do not appear here. More precisely the structure of the optimal bandwidth (ie : minimizing the maximal risk over the class $\mathcal{H}$) is very simple, we just take $C_n = \sqrt{\log n}$. This means that we have an entirely explicit estimator that achieves the optimal rate of convergence $(\log n)^{1/4}/\sqrt{n}$ either for the pointwise quadratic risk or for the risk with respect to the $\mathbb{L}_p(\mathbb{R})$-norm.

The density $\Phi_1 * g$ has obviously strong smoothness properties. We immediately see for instance that $\Phi_1 * g$ is infinitely differentiable with for every integer $k \geq 0$,

$$\parallel (\Phi_1 * g)^{(k)} \parallel_\infty \leq \parallel \Phi_1^{(k)} \parallel_\infty \tag{3.7}$$

One can go further in the analysis of the problem by considering that $\mathcal{H}$ is embedded in the totally bounded (with respect to the $\mathbb{L}_2(\mathbb{R})$-norm) set of densities h satisfying:

$$\frac{1}{2\pi} \int (\mathrm{ch}^2 \gamma t) |h^\star(t)|^2 dt = \int (\mathrm{Re} f(x + i\gamma))^2 dx \leq C_\gamma.$$

Golubev and Levit (1996) prove that the optimal rate of convergence of the quadratic risk over this class is of order $\sqrt{\log n}/\sqrt{n}$ which implies that the minimax quadratic risk on $\mathcal{H}$ is not greater than $\sqrt{\log n}/\sqrt{n}$. The class $\mathcal{A}_\gamma$ has been first introduced by Ibragimov and Hasminskii (1983) who derived the optimal minimax rates of convergence for density estimation in $\mathbb{L}_p(\mathbb{R})$-norm.

Moreover by the fact that $|(\Phi_1 * g)^\star| \leq \Phi_1^\star$ the class $\mathcal{H}$ is included into the even smaller class $\mathcal{B}_{1,1/2}(2) \subset \mathcal{A}_\gamma$ where the class $\mathcal{B}_{A,\rho}(r)$ is more generally defined as

$$\mathcal{B}_{A,\rho}(r) = \{\phi \,;\, |\phi^\star(t)| \leq Ae^{-\rho|t|^r}\}.$$

These smoothness classes have been considered by Davis (1975) and Davis (1977), and lately Levit[†] announced that the minimax quadratic risk for estimating a density over the class $\mathcal{B}_{A,\rho}(r)$ is of order $(\log n)^{1/(2r)}/\sqrt{n}$. Levit's result implies that the minimax quadratic risk on $\mathcal{H}$ is not greater than $(\log n)^{1/4}/\sqrt{n}$ and using the van Trees inequality as in Golubev and Levit (1996) for analytic densities, Taupin (1998) established a lower bound for the minimax quadratic risk over the class $\mathcal{H}$ which shows that the $(\log n)^{1/4}/\sqrt{n}$ rate cannot be further improved.

The results obtained for the estimation of h lead us to make some remarks about the convolution model.

First it is noteworthy that the class of densities $\mathcal{H}$ comes from the convolution model $Y = X + \varepsilon$, assuming that ε is normally distributed. The results we obtain are available whatever g, the density of X, is. Moreover, since $h = \Phi_1 * g$ belongs to $\mathcal{H}$ the minimax approach is naturally well adapted and we don't need to consider adaptive estimation.

The second remark is about the difference between Taupin's results with those obtained in deconvolution problems. In the last case the purpose is to estimate the density g of X. It is known that the slowest rates of convergence for estimating g are obtained for the smoothest error densities. When the errors are normally distributed, Fan (1991c) proved that the optimal rate of convergence for estimating g is of order $(\log n)^{-1/2}$ and Carroll and Hall (1988) proved that this optimal rate is of order $(\log n)^{-m/2}$ when g admits m bounded derivatives, whereas Taupin (2001) shows that the rate for estimating h is of order $(\log n)^{1/4}/\sqrt{n}$.

We now come to the problem of estimating Γ_f focusing on functions f of type $f(x) = \sum_{k=0}^{\ell} \beta_k x^k$, or $f(x) = \sum_{k=0}^{\ell} \beta_k \cos(kx)$. In the polynomial case, Taupin (2001) established the following upper bounds. Assume that $f(x) = \sum_{k=0}^{\ell} \beta_k x^k$, with $\ell \geq 1$ and $\beta = (\beta_0, ..., \beta_\ell)$ is a fixed $(\ell+1)$-tuple of real numbers. For any y_0 in $\mathbb{R}$,

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}} \mathbb{E}^{1/2}[\widehat{\Gamma}_{f,n}(y_0) - \Gamma_f(y_0)]^2 < \infty,$$

and if Z_C denotes any compact subset of $\mathbb{R}$

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{(\log n)^{(2\ell+3)/4}} \mathbb{E}[\sup_{y\in Z_c} |\widehat{\Gamma}_{f,n}(y) - \Gamma_f(y)|] < \infty.$$

Note that, due to the fact that the errors ε are Gaussian, the functional Γ_f when f is a polynomial function, can be related to the derivatives of h up to order ℓ (see Lemma 3.2.1). Therefore the problem of estimating Γ_f is related to the problem of estimating the derivatives

[†]May 1996, talk at the Ecole Normale in Paris (ULM)

of the density h up to order ℓ. As a consequence, the comments made just above about the regularity of the density h still hold for Γ_f.

We now come to the previous result about pointwise estimation of Γ_f, when f is a linear combination of cosines. Taupin (2001) proved the following upper bounds. Assume that $f(x) = \sum_{k=0}^{\ell} \beta_k \cos(kx)$, with $\ell \geq 1$ and $\beta = (\beta_0, ..., \beta_\ell)$ is a fixed $(\ell+1)$-tuple of real numbers. For any y_0 in $\mathbb{R}$,

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{\exp\{\ell\sqrt{\log n}\}} \mathbb{E}^{1/2}[\widehat{\Gamma}_{f,n}(y_0) - \Gamma_f(y_0)]^2 < \infty.$$

Moreover if Z_C denotes any compact subset of $\mathbb{R}$,

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{\sqrt{\log n}\exp\{\ell\sqrt{\log n}\}} \mathbb{E}[\sup_{y\in Z_C} |\widehat{\Gamma}_{f,n}(y) - \Gamma_f(y)|] < \infty.$$

Taupin (2001) noticed that for general f, the smoother f is, the faster the rate of convergence for the estimation of Γ_f is. In both previous cases, f admits analytic continuation in the whole complex plane and also does the functional Γ_f. Nevertheless, it is interesting to note here that, even if the function f and the functional Γ_f still admit analytic continuation on the whole complex plane when f is a linear combination of cosines, the rate of convergence is slightly slower in that case than the one in the polynomial case. We will show in this chapter that this rate of convergence can essentially (see below) not be improved.

Note that the previous two bounds with respect to the $\mathbb{L}_\infty(\mathbb{R})$-norm hold for y lying in a compact set and that there is a loss of $\sqrt{\log n}$ in the rate of convergence compared to the pointwise estimation. We improve here these two bounds, by giving a result valid for the uniform norm on $\mathbb{R}$ and with a loss equal to $\sqrt{\log\log n}$ which is optimal in the polynomial case and quite optimal in the trigonometric case. We show that if $f(x) = \sum_{k=0}^{\ell} \beta_k x^k$ then

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}\sqrt{\log\log n}} \mathbb{E} \parallel \widehat{\Gamma}_{f,n} - \Gamma_f \parallel_\infty < \infty.$$

In the same way if $f(x) = \sum_{k=0}^{\ell} \beta_k \cos(kx)$ (or the same expression with sinus), then

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{\sqrt{\log\log n}\exp\{\ell\sqrt{\log n}\}} \mathbb{E} \parallel \widehat{\Gamma}_{f,n} - \Gamma_f \parallel_\infty < \infty.$$

We also extend those results to upper bounds for risks with respect to the $\mathbb{L}_p(\mathbb{R})$-norm, for $2 \leq p < \infty$ in the polynomial case. We prove that if $f(x) = \sum_{k=0}^{\ell} \beta_k x^k$ then for $2 \leq p < \infty$,

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}} \mathbb{E} \parallel \widehat{\Gamma}_{f,n} - \Gamma_f \parallel_p < \infty.$$

These results come either from Rosenthal's inequality (see Rosenthal (1970)) with the optimal constants given by Pinelis (1995), or from results given in Ibragimov and Hasminskii (1983).

Let us describe now the main results of this chapter which concern lower bounds. First note that all the rates of convergence obtained in the polynomial case are optimal in the minimax sense. As a matter of fact we prove the following lower bounds. Fix an integer $\ell \geq 1$ and a polynomial function f of degree less or equal to ℓ. Then we have:

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}} \mathbb{E}^{1/2}[T_n - \Gamma_f(y_0)]^2 > 0, \tag{3.8}$$

where the infimum is taken over all the estimators T_n based on the observations $Y_1, \ldots Y_n$. For the $\mathbb{L}_p(\mathbb{R})$-norm risk, the lower bound becomes

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}} \mathbb{E}\|\Gamma_f - T_n\|_p > 0,$$

valid for every $1 < p < \infty$ when $\ell = 0$ and for every $1 \leq p < \infty$ when $\ell \geq 1$. And if $p = \infty$,

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}\sqrt{\log\log n}} \mathbb{E}\|\Gamma_f - T_n\|_\infty > 0,$$

where the infima are taken over all the estimators T_n based on the observations $Y_1, \ldots, Y_n$.

The lower bound (3.8) was already known for the pointwise estimation of the density h, that is to say in the case where $\ell = 0$. Indeed, Taupin (1998) proved that the rate of convergence of the pointwise quadratic risk $(\log n)^{1/4}/\sqrt{n}$ for the estimation of h over the class $\mathcal{H}$ is the optimal one.

The trigonometric case is more complicated. We prove two lower bounds : a lower bound for the pointwise minimax risk and a lower bound for the risk with respect to the $\mathbb{L}_\infty(\mathbb{R})$-norm. Concerning the pointwise quadratic risk, if $f(x) = \sum_{k=0}^{\ell} \beta_k \cos(kx)$ we prove the following lower bound

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}(\log n)^{1/4}}{\exp\{\ell\sqrt{\log n}\}} \mathbb{E}^{1/2}[T_n - \Gamma_f(y_0)]^2 > 0,$$

where the infimum is taken over all the estimators T_n based on the observations $Y_1, \ldots, Y_n$. Note that in this case, the rate achieved by our estimator is nearly minimax with a small loss of $(\log n)^{1/4}$.

Concerning the minimax risk with respect to the $\mathbb{L}_\infty(\mathbb{R})$-norm, the lower bound is

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}(\log n)^{3/4}}{\exp\{\ell\sqrt{\log n})} \mathbb{E}\|\Gamma - T_n\|_\infty > 0.$$

Also notice here that the rate achieved by our estimator is nearly minimax with a loss in $(\log n)^{3/4}/\sqrt{\log\log n}$.

The main tool for establishing lower bounds are the van Trees Inequality (see Gill and Levit (1995)) for the pointwise quadratic minimax risk and Fano's Lemma (see Fano (1952), Cover and Thomas (1991) and for the application to statiscics, Ibragimov and Has′minskiĭ (1981) and Birgé (1983, Birgé (2001)) for lower bound of the $\mathbb{L}_p(\mathbb{R})$-risk.

This chapter is organized as follows. Section 3.2 is devoted to the construction of the estimator and to the presentation of some elementary properties of Γ_f in cases we are interested in. In

Section 3.3, we give the results about pointwise estimation, starting by recalling previous results on upper bounds and thus giving lower bounds. Sections 3.4 and 3.5 consider the problem of estimation with respect to $\mathbb{L}_p(\mathbb{R})$-norm, when $1 \leq p \leq \infty$, starting with upper bounds (Section 3.4) and thus establishing the optimality properties of our estimator through lower bounds (Section 3.5). Proofs essentially follow the theorems or the propositions, except some technical proofs which are presented in Appendix C.

3.2 Construction of the estimator and preliminary results

3.2.1 Construction of the estimator

We start by introducing the notations used to recall the estimator of the functional Γ_f for general functions f. Fix y_0 in $\mathbb{R}$ and denote by T_{f,y_0} and $T^\star_{f,y_0}$ the following quantities:

$$T_{f,y_0}(x) = f(x)\Phi_1(y_0 - x) \quad \text{and} \quad T^\star_{f,y_0}(t) = \int e^{itx} T_{f,y_0}(x)dx.$$

The function f being in this chapter either a polynomial or a real valued trigonometric function, the functions T_{f,y_0} and $T^\star_{f,y_0}$ belong both to $\mathbb{L}_1(\mathbb{R})$. Denoting by $< \cdot, \cdot >$ the scalar product on $\mathbb{L}_2(\mathbb{R})$ and applying Parseval's equality, $\Gamma_f(y_0)$ can be written as:

$$\Gamma_f(y_0) = \langle T_{f,y_0}, g \rangle = \frac{1}{2\pi} \left\langle T^\star_{f,y_0}, g^\star \right\rangle = \frac{1}{2\pi} \left\langle T^\star_{f,y_0}, h^\star (\Phi_1^*)^{-1} \right\rangle.$$

Replace $h^\star(t)$ by its empirical estimator

$$\widehat{h}^\star_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itY_j} K_n^*(t) = \frac{1}{n}\sum_{j=1}^{n} \widehat{h}^*_{n,j}(t), \tag{3.9}$$

where $K_n(x) = C_n K(C_n x)$ is a kernel to be chosen and $\{C_n\}_{n\geq 0}$ is a sequence converging to infinity, generally chosen in order to balance the bias and the variance terms. By using the independence between X_n and ε_n, the function $h^\star$ equals $\Phi_1^\star g^\star$, and then we propose to estimate $g^\star$ by

$$\widehat{g^\star_n}(t) = \frac{1}{n}\sum_{j=1}^{n} \frac{\widehat{h}^\star_{n,j}(t)}{\Phi_1^\star(t)} = \frac{1}{n}\sum_{j=1}^{n} e^{itY_j} K_n^*(t) e^{t^2/2},$$

where $\widehat{h}^\star_{n,j}(t) = e^{itZ_j} K_n^\star(t)$ and $(\Phi_1^\star(t))^{-1} = e^{t^2/2}$.
Denoting by $\text{Re}(z)$ the real part of z, and following Taupin (2001), the linear functional $\Gamma_f(y_0)$ is then estimated with

$$\widehat{\Gamma}_{f,n}(y_0) = \frac{1}{2\pi}\text{Re}\left\langle T^\star_{f,y_0}, \widehat{g^\star_n} \right\rangle = \frac{1}{2\pi n}\sum_{j=1}^{n} \text{Re} \int T^\star_{f,y_0}(t) e^{t^2/2} e^{-itY_j} K_n^*(-t)dt. \tag{3.10}$$

Note that due to the presence of the term $\exp\{t^2/2\}$, the function $t \mapsto T^\star_{f,y_0}(t)\exp\{t^2/2\}$ will often be non-integrable. In order to ensure the existence of the estimator $\widehat{\Gamma}_{f,n}(y_0)$ we choose a kernel with a Fourier transform compactly supported. More precisely the kernel K is assumed to satisfy the following conditions:

[K1] The kernel K belongs to $\mathbb{L}_2(\mathbb{R})$ and is an even function.

[K2] Its Fourier transform satisfies $K^\star(t) = 1$ for any t in $[-1, 1]$.

[K3] $|K^\star(t)| \leq 1\!\!1_{[-2,2]}(t)$ for any t in $\mathbb{R}$.

Remark 3.1. Assumption [K1] ensures that the Fourier transform of the kernel is an even real valued function. Assumption [K2] allows us to control the bias term, and Assumption [K3] ensures the existence of the estimator $\widehat{\Gamma}_{f,n}(y)$. The so-called naive kernel defined by

$$\forall x \in \mathbb{R} \ , \quad S(x) = \frac{\sin(x)}{\pi x} \quad \text{with} \quad \forall t \in \mathbb{R} \ , \quad S^\star(t) = 1\!\!1_{[-1;1]}(t), \tag{3.11}$$

satisfies the previous Assumptions [K1]-[K3]. The choice of the kernel S corresponds to the truncation of the previous integral,

$$\int T^\star_{f,y_0}(t) e^{t^2/2} e^{-itY_j} S^\star_n(-t) dt = \int T^\star_{f,y_0}(t) e^{t^2/2} e^{-itY_j} 1\!\!1_{|t| \leq C_n} dt.$$

Note that this kernel is not integrable, which can be inconvenient. We may prefer the second kernel, defined by:

$$\forall x \in \mathbb{R} \ , \quad V(x) = \frac{\cos(x) - \cos(2x)}{\pi x^2}. \tag{3.12}$$

This kernel, called the analogue of the de La Vallée-Poussin kernel (see Nikol'skii (1969)) satisfies Assumptions [K1]-[K3] and is furthermore an integrable kernel. More precisely its Fourier transform satisfies:

$$V^\star(t) = \begin{cases} 1 & \text{if } t \in [-1, 1] \\ 0 & \text{if } |t| \geq 2 \\ (2 - |t|) & \text{if } |t| \in [1, 2]. \end{cases}$$

Noting that $V^\star_n(t) = V^\star(t/C_n)$, the above properties become:

$$\forall t \in \mathbb{R}, \quad |V^\star_n(t)| \leq 1\!\!1_{[-2C_n, 2C_n]}(t) \quad \text{and} \quad |1 - V^\star_n(t)| \leq 1\!\!1_{|t| \geq C_n}.$$

In this chapter, the letter S is used to denote the first kernel, the letter V is used for the de La Vallée-Poussin kernel and K is used for some non specified kernel satisfying the above conditions.

The derivatives of the density $h^{(\ell)}$ ($\ell \geq 0$) are estimated by the following kernel estimator

$$\widehat{h}^{(\ell)}_n(y) = \frac{1}{n} \sum_{j=1}^n K^{(\ell)}_n(y - Y_j) = \frac{C_n^{\ell+1}}{n} \sum_{j=1}^n K^{(\ell)}(C_n(y - Y_j)). \tag{3.13}$$

Note that the kernel used for the estimation of h is not necessary the same as the one used for the estimation of $h^{(\ell)}$ nor Γ_f.

3.2.2 Some properties of our functionals

We focus here on the case of functions f being either polynomial of the form $x \mapsto \sum_{j=0}^\ell \beta_j x^j$ or real-valued trigonometric functions (that can be viewed as linear combination of trigonometric functions of the form $\mathcal{C}_{\beta,\ell} : x \mapsto \sum_{j=0}^\ell \beta_j \cos(jx)$ or $\mathcal{S}_{\beta,\ell} : x \mapsto \sum_{j=0}^\ell \beta_j \sin(jx)$, where ℓ is a fixed integer and the parameters $(\beta_j)_{0\leq j\leq \ell}$ are real fixed numbers). These particular forms of the function f that we consider here, imply that Γ_f satisfy some formulae given here.

By the linearity of $f \mapsto \Gamma_f$, we will focus on the more specific forms defined by the following formulae:

$$\Gamma_\ell(y) = \int x^\ell g(x)\Phi_1(x-y)dx\ , \quad \forall y\in\mathbb{R},\ \ell\in\mathbb{N}, \tag{3.14}$$

$$\Gamma_{c,\ell}(y) = \int \cos(\ell x)g(x)\Phi_1(x-y)dx\ , \quad \forall y\in\mathbb{R},\ \ell\in\mathbb{N}, \tag{3.15}$$

$$\Gamma_{s,\ell}(y) = \int \sin(\ell x)g(x)\Phi_1(x-y)dx\ , \quad \forall y\in\mathbb{R},\ \ell\in\mathbb{N}. \tag{3.16}$$

There exists a relation between the functionals Γ_f where f is a polynomial function and the derivatives of the density h of the observations. More precisely, we have the following lemma.

Lemma 3.2.1. *For all fixed integer ℓ there exist polynomial functions $\{P_j\}_{1\leq j\leq \ell}$ with $deg(P_j)=j$ such that:*

$$\forall y\in\mathbb{R},\quad \Gamma_\ell(y) = h^{(\ell)}(y) + \sum_{k=0}^{\ell-1} P_{\ell-k}(y)h^{(k)}(y).$$

The proof of this lemma stands in Appendix C.

Arguing as for Lemma 3.2.1 the estimator $\widehat{\Gamma}_{\ell,n}$ of Γ_ℓ defined by (3.10) can be written in the following way

$$\widehat{\Gamma}_{\ell,n}(y) = \widehat{h}_n^{(\ell)} + \sum_{k=0}^{\ell-1} P_{\ell-k}(y)\widehat{h}_n^{(k)}(y), \tag{3.17}$$

also given by:

$$\widehat{\Gamma}_{\ell,n}(y) = \frac{1}{n}\sum_{j=1}^{n} K_n^{(\ell)}(y-Y_j) + \sum_{k=0}^{\ell-1} P_{\ell-k}(y)\frac{1}{n}\sum_{j=1}^{n} K_n^{(k)}(y-Y_j).$$

When f is a polynomial function of degree equal to ℓ a straightforward application of Lemma 3.2.1 gives the existence of a real number $a\neq 0$ and polynomial functions $\{f_j\}_{1\leq j\leq \ell}$ with $deg(f_j)=j$, such that

$$\Gamma_f(y) = ah^{(\ell)}(y) + \sum_{k=0}^{\ell-1} f_{\ell-k}(y)h^{(k)}(y)$$

and therefore,

$$\widehat{\Gamma}_{f,n}(y) = a\widehat{h}_n^{(\ell)}(y) + \sum_{k=0}^{\ell-1} f_{\ell-k}(y)\widehat{h}_n^{(k)}(y). \tag{3.18}$$

Consequently, when f is a polynomial function of degree ℓ, the study of the rates of convergence in the estimation of Γ_f is given by the study of the rates of convergence in the estimation of $f_{\ell-k}h^{(k)}$ for all $0\leq k\leq \ell$.

We now come to the trigonometric case. Using that Φ_1 is the standard Gaussian density we have the following identities.

Lemma 3.2.2.

$$\begin{aligned}\Gamma_{c,\ell}(y) &= \int \cos(\ell x)g(x)\Phi_1(x-y)dx = \frac{e^{-\ell/2}}{2}\left[e^{i\ell y}h(y+i\ell)+e^{-i\ell y}h(y-i\ell)\right],\\ \Gamma_{s,\ell}(y) &= \int \sin(\ell x)g(x)\Phi_1(x-y)dx = \frac{e^{-\ell/2}}{2i}\left[e^{i\ell y}h(y+i\ell)-e^{-i\ell y}h(y-i\ell)\right].\end{aligned}$$

The proof of this lemma stands in Appendix C.
Arguing as for Lemma 3.2.2, the estimators $\widehat{\Gamma}_{n,c,\ell}$ and $\widehat{\Gamma}_{n,s,\ell}$ of $\Gamma_{c,\ell}$ and $\Gamma_{s,\ell}$ defined by (3.10) can be written

$$\begin{aligned}\widehat{\Gamma}_{n,c,\ell}(y) &= \frac{e^{-\ell/2}}{2}\left[e^{i\ell y}\widehat{h}_n(y+i\ell)+e^{-i\ell y}\widehat{h}_n(y-i\ell)\right] \qquad (3.19)\\ \widehat{\Gamma}_{n,s,\ell}(y) &= \frac{e^{-\ell/2}}{2}\left[e^{i\ell y}\widehat{h}_n(y+i\ell)-e^{-i\ell y}\widehat{h}_n(y-i\ell)\right]\end{aligned}$$

All these formulae are useful to construct lower bounds, especially to construct sub-families related to the density h.

3.3 Pointwise estimation: upper and lower bounds

We start by recalling previous results obtained by Taupin (2001) about upper bound for the pointwise quadratic risk before giving the lower bound in the case of polynomial functions.

Theorem 3.3.1. Proposition 3.1. in Taupin (2001).
Fix an integer $\ell \geq 1$ and a polynomial function f with degree less or equal to ℓ. Consider the kernel estimator $\widehat{\widehat{\Gamma}}_{f,n}$ defined by (3.10) with the kernel V defined by (3.12) and the bandwidth $C_n = \sqrt{\log n}$.

$$\limsup_{n\to\infty} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}} \mathbb{E}^{1/2}\left[\widehat{\Gamma}_{f,n}(y_0)-\Gamma_f(y_0)\right]^2 < \infty.$$

The following theorem states that this rate of convergence is the best achievable rate of convergence and that our estimator achieves it.

Theorem 3.3.2. *Fix an integer $\ell \geq 1$ and a polynomial function f with degree less or equal to ℓ, then*

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}} \mathbb{E}^{1/2}\left[T_n - \Gamma_f(y_0)\right]^2 > 0,$$

where the infimum is taken over all the estimators T_n based on the observations $Y_1, \ldots Y_n$.

The lower bound when $\ell = 0$ was already known and proved in Taupin (1998). The main tool of the proof is the van Trees Inequality (see Gill and Levit (1995)). It follows the lines of Taupin's proof in two steps.

The first step is to construct a sub-family

$$\{h_\theta,\ |\theta| \leq \theta_n = \rho\sqrt{\log n}/\sqrt{n}\}$$

included in $\mathcal{H}$ for n large enough. Then we construct the sub-family $\{\Gamma_{\theta,\ell}\ ;\ |\theta| \leq \theta_n = \rho\sqrt{\log n}/\sqrt{n}\}$, where, according to (3.2.1), $\Gamma_{\theta,\ell} = h_\theta^{(\ell)} + \sum_{k=0}^{\ell-1} P_{\ell-k} h_\theta^{(k)}$ It is clear that, by construction, the sub-family $\{\Gamma_{\theta,\ell},\ |\theta| \leq \theta_n = \rho\sqrt{\log n}/\sqrt{n}\}$ is included into $\mathcal{G}_f$ for the polynomial function $f(x) = x^\ell$.

It follows that we have

$$R_n(f) \geq \inf_{T_n} \sup_{|\theta|\leq\theta_n} \mathbb{E}_{h_\theta}[T_n - \Gamma_\theta(y_0)]^2.$$

The second step consists in applying the van Trees inequality. The proof of this theorem can be found in Appendix C.

We now come to the trigonometric case. The upper bound for the pointwise estimation is given in the following theorem.

Theorem 3.3.3. Proposition 3.1. in Taupin (2001).
Let $\ell \geq 1$ be a fixed integer and f a trigonometric function of the form $\mathcal{C}_{\beta,\ell}$ or $\mathcal{S}_{\beta,\ell}$ where $\{\beta_j\}_{0\leq j\leq \ell}$ are real parameters. Let $\widehat{\Gamma}_{f,n}$ be defined by (3.10) or equivalently by (3.19), with the kernel S defined by (3.11) and the bandwidth $C_n = \sqrt{\log n}$. Then, for all y_0 in $\mathbb{R}$, there exists a constant $\kappa(\beta,\ell)$ such that

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{\exp\{\ell\sqrt{\log n}\}} \mathbb{E}^{1/2}[\widehat{\Gamma}_{f,n}(y_0) - \Gamma_f(y_0)]^2 \leq \kappa(\beta,\ell).$$

We give a new proof of this theorem in Appendix C, based on the relation expressed in Lemma 3.2.2. This proof is interesting as it uses techniques specifically related to this link between trigonometric functionals and the density h.
The following theorem states that the estimator defined by (3.10) or equivalently by (3.19) is nearly minimax. As a matter of fact the upper bound is of order $n^{-1/2}\exp\{\sqrt{\log n}\}$ and the lower bound of order $n^{-1/2}(\log n)^{-1/4}\exp\{\sqrt{\log n}\}$, with a loss of order $(\log n)^{1/4}$.

Theorem 3.3.4. *Let $\ell \geq 1$ be a fixed integer and f a trigonometric function of the form $\mathcal{C}_{\beta,\ell}$ or $\mathcal{S}_{\beta,\ell}$ where the $\{\beta_j\}_{0\leq j\leq\ell}$ are real parameters. Then we have*

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Gamma_f\in\mathcal{G}_f} \frac{\sqrt{n}(\log n)^{1/4}}{\exp\{\ell\sqrt{\log n}\}} \mathbb{E}^{1/2}[T_n - \Gamma_f(y_0)]^2 > 0,$$

where the infimum is taken over all the estimators T_n based on the observations $Y_1, \ldots, Y_n$.

The proof based on the van Trees Inequality follows the line of the proof of Theorem 3.3.2 and can be found in Appendix C.

3.4 Estimation in $\mathbb{L}_p(\mathbb{R})$-norm: upper bounds

The main part of our work concerns the problem of estimating Γ_f and h with respect to $\mathbb{L}_p(\mathbb{R})$-norm, for $1 \leq p \leq \infty$. We start with upper bounds for the rate of convergence of the $\mathbb{L}_p(\mathbb{R})$-risk before stating that these rates are the best achievable and that the rates of convergence of our estimator defined in (3.10) with suitable kernel, essentially achieve them.

In this section, we give upper bounds for the rate of convergence of $\widehat{\Gamma}_{f,n}$ defined by (3.10) for the risk with respect to the $\mathbb{L}_p(\mathbb{R})$-norm, for $1 \leq p \leq \infty$, starting with the density case.

3.4.1 Upper bounds in $\mathbb{L}_p(\mathbb{R})$-norm in the density case i.e. $f \equiv 1$

In this subsection, we are interested in the rate of convergence in $\mathbb{L}_p(\mathbb{R})$-norm of the estimator defined by (3.13) of the density h of the observation $Y_1, ..., Y_n$.

Theorem 3.4.1. *Let $\widehat{h}$ be the estimator defined by (3.13) with the kernel V defined by (3.12) and the bandwidth $C_n = \sqrt{\log n}$.*

If $p = \infty$, then

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{(\log n)^{1/4}\sqrt{\log\log n}} \sup_{h\in\mathcal{H}} \mathbb{E}_h \parallel \widehat{h}_n - h \parallel_\infty < \infty.$$

If $2 \le p < \infty$, then we have

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{(\log n)^{1/4}} \sup_{h\in\mathcal{H}} \mathbb{E}_h \parallel \widehat{h}_n - h \parallel_p < \infty.$$

The result for $p = \infty$ is already known and proved in Taupin (2001).

Proof. The result is proved by two different methods. More precisely, we have, according to the triangular inequality,

$$\mathbb{E}\|\widehat{h}_n - h\|_p \le \|\mathbb{E}\widehat{h}_n - h\|_p + \mathbb{E}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_p. \tag{3.20}$$

The first term is the bias term. This term is upper bounded by only one method, valid for all kernel K satisfying the property $|1 - K_n^\star| \le 1\!\!1_{|\cdot|\ge C_n}$,which is an immediate consequence of **[K2]**. The second term appearing in (3.20) is the corresponding one for the well-known "variance term" appearing in the pointwise quadratic risk. We will give two different proofs for its control. The first method we present follows from the results of Ibragimov and Hasminskii (1983). The second one is based on Rosenthal's inequality (Rosenthal 1970), with optimal constants as given in Pinelis (1995).

Let us start with the control of the bias term. We start with the case $p = \infty$. Using the properties of the kernel V_n and in particular, the inequality $|1 - V_n^\star| \le 1\!\!1_{|\cdot|\ge C_n}$, we get

$$\|\mathbb{E}\widehat{h}_n - h\|_\infty \le (\pi C_n)^{-1} e^{-C_n^2/2}. \tag{3.21}$$

We now control the bias in the case $2 \le p < \infty$. For any function u belonging to $\mathbb{L}_p(\mathbb{R})$ with $p \ge 2$, the quantity $\parallel u \parallel_p^p$ can also be written:

$$\parallel u \parallel_p^p = \frac{1}{(2\pi)^{p-2}} \int \left| \int e^{-itx} u^\star(t) dt \right|^{p-2} |u(x)|^2 dx,$$

and is thus bounded by Parseval's equality in the following way:

$$\parallel u \parallel_p^p \le \frac{1}{(2\pi)^{p-1}} \left[\int |u^\star(t)| dt \right]^{p-2} \parallel u^\star \parallel_2^2 .$$

Consequently $\parallel u \parallel_p^p$ is bounded by

$$\parallel u \parallel_p^p \le \frac{1}{(2\pi)^{p-1}} \parallel u^\star \parallel_1^{p-2} \parallel u^\star \parallel_2^2 . \tag{3.22}$$

Apply this result to the function $V_n * h - h$, to obtain that for $p \geq 2$, $\|\mathbb{E}\widehat{h}_n - h\|_p = \| V_n * h - h \|_p$ is bounded in the way

$$\|\mathbb{E}\widehat{h}_n - h\|_p \leq \frac{1}{(2\pi)^{(p-1)/p}} \parallel h^\star(1-V_n^\star) \parallel_1^{(p-2)/p} \parallel h^\star(1-V_n^\star) \parallel_2^{2/p}.$$

Note that

$$\parallel h^\star(1-V_n^\star) \parallel_2^2 \leq \int e^{-t^2} 1\!\!1_{|t|\geq C_n} dt \leq C_n^{-1} \exp\{-C_n^2\},$$

and similar arguments ensure that

$$\parallel h^\star(1-V_n^\star) \parallel_1^{(p-2)} \leq 2^{p-2} C_n^{-p+2} \exp\{-(p-2)C_n^2/2\}.$$

So that we finally get

$$\|\mathbb{E}\widehat{h}_n - h\|_p \leq \frac{1}{2^{1/p}\pi^{1-1/p}} \times \frac{e^{-C_n^2/2}}{C_n^{(p-1)/p}}. \tag{3.23}$$

Now, we are interested in the behaviour of the second term appearing in Inequality (3.20). We give two different proofs for its control.

First method: This method is based on Ibragimov and Hasminskii's results (1983) that we start by recalling. The following lemma gives a control of $\mathbb{E}_h \parallel \widehat{h}_n - \mathbb{E}_h(\widehat{h}_n) \parallel_p$ depending explicitly on p and on the bandwidth C_n.

Lemma 3.4.2. Lemma 4, Inequalities (11) and (12), Ibragimov and Hasminskii (1983).
There exists some positive constant A such that for all $2 \leq p < \infty$ and for all h in $\mathcal{H}$, we have:

$$\mathbb{E}_h \parallel \widehat{h}_n - \mathbb{E}_h(\widehat{h}_n) \parallel_p \leq A\sqrt{p}(1+ \parallel h \parallel_p)\sqrt{\frac{C_n}{n}}.$$

If $p = \infty$, there exists some positive constant A such that for all h in $\mathcal{H}$, we have:

$$\mathbb{E}_h \parallel \widehat{h}_n - \mathbb{E}_h(\widehat{h}_n) \parallel_\infty \leq A(1+ \parallel h \parallel_\infty)\sqrt{\frac{C_n \log C_n}{n}}.$$

Note that this result does not apply to the kernel $S(\cdot) = \sin(\cdot)/(\pi\cdot)$ as the constant A in the preceding lemma depends on $\|V\|_1$, which does not exist for the kernel S.

Applying this lemma and gathering Inequalities (3.20) and (3.23), we find that if $2 \leq p < \infty$, there exists a positive constant A such that:

$$\mathbb{E} \parallel \widehat{h}_n - h \parallel_p \leq A\sqrt{p}(1+ \parallel h \parallel_p)\sqrt{\frac{C_n}{n}} + \frac{1}{2^{1/p}\pi^{1-1/p}} \times \frac{e^{-C_n^2/2}}{C_n^{(p-1)/p}}, \tag{3.24}$$

and if $p = \infty$, there exists a positive constant A' such that:

$$\mathbb{E} \parallel \widehat{h}_n - h \parallel_\infty \leq A'(1+ \parallel h \parallel_\infty)\sqrt{\frac{C_n \log C_n}{n}} + \frac{1}{\pi C_n} e^{-C_n^2/2}. \tag{3.25}$$

Taking $C_n = \sqrt{\log n}$ gives the desired result.

Second method: based on Rosenthal's inequality (Rosenthal 1970).
We aim at bounding the variance term $\mathbb{E}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_p$, for $p \geq 2$. We start with the case $2 \leq p < \infty$. Write $\widehat{h}_n - \mathbb{E}\widehat{h}_n$ as a sum of independent random variables

$$\widehat{h}_n - \mathbb{E}\widehat{h}_n(x) = \frac{1}{n}\sum_{j=1}^{n} \mathcal{X}_n(x, Y_j),$$

where the $\mathcal{X}_n(x, Y_j)$ are centered random variables defined by $\mathcal{X}_n(x, Y_j) = V_n(x - Y_j) - \mathbb{E}(V_n(x - Y_j))$. We need thus to control

$$\mathbb{E}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_p = \mathbb{E}\left\|\frac{1}{n}\sum_{j=1}^{n} \mathcal{X}_n(\cdot, Y_j)\right\|_p = \mathbb{E}\left[\int\left|\frac{1}{n}\sum_{j=1}^{n} \mathcal{X}_n(x, Y_j)\right|^p dx\right]^{1/p}.$$

Using the concavity of the function $x \mapsto x^{1/p}$ for $p \geq 1$, we have

$$\mathbb{E}\left\|\frac{1}{n}\sum_{j=1}^{n} \mathcal{X}_n(\cdot, Y_j)\right\|_p \leq \left[\mathbb{E}\int\left|\frac{1}{n}\sum_{j=1}^{n} \mathcal{X}_n(x, Y_j)\right|^p dx\right]^{1/p}$$

and therefore applying Fubini's Theorem, we get that

$$\mathbb{E}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_p = \mathbb{E}\left\|\frac{1}{n}\sum_{j=1}^{n} \mathcal{X}_n(\cdot, Y_j)\right\|_p \leq \left[\int \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n} \mathcal{X}_n(x, Y_j)\right|^p dx\right]^{1/p}. \tag{3.26}$$

We now apply Rosenthal's inequality, with optimal constants as given in Pinelis (1995), to the quantity $\mathbb{E}\left|n^{-1}\sum_{j=1}^{n} \mathcal{X}_n(x, Y_j)\right|^p$. We will see later that the optimal constants are useful in the case $p = \infty$.

Lemma 3.4.3. Corollary 1 in Pinelis (1995).
Let $X_1, \ldots, X_n$, be n independent centered random variables and let $S_n = X_1 + \cdots + X_n$, and $X_n^\star = \max_{1\leq i\leq n}|X_i|$. Then there exists an absolute constant κ such that, for all $p \geq 1$,

$$\mathbb{E}^{1/p}|S_n|^p \leq \kappa\left[e\sqrt{p}\,\mathbb{E}^{1/2}|S_n|^2 + p\,\mathbb{E}^{1/p}|X_n^\star|^p\right]. \tag{3.27}$$

An immediate consequence of Inequality (3.27) is the following one:

$$\mathbb{E}^{1/p}|S_n|^p \leq e\kappa\sqrt{p}\,\mathbb{E}^{1/2}|S_n|^2 + \kappa p\mathbb{E}^{1/p}\sum_{i=1}^{n}|X_i|^p.$$

Therefore, provided that the random variables X_i are centered and identically distributed, Inequality (3.27) becomes

$$\mathbb{E}^{1/p}\left|\frac{S_n}{n}\right|^p \leq \frac{e\kappa\sqrt{p}}{\sqrt{n}}\,\mathbb{E}^{1/2}|X_1|^2 + \frac{\kappa p}{n^{1-1/p}}\,\mathbb{E}^{1/p}|X_1|^p. \tag{3.28}$$

Apply Inequality (3.28) to the quantity $\mathbb{E}^{1/p}\left|n^{-1}\sum_{j=1}^{n}\mathcal{X}_n(x,Y_j)\right|^p$, to get that

$$\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}\mathcal{X}_n(x,Y_j)\right|^p \leq 2^{p-1}\left[\frac{e\kappa\sqrt{p}}{\sqrt{n}}\,\mathbb{E}^{1/2}|\mathcal{X}_n(x,Y_1)|^2\right]^p + 2^{p-1}\left[\frac{\kappa p}{n^{1-1/p}}\,\mathbb{E}^{1/p}|\mathcal{X}_n(x,Y_1)|^p\right]^p,$$

which implies, by using (3.26) that

$$\mathbb{E}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_p \leq 2^{1-1/p}\left\{\left[\frac{e\kappa\sqrt{p}}{\sqrt{n}}\right]^p\int\mathbb{E}^{p/2}|\mathcal{X}_n(x,Y_1)|^2dx + \left[\frac{\kappa p}{n^{1-1/p}}\right]^p\int\mathbb{E}|\mathcal{X}_n(x,Y_1)|^p dx\right\}^{1/p}.$$

We control the last term $\int\mathbb{E}|\mathcal{X}_n(x,Y_1)|^p dx$, using that for $p\geq 1$,

$$\mathbb{E}|\mathcal{X}_n(x;Y_1)|^p \leq 2^p\,\mathbb{E}\,|V_n(x-Y_1)|^p,$$

and by applying Fubini's Theorem we get:

$$\int\mathbb{E}|\mathcal{X}_n(x,Y_1)|^p dx \leq 2^p\iint|V_n(x-y)|^p h(y)dy$$

and therefore

$$\int\mathbb{E}|\mathcal{X}_n(x,Y_1)|^p dx \leq 2^p C_n^{p-1}\|V\|_p^p. \tag{3.29}$$

It remains now to bound the first and main term $\int\mathbb{E}^{p/2}|\mathcal{X}_n(x,Y_1)|^2dx$. By applying Cauchy-Schwarz inequality, we have

$$\int\mathbb{E}^{p/2}|\mathcal{X}_n(x,Y_1)|^2dx \leq \int\left(\int C_nV^2(u)h(x+u/C_n)du\right)^{p/2}dx.$$

Proceeding to a Taylor expansion of $u\mapsto h(x+u/C_n)$ (the function h in $\mathcal{H}$ admits an analytic continuation to the whole complex plane) gives that the term $\int\mathbb{E}^{p/2}|\mathcal{X}_n(x,Y_1)|^2dx$ is bounded by

$$C_n^{p/2}\int\Bigg(h(x)\|V\|_2^2 + \frac{1}{C_n}h'(x)\int uV^2(u)du + \frac{1}{C_n^2}\int u^2V^2(u)\int_0^1 h''\left(x+\frac{su}{C_n}\right)dsdu\Bigg)^{p/2}dx \tag{3.30}$$

Since $\int u^2V^2(u)du<\infty$, (note that this does not hold for the kernel S) and using both Jensen's Inequality and Fubini's Theorem we obtain the bound

$$\left|\int\left(\int u^2V^2(u)\int_0^1 h''(x+su/C_n)dsdu\right)^{p/2}dx\right| \leq \left(\int u^2V^2(u)du\right)^{p/2}\int_0^1\iint u^2V^2(u)|h''(x+su/C_n)|^{p/2}dxduds,$$

and therefore, using that according to the structure of the convolution model h and all of its derivatives are integrable, we get that:

$$\left| \int \left(\int u^2 V^2(u) \int_0^1 |h''(x + su/C_n)| ds du \right)^{p/2} dx \right| \leq \left(\int u^2 V^2(u) du \right)^{1+p/2} \|h''\|_{p/2}^{p/2}.$$

This ensures the convergence to zero in $\mathbb{L}_{p/2}(\mathbb{R})$-norm of the sequence of functions

$$x \mapsto C_n^{-2} \int u^2 V^2(u) \int_0^1 h''(x + su/C_n) ds du.$$

Moreover, the dominated convergence Theorem ensures the convergence to $h\|V\|_2^2$ of the $\mathbb{L}_{p/2}(\mathbb{R})$-norm of the sequence of functions

$$x \mapsto h(x)\|V\|_2^2 + C_n^{-1} h'(x) \int u V^2(u) du.$$

Returning to the bound (3.30), we conclude on the equivalence of this bound to the quantity $C_n^{p/2} \, \|V\|_2^p \, \|h\|_{p/2}^{p/2}$. Finally, this entails that

$$\int \mathbb{E}^{p/2} |\mathcal{X}_n(x, Y_1)|^2 dx \leq C_n^{p/2} \, \|V\|_2^p \, \|h\|_{p/2}^{p/2} (1 + o(1)) \,, \qquad \text{as } n \text{ tends to infinity.} \tag{3.31}$$

Gathering (3.29) and (3.31) we find

$$\mathbb{E}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_p \leq 2\kappa \left\{ \left[\frac{e\sqrt{p}}{\sqrt{n}} \right]^p C_n^{p/2} \, \|V\|_2^p \, \|h\|_{p/2}^{p/2} (1 + o(1)) + \left[\frac{p}{n^{1-1/p}} \right]^p 2^p C_n^{p-1} \|V\|_p^p \right\}^{1/p}.$$

Finally for $p > 2$ we have:

$$\mathbb{E}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_p \leq 2\kappa \sqrt{\frac{pC_n}{n}} \, \|V\|_2 \, \|h\|_{p/2}^{1/2} \, (1 + o(1)) \,, \quad \text{as } n \text{ tends to infinity,} \tag{3.32}$$

and for $p = 2$, there exists some finite constant κ' such that

$$\mathbb{E}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_2 \leq \kappa' \sqrt{\frac{C_n}{n}}.$$

The result follows for $2 \leq p < \infty$ gathering (3.20), (3.23) and (3.32) and taking $C_n = \sqrt{\log n}$.

For $p = \infty$, the control of the analogue of the variance term follows arguing as Ibragimov and Hasminskii (1983). More precisely the bound for $\mathbb{E}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_\infty$ follows from the following lemma, due to Nikol'skii (1969). Denote by $\mathcal{M}_{C,p}$ the set of functions belonging to $\mathbb{L}_p(\mathbb{R})$ and having a Fourier transform compactly supported on $[-C; C]$.

Lemma 3.4.4. Nikol′skii (1969), p. 150.
If $1 \leq p \leq p' \leq \infty$, then for g in $\mathcal{M}_{C,p}$ we have $\|g\|_{p'} \leq 2C^{1/p - 1/p'} \|g\|_p$.

Since $\widehat{h}_n - \mathbb{E}\widehat{h}_n$ belongs to $\mathcal{M}_{2C_n,p}$, we can thus apply Lemma 3.4.4 to get that for any $p \leq \infty$,

$$\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_\infty \leq 2(2C_n)^{1/p}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_p.$$

Take $p = \log C_n$ to write that

$$\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_\infty \leq 2(2C_n)^{1/\log C_n}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_p$$

with

$$2(2C_n)^{1/\log C_n}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_p = 2e\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_p(1+o(1)),$$

and therefore by applying (3.32) with $\sup_{p\geq 2}\|h\|_p \leq A^2$, $A^2 \triangleq (2\pi)^{-1/2}\|h^\star\|_1\|h^\star\|_2^2$ (due to Inequality (3.22)), we find

$$\mathbb{E}\|\widehat{h}_n - \mathbb{E}\widehat{h}_n\|_\infty \leq 4A\kappa e^2\frac{\sqrt{C_n\log C_n}}{\sqrt{n}}\|V\|_2(1+o(1))$$

and the result follows for $p = \infty$ by taking $C_n = \sqrt{\log n}$.

This final argument shows how the optimal constants in Rosenthal's Inequality are important to get the result when $p = \infty$. □

3.4.2 Upper bounds in $\mathbb{L}_p(\mathbb{R})$-norm in the case of polynomials

In this subsection, we are interested in the rate of convergence in $\mathbb{L}_p(\mathbb{R})$-norm of estimators of the functional Γ_f when f is a polynomial function of degree less or equal to ℓ, with $\ell \geq 1$. According to Lemma 3.2.1, we will first focus on the rate of convergence of estimators of the derivatives $h^{(k)}$, $0 \leq k \leq \ell$ of the density h.

Theorem 3.4.5. *Let ℓ be a fixed integer, and $\widehat{h}_n^{(\ell)}$ the estimator defined by (3.13) with the kernel V and the bandwidth $C_n = \sqrt{\log n}$. If $p = \infty$, then*

$$\limsup_{n\to\infty}\sup_{h\in\mathcal{H}}\frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}\sqrt{\log\log n}}\,\mathbb{E}\parallel \widehat{h}_n^{(\ell)} - h^{(\ell)}\parallel_\infty < +\infty.$$

If $2 \leq p < \infty$, then

$$\limsup_{n\to\infty}\sup_{h\in\mathcal{H}}\frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}}\,\mathbb{E}\parallel \widehat{h}_n^{(\ell)} - h^{(\ell)}\parallel_p < +\infty.$$

We now give the proof of Theorem 3.4.5.

Proof. Arguing as for the density we give two methods to bound $\mathbb{E}\parallel \widehat{h}_n^{(\ell)} - h^{(\ell)}\parallel_p$. The first one is based on an extension of Lemma 4 in Ibragimov and Hasminskii (1983). The second method is based on Rosenthal's inequality as given in Pinelis (1995).
By using the triangular inequality, we get

$$\mathbb{E}\parallel h^{(\ell)} - \widehat{h}_n^{(\ell)}\parallel_p \leq \parallel h^{(\ell)} - V_n^{(\ell)} * h\parallel_p + \mathbb{E}\parallel \widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}\parallel_p . \tag{3.33}$$

We start with the control of the bias term (the first one in the right hand side of the previous inequality). Note that for all t in $\mathbb{R}$, we have the equality:

$$\left(V_n^{(\ell)} * h - h^{(\ell)}\right)^\star (t) = (it)^\ell h^\star(t)(V_n^\star(t) - 1).$$

Consider first the case $p = \infty$. Using the properties of the kernel V_n and the fact that the function $|h^*(t)|$ is bounded by $e^{-t^2/2}$, we get that :

$$\| V_n^{(\ell)} * h - h^{(\ell)} \|_\infty \leq (2\pi)^{-1} \int |t|^\ell |h^\star(t)| 1\!\!1_{|t|\geq C_n} dt$$

and therefore that

$$\| V_n^{(\ell)} * h - h^{(\ell)} \|_\infty \leq 2(\pi)^{-1} C_n^{\ell-1} e^{-C_n^2/2}, \tag{3.34}$$

for large enough n (depending on ℓ). Let us now study the case $2 \leq p < \infty$. Arguing as in the proof of Theorem 3.4.1, we infer that

$$\| V_n^{(\ell)} * h - h^{(\ell)} \|_p \leq \frac{1}{(2\pi)^{(p-1)/p}} \left(\int |t|^\ell e^{-t^2/2} 1\!\!1_{|t|>C_n} dt\right)^{\frac{p-2}{p}} \left(\int |t|^{2\ell} e^{-t^2} 1\!\!1_{|t|>C_n} dt\right)^{1/p}.$$

Note that for large enough n,

$$\int |t|^\ell e^{-t^2/2} 1\!\!1_{|t|>C_n} dt \leq 4 C_n^{\ell-1} e^{-C_n^2/2},$$

and also that

$$\int |t|^{2\ell} e^{-t^2} 1\!\!1_{|t|>C_n} dt \leq 4 C_n^{2\ell-1} e^{-C_n^2},$$

which implies finally that

$$\| V_n^{(\ell)} * h - h^{(\ell)} \|_p \quad \leq \quad \left[2(\pi)^{-1}\right]^{1-1/p} C_n^{\ell-1+1/p} e^{-C_n^2/2}. \tag{3.35}$$

We are now interested in the control of the second term in the right hand side of Inequality (3.33) for which we give two different proofs.

First method: based on Ibragimov and Hasminskii's results (1983).
The following lemma provides a bound for $\mathbb{E} \parallel \widehat{h}_n^{(\ell)} - \mathbb{E}(\widehat{h}_n^{(\ell)}) \parallel_p$. This is a straightforward application of Lemma 4 (formulae (11) and (12)) of Ibragimov and Hasminskii (1983)), and its proof is omitted.

Lemma 3.4.6. *For $2 \leq p < \infty$, there exists a positive constant κ such that:*

$$\mathbb{E}_h \parallel \widehat{h}_n^{(\ell)} - \mathbb{E}_h \widehat{h}_n^{(\ell)} \parallel_p \leq \kappa\sqrt{p} \parallel h \parallel_p \frac{C_n^{\ell+1/2}}{\sqrt{n}}.$$

For $p = \infty$, there exists a positive constant κ such that:

$$\mathbb{E}_h \parallel \widehat{h}_n^{(\ell)} - \mathbb{E}_h \widehat{h}_n^{(\ell)} \parallel_\infty \leq \kappa' \parallel h \parallel_\infty \frac{C_n^{\ell+1/2}\sqrt{\log C_n}}{\sqrt{n}}.$$

The constants κ and κ' depending on $\| V \|_1$ and $\|V\|_\infty$, the proof does not apply for the kernel S.
Combining this lemma with the control of the bias term (3.35), we get that if $2 \leq p < \infty$,

$$\mathbb{E} \parallel h^{(\ell)} - \widehat{h}_n^{(\ell)} \parallel_p \leq \kappa\sqrt{p} \parallel h \parallel_p \frac{C_n^{\ell+1/2}}{\sqrt{n}} + \left(\frac{2}{\pi}\right)^{1-1/p} C_n^{\ell-1+1/p} e^{-C_n^2/2}, \tag{3.36}$$

and if $p = \infty$,

$$\mathbb{E} \parallel h^{(\ell)} - \widehat{h}_n^{(\ell)} \parallel_\infty \leq \kappa' \parallel h \parallel_\infty \frac{C_n^{\ell+1/2}\sqrt{\log C_n}}{\sqrt{n}} + \frac{2}{\pi} C_n^{\ell-1} e^{-C_n^2/2}. \tag{3.37}$$

Take $C_n = \sqrt{\log n}$, to get

$$\begin{aligned}\mathbb{E} \parallel h^{(\ell)} - \widehat{h}_n^{(\ell)} \parallel_\infty &\leq \kappa' \parallel h \parallel_\infty \frac{(\log n)^{\frac{2\ell+1}{4}}\sqrt{\log\log n}}{\sqrt{n}} + \frac{2}{\pi}\frac{(\log n)^{\frac{\ell-1}{2}}}{\sqrt{n}} \\ &\leq \kappa'' \frac{(\log n)^{\frac{2\ell+1}{4}}\sqrt{\log\log n}}{\sqrt{n}}.\end{aligned}$$

Moreover, for $2 \leq p < \infty$,

$$\begin{aligned}\mathbb{E} \parallel h^{(\ell)} - \widehat{h}_n^{(\ell)} \parallel_p &\leq \kappa \parallel h \parallel_p \sqrt{p}\frac{(\log n)^{\frac{2\ell+1}{4}}}{\sqrt{n}} + \frac{2}{\pi}^{1-1/p} \frac{(\log n)^{\frac{p(\ell-1)+1}{2p}}}{\sqrt{n}} \\ &\leq \kappa' \sqrt{p}\frac{(\log n)^{\frac{2\ell+1}{4}}}{\sqrt{n}},\end{aligned}$$

which completes the proof.

<u>Second method:</u> based on Rosenthal's inequality (Rosenthal 1970).
We aim at bounding the term $\mathbb{E}\|\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}\|_p$. Arguing as for the density, write $\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}$ as a sum of independent random variables

$$\left(\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}\right)(x) = \frac{1}{n}\sum_{j=1}^n \mathcal{X}_n^{(\ell)}(x, Y_j),$$

where the random variables $\mathcal{X}_n^{(\ell)}(x, Y_j)$ are centered and defined by

$$\mathcal{X}_n^{(\ell)}(x, Y_j) = V_n^{(\ell)}(x - Y_j) - \mathbb{E}V_n^{(\ell)}(x - Y_j).$$

Arguing as for the density, we control

$$\left[\int \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n \mathcal{X}_n^{(\ell)}(x, Y_j)\right|^p dx\right]^{1/p},$$

by applying (3.28) to $\mathbb{E}\left|n^{-1}\sum_{j=1}^n \mathcal{X}_n^{(\ell)}(x, Y_j)\right|^p$. We deduce that there exists an absolute constant κ such that:

$$\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n \mathcal{X}_n^{(\ell)}(x, Y_j)\right|^p \leq 2^{p-1}\left[\frac{e\kappa\sqrt{p}}{\sqrt{n}}\mathbb{E}^{1/2}|\mathcal{X}_n^{(\ell)}(x, Y_1)|^2\right]^p + 2^{p-1}\left[\frac{\kappa p}{n^{1-1/p}}\mathbb{E}^{1/p}|\mathcal{X}_n^{(\ell)}(x, Y_1)|^p\right]^p$$

and therefore

$$\mathbb{E}\|\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}\|_p \leq \left\{ 2^{p-1}\left[\frac{e\kappa\sqrt{p}}{\sqrt{n}}\right]^p \int \mathbb{E}^{p/2}|\mathcal{X}_n^{(\ell)}(x,Y_1)|^2 dx \right. \\ \left. + 2^{p-1}\left[\frac{\kappa p}{n^{1-1/p}}\right]^p \int \mathbb{E}|\mathcal{X}_n^{(\ell)}(x,Y_1)|^p dx \right\}^{1/p}.$$

We control the last term $\int \mathbb{E}|\mathcal{X}_n^{(\ell)}(x,Y_1)|^p dx$ by using Fubini's Theorem

$$\int \mathbb{E}|\mathcal{X}_n^{(\ell)}(x,Y_1)|^p dx \leq 2^p \iint \left|V_n^{(\ell)}(x-y)\right|^p h(y)dydx = 2^p C_n^{(\ell+1)p-1}\|V\|_p^p. \tag{3.38}$$

It remains now to bound the first term $\int \mathbb{E}^{p/2}|\mathcal{X}_n^{(\ell)}(x,Y_1)|^2 dx$. Again apply Cauchy-Schwarz inequality to write

$$\int \mathbb{E}^{p/2}\left|\mathcal{X}_n^{(\ell)}(x,Y_1)\right|^2 dx \quad \leq \quad \int \left(\int C_n^{2\ell+1}(V^{(\ell)}(u))^2 h(x+u/C_n)du\right)^{p/2} dx.$$

The same Taylor expansion as the one used in the proof of Theorem 3.4.1, valid since the integral $\int u^2[V^{(\ell)}(u)]^2 du$ is finite, gives us

$$\int \mathbb{E}^{p/2}|\mathcal{X}_n^{(\ell)}(x,Y_1)|^2 dx = C_n^{(2\ell+1)p/2}\ \|V^{(\ell)}\|_2^p\ \|h\|_{p/2}^{p/2}(1+o(1)), \quad \text{as } n \text{ tends to infinity.} \tag{3.39}$$

Gathering (3.38) and (3.39) we find that $\mathbb{E}\|\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}\|_p$ is bounded by

$$2\kappa\left\{\left[\frac{e\sqrt{p}}{\sqrt{n}}\right]^p C_n^{p(2\ell+1)/2}\ \|V^{(\ell)}\|_2^p\ \|h\|_{p/2}^{p/2}(1+o(1)) + 2^p\left[\frac{p}{n^{1-1/p}}\right]^p C_n^{(l+1)p-1}\|V^{(\ell)}\|_p^p\right\}^{1/p}.$$

Finally for $p > 2$ we have:

$$\mathbb{E}\|\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}\|_p \quad \leq \quad 2\frac{e\kappa\sqrt{p}}{\sqrt{n}}C_n^{(2\ell+1)/2}\ \|V^{(\ell)}\|_2\ \|h\|_{p/2}^{1/2}\ (1+o(1)), \tag{3.40}$$

and for $p = 2$, there exists a constant κ' such that

$$\mathbb{E}\|\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}\|_2 \leq \kappa'\frac{C_n^{(2\ell+1)/2}}{\sqrt{n}}.$$

The result for $2 \leq p < \infty$ follows gathering (3.35) and (3.40) and taking $C_n = \sqrt{\log n}$.

We now come to the case $p = \infty$. Arguing as for the density and noting that since $\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}$ belongs to $\mathcal{M}_{2C_n,p}$, we can thus apply Lemma 3.4.4 to get that for any $p \leq \infty$,

$$\|\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}\|_\infty \leq 2(2C_n)^{1/p}\|\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}\|_p.$$

Take $p = \log C_n$ to get that

$$\|\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}\|_\infty \leq 2(2C_n)^{1/\log C_n}\|\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}\|_p,$$

and therefore applying (3.40) with $\|h\|_{p/2}^{1/2} \leq A$ for all p, we find

$$\mathbb{E}\|\widehat{h}_n^{(\ell)} - \mathbb{E}\widehat{h}_n^{(\ell)}\|_\infty \leq 4A\kappa e^2\frac{\sqrt{\log C_n}}{\sqrt{n}}C_n^{(2\ell+1)/2}\|V^{(\ell)}\|_2(1+o(1))$$

and the result follows for $p = \infty$ always by taking $C_n = \sqrt{\log n}$. □

Remark 3.2. Note that the proofs of Theorem 3.4.1 and 3.4.5 are valid for any integrable kernel K satisfying assumptions [**K1**], [**K2**] and [**K3**].

The same arguments applied to a kernel admitting a more regular Fourier transform than V (that means a kernel admitting more finite moments) will give, with the use of Lemma 3.2.1, an estimator of the functional Γ_f when f is a polynomial function of degree less or equal to ℓ, that converges at the rate obtained in the estimation of $h^{(\ell)}$.

Recall the expression of our estimator $\widehat{\Gamma}_{f,n}$ given by (3.18):

$$\widehat{\Gamma}_{f,n} = a\widehat{h}_n^{(\ell)} + \sum_{k=0}^{\ell-1} f_{\ell-k}\widehat{h}_n^{(k)}.$$

To give a meaning to this expression in $\mathbb{L}_p(\mathbb{R})$, it is necessary that the kernel satisfy the condition:

[**K4**] For all $0 \leq k \leq \ell-1$ and for all $p \geq 2$, the integral $\int |x|^{p(\ell-k)}|K^{(k)}(x)|^p dx$ is finite.

Classical analysis results give the existence of an integrable kernel satisfying [**K1**], [**K2**], [**K3**] and [**K4**] (the assumption [**K4**] is in fact valid for all $k \in \mathbb{N}$). Note that the proof of Theorem 3.4.5 also applies with this kernel.

Corollary 3.4.7. *Let ℓ be a fixed integer, and f a polynomial function of degree less or equal to ℓ. Consider the estimator $\widehat{\Gamma}_{f,n}$ defined by (3.18) with the kernel K satisfying* [**K1**], [**K2**], [**K3**] *and* [**K4**], *and the bandwidth $C_n = \sqrt{\log n}$. If $p = \infty$, then*

$$\limsup_{n\to\infty} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}\sqrt{\log\log n}} \mathbb{E} \parallel \widehat{\Gamma}_{f,n} - \Gamma_f \parallel_\infty < +\infty.$$

If $2 \leq p < \infty$, then

$$\limsup_{n\to\infty} \sup_{\Gamma_f \in \mathcal{G}_f} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}} \mathbb{E} \parallel \widehat{\Gamma}_{f,n} - \Gamma_f \parallel_p < +\infty.$$

Remark 3.3. Note that this result is an improvement of the result in Taupin (2001) since her result gives an upper bound only for a uniform norm on a compact set and the rate she gives, $\sqrt{n}(\log n)^{-(2\ell+1)/4}(\log n)^{-1/2}$ is quite slower than this one. This comes from the method that is used, which is based here on Rosenthal's Inequality combined with Inequality (3.4.4) and not based on a chaining method.

We now give the proof of this corollary.

Proof. The main point of the proof lies in stating that the rate of convergence of $\widehat{\Gamma}_{f,n}$ is given by the rate of convergence of $\widehat{h}_n^{(\ell)}$. This follows by using equality (3.18), the result of Theorem 3.4.5 and the triangular inequality. More precisely, denoting by Q_s the function $y \mapsto y^s$, we show that there exist constants C and C' such that for all $0 \leq k \leq \ell-1$ and for all $0 \leq s \leq \ell-k$, we have, when $2 \leq p < \infty$:

$$\mathbb{E}\|Q_s(\widehat{h}_n^{(k)} - h^{(k)})\|_p \leq \frac{C(\log n)^{(2\ell+1)/4}}{\sqrt{n}},$$

and when $p = \infty$:

$$\mathbb{E}\|Q_s(\widehat{h}_n^{(k)} - h^{(k)})\|_\infty \leq \frac{C'(\log n)^{(2\ell+1)/4}\sqrt{\log\log n}}{\sqrt{n}}.$$

We follow the lines of the proof of Theorem 3.4.5. The control of the bias term defined by $\|Q_s(h^{(k)} - K_n^{(k)} * h)\|_p$ uses the following identity valid for all $t \in \mathbb{R}$:

$$[Q_s(h^{(k)} - K_n^{(k)} * h)]^*(t) = \frac{1}{i^s}\frac{\partial^s}{\partial t^s}[h^{(k)} - K_n^{(k)} * h]^*(t) = i^{k-s}\frac{\partial^s}{\partial t^s}[Q_k h^*(1 - K_n^*)](t). \quad (3.41)$$

So that concerning the uniform norm, we get:

$$\begin{aligned} \|Q_s(h^{(k)} - K_n^{(k)} * h)\|_\infty &\leq \frac{1}{2\pi}\int \left|\frac{\partial^s}{\partial t^s}[Q_k h^*(1 - K_n^*)](t)\right| dt \\ &\leq \frac{1}{2\pi}\sum_{j=0}^{s}\binom{s}{j}\int \left|\frac{\partial^j}{\partial t^j}Q_k h^*\right|(t)\left|\frac{\partial^{s-j}}{\partial t^{s-j}}(1 - K_n^*)\right|(t)dt. \end{aligned}$$

Under Assumptions [**K2**] and [**K4**], the kernel K satisfies that for all $0 \leq j \leq s$, there exist a constant $\mathcal{C}(s)$ satisfying:

$$\left|\frac{\partial^{s-j}}{\partial t^{s-j}}(1 - K_n^*)\right|(t) \leq \mathcal{C}(s)\mathbb{1}_{|t|\geq C_n}. \quad (3.42)$$

Moreover, the quantity $\frac{\partial^j}{\partial t^j}Q_k h^*(t)$ is a linear combination of powers of t times derivatives of h^* at t. Note that $h^* = \Phi_1^* g^*$ so that its derivative is a linear combination of derivatives of Φ_1^* times derivatives of g^*. Since $\mathbb{E}|X|^k < \infty$ for all $k \leq \ell$, the k-th derivative of g^* exist and is uniformly bounded. We thus bound the bias term

$$\|Q_s(h^{(k)} - K_n^{(k)} * h)\|_\infty \leq \kappa \int |t|^{k+s}e^{-t^2/2}\mathbb{1}_{|t|\geq C_n}dt \leq \kappa C_n^{k+s-1}e^{-C_n^2/2}$$

where $0 \leq k \leq \ell - 1$ and $0 \leq s \leq \ell - k$, and finally:

$$\|Q_s(h^{(k)} - K_n^{(k)} * h)\|_\infty \leq \kappa C_n^{\ell-1}e^{-C_n^2/2}.$$

Arguing as for the proof of Theorem 3.4.5, we bound the bias term for the $\mathbb{L}_p(\mathbb{R})$-norm ($p < \infty$) by writing that:

$$\begin{aligned} \|Q_s(h^{(k)} - K_n^{(k)} * h)\|_p &\leq \frac{1}{(2\pi)^{(p-1)/p}}\|[Q_s(h^{(k)} - K_n^{(k)} * h)]^\star\|_1^{(p-2)/p}\|[Q_s(h^{(k)} - K_n^{(k)} * h)]^\star\|_2^{2/p} \\ &\leq \kappa\left[\int |t|^{k+s}e^{-t^2/2}\mathbb{1}_{|t|\geq C_n}dt\right]^{(p-2)/p}\left[\int |t|^{2(k+s)}e^{-t^2}\mathbb{1}_{|t|\geq C_n}dt\right]^{1/p} \\ &\leq \kappa C_n^{k+s-1}e^{-C_n^2/2} \\ &\leq \kappa C_n^{\ell-1}e^{-C_n^2/2}, \end{aligned}$$

the last inequality being valid as $0 \leq k \leq \ell - 1$ and $0 \leq s \leq \ell - k$.

Let us study the analogue of the variance term $\mathbb{E}\|Q_s(\widehat{h}_n^{(k)} - K_n^{(k)} * h)\|_p$ for $0 \leq k \leq \ell - 1$ and $0 \leq s \leq \ell - k$. We focus on the second method used in the proof of Theorem 3.4.5 based on Rosenthal's inequality. Use the notations:

$$Q_s(x)(\widehat{h}_n^{(k)} - K_n^{(k)} * h)(x) = \frac{1}{n}\sum_{j=1}^{n} x^s \mathcal{X}_n^{(k)}(x, Y_j),$$

and apply Rosenthal's inequality to get that:

$$\mathbb{E}\|Q_s(\widehat{h}_n^{(k)} - K_n^{(k)} * h)\|_p \leq \left\{2^{p-1}\left[\frac{e\kappa\sqrt{p}}{\sqrt{n}}\right]^p \int \mathbb{E}^{p/2}|\mathcal{X}_n^{(k)}(x,Y_1)|^2|x|^{2s}dx \right.$$
$$\left. + 2^{p-1}\left[\frac{\kappa p}{n^{1-1/p}}\right]^p \int \mathbb{E}|\mathcal{X}_n^{(k)}(x,Y_1)|^p|x|^{ps}dx\right\}^{1/p}.$$

Arguing as for the proof of Theorem 3.4.5, we need to control $\int \mathbb{E}^{p/2}|\mathcal{X}_n^{(k)}(x,Y_1)|^2|x|^{2s}dx$ and $\int \mathbb{E}|\mathcal{X}_n^{(k)}(x,Y_1)|^p|x|^{ps}dx$. Writing that

$$x^{2s} = \sum_{j=0}^{2s}\binom{2s}{j}(x-y)^{2s-j}y^j,$$

we bound the first term in the following way:

$$\int \mathbb{E}^{p/2}|\mathcal{X}_n^{(k)}(x,Y_1)|^2|x|^{2s}dx \leq \int\left(\int |x|^{2s}(K_n^{(k)}(x-y))^2h(y)dy\right)^{p/2}dx.$$

and thus

$$\int \mathbb{E}^{p/2}|\mathcal{X}_n^{(k)}(x,Y_1)|^2|x|^{2s}dx$$
$$\leq (2s+1)^{p/2}\sum_{j=0}^{2s}\binom{2s}{j}\int\left(\int |x-y|^{2s-j}|y|^j(K_n^{(k)}(x-y))^2h(y)dy\right)^{2/p}dx.$$

Now, we let $u = C_n(y-x)$ in the previous integral to get:

$$\int\left(\int |x-y|^{2s-j}|y|^j(K_n^{(k)}(x-y))^2h(y)dy\right)^{2/p}dx$$
$$= \int\left(\int C_n^{2k+1-2s+j}|u|^{2s-j}|x+C_n^{-1}u|^j(K^{(k)}(u))^2h(x+C_n^{-1}u)du\right)^{2/p}dx$$

and thus by using a Taylor expansion and arguing as in the proof of Theorem 3.4.5

$$\int\left(\int |x-y|^{2s-j}|y|^j(K_n^{(k)}(x-y))^2h(y)dy\right)^{2/p}dx$$
$$\leq C_n^{(2k+1)p/2}\left(\int |u|^{2s-j}(K^{(k)}(u))^2du\right)^{p/2}\left(\int |x^jh(x)|^{p/2}dx\right)(1+o(1)).$$

We finally obtain that:

$$(2s+1)^{p/2}\sum_{j=0}^{2s}\binom{2s}{j}\int\left(\int |x-y|^{2s-j}|y|^j(K_n^{(k)}(x-y))^2h(y)dy\right)^{2/p}dx \leq \kappa C_n^{(2k+1)p/2}.$$

We now come to the second term $\int \mathbb{E}|\mathcal{X}_n^{(k)}(x,Y_1)|^p|x|^{ps}dx$. By using Fubini's Theorem and Jensens's Inequality, we infer that:

$$\int \mathbb{E}|\mathcal{X}_n^{(k)}(x,Y_1)|^p|x|^{ps}dx \leq 2^p \int\int |x|^{ps}|K_n^{(k)}(x-y)|^p h(y)dydx$$

which is bounded by:

$$2^{p(s+1)}\int\int |x-y|^{ps}|K_n^{(k)}(x-y)|^p h(y)dydx + 2^{p(s+1)}\int\int |y|^{ps}|K_n^{(k)}(x-y)|^p h(y)dydx,$$

and therefore, since the first term is negligible in front of the second one for $0 \leq s \leq \ell - k$,

$$\int \mathbb{E}|\mathcal{X}_n^{(k)}(x,Y_1)|^p|x|^{ps}dx \leq \kappa C_n^{(k+1)p-1}\|K^{(k)}\|_p^p \int |y|^{ps}h(y)dy(1+o(1)).$$

One ends up the proof of the Corollary 3.4.7 in the same way as the proof of Theorem 3.4.5. □

3.4.3 Upper bounds in $\mathbb{L}_\infty(\mathbb{R})$-norm in the trigonometric case

Consider a fixed integer $\ell \geq 1$ and fixed real parameters $\{\beta_j\}_{1\leq j\leq \ell}$. Assume that f is the fixed trigonometric function $\mathcal{C}_{\beta,\ell} : x \mapsto \sum_{j=1}^{\ell}\beta_j\cos(jx)$ (the same reasoning applies for $\mathcal{S}_{\beta,\ell} : x \mapsto \sum_{j=1}^{\ell}\beta_j\sin(jx)$). Let $\widehat{\Gamma}_{f,n}$ be defined (3.10) with the kernel S defined by (3.11) and the bandwidth $C_n = \sqrt{\log n}$. According to (3.2.2), $\widehat{\Gamma}_{f,n}$ can be written

$$\widehat{\Gamma}_{f,n} = \sum_{j=1}^{\ell}\beta_j\widehat{\Gamma}_{n,c,j}(y) = \sum_{j=1}^{\ell}\beta_j\frac{e^{-j/2}}{2}\left(e^{ijy}\widehat{h}_n(y+ij) + e^{-ijy}\widehat{h}_n(y-ij)\right), \tag{3.43}$$

where $\widehat{\Gamma}_{n,c,\ell}$ is defined by (3.19).

Theorem 3.4.8. *Let $\widehat{\Gamma}_{f,n}$ be defined by (3.10) or equivalently by (3.43) with the kernel S and the bandwidth $C_n = \sqrt{\log n}$. Then, if $p=\infty$, we have*

$$\limsup_{n\to\infty}\frac{\sqrt{n}}{\exp(\ell\sqrt{\log n})\sqrt{\log\log n}}\mathbb{E}\parallel \widehat{\Gamma}_{f,n} - \Gamma_f \parallel_\infty < \infty.$$

Moreover, the same control applies when f equals the function $\mathcal{S}_{\beta,\ell}$ and $\widehat{\Gamma}_{f,n}$ is the estimator defined by: $\widehat{\Gamma}_{f,n} = \sum_{j=1}^{\ell}\beta_j\widehat{\Gamma}_{n,s,j}(y)$.

Proof. Using the linearity of $f \mapsto \Gamma_f$, we only give the proof for the function $f : x \mapsto \cos(\ell x)$, the proof being the same in the case of the sinus.

We aim at bounding $\mathbb{E}\|\widehat{\Gamma}_{f,n} - \Gamma_f\|_p$, for $p \geq 1$. According to the triangular inequality,

$$\mathbb{E}\|\widehat{\Gamma}_{f,n} - \Gamma_f\|_p \leq \mathbb{E}\|\widehat{\Gamma}_{f,n} - \mathbb{E}\widehat{\Gamma}_n\|_p + \|\mathbb{E}\widehat{\Gamma}_{f,n} - \Gamma_f\|_p.$$

Let us study the bias term. According to the definition (3.43),

$$\mathbb{E}\widehat{\Gamma}_{f,n}(y) = \frac{e^{-\ell/2}}{2}\int [e^{i\ell y}S_n(y+i\ell-x) + e^{-i\ell y}S_n(y-i\ell-x)]h(x)dx,$$

which can also be written

$$\mathbb{E}\widehat{\Gamma}_{f,n}(y) = \frac{e^{-\ell/2}}{4\pi}\left[e^{i\ell y}\int S_n^\star(t)e^{iyt}e^{-\ell t}h^\star(t)dt + e^{-i\ell y}\int S_n^\star(t)e^{iyt}e^{\ell t}h^\star(t)dt\right].$$

It leads to the expression of the bias:

$$\Gamma_f(y) - \mathbb{E}\widehat{\Gamma}_{f,n}(y) = \frac{e^{-\ell/2}}{4\pi}\left[e^{i\ell y}\int(1-S_n^\star(t))e^{iyt}e^{-\ell t}h^*(t)dt \right.$$
$$\left. + e^{-i\ell y}\int(1-S_n^\star(t))e^{iyt}e^{\ell t}h^\star(t)dt\right].$$

Upper bounding those terms gives that:

$$|\Gamma_f(y) - \mathbb{E}\widehat{\Gamma}_{f,n}(y)| \leq \frac{e^{-\ell/2}}{2\pi}\int \mathbb{1}_{|t|\geq C_n}e^{\ell|t|}e^{-t^2/2}dt \leq \frac{e^{\ell(\ell-1)/2}e^{-(C_n-\ell)^2/2}}{\pi(C_n-\ell)}.$$

So that we conclude

$$\|\Gamma_f - \mathbb{E}\widehat{\Gamma}_{f,n}\|_\infty \leq \frac{e^{\ell(\ell-1)/2}e^{-(C_n-\ell)^2/2}}{\pi(C_n-\ell)}. \tag{3.44}$$

Note that a control of the bias term for the $\mathbb{L}_p(\mathbb{R})$-norm cannot be based on the previous upper-bound. Such a control seems to be more complicated.

It remains now to control the variance term $\mathbb{E}\|\widehat{\Gamma}_{f,n} - \mathbb{E}\widehat{\Gamma}_{f,n}\|_\infty$. Write $\widehat{\Gamma}_{f,n} - \mathbb{E}\widehat{\Gamma}_{f,n}$ as a sum of independent random variables

$$\widehat{\Gamma}_{f,n}(x) - \mathbb{E}\widehat{\Gamma}_{f,n}(x) = \frac{1}{n}\sum_{j=1}^n \mathcal{X}_{n,\ell}(x, Y_j),$$

where the $\mathcal{X}_{n,\ell}(x, Y_j)$ are centered random variables defined by

$$\mathcal{X}_{n,\ell}(x, Y_j) = \frac{e^{-\ell/2}}{2}e^{i\ell x}[S_n(x+i\ell-Y_j) - \mathbb{E}S_n(x+i\ell-Y_j)]$$
$$+ \frac{e^{-\ell/2}}{2}e^{-i\ell x}[S_n(x-i\ell-Y_j) - \mathbb{E}S_n(x-i\ell-Y_j)].$$

We first control $\mathbb{E}\|\widehat{\Gamma}_{f,n} - \mathbb{E}\widehat{\Gamma}_{f,n}\|_p$ when $2 \leq p < \infty$ and use the same argument as for the density or the polynomial case to get the result when $p = \infty$. Arguing as for the density, we control

$$\left[\int \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n \mathcal{X}_{n,\ell}(x, Y_j)\right|^p dx\right]^{1/p},$$

by applying (3.28) to $\mathbb{E}\left|n^{-1}\sum_{j=1}^n \mathcal{X}_{n,\ell}(x, Y_j)\right|^p$. This gives that

$$\mathbb{E}|\frac{1}{n}\sum_{j=1}^n \mathcal{X}_{n,\ell}(x, Y_j)|^p \leq 2^{p-1}\left[\frac{e\kappa\sqrt{p}}{\sqrt{n}}\mathbb{E}^{1/2}|\mathcal{X}_{n,\ell}(x, Y_1)|^2\right]^p + 2^{p-1}\left[\frac{\kappa p}{n^{1-1/p}}\mathbb{E}^{1/p}|\mathcal{X}_{n,\ell}(x, Y_1)|^p\right]^p,$$

and therefore $\mathbb{E}\|\widehat{\Gamma}_{f,n} - \mathbb{E}\widehat{\Gamma}_{f,n}\|_p$ is bounded by

$$\left\{2^{p-1}\left[\frac{e\kappa\sqrt{p}}{\sqrt{n}}\right]^p \int \mathbb{E}^{p/2}|\mathcal{X}_{n,\ell}(x,Y_1)|^2 dx + 2^{p-1}\left[\frac{\kappa p}{n^{1-1/p}}\right]^p \int \mathbb{E}|\mathcal{X}_{n,\ell}(x,Y_1)|^p dx\right\}^{1/p}.$$

We control the last term $\int \mathbb{E}|\mathcal{X}_{n,\ell}(x,Y_1)|^p dx$ by using Fubini's Theorem. Since the following inequality holds,

$$\mathbb{E}|\mathcal{X}_n(x,Y_1)|^p \leq e^{-\ell p/2} 2^{p-1}(\mathbb{E}|S_n(x+i\ell-Y_1)|^p + \mathbb{E}|S_n(x-i\ell-Y_1)|^p),$$

we deduce the bound

$$\int \mathbb{E}|\mathcal{X}_{n,\ell}(x,Y_1)|^p dx \leq 2^{p-1}e^{-\ell p/2}\iint (|S_n(x+i\ell-y)|^p + |S_n(x-i\ell-y)|^p)h(y)dydx$$

and therefore by the definition of the kernel S_n,

$$\int \mathbb{E}|\mathcal{X}_{n,\ell}(x,Y_1)|^p dx \leq 2^{p-1}e^{-\ell p/2}C_n^{p-1}\iint (|S(u+i\ell C_n)|^p + |S(u-i\ell C_n)|^p)h(x+u/c_n)dudx.$$

Using that $\int h(x+u/c_n)dx = 1$ we obtain that

$$\int \mathbb{E}|\mathcal{X}_{n,\ell}(x,Y_1)|^p dx \leq 2^{p-1}e^{-\ell p/2}C_n^{p-1}\int (|S(u+i\ell C_n)|^p + |S(u-i\ell C_n)|^p)du.$$

Note that for all real number v:

$$|S\left(v+i\ell C_n\right)|^2 \leq \frac{e^{2\ell C_n}}{\pi^2(v^2+\ell^2 C_n^2)}, \tag{3.45}$$

and consequently

$$\int \mathbb{E}|\mathcal{X}_{n,\ell}(x,Y_1)|^p dx \leq \frac{2^{p-1}}{\pi^p}e^{-\ell p/2}C_n^{p-1}e^{\ell p C_n}\int \frac{du}{(u^2+\ell^2 C_n^2)^{p/2}}.$$

We finally get that

$$\int \mathbb{E}|\mathcal{X}_{n,\ell}(x,Y_1)|^p dx \leq \frac{2^{p-1}}{\pi^p}e^{-\ell p/2}e^{\ell p C_n}\int \frac{dv}{(v^2+\ell^2)^{p/2}} \leq A\frac{2^{p-1}}{\pi^p}e^{-\ell p/2}e^{\ell p C_n}. \tag{3.46}$$

It remains now to bound the first term $\int \mathbb{E}^{p/2}|\mathcal{X}_n(x,Y_1)|^2 dx$. Note first that we have

$$\mathbb{E}^{p/2}|\mathcal{X}_{n,\ell}(x,Y_1)|^2 \leq e^{-\ell p/4}2^{p/2-1}(\mathbb{E}^{p/2}|S_n(x+i\ell-Y_1)|^2 + \mathbb{E}^{p/2}|S_n(x-i\ell-Y_1)|^2).$$

Therefore, we only need to control $\int \mathbb{E}^{p/2}|S_n(x+il-Y_1)|^2 dx$, the control of the second term $\int \mathbb{E}^{p/2}|S_n(x-il-Y_1)|^2 dx$ being the same. We have:

$$\int \mathbb{E}^{p/2}|S_n(x+il-Y_1)|^2 dx = \int \left(C_n^2\int |S(C_n(x-y)+i\ell C_n)|^2 h(y)dy\right)^{p/2} dx.$$

Using the bound (3.45) we deduce that:

$$\int \mathbb{E}^{p/2}|S_n(x+il-Y_1)|^2 dx \leq \int \left(C_n^2 \int \frac{e^{2\ell C_n}}{\pi^2(C_n^2(x-y)^2+\ell^2 C_n^2)} h(y)dy \right)^{p/2} dx$$
$$\leq \frac{e^{\ell p C_n}}{\pi^p} \int \left(\int \frac{h(y)dy}{(x-y)^2+\ell^2} \right)^{p/2} dx.$$

Consequently for all $2 \leq p < \infty$, we have the bound

$$\int \mathbb{E}^{p/2}|\mathcal{X}_{n,\ell}(x,Y_1)|^2 dx \leq \frac{A' e^{-\ell p/4} 2^{p/2}}{\pi^p} e^{\ell p C_n}. \tag{3.47}$$

Gathering (3.46) and (3.47) we find that

$$\mathbb{E}\|\widehat{\Gamma}_{f,n} - \mathbb{E}\widehat{\Gamma}_{f,n}\|_p \leq \left\{ 2^{p-1} \left[\frac{e\kappa\sqrt{p}}{\sqrt{n}} \right]^p A' \frac{e^{-\ell p/4} 2^{p/2}}{\pi^p} e^{\ell p C_n} + 2^{p-1} \left[\frac{\kappa p}{n^{1-1/p}} \right]^p A \frac{2^{p-1}}{\pi^p} e^{-\ell p/2} e^{\ell p C_n} \right\}^{1/p},$$

and finally by using that $p \geq 2$ we have

$$\mathbb{E}\|\widehat{\Gamma}_{f,n} - \mathbb{E}\widehat{\Gamma}_{f,n}\|_p \quad \leq \quad \kappa' \left(\frac{e\sqrt{p}}{\sqrt{n}} + \frac{p}{n^{1-1/p}} \right) e^{\ell C_n}$$

and hence

$$\mathbb{E}\|\widehat{\Gamma}_{f,n} - \mathbb{E}\widehat{\Gamma}_{f,n}\|_p \quad \leq \quad \kappa'' \frac{\sqrt{p}}{\sqrt{n}} e^{\ell C_n} \tag{3.48}$$

and the result follows for $2 \leq p < \infty$, gathering (3.44) and (3.48) and by taking $C_n = \sqrt{\log n}$.

We now come to the case $p = \infty$. Note that $\widehat{\Gamma}_{n,c,\ell}(y_0) - \mathbb{E}\widehat{\Gamma}_{n,c,\ell}(y_0)$ belongs to $\mathcal{M}_{C_n,p}$ since its Fourier transform equals

$$\frac{1}{2} e^{i\ell t y_0} e^{-\ell/2} \left[e^{-t+i\ell y_0} + e^{t-i\ell y_0} \right] S_n^\star(t),$$

which is compactly supported on $[-C_n, C_n]$. Arguing as for the density or for the polynomials, we apply Lemma 3.4.4 with $p' = \infty$ and $p = \log C_n$ and thus use (3.48) to get that there exists a constant A such that

$$\mathbb{E}\|\widehat{\Gamma}_{n,c,\ell} - \mathbb{E}\widehat{\Gamma}_{n,c,\ell}\|_\infty \quad \leq \quad A \frac{\sqrt{\log C_n}}{\sqrt{n}} e^{\ell C_n}. \tag{3.49}$$

We complete the proof of Proposition 3.4.8 by taking $C_n = \sqrt{\log n}$ in (3.49). □

Remark 3.4. We controled the analogue of the variance term when $2 \leq p \leq \infty$. The only thing needed to get the rate of convergence of our estimator for the $\mathbb{L}_p(\mathbb{R})$-norm is the control of the bias term for $2 \leq p < \infty$.

3.5 Estimation in $\mathbb{L}_p(\mathbb{R})$-norm: lower bounds

We now present our lower bounds for the minimax risk with respect to the $\mathbb{L}_p(\mathbb{R})$-norm, for $1 < p \leq \infty$. We first prove that the rates found in Theorem 3.4.1 and 3.4.5 are the best achievable ones. Generalizing the method, we also prove that when f is a polynomial function, the rates obtained in Corollary 3.4.7 also are the best achievable one.

3.5.1 Lower bounds in $\mathbb{L}_p(\mathbb{R})$-norm: the density and the polynomial case

The following theorems give lower bounds for the minimax risk over $\mathcal{H}$ with respect to the $\mathbb{L}_p(\mathbb{R})$-norm, for the estimation of the density h of the observations and its derivatives $h^{(\ell)}$, when $\ell \geq 1$. We give two different theorems as the one concerning the estimation of the density h is valid only for $p > 1$.

Theorem 3.5.1. *For all $1 < p < \infty$, we have:*

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{h\in\mathcal{H}} \frac{\sqrt{n}}{(\log n)^{1/4}} \|h - T_n\|_p > 0,$$

and if $p = \infty$,

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{h\in\mathcal{H}} \frac{\sqrt{n}}{(\log n)^{1/4}\sqrt{\log\log n}} \|h - T_n\|_\infty > 0,$$

where the infima are taken over all the estimators T_n based on the observations $Y_1, \dots, Y_n$.

Theorem 3.5.2. *For all integer $\ell \geq 1$ and all $1 \leq p < \infty$, we have:*

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{h\in\mathcal{H}} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}} \|h^{(\ell)} - T_n\|_p > 0,$$

and if $p = \infty$,

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{h\in\mathcal{H}} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}\sqrt{\log\log n}} \|h^{(\ell)} - T_n\|_\infty > 0,$$

where the infima are taken over all the estimators T_n based on the observations $Y_1, \dots, Y_n$.

We give the proofs of Theorems 3.5.1 and 3.5.2. The first part of those proofs is the same for the two theorems. It concerns the construction of a finite sub-family of densities in the model, in order to apply Fano's Lemma.

Proof. The proof is based on Fano's lemma (1952), for example in its recent version due to Birgé (2001), that we start by recalling. Consider a metric space (Θ, d) and a set of probability measures $\mathcal{P}$ indexed by Θ: $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$. The problem is to give an infimum bound to the minimax risk for the estimation of θ in Θ from an observation X with law P_θ:

$$\mathrm{R}(\Theta) = \inf_{\hat\theta(X)} \sup_{\theta\in\Theta} \mathbb{E}_\theta(d(\theta; \hat\theta(X)))$$

where the infimum is over all estimators $\hat\theta(X)$ with values in Θ. Consider a finite subset Θ' of Θ such that the cardinal $|\Theta'| \geq 3$ and for all pair (θ, θ') of distinct points in Θ', we have

$d(\theta,\theta') \geq \delta > 0$. Now, we have

$$\begin{aligned}
\mathrm{R}(\Theta) &\geq \mathrm{R}(\Theta') = \inf_{\hat\theta(X)} \sup_{\theta\in\Theta'} \mathbb{E}_\theta(d(\theta;\hat\theta(X))) \\
&\geq \inf_{\hat\theta(X)} \sup_{\theta\in\Theta'} \mathbb{E}_\theta\left(d(\theta;\hat\theta(X))1\!\!1_{d(\theta;\hat\theta(X))\geq\delta/2}\right) \\
&\geq \frac{\delta}{2}\inf_{\hat\theta(X)} \sup_{\theta\in\Theta'} \mathrm{P}_\theta\left(d(\theta;\hat\theta(X)) \geq \frac{\delta}{2}\right).
\end{aligned}$$

Using the fact that the points in Θ' are δ-separated, we get the following inclusion

$$\{\omega\in\Omega/\operatorname*{argmin}_{\theta'} d(\theta',\hat\theta(X(\omega))) \neq \theta\} \subset \left\{\omega\in\Omega/d(\theta,\hat\theta(X(\omega))) \geq \frac{\delta}{2}\right\}.$$

Then

$$\begin{aligned}
\mathrm{R}(\Theta) &\geq \frac{\delta}{2}\inf_{\hat\theta(X)} \sup_{\theta\in\Theta'} \mathrm{P}_\theta\left(\operatorname*{argmin}_{\theta'} d(\theta';\hat\theta(X)) \neq \theta\right) \\
&\geq \frac{\delta}{2}\inf_{\hat T(X)} \sup_{\theta\in\Theta'} \mathrm{P}_\theta\left(\hat T(X) \neq \theta\right)
\end{aligned}$$

where the last infimum concerns estimators $\hat T(X)$ taking values only in the finite subset Θ'. In other words, we have the following infimum bound

$$\mathrm{R}(\Theta) \geq \frac{\delta}{2}\inf_{\hat T(X)}\left(1 - \inf_{\theta\in\Theta'} \mathrm{P}_\theta\left(\hat T(X) = \theta\right)\right)$$

when $\hat T(X)$ ranges over all estimators with values in Θ'. Now, we use the following result, based on Fano's inequality (1952) and proved by Birgé (2001):

Lemma 3.5.3. *Let θ_0 be a fixed point in Θ' and set*

$$\overline{\mathcal{K}} = \frac{1}{|\Theta'|}\sum_{\theta\in\Theta'} \mathcal{K}(P_\theta;P_{\theta_0})$$

where $\mathcal{K}(P_\theta;P_{\theta_0})$ denotes the Kullback-Leibler divergence between P_θ and P_{θ_0}. There exists an absolute constant α such that if $\hat T(X)$ is an estimator taking values in Θ', we have

$$\inf_{\theta\in\Theta'} P_\theta(\hat T(X) = \theta) \leq \alpha \vee \frac{\overline{\mathcal{K}}}{\log(|\Theta'|+1)}.$$

And finally

$$\mathrm{R}(\Theta) \geq \frac{\delta}{2}\left(1 - \alpha \vee \frac{\overline{\mathcal{K}}}{\log(|\Theta'|+1)}\right).$$

We are going to apply this result for $\mathcal{P} = \{hd\lambda\ ;\ h\in\mathcal{H}\}^{\otimes n}$ where $d\lambda$ is the Lebesgue measure on $\mathbb{R}$, and $(\Theta,d) = (\mathcal{H}_\ell = \{h^{(\ell)}\ ;\ h\in\mathcal{H}\}, \|\cdot\|_p)$ when $1\leq p\leq\infty$ and ℓ is a positive integer. In other words, we estimate the density of the observations and its derivatives, and compute an infimum bound for the minimax risk expressed with the $\mathbb{L}_p(\mathbb{R})$-norm

$$\mathrm{R}_{n,p}(\mathcal{H}_\ell) = \inf_{T_n}\sup_{h\in\mathcal{H}} \mathbb{E}_h\|T_n - h^{(\ell)}\|_p,$$

where the infimum is taken over all the estimators T_n based on the observations $Y_1, \dots Y_n$. Fix an integer $m \geq 3$. In order to apply Fano's Lemma, we will

- define a family of probability densities belonging to $\mathcal{H}$: $(\varphi_{m,a})_{a\in\mathcal{A}}$ indexed by some finite set $\mathcal{A}$ whose cardinal depends on m,

- calculate the minimum distance between two points in this family: $\inf_{a,a'\in\mathcal{A}} \|\varphi^{(\ell)}_{m,a} - \varphi^{(\ell)}_{m,a'}\|_p \geq \delta_m$,

- check that there exists a constant $C_1 \leq \alpha$ and a point $\varphi_{m,0}$ in the family such that

$$\frac{n\sum_{a\in\mathcal{A}} \mathcal{K}(\varphi_{m,a};\varphi_{m,0})}{|\mathcal{A}|\log(|\mathcal{A}|+1)} \leq C_1,$$

(note that the Kullback-Leibler distance between points in this family of probabilities satisfies $\mathcal{K}(P^{\otimes n}_{m,a}; P^{\otimes n}_{m,0}) = n\mathcal{K}(\varphi_{m,a};\varphi_{m,0})$, where $P_{m,a}$ denotes the probability $\varphi_{m,a}d\lambda$),

- and then conclude that

$$\mathrm{R}_{n,p}(\mathcal{H}_\ell) \geq \frac{\alpha}{2}\delta_m.$$

We start with the construction of the family. Define the following functions on $\mathbb{R}$:

$$\alpha(x) = \frac{\sin^2(x)}{\pi x^2}\ ; \qquad \alpha_0 = \alpha * \alpha\ ; \qquad f_0 = \alpha_0 * \Phi_1$$

$$V(x) = \frac{\cos(x)-\cos(2x)}{\pi x^2}\ ; \qquad V_{m,j}(x) = \theta_m V(mx - j)\ ,\ \ 1 \leq j \leq m-1\ ; \qquad V_{m,0} = 0,$$

where m is an integer greater or equal to 3, and the sequence $\{\theta_m\}_{m\geq 0}$ is positive and converges to zero as m tends to infinity. Note that α and then also α_0 and f_0 are probability densities on $\mathbb{R}$. We denote by $\bar{V}_{m,j}$ the normalizing constant:

$$\bar{V}_{m,j} = \int f_0(x)V_{m,j}(x)dx, \quad \forall 0 \leq j \leq m-1. \tag{3.50}$$

Now, let us consider the following family of probability densities:

$$\begin{aligned}\varphi_{m,a}(x) &= f_0(x) + \sum_{j=1}^{m-1} a(j) f_0(x)(V_{m,j}(x) - \bar{V}_{m,j}) \qquad (3.51)\\ &= \left(1 - \sum_{j=1}^{m-1} a(j)\bar{V}_{m,j}\right)\alpha_0 * \Phi_1(x) + \sum_{j=1}^{m-1} a(j) f_0(x) V_{m,j}(x)\end{aligned}$$

where a ranges the set $\mathcal{A}$ of applications of $\{1;\cdots;m-1\}$ to $\{0;1\}$. Note that the constant $\bar{V}_{m,j}$ ensures that $\int \varphi_{m,a}(x)dx = 1$. The following lemma states that the set of functions $\{\varphi_{m,a}\}_{a\in\mathcal{A}}$ is a family of probability densities included in $\mathcal{H}$ under suitable assumption on the parameter θ_m.

Assumption 3.1. $m\theta_m e^{2m^2+4m} \underset{\mathrm{m}\to\infty}{\longrightarrow} 0.$

Lemma 3.5.4. *Under Assumption 3.1 and for large enough m, the family $\{\varphi_{m,a}\}_{a\in\mathcal{A}}$ is included into $\mathcal{H}$.*

The proof of Lemma 3.5.4 can be found in Appendix C.

First case: the uniform norm

We first consider the case of the uniform norm. Consider the subset $\mathcal{A}_1$ of $\mathcal{A}$ of the applications of $\{1;\cdots;m-1\}$ to $\{0;1\}$ that take value 0 everywhere, except at one point. For simplicity of notations, we denote by

$$\varphi_{m,j}(x) = f_0(x)(1 + V_{m,j}(x) - \bar{V}_{m,j}), \quad \forall\, 0 \le j \le m-1. \tag{3.52}$$

The resulting family $(\varphi_{m,a})_{a \in \mathcal{A}_1 \cup \{0\}}$ is exactly the family $(\varphi_{m,j})_{0 \le j \le m-1}$. This is the family we will work with.
Now that our family is defined, the second step of the proof consists in lower bounding the distance between the different parameters in this family that we wish to estimate. We compute this minimum distance in the following lemma.

Lemma 3.5.5. *For all fixed integer ℓ, there exists a positive constant C such that for all j, k in $\{0;\cdots;m-1\}$, we have*

$$\|\varphi^{(\ell)}_{m,j} - \varphi^{(\ell)}_{m,k}\|_\infty \ge C m^\ell \theta_m.$$

Proof. We will consider the different cases $\ell = 0$ (concerning the estimation of the density of the observations) and $\ell \ge 1$ (concerning the derivatives of this density).
<u>First case:</u> $\ell = 0$.
By definition, we have:

$$\|\varphi_{m,j} - \varphi_{m,k}\|_\infty = \|f_0(V_{m,j} - V_{m,k} - \bar{V}_{m,j} + \bar{V}_{m,k})\|_\infty$$

and therefore satisfies the inequalities

$$\begin{aligned} \|\varphi_{m,j} - \varphi_{m,k}\|_\infty &\ge \theta_m f_0\left(\frac{j}{m}\right)\left(|V(0) - V(j-k)| - \frac{|\bar{V}_{m,j} - \bar{V}_{m,k}|}{\theta_m}\right) \\ &\ge \theta_m \inf_{[0;1]} f_0 \times \left(|V(0) - V(j-k)| - \frac{|\bar{V}_{m,j} - \bar{V}_{m,k}|}{\theta_m}\right). \end{aligned}$$

Using the expression (3.50), we get

$$|\bar{V}_{m,j}| \le \theta_m \|f_0\|_\infty \int |V(mx - j)| dx$$

and hence

$$|\bar{V}_{m,j}| \le \frac{\theta_m}{m} \|f_0\|_\infty \|V\|_1. \tag{3.53}$$

It follows that

$$\frac{|\bar{V}_{m,j} - \bar{V}_{m,k}|}{\theta_m} \le \frac{2}{m} \|f_0\|_\infty \|V\|_1 \underset{\mathrm{m}\to\infty}{\longrightarrow} 0.$$

Since the function V is an even function that converges to zero at infinity, we obtain

$$|V(0) - V(j-k)| \ge \inf_{l \ge 1} \left|\frac{3}{2\pi} - V(l)\right| \ge \inf\left(\frac{3}{4\pi}; \inf_{1 \le l \le m_0} \left|\frac{3}{2\pi} - V(l)\right|\right) > 0,$$

for some positive fixed integer m_0. Finally, there exists a non-negative constant C such that the uniform distance between two points of this family is lower bounded in the following way:

$$\|\varphi_{m,j} - \varphi_{m,k}\|_\infty \geq C\theta_m.$$

Second case: $\ell \geq 1$.
Now, let us consider the case $\ell \geq 1$. We have the following identity, valid for all $0 \leq j \leq m-1$:

$$\varphi_{m,j}^{(\ell)}(x) = f_0^{(\ell)}(x)(1 + V_{m,j}(x) - \bar{V}_{m,j}) + \sum_{p=0}^{\ell-1} \binom{\ell}{p} f_0^{(p)}(x) V_{m,j}^{(\ell-p)}(x).$$

Then by definition,

$$\|\varphi_{m,j}^{(\ell)} - \varphi_{m,k}^{(\ell)}\|_\infty = \left\| f_0^{(\ell)}(V_{m,j} - V_{m,k} - \bar{V}_{m,j} + \bar{V}_{m,k}) + \sum_{p=0}^{\ell-1} \binom{\ell}{p} f_0^{(p)}(V_{m,j}^{(\ell-p)} - V_{m,k}^{(\ell-p)}) \right\|_\infty .$$

Use the triangular inequality to get that:

$$\begin{aligned}\left\|\varphi_{m,j}^{(\ell)} - \varphi_{m,k}^{(\ell)}\right\|_\infty \geq \left\| f_0(V_{m,j}^{(\ell)} - V_{m,k}^{(\ell)})\right\|_\infty - \sum_{p=1}^{\ell-1} \binom{\ell}{p} \left\| f_0^{(p)}\right\|_\infty \left\| V_{m,j}^{(\ell-p)} - V_{m,k}^{(\ell-p)}\right\|_\infty \\ - \left\| f_0^{(\ell)}(V_{m,j} - V_{m,k} + \bar{V}_{m,k} - \bar{V}_{m,j})\right\|_\infty . \quad (3.54)\end{aligned}$$

The first term is lower bounded in the following way:

$$\begin{aligned}\|f_0(V_{m,j}^{(\ell)} - V_{m,k}^{(\ell)})\|_\infty &= m^\ell \theta_m \sup_{u\in\mathbb{R}} f_0\left(\frac{u+j}{m}\right) \left|V^{(\ell)}(u) - V^{(\ell)}(u+j-k)\right| \\ &\geq m^\ell \theta_m \sup_{u\in[0;1]} f_0\left(\frac{u+j}{m}\right) \left|V^{(\ell)}(u) - V^{(\ell)}(u+j-k)\right| \\ &\geq m^\ell \theta_m \inf_{u\in[0;1]} f_0\left(\frac{u+j}{m}\right) \sup_{u\in[0;1]} \left|V^{(\ell)}(u) - V^{(\ell)}(u+j-k)\right|.\end{aligned}$$

Note that for all u in $[0;1]$ and all $0 \leq j \leq m-1$, the real number $(u+j)/m$ belongs to $[0;1]$. Then we get

$$\begin{aligned}\|f_0(V_{m,j}^{(\ell)} - V_{m,k}^{(\ell)})\|_\infty &\geq m^\ell \theta_m \left(\inf_{[0;1]} f_0\right) \inf_{|n|\geq 1} \sup_{u\in[0;1]} \left|V^{(\ell)}(u) - V^{(\ell)}(u+n)\right| \\ &\geq C_1 m^\ell \theta_m,\end{aligned}$$

where C_1 is a positive constant. Now upper bounding the remainder terms appearing in (3.54), we get:

$$\begin{aligned}\left\|\varphi_{m,j}^{(\ell)} - \varphi_{m,k}^{(\ell)}\right\|_\infty \geq m^\ell \theta_m \left(\inf_{[0;1]} f_0\right) \inf_{|n|\geq 1} \sup_{u\in[0;1]} |V^{(\ell)}(u) - V^{(\ell)}(u+n)| \\ - 2^\ell m^{\ell-1}\theta_m \max_{1\leq p\leq \ell-1} \left(\|V^{(p)}\|_\infty \|f_0^{(p)}\|_\infty\right) \\ - \|f_0^{(\ell)}\|_\infty \left(\|V_{m,j} - V_{m,k}\|_\infty + |\bar{V}_{m,j} - \bar{V}_{m,k}|\right).\end{aligned}$$

Use the definition of $V_{m,j}$ and the inequality (3.53) to get:

$$\|\varphi_{m,j}^{(\ell)} - \varphi_{m,k}^{(\ell)}\|_\infty \geq m^\ell \theta_m \left(\inf_{[0;1]} f_0\right) \inf_{|n|\geq 1} \sup_{u\in[0;1]} |V^{(\ell)}(u) - V^{(\ell)}(u+n)|$$
$$- 2^\ell m^{\ell-1}\theta_m \max_{1\leq p\leq \ell-1} \left(\|V^{(p)}\|_\infty \|f_0^{(p)}\|_\infty\right)$$
$$- \|f_0^{(\ell)}\|_\infty \left(2\theta_m\|V\|_\infty + 2\frac{\theta_m}{m}\|f_0\|_\infty\|V\|_1\right).$$

For large enough m, and using that $\ell \geq 1$, we have

$$\|\varphi_{m,j}^{(\ell)} - \varphi_{m,k}^{(\ell)}\|_\infty \geq \frac{1}{2} m^\ell \theta_m \left(\inf_{[0;1]} f_0\right) \inf_{|n|\geq 1} \sup_{u\in[0;1]} |V^{(\ell)}(u) - V^{(\ell)}(u+n)|.$$

Finally, the uniform distance between two points of this family is lower bounded in the following way

$$\|\varphi_{m,j}^{(\ell)} - \varphi_{m,k}^{(\ell)}\|_\infty \geq C m^\ell \theta_m,$$

where C is a positive constant. This concludes the proof of Lemma 3.5.5. □

The following step of the proof is to check that there exists a constant $C_1 \leq \alpha$ such that

$$\frac{n\sum_{j=1}^{m-1} \mathcal{K}(\varphi_{m,j};\varphi_{m,0})}{m\log(m+1)} \leq C_1.$$

We first compute the Kullback-Leibler divergence involved in our family:

$$\mathcal{K}(\varphi_{m,j};\varphi_{m,0}) = \mathcal{K}(\varphi_{m,j};f_0) = \int \log\left(\frac{\varphi_{m,j}(x)}{f_0(x)}\right)\varphi_{m,j}(x)dx.$$

By definition, and using the inequality $\log(1+u) \leq u$, we get that $\mathcal{K}(\varphi_{m,j};f_0)$ given by

$$\mathcal{K}(\varphi_{m,j};f_0) = \int \log\left(1 + V_{m,j}(x) - \bar{V}_{m,j}\right) f_0(x)(1 + V_{m,j}(x) - \bar{V}_{m,j})dx$$

is bounded in the following way

$$\mathcal{K}(\varphi_{m,j};f_0) \leq \int \left(V_{m,j}(x) - \bar{V}_{m,j}\right) f_0(x)(1 + V_{m,j}(x) - \bar{V}_{m,j})dx.$$

Since

$$\int (V_{m,j}(x) - \bar{V}_{m,j}) f_0(x)dx = 0,$$

then

$$\mathcal{K}(\varphi_{m,j};f_0) \leq \int (V_{m,j}(x) - \bar{V}_{m,j})^2 f_0(x)dx$$

which combined with the fact that

$$\int (V_{m,j}(x) - \bar{V}_{m,j})^2 f_0(x)dx = \int V_{m,j}^2(x) f_0(x)dx - \bar{V}_{m,j}^2$$

give the bound

$$\mathcal{K}(\varphi_{m,j}; f_0) \leq \int V_{m,j}^2(x) f_0(x) dx \leq \frac{\theta_m^2}{m} \|f_0\|_\infty \|V\|_2^2.$$

Finally, Fano's Lemma tells us that there exists a positive constant C such that

$$\inf_{\hat{h}_n} \sup_{h \in \mathcal{H}} \|h^{(\ell)} - \hat{h}_n\|_\infty \geq C m^\ell \theta_m^\star(n),$$

where $\theta_m^\star(n)$ is the supremum over all the parameters θ_m satisfying the following conditions

$$\|f_0\|_\infty \|V\|_2^2 \frac{n\theta_m^2}{m \log(m+1)} \leq \alpha \qquad \text{and} \qquad m\theta_m e^{2m^2+4m} \underset{\mathrm{m}\to\infty}{\longrightarrow} 0.$$

Choose

$$\theta_m = e^{-3m^2} \qquad \text{and} \qquad m = \sqrt{\frac{1}{6}\log n - \frac{1}{12}\log\log n - \frac{1}{6}\log\log\log n}.$$

We finally have $\theta_m = e^{-3m^2} = (\log n)^{1/4} (\log\log n)^{1/2}/\sqrt{n}$ and we obtain the result for $p = \infty$:

$$\inf_{\hat{h}_n} \sup_{h \in \mathcal{H}} \|h^{(\ell)} - \hat{h}_n\|_\infty \geq C \frac{(\log n)^{(2\ell+1)/4}\sqrt{\log\log n}}{\sqrt{n}}.$$

This concludes the proof of Theorem 3.5.1.

Second case: the $\mathbb{L}_p(\mathbb{R})$-norm

We now come to lower bound the risk with respect to the $\mathbb{L}_p(\mathbb{R})$-norm, when $1 \leq p < \infty$. Consider the subset $\mathcal{A}_2$ of applications of $\{1; \cdots; m-1\}$ to $\{0; 1\}$ containing the application identically equal to zero and such that for all $a \neq a'$ in $\mathcal{A}_2$, we have

$$\sum_{j=1}^{m-1} |a(j) - a'(j)| > \frac{m-1}{4}.$$

The cardinal of this set of applications satisfies

$$|\mathcal{A}_2| \geq e^{\frac{m-1}{8}}$$

(the proof of this result stands in Ibragimov and Hasminskii (1983)). We will work with this family of densities $\{\varphi_{m,a}\}_{a \in \mathcal{A}_2}$.
The first step consists in computing the minimum distance in $\mathbb{L}_p(\mathbb{R})$-norm between two points in our family of parameters $\{\varphi_{m,a}^{(\ell)}\}_{a \in \mathcal{A}_2}$.

Lemma 3.5.6. *For all $1 < p < \infty$, there exists a positive constant C such that for all a, a' in $\mathcal{A}_2$:*

$$\|\varphi_{m,a} - \varphi_{m,a'}\|_p \geq C\theta_m.$$

Moreover, for all fixed integer ℓ and all $1 \leq p < \infty$, there exists a positive constant C' such that for all a, a' in $\mathcal{A}_2$:

$$\|\varphi_{m,a}^{(\ell)} - \varphi_{m,a'}^{(\ell)}\|_p \geq C' m^\ell \theta_m.$$

The proof of this lemma stands in Appendix C. It is a careful generalization of the methods used to prove Lemma 3.5.5. Note that in this proof, we do not work with the function V but with the function $V_\lambda = V(\lambda\cdot)$ for some fixed well-chosen $\lambda > 0$. In the following of the proof, we will assume with no loss of generality and in order to simplify notations, that $\lambda = 1$ fits.

Now, we apply Fano's lemma and claim that for all $1 < p < \infty$, there exists a positive constant C such that

$$\inf_{\hat{h}_n} \sup_{h \in \mathcal{H}} \|h - \hat{h}_n\|_p \geq C\theta_m^\star(n),$$

and for all integer $\ell \geq 1$ and all $1 \leq p < \infty$, there exists a positive constant C' such that

$$\inf_{\hat{h}_n} \sup_{h \in \mathcal{H}} \|h^{(\ell)} - \hat{h}_n\|_p \geq C' m^\ell \theta_m^\star(n),$$

where $\theta_m^\star(n)$ is the supremum over all the parameters θ_m satisfying the following conditions

$$\frac{n \sum_{a \in \mathcal{A}_2} \mathcal{K}(\varphi_{m,a}; f_0)}{|\mathcal{A}_2| \log(|\mathcal{A}_2| + 1)} \leq \alpha \qquad \text{and} \qquad m\theta_m e^{2m^2+4m} \underset{\mathrm{m}\to\infty}{\longrightarrow} 0.$$

Compute the Kullback-Leibler divergence between $\varphi_{m,a}$ and f_0:

$$\mathcal{K}(\varphi_{m,a}; f_0)$$
$$= \int \log\left(1 + \sum_{j=1}^{m-1} a(j)(V_{m,j}(x) - \bar{V}_{m,j})\right) f_0(x) \left(1 + \sum_{j=1}^{m-1} a(j)(V_{m,j}(x) - \bar{V}_{m,j})\right) dx$$

Using that $\log(1+u) \leq u$ combined with the definition of $\bar{V}_{m,j}$, we get the bound

$$\mathcal{K}(\varphi_{m,a}; f_0) \leq \sum_{j=1}^{m-1} a(j) \int (V_{m,j}(x) - \bar{V}_{m,j}) f_0(x) \left(1 + \sum_{k=1}^{m-1} a(k)(V_{m,k}(x) - \bar{V}_{m,k})\right) dx$$

and using that

$$\sum_{j=1}^{m-1} a(j) \int (V_{m,j}(x) - \bar{V}_{m,j}) f_0(x) \left(1 + \sum_{k=1}^{m-1} a(k)(V_{m,k}(x) - \bar{V}_{m,k})\right) dx$$

also equals

$$\sum_{j=1}^{m-1}\sum_{k=1}^{m-1} a(j)a(k) \int (V_{m,j}(x) - \bar{V}_{m,j})(V_{m,k}(x) - \bar{V}_{m,k}) f_0(x) dx$$

we get the bound

$$\mathcal{K}(\varphi_{m,a}; f_0) \leq \sum_{j=1}^{m-1}\sum_{k=1}^{m-1} a(j)a(k) \left[\int V_{m,j}(x)V_{m,k}(x) f_0(x) dx - \bar{V}_{m,j}\bar{V}_{m,k}\right].$$

Apply Cauchy-Schwarz inequality to

$$\left|\int V_{m,j}(x)V_{m,k}(x) f_0(x) dx\right| = \theta_m^2 \left|\int V(mx - j)V(mx - k) f_0(x) dx\right|$$

to obtain that

$$\left|\int V_{m,j}(x)V_{m,k}(x)f_0(x)dx\right| \leq \theta_m^2 \left(\int V^2(mx-j)f_0(x)dx\right)^{1/2} \left(\int V^2(mx-k)f_0(x)dx\right)^{1/2}$$

and therefore

$$\left|\int V_{m,j}(x)V_{m,k}(x)f_0(x)dx\right| \leq \frac{\theta_m^2}{m}\|f_0\|_\infty\|V\|_2^2.$$

The use of Inequality (3.53) provides

$$\mathcal{K}(\varphi_{m,a}; f_0) \leq (m-1)^2\frac{\theta_m^2}{m}\|f_0\|_\infty\left(\|V\|_2^2 + \frac{1}{m}\|f_0\|_\infty\|V\|_1^2\right) \leq 2m\theta_m^2\|f_0\|_\infty\|V\|_2^2.$$

Now, remember that the cardinal of $\mathcal{A}_2$ satisfies $\log(|\mathcal{A}_2|) \geq (m-1)/8$. We have to find the supremum of the parameters θ_m satisfying the conditions

$$16\|f_0\|_\infty\|V\|_2^2\frac{m\theta_m^2}{m-1} \leq \frac{\alpha}{n} \qquad \text{and} \qquad m\theta_m e^{2m^2+4m} \underset{m\to\infty}{\longrightarrow} 0.$$

Choose

$$\theta_m = e^{-3m^2} \qquad \text{and} \qquad m = \sqrt{\frac{1}{6}\log n - \frac{1}{12}\log\log n}$$

to get that there exists a positive constant C such that

$$\inf_{\hat{h}_n}\sup_{h\in\Sigma}\|h^{(\ell)} - \hat{h}_n\|_p \geq Cm^\ell\theta_m = C\frac{(\log n)^{(2\ell+1)/4}}{\sqrt{n}},$$

for all $1 \leq p < \infty$ and all $\ell \geq 1$, and that there exists a positive constant C' such that

$$\inf_{\hat{h}_n}\sup_{h\in\Sigma}\|h - \hat{h}_n\|_p \geq C\theta_m = C\frac{(\log n)^{3/4}}{\sqrt{n}},$$

for all $1 < p < \infty$, which entails the result. □

The following corollary of Theorem 3.5.2 gives lower bounds for the estimation of Γ_f when f is a polynomial function. It is based on the use of Lemma 3.2.1 and its proof is a generalization of the methods used in the proof of Theorem 3.5.2.

Corollary 3.5.7. *Fix an integer $\ell \geq 1$ and a polynomial function f of degree less or equal to ℓ. Then, for all $1 \leq p < \infty$, we have:*

$$\liminf_{n\to\infty}\inf_{T_n}\sup_{\Gamma_f\in\mathcal{G}_f}\frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}}\|\Gamma_f - T_n\|_p > 0,$$

and if $p = \infty$,

$$\liminf_{n\to\infty}\inf_{T_n}\sup_{\Gamma_f\in\mathcal{G}_f}\frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}\sqrt{\log\log n}}\|\Gamma_f - T_n\|_\infty > 0,$$

where the infima are taken over all the estimators T_n based on the observations $Y_1, \ldots, Y_n$.

Proof. The family used in the preceding proof is not adapted here in the sense that the product of a polynomial function with a derivative of f_0 does not always belong to $\mathbb{L}_p(\mathbb{R})$. We use a new function $f_{0,\ell}$ for which we will prove that the induced family $(\varphi_{m,a,\ell})_{a\in\mathcal{A}}$ still belongs to $\mathcal{H}$, and such that the Kullback-Leibler distance between the densities behaves in the same way. More precisely, when f is a polynomial function of degree equal to ℓ, define the densities:

$$\alpha_\ell(x) = C_\ell \left(\frac{\sin x}{\pi x}\right)^{2\ell+2},$$

where C_ℓ is a normalising constant, and

$$\alpha_{0,\ell} = \alpha_\ell * \alpha_\ell\ , \qquad f_{0,\ell} = \alpha_{0,\ell} * \Phi_1.$$

The family of functions $\{\varphi_{m,a,\ell}\}_{a\in\mathcal{A}}$ is defined in the same way as $\{\varphi_{m,a}\}_{a\in\mathcal{A}}$, replacing f_0 by $f_{0,\ell}$. This new choice of $f_{0,\ell}$ changes the proof of Lemma 3.5.4. Let us consider the assumption needed to generalize this lemma.

Assumption 3.2. $m\theta_m e^{2m^2+4(\ell+1)m} \longrightarrow_{m\to\infty} 0.$

Lemma 3.5.8. *Under Assumption 3.2 and for large enough m, the family $\{\varphi_{m,a,\ell}\}_{a\in\mathcal{A}}$ is included into $\mathcal{H}$.*

The proof of this lemma stands in Appendix C.
The following lines of the proof of Theorem 3.5.2 are not modified by the change of f_0. In particular, we still have the property that for all $a \neq a'$ in $\mathcal{A}$,

$$\|\varphi^{(\ell)}_{m,a,\ell} - \varphi^{(\ell)}_{m,a',\ell}\|_p \geq C m^\ell \theta_m,$$

and the Kullback-Leibler distance of the family is bounded in the same way (replacing the uniform norms of f_0 and its derivatives by the uniform norms of $f_{0,\ell}$ and its derivatives).
Using Lemma 3.2.1, this family of densities induces a family of functionals $\{\Gamma_{f,m,a}\}_{a\in\mathcal{A}}$ defined by

$$\Gamma_{f,m,a}(y) = \gamma\varphi^{(\ell)}_{m,a,\ell}(y) + \sum_{k=0}^{\ell-1} f_{\ell-k}(y)\varphi^{(k)}_{m,a,\ell}(y),$$

where $\gamma \neq 0$ and f_j is a polynomial function of degree j related to f.

The only thing to care about is that the minimum distance between two functionals $\|\Gamma_{f,m,a} - \Gamma_{f,m,a'}\|_p$ in the family is still lower bounded by a constant times $m^\ell\theta_m$. Then the conclusion follows arguing as in the proofs of Theorem 3.5.1 and Theorem 3.5.2.

Let us lower bound the distance $\|\Gamma_{f,m,a} - \Gamma_{f,m,a'}\|_p$ when $a \neq a'$ belong to $\mathcal{A}$. Using the triangular inequality:

$$\begin{aligned}\|\Gamma_{f,m,a} - \Gamma_{f,m,a'}\|_p &\geq \gamma\|\varphi^{(\ell)}_{m,a,\ell} - \varphi^{(\ell)}_{m,a',\ell}\|_p - \sum_{k=0}^{\ell-1}\|f_{\ell-k}(\varphi^{(k)}_{m,a,\ell} - \varphi^{(k)}_{m,a',\ell})\|_p \\ &\geq C m^\ell\theta_m - \sum_{k=0}^{\ell-1}\|f_{\ell-k}(\varphi^{(k)}_{m,a,\ell} - \varphi^{(k)}_{m,a',\ell})\|_p.\end{aligned}$$

When $p=\infty, k=0$, we immediatly deduce the bound:

$$\begin{aligned}
\|f_\ell(\varphi_{m,s,\ell}-\varphi_{m,t,\ell})\|_\infty &= \|f_\ell f_{0,\ell}(V_{m,s}-V_{m,t}-\bar V_{m,s}+\bar V_{m,t})\|_\infty \\
&\leq \|f_\ell f_{0,\ell}\|_\infty\left(2\theta_m\|V\|_\infty+\frac{\theta_m}{m}\|f_{0,\ell}\|_\infty\|V\|_1\right) \\
&\leq C\theta_m.
\end{aligned}$$

When $p=\infty, 1\leq k\leq \ell-1$, we have the bound:

$$\begin{aligned}
\|f_{\ell-k}(\varphi^{(k)}_{m,s,\ell}-\varphi^{(k)}_{m,t,\ell})\|_\infty = \|(f_{\ell-k}f_{0,\ell})^{(k)}&(V_{m,s}-V_{m,t}-\bar V_{m,s}+\bar V_{m,t}) \\
&+\sum_{j=0}^{k-1}\binom{k}{j}(f_{\ell-k}f_0)^{(j)}(V^{(k-j)}_{m,s}-V^{(k-j)}_{m,t})\|_\infty
\end{aligned}$$

and is thus bounded by

$$\begin{aligned}
\|f_{\ell-k}(\varphi^{(k)}_{m,s,\ell}-\varphi^{(k)}_{m,t,\ell})\|_\infty &\leq \|(f_{\ell-k}f_{0,\ell})^{(k)}\|_\infty\left(2\theta_m\|V\|_\infty+\frac{\theta_m}{m}\|f_{0,\ell}\|_\infty\|V\|_1\right) \\
&\quad +2^{k+1}\max_{0\leq j\leq k-1}\|(f_{\ell-k}f_0)^{(j)}\|_\infty\theta_m m^k\max_{1\leq j\leq k}\|V^{(j)}\|_\infty \\
&\leq Cm^{\ell-1}\theta_m.
\end{aligned}$$

When $1\leq p<\infty, k=0$, the bounds are the following ones:

$$\|f_\ell(\varphi_{m,a,\ell}-\varphi_{m,a',\ell})\|_p\leq\sum_{j=1}^{m-1}\|f_\ell f_0\|_\infty\|V_{m,j}-\bar V_{m,j}\|_p\leq Cm^{1-1/p}\theta_m$$

which is still negligible with respect to $m^\ell\theta_m$.
When $1\leq p<\infty, 1\leq k\leq\ell-1$,

$$\begin{aligned}
\|f_{\ell-k}(\varphi^{(k)}_{m,a,\ell}-\varphi^{(k)}_{m,a',\ell})\|_p &\leq \sum_{j=1}^{m-1}\left[\|(f_{\ell-k}f_{0,\ell})^{(k)}(V_{m,j}-\bar V_{m,j})\|_p+\sum_{j=0}^{k-1}\binom{k}{j}\|(f_{\ell-k}f_0)^{(j)}V^{(k-j)}_{m,j}\|_p\right] \\
&\leq Cm^{1+k-1/p}\theta_m,
\end{aligned}$$

which is still negligible with respect to $m^\ell\theta_m$.
The proof of Corollary 3.5.7 is completed by using the same arguments as in the proof of Theorem 3.5.2.

□

3.5.2 Lower bounds in $\mathbb{L}_\infty(\mathbb{R})$-norm: the trigonometric case

We prove the following theorem.

Theorem 3.5.9. *Fix an integer $\ell\geq 1$ and a function f of the form $x\mapsto\sum_{j=1}^{\ell}\beta_j\cos(jx)$ or $x\mapsto\sum_{j=1}^{\ell}\beta_j\sin(jx)$ where the $\{\beta_j\}_{1\leq j\leq\ell}$ are real fixed parameters. Then we have:*

$$\liminf_{n\to\infty}\inf_{T_n}\sup_{\Gamma_f\in\mathcal{G}_f}\frac{\sqrt{n}(\log n)^{3/4}}{\exp(\ell\sqrt{\log n})}\|\Gamma_f-T_n\|_\infty>0,$$

where the infimum is taken over all the estimators T_n based on the observations $Y_1,\dots,Y_n$.

Proof. We only give the proof for the function $f : x \mapsto \cos(\ell x)$ the one for $x \mapsto \sin(\ell x)$ is proved exactly in the same way, and by linearity of $f \mapsto \Gamma_f$, we obtain the general case.
The idea is to use Fano's Lemma with the same family $\{\varphi_{m,j}\}_{0\le j\le m-1}$ as in the density case (Theorem 3.5.1). The only thing to care about is the minimum distance between two points in this family. Using Lemma 3.2.2, we have the following identities on the functionals induced by the family of densities $\{\varphi_{m,j}\}_{0\le j\le m-1}$:

$$\Gamma_{m,j}(y) \triangleq \int \cos(\ell x) g_{m,j}(x)\Phi_1(x-y)dx = \frac{e^{-\ell/2}}{2}\left(e^{i\ell y}\varphi_{m,j}(y+i\ell) + e^{-i\ell y}\varphi_{m,j}(y-i\ell)\right).$$

Now, the following result gives the minimum distance between two parameters to be estimated in our family.

Lemma 3.5.10. *There exists a positive constant C such that for all $j,k \in [0; m-1]$, we have:*

$$\|\Gamma_{m,j} - \Gamma_{m,k}\|_\infty \ge C\frac{\theta_m e^{2m\ell}}{m^2}.$$

We now give the proof of the lemma. By definition, we have:

$$\begin{aligned}\|\Gamma_{m,j} - \Gamma_{m,k}\|_\infty = \frac{e^{-\ell/2}}{2}\sup_{y\in\mathbb{R}}\Big|&e^{i\ell y} f_0(y+i\ell)(V_{m,j}(y+i\ell) - V_{m,k}(y+i\ell) - \bar V_{m,j} - \bar V_{m,k})\\ &+e^{-i\ell y} f_0(y-i\ell)(V_{m,j}(y-i\ell) - V_{m,k}(y-i\ell) - \bar V_{m,j} - \bar V_{m,k})\Big|\end{aligned}$$

And using the triangular Inequality, we get:

$$\begin{aligned}\|\Gamma_{m,j} - \Gamma_{m,k}\|_\infty \ge \frac{e^{-\ell/2}}{2}\sup_{y\in\mathbb{R}}\Big(\Big|&e^{i\ell y} f_0(y+i\ell)(V_{m,j}(y+i\ell) - V_{m,k}(y+i\ell))\\ &+e^{-i\ell y} f_0(y-i\ell)(V_{m,j}(y-i\ell) - V_{m,k}(y-i\ell))\Big|\\ &- |\bar V_{m,j} - \bar V_{m,k}||f_0(y+i\ell) + f_0(y-i\ell)|\Big).\end{aligned}$$

Now, using (3.53) we get that:

$$\begin{aligned}&|\bar V_{m,j} - \bar V_{m,k}||f_0(y+i\ell) + f_0(y-i\ell)|\\ &\qquad \le \frac{2\theta_m}{m}\|f_0\|_\infty\|V\|_1\sup_{y\in\mathbb{R}}|f_0(y+i\ell) + f_0(y-i\ell)| = O\left(\frac{\theta_m}{m}\right).\end{aligned}$$

So that finally:

$$\begin{aligned}\|\Gamma_{m,j} - \Gamma_{m,k}\|_\infty \ge \frac{e^{-\ell/2}}{2}\sup_{y\in\mathbb{R}}\Big|&e^{i\ell y} f_0(y+i\ell)(V_{m,j}(y+i\ell) - V_{m,k}(y+i\ell))\\ &+e^{-i\ell y} f_0(y-i\ell)(V_{m,j}(y-i\ell) - V_{m,k}(y-i\ell))\Big| + O\left(\frac{\theta_m}{m}\right).\end{aligned}$$

We now come to the study of the main term. Straightforward calculations provide that

$$V_{m,j}(y+i\ell) = \frac{-\theta_m e^{2m\ell} e^{-2imy}(e^{2ij} + r_{m,j}(y))}{2\pi(my + mi\ell - j)^2},$$

with $|r_{m,j}(y)| \leq 3e^{-m\ell}$, and also that

$$V_{m,j}(y - i\ell) = \frac{-\theta_m e^{2m\ell} e^{2imy}(e^{-2ij} + \tilde{r}_{m,j}(y))}{2\pi(my - mi\ell - j)^2},$$

with $|\tilde{r}_{m,j}(y)| \leq 3e^{-m\ell}$. It follows that

$$\begin{aligned}\|\Gamma_{m,j} - \Gamma_{m,k}\|_\infty \geq \frac{e^{-\ell/2} e^{2m\ell}\theta_m}{4\pi m^2} \\ \sup_{y\in\mathbb{R}} \Bigg| e^{i(\ell-2m)y} f_0(y+i\ell) \left(\frac{e^{2ij} + r_{m,j}(y)}{(y + i\ell - j/m)^2} - \frac{e^{2ik} + r_{m,k}(y)}{(y + i\ell - k/m)^2} \right) \\ + e^{-i(\ell-2m)y} f_0(y - i\ell) \left(\frac{e^{-2ij} + \tilde{r}_{m,j}(y)}{(y - i\ell - j/m)^2} - \frac{e^{-2ik} + \tilde{r}_{m,k}(y)}{(y - i\ell - k/m)^2} \right) \Bigg| + O\left(\frac{\theta_m}{m}\right).\end{aligned}$$

Now consider the particular point $y = j/m$ and lower bound this quantity in the following way:

$$\begin{aligned}\|\Gamma_{m,j} - \Gamma_{m,k}\|_\infty \geq \frac{e^{-\ell/2} e^{2m\ell}\theta_m}{4\pi m^2} \inf_{j\neq k} \Bigg| e^{i(\ell/m-2)j} f_0(j/m + i\ell) \left(\frac{e^{2ij}}{\ell^2} + \frac{e^{2ik}}{(i\ell + (j-k)/m)^2} \right) \\ + e^{-i(\ell/m-2)j} f_0(j/m - i\ell) \left(\frac{e^{-2ij}}{\ell^2} + \frac{e^{-2ik}}{(-i\ell + (j-k)/m)^2} \right) + O(e^{-m\ell}) \Bigg| + O\left(\frac{\theta_m}{m}\right).\end{aligned}$$

This leads to:

$$\|\Gamma_{m,j} - \Gamma_{m,k}\|_\infty \geq \frac{e^{-\ell/2} f_0(i\ell) e^{2m\ell}\theta_m}{4\pi m^2} \left[\inf_{j\neq k} \left| \frac{1 - e^{2i(k-j)}}{\ell^2} + \frac{1 - e^{-2i(k-j)}}{\ell^2} \right| + o(1) \right].$$

And finally,

$$\|\Gamma_{m,j} - \Gamma_{m,k}\|_\infty \geq \frac{e^{-\ell/2} f_0(i\ell) e^{2m\ell}\theta_m}{2\pi m^2} \left[\inf_{j\neq k} \left| \frac{1 - \cos(2(k-j))}{\ell^2} \right| + o(1) \right].$$

Consequently we get that there exists a positive constant C such that

$$\|\Gamma_{m,j} - \Gamma_{m,k}\|_\infty \geq C \frac{e^{2m\ell}\theta_m}{m^2}.$$

Then we conclude using Fano's Lemma, in the same way as we did in the case of the density. More precisely, there exists a positive constant C such that:

$$\inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \|\Gamma_f - T_n\|_\infty \geq C \frac{e^{2m\ell}\theta^\star_m(n)}{m^2},$$

where $\theta^\star_m(n)$ is the supremum over all the parameters θ satisfying the following conditions:

$$\|f_0\|_\infty \|V\|_2 \frac{n\theta_m^2}{m\log(m+1)} \leq \alpha \quad \text{and} \quad m\theta_m e^{2m^2+4m} \underset{m\to\infty}{\longrightarrow} 0.$$

Choose the parameters

$$\theta_m = \frac{(\log n)^{1/4}}{\sqrt{n}} \quad \text{and} \quad m = \frac{1}{2}\sqrt{\log n} - C, \text{ where } C > 1$$

in order to get that

$$m\theta_m e^{2m^2+4m} = e^{C^2-4C}\Big(\frac{1}{2}\sqrt{\log n} - C\Big)(\log n)^{1/4} e^{-2(C-1)\sqrt{\log n}} \underset{n\to\infty}{\longrightarrow} 0,$$

and there exists a positive constant κ such that

$$\frac{n\theta_m^2}{m\log(m+1)} \leq \frac{\kappa}{\log\log n} \underset{n\to\infty}{\longrightarrow} 0.$$

Finally there exists a positive constant C such that

$$\inf_{T_n} \sup_{\Gamma_f \in \mathcal{G}_f} \|\Gamma - T_n\|_\infty \geq C \frac{e^{\ell\sqrt{\log n}}}{\sqrt{n}(\log n)^{3/4}},$$

which completes the proof.

□

Annexe A

Annexe à l'Introduction

Preuve. Preuve de la Proposition I.1.
La variance σ^2 du bruit est un paramtre identifiable si pour toutes chanes de Markov homognes $\{X_n\}_{n\geq 0}$, $\{X'_n\}_{n\geq 0}$ de transitions respectives q et q' vrifiant les Hypothses I.1 et I.2, et pour toutes suites de bruits blancs $\{\varepsilon_n\}_{n\geq 0}$ et $\{\varepsilon'_n\}_{n\geq 0}$ de lois normales centres et de variances respectives σ^2 et $(\sigma')^2$, on a l'implication suivante :

$$\{X_n+\varepsilon_n\}_{n\geq 0} \overset{\mathcal{L}\text{oi}}{=} \{X'_n+\varepsilon'_n\}_{n\geq 0} \qquad \Longrightarrow \qquad \sigma=\sigma' \text{ et } \{X_n\}_{n\geq 0} \overset{\mathcal{L}\text{oi}}{=} \{X'_n\}_{n\geq 0}. \tag{A.1}$$

Pour montrer cela, nous allons prouver la proprit suivante :

Lemme A.1. *Si $\{X_n\}_{n\geq 0}$ est une chane de Markov homogne dont la transition q vrifie les Hypothses I.1 et I.2 et $\{\varepsilon_n\}_{n\geq 0}$ est un bruit blanc de loi normale centre, de variance $\sigma^2>0$ et indpendant de $\{X_n\}_{n\geq 0}$, alors la suite $\{Y_n\}_{n\geq 0}$ dfinie par $Y_n=X_n+\varepsilon_n$ pour tout entier n positif, n'est pas une chane de Markov vrifiant l'Hypothse I.1 (i.e n'est pas une vraie chane de Markov).*

Ce lemme permet de conclure de la faon suivante : si la premire galit dans l'implication (A.1) est vraie, alors supposons par exemple $\sigma'\geq\sigma$, ce qui permet d'crire que

$$\{X_n\}_{n\geq 0} \overset{\mathcal{L}\text{oi}}{=} \{X'_n+\eta_n\}_{n\geq 0},$$

o $\{\eta_n\}_{n\geq 0}$ est un bruit blanc de loi normale centre et de variance $(\sigma')^2-\sigma^2$. Mais alors si $\sigma'>\sigma$, la suite de variables alatoires dfinie par $Y_n=X'_n+\eta_n$ pour tout n positif, vrifie les hypothses du Lemme A.1 donc n'est pas une vraie chane de Markov. Ceci est en contradiction avec l'hypothse faite sur la loi la chane $\{X_n\}_{n\geq 0}$, nous en concluons donc que $\sigma=\sigma'$, et donc que $\{X_n\}_{n\geq 0} =^{\mathcal{L}\text{oi}} \{X'_n\}_{n\geq 0}$.

Prouvons prsent le Lemme A.1. Nous noterons ν une mesure dominante pour la loi de la chaîne (ν peut être la mesure de Lebesgue ou bien une mesure à support dénombrable ou fini) sur l'espace des états $\mathcal{X}$ (fini ou non). Notons que les variables aléatoires Y_n ont une loi à densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$. Dans la suite, la lettre f désignera de faon générique une densité de loi, soit par rapport à la mesure de Lebesgue, soit par rapport à la mesure ν, ou encore par rapport à un produit tensoriel de ces deux mesures. Supposons que $\{Y_n\}_{n\geq 0}$ est une vraie chane de Markov, ce qui équivaut à :

$$\forall y_1,y_2,y_3\in\mathbb{R}\ ,\ f(y_3|Y_1=y_1;Y_2=y_2)=f(y_3|Y_2=y_2),$$

ce qui quivaut encore :

$$\begin{aligned}\forall y_1,y_2,y_3\in\mathbb{R}\ ,\ \int_{\mathcal{X}} f(y_3|X_3=x_3)\ f(x_3|Y_1=y_1;Y_2=y_2)d\nu(x_3)\\ =\int_{\mathcal{X}} f(y_3|X_3=x_3)\ f(x_3|Y_2=y_2)d\nu(x_3).\end{aligned}$$

Or, nous avons l'galit $f(y_3|X_3=x_3)=\Phi_\sigma(y_3-x_3)$, et la famille $\{\Phi_\sigma(y-x_3)\}_{y\in\mathbb{R}}$ forme un systme total dans $\mathbb{L}_2(\mathbb{R})$. L'identit prcdente quivaut donc encore

$$\forall y_1,y_2\in\mathbb{R}\ ,\forall x_3\in\mathcal{X},\ f(x_3|Y_1=y_1;Y_2=y_2)=f(x_3|Y_2=y_2),$$

ce qui s'crit encore

$$\forall y_1, y_2 \in \mathbb{R}\ , \forall x_3 \in \mathcal{X},\ \int_{\mathcal{X}} q(x_2; x_3) f(x_2 | Y_1 = y_1; Y_2 = y_2) d\nu(x_2)$$
$$= \int_{\mathcal{X}} q(x_2; x_3) f(x_2 | Y_2 = y_2) d\nu(x_2).$$

On utilise alors l'Hypothse I.2 sur le noyau de transition q pour crire que ceci quivaut :

$$\forall y_1, y_2 \in \mathbb{R}\ ,\ \forall x_2 \in \mathcal{X},\ f(x_2 | Y_1 = y_1; Y_2 = y_2) = f(x_2 | Y_2 = y_2).$$

Ainsi, nous obtenons les quivalences suivantes :

$$\begin{aligned}
&\{Y_n\}_{n \geq 0} \text{ est une vraie chane de Markov} \\
&\iff \forall y_1, y_2 \in \mathbb{R}\ , \forall x_2 \in \mathcal{X},\ f(x_2 | Y_1 = y_1; Y_2 = y_2) = f(x_2 | Y_2 = y_2) \\
&\iff \forall y_1, y_2 \in \mathbb{R}\ , \forall x_2 \in \mathcal{X},\ f(x_2; y_2)\ f(y_1; y_2) = f(x_2; y_1; y_2)\ f(y_2) \\
&\iff \forall y_1, y_2 \in \mathbb{R}\ , \forall x_2 \in \mathcal{X},\ f(y_1 | Y_2 = y_2) = \frac{f(x_2; y_1; y_2)}{f(x_2; y_2)}.
\end{aligned}$$

En notant μ_1 la densité de la loi initiale de la chane de Markov par rapport à ν (c'est dire la loi de X_1), on obtient l'galit suivante

$$\frac{f(x_2; y_1; y_2)}{f(x_2; y_2)} = \frac{\int \mu_1(x_1) q(x_1; x_2) f(y_1 | X_1 = x_1) f(y_2 | X_2 = x_2) d\nu(x_1)}{f(y_2 | X_2 = x_2) f(x_2)}$$
$$= \frac{\int \mu_1(x_1) q(x_1; x_2) f(y_1 | X_1 = x_1) d\nu(x_1)}{f(x_2)},$$

ce qui donne finalement

$$\begin{aligned}
&\{Y_n\}_{n \geq 0} \text{ est une vraie chane de Markov} \\
&\iff \forall y_1, y_2 \in \mathbb{R}\ , \forall x_2 \in \mathcal{X},\ f(y_1 | Y_2 = y_2) = \frac{\int \mu_1(x_1) q(x_1; x_2) f(y_1 | X_1 = x_1) d\nu(x_1)}{f(x_2)}.
\end{aligned}$$

Il suffit alors de remarquer que le membre de droite de cette galit ne dpend pas de y_2, ce qui signifie que la loi de Y_1 est indpendante de la loi de Y_2. Comme la chane de Markov $\{Y_n\}_{n \geq 0}$ est homogne, il s'ensuit que toutes les variables sont indpendantes, d'o une contradiction avec l'hypothse de vraie chane de Markov.
Ceci achve la preuve du Lemme A.1 ainsi que la preuve de la Proposition I.1. □

Annexe B

Technical proofs for the Hidden Markov model

B.1 Exponential forgetting for the prediction filter

Proof. (Proposition 1.2.1) Only the main steps of the proof will be described here. It is adapted from Mevel (1997) and uses results from Seneta (1981). The original idea comes from a work of Arapostathis and Marcus (1990). Proofs which are not easily deduced from the finite state space case are entirely described here.

1st step: Let p be a transition density function on $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \gamma)$. Define

$$\tau_1(p) \triangleq \frac{1}{2} \operatorname*{esssup}_{x,x'} \int_v |p(x,v) - p(x',v)| d\gamma(v)$$

and, for all u, v in $\mathbf{K}$, for all f measurable and non-negative, denote, for $n \geq m$

$$\begin{aligned}
A^\theta_{n,m}(u,v) &\triangleq \int_{u_n} q_\theta(u_n, v) g_\theta(Y_n|u_n) \dots \int_{u_{m+1}} q_\theta(u_{m+1}, u_{m+2}) g_\theta(Y_{m+1}|u_{m+1}) \\
&\quad \times q_\theta(u, u_{m+1}) g_\theta(Y_m|u) d\gamma(u_{m+1}) \dots d\gamma(u_n) \\
N(A^\theta_{n,m})(u) &\triangleq \int_v A^\theta_{n,m}(u,v) d\gamma(v) \\
N(A^\theta_{n,m} f) &\triangleq \int_u N(A^\theta_{n,m})(u) f(u) d\gamma(u) = \int_{u,v} A^\theta_{n,m}(u,v) f(u) d\gamma(u) d\gamma(v)
\end{aligned}$$

Lemma B.1.1.

$$\begin{aligned}
&\|\Phi_{n-m+1}(Y_n^m, f; \theta) - \Phi_{n-m+1}(Y_n^m, f'; \theta)\|_1 \\
&\qquad \leq \tau_1\left(\frac{A^\theta_{n,m}}{N(A^\theta_{n,m})}\right) \times \left\| \frac{N(A^\theta_{n,m})}{N(A^\theta_{n,m} f)} f - \frac{N(A^\theta_{n,m})}{N(A^\theta_{n,m} f')} f' \right\|_1
\end{aligned}$$

Proof. With the previous definitions, we may write

$$\begin{aligned}
&\|\Phi_{n-m+1}(Y_n^m, f; \theta) - \Phi_{n-m+1}(Y_n^m, f'; \theta)\|_1 \\
&\qquad = \int_v \left| \int_u \frac{A^\theta_{n,m}(u,v)}{N(A^\theta_{n,m})(u)} \left[\frac{N(A^\theta_{n,m})(u)}{N(A^\theta_{n,m} f)} f(u) - \frac{N(A^\theta_{n,m})(u)}{N(A^\theta_{n,m} f')} f'(u) \right] d\gamma(u) \right| d\gamma(v)
\end{aligned}$$

We then use the following general result: let Ψ be a measurable function such that $\int_u \Psi(u) d\gamma(u) = 0$, then we have $\int \Psi^+ = \int \Psi^- = 1/2 \int |\Psi|$, where $\Psi^+ = \max(\Psi, 0)$, $\Psi^- = \max(-\Psi, 0)$. We may also write

$$\Psi(x) = \int a(s,x) d\gamma(s) - \int a(x,t) d\gamma(t) \;\text{ where }\; a(s,t) = \frac{\Psi^-(s)\Psi^+(t)}{\int \Psi^+}$$

Note that $\int \int a(s,t) d\gamma(s) d\gamma(t) = 1/2 \int |\Psi(x)| d\gamma(x)$. Applying this result to the function

$$\Psi(u) = \frac{N(A^\theta_{n,m})(u)}{N(A^\theta_{n,m} f)} f(u) - \frac{N(A^\theta_{n,m})(u)}{N(A^\theta_{n,m} f')} f'(u)$$

(with mean zero), we get

$$\begin{aligned}
& \int_v \left| \int_u \frac{A^\theta_{n,m}(u,v)}{N(A^\theta_{n,m})(u)} \Psi(u) d\gamma(u) \right| d\gamma(v) \\
= & \int_v \left| \int_u \frac{A^\theta_{n,m}(u,v)}{N(A^\theta_{n,m})(u)} \left(\int_s a(s,u) d\gamma(s) - \int_t a(u,t) d\gamma(t) \right) d\gamma(u) \right| d\gamma(v) \\
= & \int_v \left| \int_{x,y} \frac{A^\theta_{n,m}(x,v) a(y,x)}{N(A^\theta_{n,m})(x)} d\gamma(x) d\gamma(y) - \int_{x,y} \frac{A^\theta_{n,m}(y,v) a(y,x)}{N(A^\theta_{n,m})(y)} d\gamma(x) d\gamma(y) \right| d\gamma(v) \\
\leq & \int_{v,x,y} \left| \frac{A^\theta_{n,m}(x,v)}{N(A^\theta_{n,m})(x)} - \frac{A^\theta_{n,m}(y,v)}{N(A^\theta_{n,m})(y)} \right| a(y,x) d\gamma(x) d\gamma(y) d\gamma(v) \\
\leq & \frac{1}{2} \left(\underset{x,y}{\text{esssup}} \int_v \left| \frac{A^\theta_{n,m}(x,v)}{N(A^\theta_{n,m})(x)} - \frac{A^\theta_{n,m}(y,v)}{N(A^\theta_{n,m})(y)} \right| d\gamma(v) \right) \\
& \times \int_u \left| \frac{N(A^\theta_{n,m})(u)}{N(A^\theta_{n,m} f)} f(u) - \frac{N(A^\theta_{n,m})(u)}{N(A^\theta_{n,m} f')} f'(u) \right| d\gamma(u) \\
= & \tau_1 \left(\frac{A^\theta_{n,m}}{N(A^\theta_{n,m})} \right) \left\| \frac{N(A^\theta_{n,m})}{N(A^\theta_{n,m} f)} f - \frac{N(A^\theta_{n,m})}{N(A^\theta_{n,m} f')} f' \right\|_1
\end{aligned}$$

which concludes the proof of Lemma B.1.1. □

2^{nd} step: it consists in bounding from above the last term of the right hand side of the previous inequality.

Lemma B.1.2.

$$\left\| \frac{N(A^\theta_{n,m})}{N(A^\theta_{n,m} f)} f - \frac{N(A^\theta_{n,m})}{N(A^\theta_{n,m} f')} f' \right\|_1 \leq 2\omega^\theta \|f - f'\|_1 \quad \text{where} \quad \omega^\theta = \sup \frac{N(A^\theta_{n,m} f)}{N(A^\theta_{n,m} g)},$$

The supremum is taken over the set of measurable and non-negative functions f and g such that $\int f d\gamma = \int g d\gamma = 1$.

3^{rd} step: we prove that

Lemma B.1.3. $\omega^\theta \leq {\epsilon_\theta}^{-1} \delta_\theta(Y_m)$.

The proofs of Lemmas B.1.2 and B.1.3 are straightforward adaptations from Mevel (1997, Propositions A.1 and A.2), and are omitted here for brevity. Then we conclude

$$\|\Phi_{n-m+1}(Y^m_n, f; \theta) - \Phi_{n-m+1}(Y^m_n, f'; \theta)\|_1 \leq 2{\epsilon_\theta}^{-1} \delta_\theta(Y_m) \times \tau_1 \left(\frac{A^\theta_{n,m}}{N(A^\theta_{n,m})} \right) \times \|f - f'\|_1$$

Let us examine now the behavior of $\tau_1(A^\theta_{n,m}/N(A^\theta_{n,m}))$. We want to proceed by analogy between $A^\theta_{n,m}$ and a product of random matrices (which appears in finite state space models): this technique will bring out the term $(1-\epsilon_\theta)^{n-m+1}$. We obtain an upper bound for τ_1 by using another coefficient, denoted τ_B (Birkhoff's coefficient), which does not depend on the normalization $N(A^\theta_{n,m})$.

4^{th} step: All these results are generalizations of those obtained by Seneta (1981, Theorems 3.12 and 3.13, p.106-110) and the proofs will be omitted. Let us define Birkhoff's coefficient by

$$\tau_B(p) = \sup \frac{d(p^*f, p^*g)}{d(f,g)} \quad \text{where} \quad \begin{cases} d(f,g) = \text{esssup}_{x,x'} \log \left[\frac{f(x)g(x')}{g(x)f(x')}\right] \\ p^*f(v) = \int_u p^*(v,u)f(u)d\gamma(u) = \int_u p(u,v)f(u)d\gamma(u) \end{cases}$$

First supremum is taken over the set of measurable and non-negative functions f and g such that $f \neq \lambda g$ ($\lambda \in \mathbb{R}$), and p^* is the transpose density function associated with p (i.e $p^*(x,y) = p(y,x)$). With this definition, τ_B does not depend on the normalization, and is, moreover, sub-multiplicative. We now give another expression of τ_B, and then the upper bound of τ_1 by τ_B.

Lemma B.1.4.

$$\tau_B(p) = \frac{1-\alpha(p)^{\frac{1}{2}}}{1+\alpha(p)^{\frac{1}{2}}} \geq \tau_1(p) \quad \text{where } \alpha(p) = \operatorname*{essinf}_{x,y,z,t} \frac{p(x,z)p(y,t)}{p(x,t)p(y,z)} = \alpha(p^*)$$

We now apply these results with $p = A^\theta_{n,m}/(N(A^\theta_{n,m}))$, we have

$$\tau_1\left(\frac{A^\theta_{n,m}}{N(A^\theta_{n,m})}\right) \leq \tau_B\left(\frac{A^\theta_{n,m}}{N(A^\theta_{n,m})}\right)$$

It follows from the definition that the quantity $\alpha(A^\theta_{n,m}/N(A^\theta_{n,m}))$ equals $\alpha(A^\theta_{n,m})$, and then $\tau_B(A^\theta_{n,m}/N(A^\theta_{n,m})) = \tau_B(A^\theta_{n,m})$ by application of Lemma B.1.4. Since τ_B is sub-multiplicative,

$$\tau_B(A^\theta_{n,m}) \leq \Pi^n_{k=m}\tau_B[q_\theta(.,.)g_\theta(Y_k|.)]$$

Finally, using again Lemma B.1.4, $\tau_B[q_\theta(.,.)g_\theta(Y_k|.)] = \tau_B(q_\theta)$, and under assumption 1.1

$$\alpha(q_\theta) = \operatorname*{essinf}_{x,y,z,t} \frac{q_\theta(x,z)q_\theta(y,t)}{q_\theta(x,t)q_\theta(y,z)} \geq {\epsilon_\theta}^2.$$

Thus $\tau_B(q_\theta) \leq 1-\epsilon_\theta$, which finally implies that $\tau_B(A^\theta_{n,m}) \leq (1-\epsilon_\theta)^{n-m+1}$.

□

B.2 Geometric ergodicity of the extended chain

Proof. (Proposition 1.2.2) Let h in Lip($\mathbf{E}$), and $z = (x,y,f,f^*), z' = (x',y',f',(f^*)')$ in $\mathbf{E}$. In order to get an upper bound for $|\Pi^n_\theta h(z) - \Pi^n_\theta h(z')|$, use the decomposition

$$\Pi^n_\theta h(z) - \Pi^n_\theta h(z') = \left[\Pi^n_\theta h(x,y,f,f^*) - \Pi^n_\theta h(x,y',f',(f^*)')\right] + \left[\Pi^n_\theta h(x,y',f',(f^*)') - \Pi^n_\theta h(x',y',f',(f^*)')\right]$$

Denote A (resp. B) the first (resp. second) term of the right hand side in the previous equation. A and B will be treated separately.

1st term: denote for $i \leq j$, $y_j^i = (y_j, \dots y_i)$ and with an abuse of notation $d\gamma(x_j^i) \triangleq d\gamma(x_i)\dots d\gamma(x_j)$, and $d\nu(y_j^i) \triangleq d\nu(y_i)\dots d\gamma(y_j)$. Then,

$$A = \int_{x_1 \cdots y_n} \Big[h(x_n, y_n, \Phi_n(y_{n-1}^1, y, f; \theta), \Phi_n(y_{n-1}^1, y, f^*; \theta^*)) \\ - h(x_n, y_n, \Phi_n(y_{n-1}^1, y', f'; \theta), \Phi_n(y_{n-1}^1, y', (f^*)'; \theta^*)) \Big] \\ \times q_{\theta^*}(x, x_1) \cdots q_{\theta^*}(x_{n-1}, x_n) g_{\theta^*}(y_1|x_1) \cdots g_{\theta^*}(y_n|x_n) d\gamma(x_n^1) d\nu(y_n^1)$$

Then, using Definition 1 of Lip($\mathbf{E}$), and the fact that for all θ in Θ,
$\Phi_n(y_{n-1}^1, y, f; \theta) = \Phi_{n-1}(y_{n-1}^1, \Phi_1(y, f; \theta); \theta)$, we have

$$|A| \leq \int_{x_1 \cdots y_n} \text{lip}(h, x_n, y_n) \Big[\|\Phi_{n-1}(y_{n-1}^1, \Phi_1(y, f; \theta); \theta) - \Phi_{n-1}(y_{n-1}^1, \Phi_1(y', f'; \theta); \theta)\|_1 \\ + \|\Phi_{n-1}(y_{n-1}^1, \Phi_1(y, f^*; \theta^*); \theta^*) - \Phi_{n-1}(y_{n-1}^1, \Phi_1(y', (f^*)'; \theta^*); \theta^*)\|_1 \Big] \\ \times q_{\theta^*}(x, x_1) \cdots q_{\theta^*}(x_{n-1}, x_n) g_{\theta^*}(y_1|x_1) \cdots g_{\theta^*}(y_n|x_n) d\gamma(x_n^1) d\nu(y_n^1)$$

Exponential forgetting of the prediction filter yields

$$\begin{aligned} & \|\Phi_{n-1}(y_{n-1}^1, \Phi_1(y, f; \theta); \theta) - \Phi_{n-1}(y_{n-1}^1, \Phi_1(y', f'; \theta); \theta)\|_1 \\ \leq \;& 2{\epsilon_\theta}^{-1} \delta_\theta(y_1)(1-\epsilon_\theta)^{n-1} \|\Phi_1(y, f; \theta) - \Phi_1(y', f'; \theta)\|_1 \\ \leq \;& 4{\epsilon_\theta}^{-1} \delta_\theta(y_1)(1-\epsilon_\theta)^{n-1} \end{aligned}$$

The same inequality holds at the point θ^*. Integrating the above expression and using assuption 1.1 yields

$$\begin{aligned} |A| \;&\leq\; 4\text{lip}(h) \Big[{\epsilon_\theta}^{-1} \Delta_1 (1-\epsilon_\theta)^{n-1} + {\epsilon_{\theta^*}}^{-1} \Delta_1 (1-\epsilon_{\theta^*})^{n-1} \Big] \\ &\leq\; 8\text{lip}(h) \epsilon^{-1} \Delta_1 (1-\epsilon)^{n-1} \end{aligned} \tag{B.1}$$

Remark B.1. We may prove along the same line that if h is in Lip($\mathbf{E}$) then for any integer n, $\Pi_\theta^n h$ belongs to Lip($\mathbf{E}$) too, with $\text{lip}(\Pi_\theta^n h, x, y) \leq \text{lip}(h)\delta_\theta(y)$ and $\text{k}(\Pi_\theta^n h, x, y) \leq \text{k}(h)$.

2nd term:

$$B = \int_{x_1 \cdots y_n} h(x_n, y_n, \Phi_n(y_{n-1}^1, y', f'; \theta), \Phi_n(y_{n-1}^1, y', (f^*)'; \theta^*)) \\ \times [q_{\theta^*}(x, x_1) - q_{\theta^*}(x', x_1)] \, q_{\theta^*}(x_1, x_2) \cdots q_{\theta^*}(x_{n-1}, x_n) \\ \times g_{\theta^*}(y_1|x_1) \cdots g_{\theta^*}(y_n|x_n) d\gamma(x_n^1) d\nu(y_n^1)$$

Let $m = [n/2]$ and F, F^* be fixed functions in $\mathbf{S}^+$. Note that

$$h(x_n, y_n, \Phi_n(y_{n-1}^1, y', f'; \theta), \Phi_n(y_{n-1}^1, y', (f^*)'; \theta^*)) \\ = h(x_n, y_n, \Phi_{n-m}(y_{n-1}^m, \Phi_m(y_{m-1}^1, y', f'; \theta); \theta), \Phi_{n-m}(y_{n-1}^m, \Phi_m(y_{m-1}^1, y', (f^*)'; \theta^*); \theta^*)) \\ - h(x_n, y_n, \Phi_{n-m}(y_{n-1}^m, F; \theta), \Phi_{n-m}(y_{n-1}^m, F^*; \theta^*)) \\ + h(x_n, y_n, \Phi_{n-m}(y_{n-1}^m, F; \theta), \Phi_{n-m}(y_{n-1}^m, F^*; \theta^*))$$

Insert this form into B and use that h is in Lip($\mathbf{E}$) for the first part and take the integral with respect to y_{m-1}^1 and x_{m-1}^1 in the second; it yields

$$
\begin{aligned}
|B| \leq \int_{x_1 \cdots y_n} & \operatorname{lip}(h, x_n, y_n) \times \big(\|\Phi_{n-m}(y_{n-1}^m, \Phi_m(y_{m-1}^1, y', f'; \theta); \theta) - \Phi_{n-m}(y_{n-1}^m, F; \theta)\|_1 \\
& + \|\Phi_{n-m}(y_{n-1}^m, \Phi_m(y_{m-1}^1, y', (f^*)'; \theta^*); \theta^*) - \Phi_{n-m}(y_{n-1}^m, F^*; \theta^*)\|_1\big) \\
& \times |q_{\theta^*}(x, x_1) - q_{\theta^*}(x', x_1)|\, q_{\theta^*}(x_1, x_2) \cdots q_{\theta^*}(x_{n-1}, x_n) \\
& \times g_{\theta^*}(y_1|x_1) \cdots g_{\theta^*}(y_n|x_n) d\gamma(x_n^1) d\nu(y_n^1) \\
& + \int_{x_m \cdots x_n} \int_{y_m \cdots y_n} |h(x_n, y_n, \Phi_n(y_{n-1}^m, F; \theta), \Phi_n(y_{n-1}^m, F^*; \theta^*))| \\
& \times |q_{\theta^*}^m(x, x_m) - q_{\theta^*}^m(x', x_m)| q_{\theta^*}(x_m, x_{m+1}) \cdots q_{\theta^*}(x_{n-1}, x_n) \\
& \times g_{\theta^*}(y_m|x_m) \cdots g_{\theta^*}(y_n|x_n) d\gamma(x_n^m) d\nu(y_n^m)
\end{aligned}
$$

Using Proposition 1.2.1 and inequality (1.9) yields

$$
\begin{aligned}
|B| \leq \int_{x_1 \cdots y_n} & \operatorname{lip}(h, x_n, y_n) \times 2\epsilon^{-1}(1-\epsilon)^{n-m} \\
& \times \left[\|\Phi_m(y_{m-1}^1, y', f'; \theta) - F\|_1 \delta_\theta(y_m) + \|\Phi_m(y_{m-1}^1, y', (f^*)'; \theta^*) - F^*\|_1 \delta_{\theta^*}(y_m)\right] \\
& \times [q_{\theta^*}(x, x_1) + q_{\theta^*}(x', x_1)]\, q_{\theta^*}(x_1, x_2) \cdots q_{\theta^*}(x_{n-1}, x_n) \\
& \times g_{\theta^*}(y_1|x_1) \cdots g_{\theta^*}(y_n|x_n) d\gamma(x_n^1) d\nu(y_n^1) \\
& + \int_{x_m \cdots x_n} \int_{y_m \cdots y_n} \mathrm{k}(h, x_n, y_n) \rho_0^m q_{\theta^*}(x_m, x_{m+1}) \cdots q_{\theta^*}(x_{n-1}, x_n) \\
& \qquad \times g_{\theta^*}(y_m|x_m) \cdots g_{\theta^*}(y_n|x_n) d\gamma(x_n^m) d\nu(y_n^m)
\end{aligned}
$$

$$
\text{so that} \qquad |B| \leq 16 \times \operatorname{lip}(h) \epsilon^{-1} \Delta_1 (1-\epsilon)^{n-m} + \mathrm{k}(h) \rho_0^m \tag{B.2}
$$

Inequalities (B.1) and (B.2) complete the proof. □

Proof. (Corollary 1.2.3 and 1.2.4) The existence of $\Lambda_\theta(h)$ follows from Proposition 1.2.2, using the same ideas as in Sunyach (1975, Theorem 1). Then we can easily check that the function defined on $\mathbf{E}$ by

$$
V(z) = \sum_{n \geq 0} [\Pi_\theta^n h(z) - \Lambda_\theta(h)] \tag{B.3}
$$

is a solution to the Poisson equation. As the space $\mathbf{E}$ is not locally compact, we cannot use the Riesz representation theorem to show that $\Lambda_\theta(h)$ is the expectation of h under a suitable measure. Let us show that the family of probability measures $(\Pi_\theta^n(z, .))_{n \geq 0}$ is asymptotically tight (see for definition Van der Vaart and Wellner (1996), p.20), using the same basic ideas as in Sunyach ((1975), Theorem 1).

Let $d(.,.)$ be a distance in $\mathbf{E}$. For $\mathbf{D}$ a compact set of $\mathbf{E}$ and $\delta > 0$, define $\varphi_{\delta,\mathbf{D}}(z) \triangleq \inf\left(1, \frac{1}{\delta} d(z, \mathbf{D})\right)$. It is straightforward that $\varphi_{\delta,\mathbf{D}}(z)$ is in Lip($\mathbf{E}$), with $\operatorname{lip}(\varphi_{\delta,\mathbf{D}}) \leq \delta^{-1}$ and $\mathrm{k}(\varphi_{\delta,\mathbf{D}}) \leq 1$. Moreover, for any set $\mathcal{A}$, denote $\mathcal{A}_\delta \triangleq \{z \in \mathbf{E}; d(z, \mathcal{A}) \leq \delta\}$ the δ-enlargement of $\mathcal{A}$.

Let $\epsilon, \delta > 0$ and z in $\mathbf{E}$, there exists p sufficiently large such that $C\left[\delta^{-1}+1\right]\rho^p(1-\rho)^{-1} < \epsilon/2$ (where C is the constant appearing in (1.12)). Now, choose a compact $\mathbf{D}^\epsilon$ such that $\Pi_\theta^k\left[z,(\mathbf{D}^\epsilon)^c\right] < \epsilon/2$ for $k = 1, ..., p$ which is possible since any probability measure in a separable and complete space is tight (the superscript "c"denotes the complement in the space $\mathbf{E}$). Note that we have $\mathbb{1}_{((D^\epsilon)_\delta)^c} \leq \varphi_{\delta,\mathbf{D}^\epsilon} \leq \mathbb{1}_{(D^\epsilon)^c}$ so we get, for any $n > p$,

$$
\begin{aligned}
\Pi_\theta^n\left[z,((\mathbf{D}^\epsilon)_\delta)^c\right] &\leq \Pi_\theta^n(\varphi_{\delta,\mathbf{D}^\epsilon})(z) \\
&\leq \Pi_\theta^p(\varphi_{\delta,\mathbf{D}^\epsilon})(z) + C\left[\frac{1}{\delta}+1\right]\frac{\rho^p}{1-\rho} \\
&\leq \Pi_\theta^p[z,(\mathbf{D}^\epsilon)^c] + \frac{\epsilon}{2} \\
&\leq \epsilon
\end{aligned}
\tag{B.4}
$$

So, $(\Pi_\theta^n(z,.))_{n\geq 0}$ is asymptotically tight. Hence, by Prohorov's theorem (Van der Vaart and Wellner 1996, p.21), there exists a probability measure λ_θ and a subsequence $(n_i)_{i\geq 0}$ such that $\Pi_\theta^{n_i}(z,.)$ converges weakly to λ_θ. This proves that for any continuous and bounded function h in Lip($\mathbf{E}$), $\Lambda_\theta(h) = \int h(z)d\lambda_\theta(z)$. Using the result of Corollary 1.2.3, we obtain that for any continuous and bounded function h in Lip($\mathbf{E}$),

$$
\lambda(\zeta)\Pi_\theta^n h \triangleq \int \Pi_\theta^n h(z)\lambda(\zeta)(dz) = \mathbb{E}_{\theta,\lambda(\zeta)}(h(Z_n)) \underset{n\to\infty}{\longrightarrow} \int h(z)d\lambda_\theta(z) = \Lambda_\theta(h) \tag{B.5}
$$

Then, using the fact that for any h in Lip($\mathbf{E}$), $\Pi_\theta h$ is still in Lip($\mathbf{E}$) (see Remark 1), λ_θ is the unique invariant probability measure associated with the kernel Π_θ.
Now let h be a function in Lip($\mathbf{E}$) (h is not necessarily continuous and bounded). Define

$$
G_h(x,y,f,f^*) \triangleq \int h(x_0,y_0,f,f^*)g_{\theta^*}(y_0|x_0)f^*(x_0)d\nu(y_0)d\gamma(x_0) \tag{B.6}
$$

It can be easily checked that G_h is a bounded and continuous function in Lip($\mathbf{E}$). Applying the previous result to the function G_h, we obtain

$$
\begin{aligned}
\lim_n \lambda(\zeta)\Pi_\theta^n G_h &= \int G_h(z)d\lambda_\theta(z) \\
&= \int h(x_0,y_0,f,f^*)g_{\theta^*}(y_0|x_0)f^*(x_0)d\nu(y_0)d\gamma(x_0)\lambda_\theta(\mathbf{K},\mathcal{Y},df,df^*)
\end{aligned}
\tag{B.7}
$$

Using the remark 2,

$$
\begin{aligned}
\lambda(\zeta)\Pi_\theta^n h &= \mathbb{E}_{\theta,\lambda(\zeta)}(h(Z_n)) \\
&= \mathbb{E}^*\left[\mathbb{E}^*\left(h(X_n,Y_n,f_{\theta,n}^\zeta,f_{\theta^*,n}^{\pi^*})\middle|Y_{n-1}^0\right)\right] \\
&= \mathbb{E}^*\left(G_h(X_n,Y_n,f_{\theta,n}^\zeta,f_{\theta^*,n}^{\pi^*})\right) \\
&= \mathbb{E}_{\theta,\lambda(\zeta)}(G_h(Z_n)) \\
&= \lambda(\zeta)\Pi_\theta^n G_h
\end{aligned}
\tag{B.8}
$$

But we already know by (B.5) that

$$
\lim_n \lambda(\zeta)\Pi_\theta^n h = \Lambda_\theta(h) \tag{B.9}
$$

Combining with (B.7), (B.8) and (B.9),

$$\Lambda_\theta(h) = \int h(x_0, y_0, f, f^*) g_{\theta^*}(y_0|x_0) f^*(x_0) d\nu(y_0) d\gamma(x_0) \lambda_\theta(\mathbf{K}, \mathcal{Y}, df, df^*)$$

But we have,

$$g_{\theta^*}(y_0|x_0) f^*(x_0) d\nu(y_0) d\gamma(x_0) \lambda_\theta(\mathbf{K}, \mathcal{Y}, df, df^*) = \lambda_\theta(dx_0, dy_0, df, df^*)$$

as these measures are entirely identified by their integrals on Lipschitz and bounded functions on $\mathbf{E}$. Then we obtain, for all h in $\mathrm{Lip}(\mathbf{E})$

$$\Lambda_\theta(h) = \int h(z) d\lambda_\theta(z).$$

□

B.3 Consistency of the maximum likelihood estimator

Proof. (Proposition 1.2.5)

i) Denote V, the solution of the Poisson equation associated to the function h (see Corollary 1.2.3) and $\mathcal{F}_n$ the σ-algebra generated by $\{Z_0, \ldots, Z_n\}$. It is easy to show, using inequality (1.12) for $n \geq 1$ that

$$|V(z)| \leq C\rho \frac{\mathrm{lip}(h) + \mathrm{k}(h)}{(1-\rho)^2} + |\Lambda_\theta(h)| + \mathrm{k}(h, x, y) \tag{B.10}$$

and thus there exist $s > 1$ (cf assumption 1.3) such that $\Pi_\theta |V|^s$ is bounded which implies that $\mathbb{E}_\theta(|V(Z_{n+1}) - \Pi_\theta V(Z_n)|^s \big| \mathcal{F}_n)$ is bounded independently from n. The proof follows from the classical identity (Meyn and Tweedie, 1996, 17.4.3)

$$\frac{1}{n} \sum_{m=0}^{n-1} h(Z_m) = \Lambda_\theta(h) + \frac{1}{n} \sum_{m=1}^{n} [V(Z_m) - \Pi_\theta V(Z_{m-1})] + \frac{1}{n} [\Pi_\theta V(Z_0) - \Pi_\theta V(Z_n) + h(Z_0) - h(Z_n)]$$

by applying the corollary of Chow's theorem (Hall and Heyde, (1980), p.36).

ii) This is a direct consequence of i) by applying the Remark 2.

□

Proof. (Proposition 1.2.6) It is easy to show that

$$\begin{aligned} |h_\theta(x, y, f_1, f_2) - h_\theta(x, y, f_1', f_2')| &= \left| \log \left(1 + \frac{\int_u g_\theta(y|u)(f_1 - f_1')(u) d\gamma(u)}{\int_u g_\theta(y|u) f_1'(u) d\gamma(u)} \right) \right| \\ &\leq \delta_\theta(y) \|f_1 - f_1'\|_1 \end{aligned} \tag{B.11}$$

and

$$|h_\theta(x, y, f_1, f_2)| \leq k_\theta(y) \tag{B.12}$$

which proves that h is in $\mathrm{Lip}(\mathbf{E})$ and verifies equation (1.15), using assumptions 1.2 and H3. □

Proof. (Lemma 1.2.8) Let $\eta > 0$ and $\|\theta - \theta'\| \leq \eta$. By Proposition 1.2.1,

$$\begin{aligned}\|\Phi_{n+1}(Y_n^0, \zeta; \theta) - \Phi_{n+1}(Y_n^0, \zeta; \theta')\|_1 &= \|\Phi_n(Y_n^1, \Phi_1(Y_0, \zeta; \theta); \theta) - \Phi_n(Y_n^1, \Phi_1(Y_0, \zeta; \theta'); \theta')\|_1 \\ &\leq \|\Phi_n(Y_n^1, \Phi_1(Y_0, \zeta; \theta); \theta) - \Phi_n(Y_n^1, \Phi_1(Y_0, \zeta; \theta'); \theta)\|_1 \\ &+ \|\Phi_n(Y_n^1, \Phi_1(Y_0, \zeta; \theta'); \theta) - \Phi_n(Y_n^1, \Phi_1(Y_0, \zeta; \theta'); \theta')\|_1 \\ &\leq 2\epsilon_\theta^{-1}\delta_\theta(Y_1)\rho^n \|\Phi_1(Y_0, \zeta; \theta) - \Phi_1(Y_0, \zeta; \theta')\|_1 \\ &\quad + \sup_f \|\Phi_n(Y_n^1, f; \theta) - \Phi_n(Y_n^1, f; \theta')\|_1 \end{aligned} \tag{B.13}$$

It is straightforward that

$$\begin{aligned}\|\Phi_1(Y_0, \zeta; \theta) - \Phi_1(Y_0, \zeta; \theta')\|_1 \leq & \left\| \frac{\int (q_\theta(u,.) - q_{\theta'}(u,.)) g_\theta(Y_0|u)\zeta(u)d\gamma(u)}{\int g_\theta(Y_0|u)\zeta(u)d\gamma(u)} \right\|_1 \\ & + \left\| \frac{\int q_{\theta'}(u,.) g_\theta(Y_0|u)\zeta(u)d\gamma(u)}{\int g_\theta(Y_0|u)\zeta(u)d\gamma(u)} - \frac{\int q_{\theta'}(u,.) g_{\theta'}(Y_0|u)\zeta(u)d\gamma(u)}{\int g_{\theta'}(Y_0|u)\zeta(u)d\gamma(u)} \right\|_1\end{aligned}$$

The first term of the right hand side of this expression is upper-bounded by $\omega^q(\eta)$. The second term of the right hand side is upper-bounded by $2\delta'(Y_0, \eta)$, by breaking the difference into two terms including $\frac{\int q_{\theta'}(u,.) g_\theta(Y_0|u)\zeta(u)d\gamma(u)}{\int g_{\theta'}(Y_0|u)\zeta(u)d\gamma(u)}$. Combining these two bounds with (B.13),

$$\begin{aligned}\sup_\zeta \|\Phi_{n+1}(Y_n^0, \zeta; \theta) - \Phi_{n+1}(Y_n^0, \zeta; \theta')\|_1 \leq & \, 2\epsilon^{-1}\delta_\theta(Y_1)(\omega^q(\eta) + 2\delta'(Y_0, \eta))\rho^n \|\theta - \theta'\| \\ & + \sup_\zeta \|\Phi_n(Y_n^1, \zeta; \theta) - \Phi_n(Y_n^1, \zeta; \theta')\|_1\end{aligned}$$

A straightforward recurrence completes the proof. □

B.4 Asymptotic normality of the maximum likelihood estimator

B.4.1 Exponential forgetting for the gradient of the prediction filter and consequences

Iterating equation (1.35) at order m, $m \leq n$ yields, for all v in $\mathbf{K}$,
$\nabla f_{\theta,n+1}^\zeta(v) = \Psi_{n-m+1}(Y_n^m, f_{\theta,m}^\zeta, \nabla f_{\theta,m}^\zeta; \theta)(v)$, where Ψ_n is the vector function $(\Psi_n^k)_{1 \leq k \leq p}$ defined by

$$\begin{aligned}\Psi_{n-m+1}^k(Y_n^m, f, \sigma; \theta)(v) = & \\ & \int_{u_n} a_\theta(Y_n, f_n)(u_n, v) \ldots \int_{u_m} a_\theta(Y_m, f)(u_m, u_{m+1}) \, \sigma_k(u_m) d\gamma(u_m) \ldots d\gamma(u_n) \\ & + \sum_{l=m}^{n-1} \int_{u_n} a_\theta(Y_n, f_n)(u_n, v) \ldots \\ & \ldots \int_{u_{l+1}} a_\theta(Y_{l+1}, f_{l+1})(u_{l+1}, u_{l+2}) \times U_{\theta,k}(Y_l, f_l)(u_{l+1}) d\gamma(u_{l+1}) \ldots d\gamma(u_n) \\ & + U_{\theta,k}(Y_n, f_n)(v)\end{aligned}$$

where we use the notation : for all $l \leq m+1$, $f_l \triangleq \Phi_{l-m}(Y_{l-1}^m, f; \theta)$.
In this part, it is proved the exponential forgetting for Ψ of the initial conditions. It is straightforward that for any k in $\{1, \dots, p\}, U_{\theta,k}$ verifies a Lipschitz condition :
For all y in $\mathcal{Y}$, for all f, f' in $\mathbf{S}^+$,

$$\begin{aligned} \|U_{\theta,k}(y,f) - U_{\theta,k}(y,f')\|_1 &\leq \mathrm{lip}(U_{\theta,k}, y)\|f - f'\|_1 \\ \|U_{\theta,k}(y,f)\|_1 &\leq \mathrm{k}(U_{\theta,k}, y) \end{aligned}$$

where

$$\begin{aligned} \mathrm{lip}(U_{\theta,k}, y) = \delta_\theta(y)(1+\delta_\theta(y)) &\left(\max_{1\leq k\leq p} \int_{x'} \operatorname*{esssup}_x |\partial_k q_\theta(x,x')| d\gamma(x') \right) + \delta'_\theta(y)(1+\delta_\theta(y)) \\ &+ 2\delta_\theta(y)\delta'_\theta(y)(1+\delta_\theta(y)^2) \\ \mathrm{k}(U_{\theta,k}, y) = \delta_\theta(y) &\left(\max_{1\leq k\leq p} \int_{x'} \operatorname*{esssup}_x |\partial_k q_\theta(x,x')| d\gamma(x') \right) + \delta'_\theta(y)(1+\delta_\theta(y)) \end{aligned}$$

Note that $\mathrm{lip}(U_{\theta,k}) \triangleq \mathrm{esssup}_x \int_y \mathrm{lip}(U_{\theta,k}, y) g_{\theta^*}(y|x) d\nu(y) < \infty$ as soon as Δ_2 and $(\Delta\Delta')_{3,1}$ are finite, and $\mathrm{k}(U_{\theta,k}) \triangleq \mathrm{esssup}_x \int_y \mathrm{k}(U_{\theta,k}, y) g_{\theta^*}(y|x) d\nu(()y) < \infty$ as soon as Δ_1 and $(\Delta\Delta')_{11}$ are finite.

Proposition B.4.1. *Under assumptions 1.1 and 1.7, and for all f, f' in $\mathbf{S}^+$, σ, σ' in $\mathbf{\Sigma}$, θ in Θ, and all $1 \leq k \leq p$,*

$$\begin{aligned} &\|\Psi_{n-m+1}^k(Y_n^m, f, \sigma; \theta) - \Psi_{n-m+1}^k(Y_n^m, f', \sigma'; \theta)\|_1 \\ &\quad\leq 6\delta_\theta(Y_m)^3 {\epsilon_\theta}^{-1}(1-\epsilon_\theta)^{n-m+1} \left[\|\sigma_k - \sigma'_k\|_1 + \|f - f'\|_1(1 + \|\sigma_k\|_1 + \|\sigma'_k\|_1)\right] \\ &\qquad\qquad + 10\delta_\theta(Y_m){\epsilon_\theta}^{-1}(1-\epsilon_\theta)^{n-m+1}\|f - f'\|_1 \\ &\qquad\qquad [\mathrm{lip}(U_{\theta,k}, Y_n) + \sum_{l=m}^{n-1} \mathrm{lip}(U_{\theta,k}, Y_l)\delta_\theta(Y_{l+1})^2 + \sum_{l=m}^{n-1} \delta_\theta(Y_{l+1})^3 \mathrm{k}(U_{\theta,k}, Y_l)] \end{aligned}$$

The proof is a straightforward adaptation from Mevel (1997, Part I, Proposition 4.4) and is omitted here.

Now, define a class of Lipschitz functions on **E'**

Definition B.4.1. Lip(**E'**) *is the set of vector-valued functions $h = (h_k)_{1\leq k\leq p}$ such that each h_k is a real valued measurable function on* **E'** *and for all (x, y) in $\mathbf{K} \times \mathcal{Y}$, there exists* lip'$(h_k, x, y)$ *and* $\mathrm{k}'(h_k, x, y)$ *such that for all f, f' in $\mathbf{S}^+$, for all σ, σ' in $\mathbf{\Sigma}$,*

$$\begin{aligned} &|h_k(x,y,f,\sigma) - h_k(x,y,f',\sigma')| \leq \\ &\qquad\qquad \mathrm{lip}'(h_k,x,y)\left[\|\sigma_k - \sigma'_k\|_1 + \|f - f'\|_1 \times (1 + \|\sigma_k\|_1 + \|\sigma'_k\|_1)\right] \end{aligned}$$

$|h_k(x,y,f,\sigma)| \leq \mathrm{k}'(h_k,x,y)(1 + \|\sigma_k\|_1)$

and $\begin{cases} \mathrm{lip}'(h_k) = \mathrm{esssup}_x \int \mathrm{lip}'(h_k,x,y)\, g_{\theta^*}(y|x) d\nu(y) < \infty \\ \mathrm{k}'(h_k) = \mathrm{esssup}_x \int \mathrm{k}'(h_k,x,y)\, g_{\theta^*}(y|x) d\nu(y) < \infty \end{cases}$

We prove the analog of Proposition 1.2.2.

Proposition B.4.2. *Under assumptions 1.1 and 1.7, there exists constants $C > 0$ and $\tilde{\rho} \in]\rho; 1[$ such that, for all z, z' in* **E'***, for all h in* Lip(**E'**)*, for all $n \geq 1$, and for all $1 \leq k \leq p$, we have*

$$|\tilde{\Pi}_\theta^n h_k(z) - \tilde{\Pi}_\theta^n h_k(z')| \leq C(\mathrm{k}'(h_k) + \mathrm{lip}'(h_k))\tilde{\rho}^n (1 + \|\sigma_k\|_1 + \|\sigma'_k\|_1) \\ (\delta_\theta(y)^2 + \delta_\theta(y')^2 + \mathrm{k}(U_{\theta,k}, y) + \mathrm{k}(U_{\theta,k}, y'))$$

Corollary B.4.3. *Under assumptions 1.1 and 1.7, there exists constants $C > 0$ and $\tilde{\rho} \in]\rho; 1[$, such that for all h in* Lip(**E'**)*, there exists a family of constants $(\tilde{\Lambda}_\theta(h_k))_{1 \leq k \leq p}$, such that for all z in* **E'***, for all $n \geq 1$, and for all $1 \leq k \leq p$,*

$$|\tilde{\Pi}_\theta^n h_k(z) - \tilde{\Lambda}_\theta(h_k)| \leq C(\mathrm{lip}'(h_k) + \mathrm{k}'(h_k))\tilde{\rho}^n (1 + \|\sigma_k\|_1)(\delta_\theta(y)^2 + \mathrm{k}(U_{\theta,k}, y))$$

For any initial probability measure $\tilde{\lambda}$ on $(\mathbf{E'}, \mathcal{B}(\mathbf{E'}))$*, and for any function h in* Lip(**E'**)

$$\tilde{\Lambda}_\theta(h) = \lim_{n \to \infty} \tilde{\mathbb{E}}_{\theta, \tilde{\lambda}}(h(\tilde{Z}_n)) \tag{B.14}$$

and, for all $1 \leq k \leq p$, there exists a solution in **E'** *to the Poisson equation $(I - \tilde{\Pi}_\theta)V(.) = h_k(.) - \tilde{\Lambda}_\theta(h_k)$.*

The proofs are similar to the proofs of Proposition 1.2.2 and its Corollary, and are omitted here (note that they assume that Δ_3 is finite).
Now, we can prove the convergence of the gradient of the normalized log-likelihood.

B.4.2 Proof of Proposition 1.3.1

Proof. Differentiating the expression of the log-likelihood, we get for all k in $\{1, \dots p\}$,

$$\begin{aligned} \frac{1}{n}\partial_k \ell_n(\theta, \zeta) &= \frac{1}{n}\sum_{m=0}^{n-1} \frac{\int \partial_k g_\theta(Y_m|u) f_{\theta,m}^\zeta(u) d\gamma(u) + \int g_\theta(Y_m|u)\partial_k f_{\theta,m}^\zeta(u) d\gamma(u)}{\int g_\theta(Y_m|u) f_{\theta,m}^\zeta(u) d\gamma(u)} \quad P^* - a.s \\ &= \frac{1}{n}\sum_{m=0}^{n-1} j_{\theta,k}(\tilde{Z}_m) \;\; P_{\theta,\lambda(\zeta)} - a.s \end{aligned}$$

where

$$j_{\theta,k}(x, y, f, \sigma) \triangleq \frac{\int \partial_k g_\theta(y|u) f(u) d\gamma(u) + \int g_\theta(y|u)\sigma_k(u) d\gamma(u)}{\int g_\theta(y|u) f(u) d\gamma(u)}$$

We check (with simple calculus) that j_θ is in Lip(**E'**), assuming that Δ_2 and $(\Delta\Delta')_{1,1}$ are finite. (In fact, lip'$(j_\theta, x, y) = \delta'_\theta(y)(1 + \delta_\theta(y)) + \delta_\theta(y)^2$ and k'$(j_\theta, x, y) = \delta_\theta(y) + \delta'_\theta(y)$). Assuming moreover that for some $s > 1$, Δ_s and Δ'_s are finite (in order to get a condition like (1.15)), we prove, just as in Proposition 1.2.6, that for all k in $\{1 \dots p\}$,

$$\frac{1}{n}\partial_k \ell_n(\theta, \zeta) \longrightarrow \tilde{\Lambda}_\theta(j_{\theta,k}) \quad P^* - a.s.$$

But the score function $\tilde{\Lambda}_\theta(j_\theta)$ vanishes at θ^*: Propositions 1.3.2 and 1.3.3 show that the convergence of $(1/n)\nabla^2 \ell_n(\theta, \zeta)$ happens P^* almost surely and uniformly for θ in a neighborhood of θ^*, for some sub-sequence; and then for the whole family, as it is yet P^* almost surely convergent. Thus, $(1/n)\nabla \ell_n(\theta, \zeta)$ is uniformly convergent, vanishes at $\hat{\theta}_n(\zeta)$, and $\hat{\theta}_n(\zeta)$ converges to

θ^*, which proves the result.
Let V_k^* denote the solution to the Poisson equation $(I-\tilde{\Pi}_{\theta^*})V = j_{\theta^*,k} - \tilde{\Lambda}_{\theta^*}(j_{\theta^*,k})$ (which exists by Corollary B.4.3). We have, $\tilde{P}_{\theta^*,\tilde{\lambda}(\zeta)}$ almost surely

$$\begin{aligned}
\frac{1}{n^{\frac{1}{2}}}\partial_k \ell_n(\theta^*,\zeta) &= \frac{1}{n^{\frac{1}{2}}}\sum_{m=0}^{n-1} j_{\theta^*,k}(\tilde{Z}_m) - \tilde{\Lambda}_{\theta^*}(j_{\theta^*,k}) \\
&= \frac{1}{n^{\frac{1}{2}}}\sum_{m=1}^{n} V_k^*(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*}V_k^*(\tilde{Z}_{m-1}) \\
&\quad + \frac{1}{n^{\frac{1}{2}}}\Big(\tilde{\Pi}_{\theta^*}V_k^*(\tilde{Z}_0) - \tilde{\Pi}_{\theta^*}V_k^*(\tilde{Z}_n) + j_{\theta^*,k}(\tilde{Z}_0) - j_{\theta^*,k}(\tilde{Z}_n)\Big)
\end{aligned}$$

The second term converges to zero as the quantities $\tilde{\Pi}_{\theta^*}V_k^*(.)$ and $j_{\theta^*,k}$ are bounded (Recall that an inequality like (B.10) still holds. In fact, $V_k^*(z) \le C(\delta_\theta(y) + \mathrm{k}(U_{\theta,k},y)) + \mathrm{k}(j_{\theta^*,k},x,y))$. So that, for all $1 \le k \le p$,

$$\lim_{n\to\infty}\left[\frac{1}{n^{\frac{1}{2}}}\partial_k \ell_n(\theta^*,\zeta) - \frac{1}{n^{\frac{1}{2}}}\sum_{m=1}^{n}\Big(V_k^*(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*}V_k^*(\tilde{Z}_{m-1})\Big)\right] = 0, \;\; \tilde{P}_{\theta^*,\tilde{\lambda}(\zeta)} - a.s$$

Now, since $\Big(\tilde{\Pi}_{\theta^*}V_k^*(\tilde{Z}_n)\Big)_n$ is bounded, the convergence holds in $\mathbf{L}_2(\tilde{P}_{\theta^*,\tilde{\lambda}(\zeta)})$:

$$\tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}\left[\frac{1}{n^{\frac{1}{2}}}\partial_k \ell_n(\theta^*,\zeta) - \frac{1}{n^{\frac{1}{2}}}\sum_{m=1}^{n} V_k^*(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*}V_k^*(\tilde{Z}_{m-1})\right]^2 \underset{n\to\infty}{\longrightarrow} 0 \quad \text{(B.15)}$$

Let u be a fixed vector of $\mathbb{R}^p$; $\sum_{m=1}^n u^t B_m \triangleq \sum_{m=1}^n u^t(V^*(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*}V^*(\tilde{Z}_{m-1}))$ is a real valued Martingale adapted to the Borel σ-field $\{\tilde{\mathcal{Z}}_n = \sigma((\tilde{Z}_k)_{k\le n})\}_{n\ge 0}$ (the superscript "t" denotes the transpose vector in $\mathbb{R}^p$). We shall use Theorem 3.2 in Hall and Heyde (1980). We study

$$\begin{aligned}
\sigma_n^2 &\triangleq \sum_{m=1}^{n} \tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}(u^t.B_m)^2 \\
&= \sum_{1\le k,l\le p} u_k u_l \sum_{m=1}^{n} \tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)} \\
&\qquad \tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}\left[(V_k^*(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*}V_k^*(\tilde{Z}_{m-1}))(V_l^*(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*}V_l^*(\tilde{Z}_{m-1}))\Big|\tilde{\mathcal{Z}}_{m-1}\right] \\
&= \sum_{1\le k,l\le p} u_k u_l \sum_{m=1}^{n} \tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}\left[(V_k^*V_l^*)(\tilde{Z}_m) - \tilde{\Pi}_{\theta^*}V_k^*(\tilde{Z}_{m-1})\tilde{\Pi}_{\theta^*}V_l^*(\tilde{Z}_{m-1})\right]
\end{aligned}$$

Consider the function defined by:

$$z \to (V_k^*V_l^*)(z) \triangleq \mathcal{V}_{k,l}(z)$$

We check that this function is in Lip(**E'**) and verifies the condition (1.15). First

$$|\mathcal{V}_{k,l}(z)| \le C(\delta_{\theta^*}(y) + \delta'_{\theta^*}(y))^2$$

(where C depends on Δ_2 and Δ_2'), with Δ_{2s} and Δ_{2s}' finite for some $s > 1$, by assumption. The Lipschitz criteria is verified studying the quantity $\mathcal{V}_{k,l}(x, y, f, \sigma) - \mathcal{V}_{k,l}(x, y, f', \sigma')$ where $z = (x, y, f, \sigma)$ and $z' = (x, y, f', \sigma')$ are elements of the space **E'**. We have:

$$\begin{aligned}
&|V_k^* V_l^*(z) - V_k^* V_l^*(z')| \\
&\leq |V_k^*(z)| \times |V_l^*(z) - V_l^*(z')| + |V_l^*(z)| \times |V_k^*(z) - V_k^*(z')| \\
&\leq C(\delta_{\theta^*}(y) + \delta'_{\theta^*}(y)) \max_{1\leq l\leq p} |V_l^*(z) - V_l^*(z')|
\end{aligned}$$

and by definition :

$$V_l^*(z) = \sum_{n\geq 0} (\tilde{\Pi}^n_{\theta^*} j_{\theta^*,l}(z) - \tilde{\Lambda}_{\theta^*}(j_{\theta^*,l})) = \sum_{n\geq 0} \tilde{\Pi}^n_{\theta^*} j_{\theta^*,l}(z)$$

But $j_{\theta^*,l}$ is in Lip(**E'**) and for $n \geq 1$, the function $\tilde{\Pi}^n_{\theta^*} j_{\theta^*,l}$ belongs to Lip(**E'**) too, with $\text{lip}'(\Pi^n_{\theta^*} j_{\theta^*,l}, x, y) \leq C(\delta_\theta(y)^3 + \text{lip}(U_{\theta,k}, y) + \text{k}(U_{\theta,k}, y))\tilde{\rho}^n$ (the same argument as in remark 1 can be used to see this result). Then, assuming that $\Delta_4, (\Delta\Delta')_{4,1}$ and $(\Delta\Delta')_{3,2}$ are finite, we obtain that $\mathcal{V}_{k,l}$ belongs to Lip(**E'**).
The same argument shows that $z \to \tilde{\Pi}_{\theta^*} V_k^*(z) \tilde{\Pi}_{\theta^*} V_l^*(z)$ belongs to Lip(**E'**) and then

$$\begin{aligned}
&\frac{1}{n}\sum_{m=1}^{n} (V_k^* V_l^*)(\tilde{Z}_m) - (\tilde{\Pi}_{\theta^*} V_k^* \tilde{\Pi}_{\theta^*} V_l^*)(\tilde{Z}_{m-1}) \underset{n\to\infty}{\longrightarrow} \\
&\qquad \tilde{\Lambda}_{\theta^*}(V_k^* V_l^*)(z) - \tilde{\Lambda}_{\theta^*}(\tilde{\Pi}_{\theta^*} V_k^* \times \tilde{\Pi}_{\theta^*} V_l^*) \quad \tilde{P}_{\theta^*,\tilde{\lambda}(\zeta)} - a.s \\
&\text{i.e. } \frac{1}{n} \sum_{1\leq k,l\leq p} \sum_{m=1}^{n} u_k u_l \left[(V_k^* V_l^*)(\tilde{Z}_m) - (\tilde{\Pi}_{\theta^*} V_k^* \tilde{\Pi}_{\theta^*} V_l^*)(\tilde{Z}_{m-1})\right] \\
&\qquad \underset{n\to\infty}{\longrightarrow} \tilde{\Lambda}_{\theta^*}(V_k^* V_l^*) - \tilde{\Lambda}_{\theta^*}(\tilde{\Pi}_{\theta^*} V_k^* \times \tilde{\Pi}_{\theta^*} V_l^*) \quad \tilde{P}_{\theta^*,\tilde{\lambda}(\zeta)} - a.s
\end{aligned}$$

And, as this family is uniformly integrable under assumption 1.7, the convergence also happens for its first moment so that :

$$\frac{1}{\sigma_n^2} \sum_{m=1}^{n} \tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}[(u^t.B_m)^2 | \tilde{\mathcal{Z}}_{m-1}] \underset{n\to\infty}{\longrightarrow} 1 \quad \textit{in } \tilde{P}_{\theta^*,\tilde{\lambda}(\zeta)} - \textit{probability}.$$

We prove now that for any positive ϵ,

$$\frac{1}{n} \sum_{m=1}^{n} \tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)} \left((u^t.B_m)^2 . \mathbb{1}_{|u^t.B_m| \geq \sigma_n \epsilon} \middle| \tilde{\mathcal{Z}}_{m-1}\right) \underset{n\to\infty}{\longrightarrow} 0 \quad \textit{in } \tilde{P}_{\theta^*,\tilde{\lambda}(\zeta)} - \textit{probability}.$$

For $\alpha > 0$,

$$\begin{aligned}
& \tilde{P}_{\theta^*,\tilde{\lambda}(\zeta)}\left(\frac{1}{n}\sum_{m=1}^{n}\tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}\left[(u^t.B_m)^2 \mathbb{1}_{|u^t.B_m|\geq \sigma_n\epsilon}|\tilde{\mathcal{Z}}_{m-1}\right] \geq \alpha\right) \\
= \; & \tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}\left(\mathbb{1}_{\sum_{m=1}^{n}\tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}((u^t.B_m)^2\mathbb{1}_{|u^t.B_m|\geq\sigma_n\epsilon}|\tilde{\mathcal{Z}}_{m-1})\geq n\alpha}\right) \\
\leq \; & \frac{1}{n\alpha}\sum_{m=1}^{n}\tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}\,\tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}[(u^t.B_m)^2\mathbb{1}_{|u^t.B_m|\geq\sigma_n\epsilon}|\tilde{\mathcal{Z}}_{m-1}] \\
\leq \; & \frac{1}{\alpha}\max_{1\leq m\leq n}\tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}[(u^t.B_m)^2\mathbb{1}_{|u^t.B_m|^{2s-2}\geq(\sigma_n\epsilon)^{2s-2}}] \\
\leq \; & \frac{1}{\alpha(\epsilon\sigma_n)^{2s-2}}\max_{1\leq m\leq n}\tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}(|u^t.B_m|^{2s})
\end{aligned}$$

as $(\tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}(|u^t.B_m|^{2s}))_m$ is bounded and $1/\sigma_n$ tends to zero, we get the result. We conclude that

$$\frac{1}{\sigma_n}\sum_{m=1}^{n}u^t.B_m \to \mathcal{N}(0,1) \quad \textit{weakly under } \tilde{P}_{\theta^*,\tilde{\lambda}(\zeta)}.$$

which means

$$\frac{1}{n^{\frac{1}{2}}}\sum_{m=1}^{n}u^t.B_m \to \mathcal{N}(0,u^t.J_{\theta^*}.u) \quad \textit{weakly under } \tilde{P}_{\theta^*,\tilde{\lambda}(\zeta)}.$$

where $(J_{\theta^*})_{k,l} = \tilde{\Lambda}_{\theta^*}(V_k^*V_l^*) - \Lambda_{\theta^*}(\tilde{\Pi}_{\theta^*}V_k^* \times \tilde{\Pi}_{\theta^*}V_l^*)$. Moreover,

$$\frac{1}{n}\sum_{m=1}^{n}\tilde{\mathbb{E}}_{\theta^*,\tilde{\lambda}(\zeta)}\left(B_m^t.B_m\right) \underset{n\to\infty}{\longrightarrow} J_{\theta^*}$$

Then, equation (B.15) shows that:

$$\begin{aligned}
& n^{-\frac{1}{2}}\nabla\ell_n(\theta^*,\zeta) \longrightarrow \mathcal{N}(0,J_\theta^*) \quad \text{weakly under P}^* \\
& \text{and } \lim_{n\to\infty}\frac{1}{n}\mathbb{E}^*\left(\nabla\ell_n(\theta^*,\zeta)^t.\nabla\ell_n(\theta^*,\zeta)\right) = J_{\theta^*}
\end{aligned}$$

□

B.4.3 Exponential forgetting for the Hessian of the prediction filter and consequences

The chain $\{\tilde{Z}_n\}$ lies in the space $\mathbf{E"} = \mathbf{K}\times\mathcal{Y}\times\mathbf{S}^+\times\mathbf{\Sigma}\times\mathbf{\Sigma_2}$ (where $\mathbf{\Sigma_2} = \{\tau = (\tau_{k,l})/\tau_{k,l} : \mathbf{K}\to\mathbb{R}$ *and for all* v *in* $\mathbf{K}, \tau(v)$ *is symetric and positive*$\}$).
We first need to find a recurrence relation between $\nabla^2 f_{\theta,n+1}^{\zeta}$ and $\nabla^2 f_{\theta,n}^{\zeta}$. Using the recurrence relation (1.35) between $\nabla f_{\theta,n+1}^{\zeta}$ and $\nabla f_{\theta,n}^{\zeta}$, we obtain that $\{\nabla^2 f_{\theta,n}^{\zeta}\}$ is a functional auto-regressive process

$$\partial_{k,l}^2 f_{\theta,n+1}^{\zeta}(v) = \int_u a_\theta(Y_n, f_{\theta,n}^{\zeta})(u,v)\partial_{k,l}^2 f_{\theta,n}^{\zeta}(u)d\gamma(u) + T_{\theta,k,l}(Y_n, f_{\theta,n}^{\zeta}, \nabla f_{\theta,n}^{\zeta})(v) \qquad \text{(B.16)}$$

where

$$
\begin{aligned}
T_{\theta,k,l}(y,f,\sigma)(v) &= \frac{\int \partial^2_{k,l} q_\theta(u,v) g_\theta(y|u) f(u) d\gamma(u)}{\int g_\theta(y|w) f(w) d\gamma(w)} \\
&+ \int \frac{\partial_l g_\theta(y|u) f(u) + g_\theta(y|u)\sigma_l(u)}{\int g_\theta(y|u) f(u) d\gamma(u)} \times \left(\partial_k q_\theta(u,v) - \frac{\int \partial_k q_\theta(w,v) g_\theta(y|w) f(w) d\gamma(w)}{\int g_\theta(y|w) f(w) d\gamma(w)} \right) d\gamma(u) \\
&+ \int \frac{\partial_k g_\theta(y|u) f(u) + g_\theta(y|u)\sigma_k(u)}{\int g_\theta(y|w) f(w) d\gamma(w)} \times \left(\partial_l q_\theta(u,v) - \frac{\int \partial_l q_\theta(w,v) g_\theta(y|w) f(w) d\gamma(w)}{\int g_\theta(y|w) f(w) d\gamma(w)} \right) d\gamma(u) \\
&+ \int \Bigg[\frac{\partial_k g_\theta(y|u)\sigma_l(u) + \partial_l g_\theta(y|u)\sigma_k(u)}{\int g_\theta(y|w) f(w) d\gamma(w)} \\
&- \left(\frac{\int \partial_l g_\theta(y|w) f(w) d\gamma(w) + \int g_\theta(y|w)\sigma_l(w) d\gamma(w)}{\int g_\theta(y|w) f(w) d\gamma(w)} \right) \times \left(\frac{g_\theta(y|u)\sigma_k(u) + \partial_k g_\theta(y|u) f(u)}{\int g_\theta(y|w) f(w) d\gamma(w)} \right) \\
&- \left(\frac{\int \partial_k g_\theta(y|w) f(w) d\gamma(w) + \int g_\theta(y|w)\sigma_k(w) d\gamma(w)}{\int g_\theta(y|w) f(w) d\gamma(w)} \right) \times \left(\frac{g_\theta(y|u)\sigma_l(u) + \partial_l g_\theta(y|u) f(u)}{\int g_\theta(y|w) f(w) d\gamma(w)} \right) \Bigg] \\
&\times \left[q_\theta(u,v) - \frac{\int q_\theta(w,v) g_\theta(y|w) f(w) d\gamma(w)}{\int g_\theta(y|w) f(w) d\gamma(w)} \right] d\gamma(u) \\
&+ \int \frac{\partial^2_{k,l} g_\theta(y|u) f(u)}{\int g_\theta(y|w) f(w) d\gamma(w)} \left[q_\theta(u,v) - \frac{\int q_\theta(w,v) g_\theta(y|w) f(w) d\gamma(w)}{\int g_\theta(y|w) f(w) d\gamma(w)} \right] d\gamma(u)
\end{aligned}
$$

Relation (B.16) can be iterated to obtain $\nabla^2 f^\zeta_{\theta,n+1} = \alpha_{n-m+1}(Y_n^m, f^\zeta_{\theta,m}, \nabla f^\zeta_{\theta,m}, \nabla^2 f^\zeta_{\theta,m}; \theta)$. Such as in the preceding sections, we prove an exponential forgetting of the initial conditions for $\alpha_{n-m}(.;\theta)$.

Proposition B.4.4. *Under assumptions 1.1 and 1.9, there exists $C > 0$ such that,*

$$
\begin{aligned}
&\sup_{x_m,\ldots,x_n} \int \|\alpha^{k,l}_{n-m+1}(y_n^m, f, \sigma, \tau; \theta) - \alpha^{k,l}_{n-m+1}(y_n^m, f', \sigma', \tau'; \theta)\|_1 \\
&\qquad g_{\theta^*}(y_m|x_m) \ldots g_{\theta^*}(y_n|x_n) d\nu(y_m) \ldots d\nu(y_n) \\
\leq\ & C(n-m)^2(1-\epsilon)^{n-m+1}[\|\tau_{k,l} - \tau'_{k,l}\|_1 + [\|\sigma_k - \sigma'_k\|_1 + \|\sigma_l - \sigma'_l\|_1] \\
&(1 + \|\sigma_k\|_1 + \|\sigma_l\|_1 + \|\sigma'_k\|_1 + \|\sigma'_l\|_1) \\
&+ \|f - f'\|_1 \left[\|\tau_{k,l}\|_1 + \|\tau'_{k,l}\|_1 + (1 + \|\sigma_k\|_1 + \|\sigma_l\|_1 + \|\sigma'_k\|_1 + \|\sigma'_l\|_1)^2\right]]
\end{aligned}
$$

Then we define a set of Lipschitz functions on **E"**

Definition B.4.2. Lip(**E"**) *is the set of vector-valued fuctions $h = (h_{k,l})_{1 \leq k,l \leq p}$ with each $h_{k,l}$ being a measurable function on* **E"** *such that for all k, l in $\{1 \ldots p\}$, for all (x, y) in $\mathbf{K} \times \mathcal{Y}$, there exists* lip"$(h, x, y)$ *and* k"(h, x, y) *such that: for any f, f' in $\mathbf{S}^+$, σ, σ' in $\mathbf{\Sigma}$, and τ, τ' in $\mathbf{\Sigma_2}$,*

$$
\begin{aligned}
&|h_{k,l}(x,y,f,\sigma,\tau) - h_{k,l}(x,y,f',\sigma',\tau')| \\
&\quad \leq \mathrm{lip}''(h,x,y)\, [\|\tau_{k,l} - \tau'_{k,l}\|_1 + (\|\sigma_k - \sigma'_k\|_1 + \|\sigma_l - \sigma'_l\|_1) \times (1 + \|\sigma_k\|_1 + \|\sigma'_k\|_1 + \|\sigma_l\|_1 + \|\sigma'_l\|_1) \\
&\quad + \|f - f'\|_1 [\|\tau_{k,l}\|_1 + \|\tau'_{k,l}\|_1 + (1 + \|\sigma_k\|_1 + \|\sigma'_k\|_1 + \|\sigma_l\|_1 + \|\sigma'_l\|_1)^2]]
\end{aligned}
$$

$$
|h_{k,l}(x,y,f,\sigma,\tau)| \leq \mathrm{k}''(h,x,y)[\|\tau_{k,l}\|_1 + (1 + \|\sigma_k\|_1 + \|\sigma_l\|_1)^2]
$$

$$and \left\{ \begin{array}{l} \text{lip}''(h) = \text{esssup}_x \int \text{lip}''(h,x,y)\, g_{\theta^*}(y|x) d\nu(y) < \infty \\ \text{k}''(h) = \text{esssup}_x \int \text{k}''(h,x,y)\, g_{\theta^*}(y|x) d\nu(y) < \infty \end{array} \right.$$

The transition kernel $\check{\Pi}$ satisfies a condition of contraction for those functions, and then we prove the ergodicity of the chain.

Proposition B.4.5. *Under assumptions 1.1 and 1.9, there exists finite constants $\check{\rho} \in]\rho;1[$ and $C > 0$, and for all z, z' in* **E"**, *h in Lip(***E"***), and $n \geq 1$,*

$$\begin{aligned} |\check{\Pi}^n_\theta h_{k,l}(z) - \check{\Pi}^n_\theta h_{k,l}(z')| \leq C(\text{k}''(h) + \text{lip}''(h))\check{\rho}^n \big[\|\alpha_1^{k,l}(y,f,\sigma,\tau;\theta)\|_1 \\ + \|\alpha_1^{k,l}(y',f',\sigma',\tau';\theta)\|_1 + [1 + \|\Psi_1^k(y,f,\sigma;\theta)\|_1 + \|\Psi_1^l(y,f,\sigma;\theta)\|_1 \\ + \|\Psi_1^k(y',f',\sigma';\theta)\|_1 + \|\Psi_1^l(y',f',\sigma';\theta)\|_1]^2 \big] \end{aligned}$$

Corollary B.4.6. *Under assumptions 1.1 and 1.9, the Markov Chain $\check{Z}_n$ is geometrically ergodic.*
*There exists finite constants $\check{\rho} \in]\rho;1[$ and $C > 0$, such that for all h in Lip(***E"***), there exists a finite constant $\check{\Lambda}_\theta(h)$, such that for all z in* **E"** *and, for all $n \geq 1$,*

$$\begin{aligned} |\check{\Pi}^n_\theta h_{k,l}(z) - \check{\Lambda}_\theta(h)| \leq C \times (\text{lip}''(h) + \text{k"}(h))\check{\rho}^n \\ [\|\alpha_1^{k,l}(y,f,\sigma,\tau;\theta)\|_1 + (1 + \|\Psi_1^k(y,f,\sigma;\theta)\|_1 + \|\ \Psi_1^l(y,f,\sigma;\theta)\|_1)^2] \end{aligned} \quad \text{(B.17)}$$

For any initial probability measure $\check{\lambda}$ on (**E"**, $\mathcal{B}$(**E"**)), *and for any function h in* Lip(**E"**)

$$\check{\Lambda}_\theta(h) = \lim_{n\to\infty} \check{\mathbb{E}}_{\theta,\check{\lambda}}(h(\check{Z}_n)) \quad \text{(B.18)}$$

and there exists a solution in **E"** *to the Poisson equation*

$$(I - \check{\Pi}_\theta)V = h - \check{\Lambda}_\theta(h)$$

Proofs of Proposition B.4.5 and Corollary B.4.6 follow the same lines as Proposition B.4.2 and Corollary B.4.3.

Proof. (Proposition 1.3.2) Differentiating the expression of the gradient of the log-likelihood, we get for all k,l in $\{1 \cdots p\}$,

$$\frac{1}{n}\partial^2_{k,l}\ell_n(\theta,\zeta) = \frac{1}{n}\sum_{m=0}^{n-1} i_{\theta,kl}(\check{Z}_m) \quad \check{P}_{\theta,\check{\lambda}(\zeta)} - a.s. \quad \text{(B.19)}$$

where,

$$\begin{aligned} i_{\theta,kl}(x,y,f,\sigma,\tau) \triangleq \frac{\int g_\theta(y|u)\tau^{kl}(u)d\gamma(u)}{\int g_\theta(y|u)f(u)d\gamma(u)} + \frac{\int \partial^2_{kl}g_\theta(y|u)f(u)d\gamma(u)}{\int g_\theta(y|u)f(u)d\gamma(u)} \\ + \frac{\int \partial_k g_\theta(y|u)\sigma_l(u)d\gamma(u)}{\int g_\theta(y|u)f(u)d\gamma(u)} + \frac{\int \partial_l g_\theta(y|u)\sigma_k(u)d\gamma(u)}{\int g_\theta(y|u)f(u)d\gamma(u)} \\ - \left(\frac{\int \partial_l g_\theta(y|u)f(u)d\gamma(u)}{\int g_\theta(y|u)f(u)d\gamma(u)} + \frac{\int g_\theta(y|u)\sigma_l(u)d\gamma(u)}{\int g_\theta(y|u)f(u)d\gamma(u)}\right) \\ \times \left(\frac{\int \partial_k g_\theta(y|u)f(u)d\gamma(u)}{\int g_\theta(y|u)f(u)d\gamma(u)} + \frac{\int g_\theta(y|u)\sigma_k(u)d\gamma(u)}{\int g_\theta(y|u)f(u)d\gamma(u)}\right) \end{aligned}$$

With long but simple calculus, we check that $i_{\theta,kl}$ is in Lip(**E"**) and verifies (1.15), assuming that Δ_{2s}, Δ'_{2t}, Δ''_r $(\Delta\Delta'')_{1,1}$, and $(\Delta\Delta')_{2,2}$ are finite for some $r, s, t > 1$.
Using similar arguments as for Proposition 1.2.5, we obtain

$$\frac{1}{n}\nabla^2\ell_n(\theta^*,\zeta) \longrightarrow \check{\Lambda}_{\theta^*}(i_{\theta^*}) \triangleq -I_{\theta^*}$$

Then,

$$I_{\theta^*} = -\lim_{n\to\infty}\mathbb{E}^*\left(\frac{1}{n}\nabla^2\ell_n(\theta^*,\zeta)\right)$$

And as permutation between differentiation and integration is valid in our case,

$$J_{\theta^*} = \lim_{n\to\infty}\mathbb{E}^*\left(\frac{1}{n}\nabla\ell_n(\theta^*,\zeta)^t.\nabla\ell_n(\theta^*,\zeta)\right) = -\lim_{n\to\infty}\mathbb{E}^*\left(\frac{1}{n}\nabla^2\ell_n(\theta^*,\zeta)\right) = I_{\theta^*}$$

□

Proof. (Proposition 1.3.3) Same proof as for the quantity $\ell_n(\theta,\zeta)-\ell_n(\theta',\zeta)$, generalizing Lemma 1.2.8 to the gradient and the Hessian of the prediction filter. □

Annexe C

Technical proofs for the estimation of linear functionals in the convolution model

Proof. Proof of Lemma 3.2.1.
We write the functional Γ_ℓ in the following way:

$$\Gamma_\ell(y) \;=\; \int (x-y+y)^\ell \Phi_1(x-y)g(x)dx = \sum_{j=0}^{\ell} \binom{\ell}{j} y^{\ell-j} \int (x-y)^j \Phi_1(x-y)g(x)dx,$$

where

$$(x-y)^j\Phi_1(x-y) = \frac{1}{\sqrt{2\pi}}(x-y)^j e^{-(x-y)^2/2} = \sum_{k=0}^{j} \gamma(k,j)\frac{d^k}{dy^k}\Phi_1(x-y) \tag{C.1}$$

with known constants $(\gamma(k,j))_{0\le k\le j}$, and $\gamma(j,j)=1$. It leads to the equalities

$$\begin{aligned}\Gamma_\ell(y) &= \sum_{j=0}^{\ell}\binom{\ell}{j} y^{\ell-j}\int \sum_{k=0}^{j}\gamma(k,j)\frac{d^k}{dy^k}\Phi_1(x-y)g(x)dx \\ &= \sum_{j=0}^{\ell}\binom{\ell}{j} y^{\ell-j}\sum_{k=0}^{j}\gamma(k,j)\frac{d^k}{dy^k}\int \Phi_1(x-y)g(x)dx\end{aligned}$$

And finally, we get

$$\Gamma_\ell(y) = \sum_{j=0}^{\ell}\binom{\ell}{j} y^{\ell-j}\sum_{k=0}^{j}\gamma(k,j)h^{(k)}(y) = h^{(\ell)}(y) + \sum_{k=0}^{\ell-1} P_{\ell-k}(y)h^{(k)}(y), \tag{C.2}$$

for some polynomial functions $\{P_j\}_{1\le j\le \ell}$ with $deg(P_j)=j$, which concludes the proof. □

Proof. Proof of Lemma 3.2.2
We only give the proof for the functional $\Gamma_{c,\ell}$, the proof concerning $\Gamma_{s,\ell}$ being exactly the same. Use $\cos(\ell x) = [\exp(i\ell x)+\exp(-i\ell x)]/2$, to write $\Gamma_{c,\ell}$ as

$$\begin{aligned}\Gamma_{c,\ell}(y) = \frac{1}{2\sqrt{2\pi}}\int e^{i\ell x}e^{-(y-x)^2/2}g(x)dx + \frac{1}{2\sqrt{2\pi}}\int e^{-i\ell x}e^{-(y-x)^2/2}g(x)dx \\ = \frac{e^{-\ell/2}}{2\sqrt{2\pi}}\left[\int e^{i\ell y}e^{-(y+i\ell-x)^2/2}g(x)dx + \int e^{-i\ell y}e^{-(y-i\ell-x)^2/2}g(x)dx\right].\end{aligned}$$

We finish the proof by using that the function h defined as $z\mapsto \int e^{-(z-x)^2/2}g(x)dx$ admits analytic continuation on the whole complex plane. □

Proof. Proof of Theorem 3.3.2.
The use of the van Trees inequality (Gill and Levit 1995) requires to construct a one-dimensional family of densities included into $\mathcal{H}$. Consider the kernel $S_n(x) = C_nS(C_nx)$ with $S(x) = \sin(x)/(\pi x)$, a sequence $\{C_n\}_{n\ge 0}$ of real positive numbers tending to infinity. Denote by g_2 the function $x\mapsto g_2(x) = \sin^2(x)/(\pi x^2)$, and define g_0 and h_0 by:

$$g_0 = g_2 * g_2 \quad \text{and} \quad h_0 = \Phi_1 * g_0\,. \tag{C.3}$$

Note that g_2 and then g_0 and h_0 are probability densities. Moreover define the normalising constant:

$$\overline{S}_n(y_0) = \int S_n\left(y_0 - u\right) h_0(u) du. \tag{C.4}$$

For some fixed sequence of parameters $\{\theta_n\}_{n\geq 0}$ decreasing to zero (to be specified later), we define the parametric path $\{h_\theta\}_{|\theta|\leq\theta_n}$ by the formula:

$$\forall y \in \mathbb{R}, \quad h_\theta(y) = h_0(y)\left[1 + \theta\left(S_n(y_0 - y) - \overline{S}_n(y_0)\right)\right], \quad \forall\, |\theta| \leq \theta_n, \tag{C.5}$$

where $\overline{S}_n(y_0)$ is defined by (C.4). Note that the constant $\overline{S}_n(y_0)$ ensures that $\int h_\theta = 1$.
Denote by $I(\theta)$ the Fisher information for the family of probability densities $\{h_\theta; |\theta| \leq \theta_n\}$:

$$I(\theta) = \int \left[\frac{\partial \log h_\theta(x)}{\partial\theta}\right]^2 h_\theta(x) dx = \int \frac{\left(S_n(y_0 - x) - \overline{S}_n(y_0)\right)^2}{1 + \theta(S_n(y_0 - x) - \overline{S}_n(y_0))} h_0(x) dx. \tag{C.6}$$

The following result due to Taupin (1998) ensures that for an appropriate choice of the parameters $\{C_n\}_{n\geq 0}$ and $\{\theta_n\}_{n\geq 0}$ the family of densities defined by (C.5) is included into the set $\mathcal{H}$. Moreover, it gives an upper-bound on the Fisher information of this family.

Assumption C.1. $\theta_n c_n^2 e^{(c_n+2)^2/2} \underset{\text{n}\to\infty}{\longrightarrow} 0.$

Lemma C.0.7. *Under Assumption C.1 we have, for large enough n,*

$$\{h_\theta\}_{|\theta|\leq\theta_n} \textit{ is included in } \mathcal{H}.$$

Moreover, there exists a positive constant κ such that for large enough n:

$$I(\theta) \leq \kappa C_n, \quad \forall |\theta| \leq \theta_n.$$

We now use the van Trees inequality to get an infimum bound for the minimax quadratic risk for the estimation of polynomial functionals Γ_f. Using Lemma 3.2.1, we turn this problem to the estimation of the derivatives $h^{(k)}$ of the density h for $0 \leq k \leq \ell$. In fact, the point y_0 is fixed so that in this context, $\Gamma_f(y_0)$ is exactly a linear combination of the derivatives $h^{(k)}(y_0)$ for $0 \leq k \leq \ell$, when f is a polynimial function of degree equal to ℓ. Consider $\theta \mapsto \lambda_0(\theta)$ a probability density on $[-1, 1]$ verifying $\lambda_0(-1) = \lambda_0(1) = 0$ and λ_0 is continuously differentiable on $]-1;1[$. Its Fisher information is defined by

$$I_0 = \int_{-1}^{1} \frac{{\lambda_0'}^2(\theta)}{\lambda_0(\theta)} d\theta.$$

Rescaling this probability density on the interval $[-\theta_n; \theta_n]$, we define: $\lambda(\theta) = \theta_n^{-1}\lambda_0(\theta_n^{-1}\theta)$ with Fisher information $I(\lambda) = \theta_n^{-2} I_0$.

Fix y_0 in $\mathbb{R}$ and an integer $0 \leq k \leq \ell$. We have:

$$\begin{aligned}
\inf_{\widehat{h}_n} \sup_{h\in\mathcal{H}} \mathbb{E}[\widehat{h}_n - h_\theta^{(k)}(y_0)]^2 &\geq \inf_{\widehat{h}_n} \sup_{|\theta|\leq\theta_n} \mathbb{E}_{h_\theta}[\widehat{h}_n - h_\theta^{(k)}(y_0)]^2 \\
&\geq \inf_{\widehat{h}_n} \int_{-\theta_n}^{+\theta_n} \mathbb{E}_{h_\theta}[\widehat{h}_n - h_\theta^{(k)}(y_0)]^2 \lambda(\theta) d\theta,
\end{aligned}$$

where the infima are taken over all the estimators $\widehat{h}_n$ based on the observations $Y_1, \ldots, Y_n$. Applying the van Trees inequality (Gill and Levit 1995), we get:

$$\inf_{\widehat{h}_n} \sup_{h \in \mathcal{H}} \mathbb{E}[\widehat{h}_n - h^{(k)}(y_0)]^2 \geq \left[\int_{-\theta_n}^{\theta_n} \frac{\partial h_\theta^{(k)}(y_0)}{\partial \theta} \lambda(\theta) d\theta \right]^2 \left[n \int_{-\theta_n}^{\theta_n} I(\theta)\lambda(\theta)d\theta + I(\lambda) \right]^{-1}.$$

By definition of $h_\theta^{(k)}$, we have the equalities

$$\begin{aligned} \frac{\partial h_\theta^{(k)}(y_0)}{\partial \theta} &= \frac{\partial}{\partial \theta} \frac{d^k}{dz^k} [h_0(z)(1 + \theta(S_n(y_0 - z) - \overline{S}_n(y_0)))]_{|z=y_0} \\ &= \frac{d^k}{dz^k} [h_0(z) S_n(y_0 - z)]_{|z=y_0} - h_0^{(k)}(y_0)\overline{S}_n(y_0). \end{aligned}$$

Since $S_n^{(k)}(0) = C_n^{k+1} S^{(k)}(0)$ we obtain that

$$\begin{aligned} \left[\frac{d^k}{dz^k} h_0(z) S_n(y_0 - z) \right]_{|z=y_0} &= \sum_{m=0}^{k} \binom{k}{m} (-1)^{k-m} h_0^{(m)}(y_0) S_n^{(k-m)}(0) \\ &= h_0(y_0)[(-1)^k S_n^{(k)}(0) + r_n^{(k)}(y_0)], \end{aligned}$$

where $r_n^{(k)}(y_0) = O(C_n^k)$. This leads to the equality

$$\frac{\partial h_\theta^{(k)}(y_0)}{\partial \theta} = h_0(y_0)[(-1)^k S_n^{(k)}(0) + O(C_n^k)] - h_0^{(k)}(y_0)\overline{S}_n(y_0). \tag{C.7}$$

We now use a result proved by Taupin (1998).

Lemma C.0.8.

$$\overline{S}_n(y_0) = h_0(y_0) + r_n(y_0),$$

where $r_n(y_0)$ converges to zero as n tends to infinity.

Combining this result with equation (C.7) leads to

$$\left[\int_{-\theta_n}^{\theta_n} \frac{\partial h_\theta^{(k)}(y_0)}{\partial \theta} \lambda(\theta) d\theta \right]^2 = \left[h_0(y_0) C_n^{k+1} S^{(k)}(0) + O(C_n^k) \right]^2,$$

and finally

$$\left[\int_{-\theta_n}^{\theta_n} \frac{\partial h_\theta^{(k)}(y_0)}{\partial \theta} \lambda(\theta) d\theta \right]^2 \geq \kappa' C_n^{2(k+1)} h_0^2(y_0) [S^{(k)}(0)]^2$$

where κ' is a positive constant. Using that for large enough n, $I(\theta) \leq \kappa C_n$ (Lemma C.0.7), we obtain the following lower bound:

$$\inf_{\widehat{h}_n} \sup_{h \in \mathcal{H}} \mathbb{E}[\widehat{h}_n - h^{(k)}(y_0)]^2 \geq \frac{\kappa' C_n^{2(k+1)} h_0^2(y_0)[S^{(k)}(0)]^2}{n C_n \kappa + \theta_n^{-2} I_0}.$$

Maximise this bound under the constraint given by Assumption C.1 (ensuring that the family $\{h_\theta, |\theta| \leq \theta_n\}$ is included into to $\mathcal{H}$):

$$\theta_n C_n^2 e^{(C_n+2)^2/2} \underset{n \to \infty}{\longrightarrow} 0.$$

Choose

$$C_n = \sqrt{\alpha \log n} \text{ with } \alpha < 1; \qquad \theta_n = \frac{\sqrt{\log n}}{\sqrt{n}},$$

to get that there exists a positive constant κ'' such that

$$\inf_{\widehat{h}_n} \sup_{h \in \mathcal{H}} \mathbb{E}[\widehat{h}_n - h^{(k)}(y_0)]^2 \geq \kappa'' \frac{C_n^{2k+1}}{n}.$$

Finally, we obtain the desired result:

$$\liminf_{n\to\infty} \frac{\sqrt{n}}{(\log n)^{(2k+1)/4}} \inf_{\widehat{h}_n} \sup_{h \in \mathcal{H}} \mathbb{E}[\widehat{h}_n - h^{(k)}(y_0)]^2 > 0,$$

and its corollary (with the use of Lemma 3.2.1):

$$\liminf_{n\to\infty} \frac{\sqrt{n}}{(\log n)^{(2\ell+1)/4}} \inf_{\widehat{\Gamma}_n} \sup_{\Gamma \in \mathcal{G}_{poly,\ell}} \mathbb{E}[\widehat{\Gamma}_n - \Gamma_f(y_0)]^2 > 0.$$

□

Proof. Proof of Theorem 3.3.3.
For sake of simplicity, we only study the rate of convergence for the estimation of

$$\Gamma_c(y_0) = \int \cos(x)\Phi_1(y_0 - x)g(x)dx.$$

with the estimator

$$\widehat{\Gamma}_{c,n}(y_0) = \frac{e^{-1/2}}{2n} \sum_{k=1}^{n} [e^{iy_0} S_n(y_0 + i - Y_k) + e^{-iy_0} S_n(y_0 - i - Y_k)].$$

Using the triangular inequality, the term $\mathbb{E}[\widehat{\Gamma}_{c,n}(y_0) - \Gamma_c(y_0)]^2$ is bounded by the sum of the square of the bias term $\mathbb{E}[\widehat{\Gamma}_{c,n}(y_0)] - \Gamma_c(y_0)$ and the variance term Var $(\widehat{\Gamma}_{c,n})$. We start with the bound of the bias $\mathbb{E}[\widehat{\Gamma}_{c,n}(y_0)] - \Gamma_c(y_0)$, given by

$$\frac{e^{-1/2}}{2} \left[e^{iy_0} \left(\mathbb{E}S_n(y_0 + i - Y_1) - h(y_0 + i) \right) + e^{-iy_0} \left(\mathbb{E}S_n(y_0 - i - Y_1) - h(y_0 - i) \right) \right].$$

Use that

$$\int e^{ity - t} S_n^\star(t)dt = S_n(y + i),$$

to get the bounds

$$\begin{aligned} |\mathbb{E}S_n(y_0 + i - Y_1) - h(y_0 + i)| &\leq \frac{1}{2\pi} \int e^{t} |h^\star(t)||S_n^\star(t) - 1|dt, \\ \text{and} \quad |\mathbb{E}S_n(y_0 - i - Y_1) - h(y_0 - i)| &\leq \frac{1}{2\pi} \int e^{-t} |h^\star(t)||S_n^\star(t) - 1|dt, \end{aligned}$$

which provide, using the Fourier transform properties of S, [K1]-[K3], the bound of the bias

$$\sup_{y_0} \left| \mathbb{E}[\widehat{\Gamma}_{c,n}(y_0)] - \Gamma_c(y_0) \right| \leq \frac{e^{1/2}}{2\pi} \int e^{|t| - t^2/2} \mathbb{1}_{|t| \geq C_n} dt.$$

Finally we have,

$$\sup_{y_0}\left|\mathbb{E}[\widehat{\Gamma}_{c,n}(y_0)]-\Gamma_c(y_0)\right|^2 \leq \frac{e^{-1}}{\pi^2}\times\frac{e^{-(C_n-1)^2}}{(C_n-1)^2}. \tag{C.8}$$

We now come to the variance term. Combining both that Var $(X)\leq\mathbb{E}[X^2]$ and that $e^{iy_0}S_n(y_0+i-Y_1)+e^{-iy_0}S_n(y_0-i-Y_1)$ belongs to $\mathbb{R}$, to get

$$\text{Var }(\widehat{\Gamma}_{c,n}(y_0))\leq\frac{e^{-1}}{4n}\mathbb{E}\left|e^{iy_0}S_n(y_0+i-Y_1)+e^{-iy_0}S_n(y_0-i-Y_1)\right|^2.$$

Since

$$\mathbb{E}\left|e^{iy_0}S_n(y_0+i-Y_1)+e^{-iy_0}S_n(y_0-i-Y_1)\right|^2\leq \\ 2\mathbb{E}\left|S_n(y_0+i-Y_1)\right|^2+2\mathbb{E}\left|S_n(y_0-i-Y_1)\right|^2,$$

we have

$$\text{Var }(\widehat{\Gamma}_{c,n}(y_0))\leq\frac{e^{-1}}{2n}\left[\mathbb{E}|S_n(y_0+i-Y_1)|^2+\mathbb{E}|S_n(y_0-i-Y_1)|^2\right].$$

By definition

$$\begin{aligned}\mathbb{E}\left|S_n(y_0+i-Y_1)\right|^2 &= C_n\int|S(u+iC_n)|^2h(y_0+u/C_n)du,\\ \text{and}\quad \mathbb{E}\left|S_n(y_0-i-Y_1)\right|^2 &= C_n\int|S(u-iC_n)|^2h(y_0+u/C_n)du.\end{aligned}$$

Since $|h(y_0+u/C_n)|\leq\| h\|_\infty<\infty$, the variance is bounded by

$$\text{Var }(\widehat{\Gamma}_{c,n}(y_0))\leq\frac{e^{-1}}{2n}\|h\|_\infty C_n\int\left(|S(u+iC_n)|^2+|S(u-iC_n)|^2\right)du,$$

with

$$\begin{aligned}|S(u+iC_n)|^2 &= \frac{e^{-2C_n}+e^{2C_n}-2e^{-2C_n}\cos(2u)}{4\pi^2(u^2+C_n^2)},\\ \text{and}\quad |S(u-iC_n)|^2 &= \frac{e^{-2C_n}+e^{2C_n}-2e^{2C_n}\cos(2u)}{4\pi^2(u^2+C_n^2)},\end{aligned}$$

and therefore

$$\begin{aligned}|S(u+iC_n)|^2 &\leq \frac{e^{2C_n}+3}{4\pi^2(u^2+C_n^2)},\\ \text{and}\quad |S(u-iC_n)|^2 &\leq \frac{3e^{2C_n}+1}{4\pi^2(u^2+C_n^2)}.\end{aligned}$$

It follows that

$$\text{Var }(\widehat{\Gamma}_{c,n}(y_0))\leq\frac{e^{-1}\|h\|_\infty C_n(3e^{2C_n}+1)}{4\pi^2n}\int\frac{du}{u^2+C_n^2}\leq\frac{e^{-1}\|h\|_\infty(3e^{2C_n}+1)}{4\pi^2n}\int\frac{dv}{v^2+1},$$

and therefore there exists a constant A such that

$$\sup_{y_0}\text{Var }(\widehat{\Gamma}_c(y_0)) \quad\leq\quad A\frac{e^{2C_n}}{n}. \tag{C.9}$$

Combining inequalities (C.8) and (C.9) entails the result with the choice $C_n = \sqrt{\log n}$.

□

Proof. Proof of Theorem 3.3.4.
We only give the proof for the function $f : x \mapsto \cos(\ell x)$. By linearity of the functional $f \mapsto \Gamma_f$, it induces the result for linear combinations of this type. The proof concerning the sinus is a simple adaptation of this one.
To prove this theorem, we consider the same path $\{h_\theta\}_{|\theta|\leq\Theta}$ as the one used in the proof of Theorem 3.3.2 and the path $\{\Gamma_\theta\}_{\theta\in\Theta}$ induced by the equation of Lemma 3.2.2. More precisely:

$$\Gamma_\theta(z) = \frac{e^{-\ell/2}}{2}[e^{i\ell z}h_\theta(z+i\ell) + e^{-i\ell z}h_\theta(z-i\ell)] \tag{C.10}$$

where h_θ is defined by (C.5). Assume that the parameters θ_n and C_n satisfy Assumption C.1, so that the family $\{h_\theta, |\theta| \leq \theta_n\}$ is included in $\mathcal{H}$. By construction, we get the inclusion:

$$\{\Gamma_\theta, |\theta| \leq \theta_n\} \subset \mathcal{G}_f.$$

We now use the van Trees inequality to obtain an infimum bound on the pointwise minimax quadratic risk. Proceeding as in the polynomial case, we consider $\lambda_0(\theta)$ a probability density on the interval $[-1,1]$ with Fisher information given by

$$I_0 = \int_1^1 \frac{{\lambda_0'}^2(\theta)}{\lambda_0(\theta)}d\theta,$$

where $\lambda_0(-1) = \lambda_0(1) = 0$ and λ_0 is continuously differentiable on $]-1;1[$. Using the fact that the sub-family $\{\Gamma_\theta, |\theta| \leq \theta_n\}$ is included $\mathcal{G}_f$ we get that the pointwise minimax quadratic risk is lower bounded in the following way:

$$\inf_{\widehat{\Gamma}_n}\sup_{\Gamma_f\in\mathcal{G}_f}\mathbb{E}[\widehat{\Gamma}_n - \Gamma_f(y_0)]^2 \geq \inf_{\widehat{\Gamma}_n}\sup_{|\theta|\leq\theta_n}\mathbb{E}_\theta[\widehat{\Gamma}_n - \Gamma_\theta(y_0)]^2 \geq \inf_{\widehat{\Gamma}_n}\int_{-\theta_n}^{+\theta_n}\mathbb{E}_\theta[\widehat{\Gamma}_n - \Gamma_{\theta,f}(y_0)]^2\lambda(\theta)d\theta.$$

Now, we apply the van Trees inequality and get that

$$\inf_{\widehat{\Gamma}_n}\sup_{\Gamma_f\in\mathcal{G}_f}\mathbb{E}[\widehat{\Gamma}_n - \Gamma_f(y_0)]^2 \geq \left[\int_{-\theta_n}^{\theta_n}\frac{\partial\Gamma_\theta(y_0)}{\partial\theta}\lambda(\theta)d\theta\right]^2\left[n\int_{-\theta_n}^{\theta_n}I(\theta)\lambda(\theta)d\theta + I(\lambda)\right]^{-1}.$$

The denominator of this expression has already been studied in the proof of Theorem 3.3.2. The purpose is then to study the behaviour of the numerator of this expression. Using Definitions (C.5) and (C.10) of the families $\{h_\theta; |\theta| \leq \theta_n\}$ and $\{\Gamma_\theta; \leq \theta_n\}$, we get

$$\frac{\partial}{\partial\theta}\Gamma_\theta(y_0) = \frac{e^{-\ell/2}}{2}e^{i\ell y_0}h_0(y_0+i\ell)(S_n(-i\ell)-\overline{S_n}(y_0)) + \frac{e^{-\ell/2}}{2}e^{-i\ell y_0}h_0(y_0-i\ell)(S_n(i\ell)-\overline{S_n}(y_0)),$$

and therefore

$$\frac{\partial\Gamma_\theta(y_0)}{\partial\theta} = \frac{e^{-\ell/2}}{2}\left[e^{i\ell y_0}h_0(y_0+i\ell)S_n(-i\ell) + e^{-i\ell y_0}h_0(y_0-i\ell)S_n(i\ell)\right] - \Gamma_0(y_0)\overline{S_n}(y_0).$$

where

$$\Gamma_0(y_0) = e^{-\ell/2}/2[e^{i\ell y_0}h_0(y_0+i\ell) + e^{-i\ell y_0}h_0(y_0-i\ell)].$$

Using that the kernel S is even, $S_n(i\ell) = S_n(-i\ell) = C_n S_n(i\ell C_n)$ which implies that

$$\begin{aligned}\frac{\partial \Gamma_\theta(y_0)}{\partial\theta} &= \frac{e^{-\ell/2}}{2} C_n S(i\ell C_n)\left[e^{iy_0}h_0(y_0+i\ell) + e^{-iy_0}h_0(y_0-i\ell)\right] - \Gamma_0(y_0)\overline{S_n}(y_0) \\ &= C_n S(i\ell C_n)\Gamma_0(y_0) - \Gamma_0(y_0)\overline{S_n}(y_0).\end{aligned}$$

Finally

$$\frac{\partial \Gamma_\theta(y_0)}{\partial\theta} = \Gamma_0(y_0)\left(C_n S(i\ell C_n) - \overline{S_n}(y_0)\right),$$

and using (C.4)

$$\left[\int_{-\theta_n}^{\theta_n} \frac{\partial \Gamma_\theta(y_0)}{\partial\theta}\lambda(\theta)d\theta\right]^2 = \Gamma_0(y_0)^2 C_n^2 S^2(i\ell C_n)[1+O(1)].$$

Using the definition of the kernel S, we get

$$C_n S(i\ell C_n) = \frac{e^{\ell C_n} - e^{-\ell C_n}}{2\pi\ell},$$

which leads to

$$\left[\int_{-\theta_n}^{\theta_n} \frac{\partial \Gamma_\theta(y_0)}{\partial\theta}\lambda(\theta)d\theta\right]^2 = \frac{\Gamma_0(y_0)^2}{4\pi^2\ell^2}e^{2\ell C_n}[1+O(1)].$$

Noting that there exists a positive constant κ_1 such that $I(\theta) \le \kappa_1 c_n$ (see the proof of Theorem 3.3.2), we get that for large enough n:

$$\inf_{\widehat{\Gamma}_n} \sup_{\Gamma_f \in \mathcal{G}_f} \mathbb{E}[\widehat{\Gamma}_n - \Gamma_f(y_0)]^2 \ge \frac{\kappa_2\Gamma_0(y_0)^2 e^{2\ell C_n}}{nc_n\kappa_1 + \theta_n^{-2}I_0}.$$

We have to optimize this lower bound under the constraint imposed by Assumption C.1:

$$\theta_n C_n^2 e^{(C_n+2)^2/2} \underset{n\to\infty}{\longrightarrow} 0.$$

Choose

$$C_n = \sqrt{\log n} - 2\sqrt{2} \quad \text{and} \quad \theta_n = \frac{\sqrt{\log n}}{\sqrt{n}},$$

which ensures the existence of a positive constant κ such that

$$\inf_{\widehat{\Gamma}_n} \sup_{\Gamma_f \in \mathcal{G}_f} \mathbb{E}[\widehat{\Gamma}_n - \Gamma_f(y_0)]^2 \ge \kappa\frac{e^{2\ell\sqrt{\log n}}}{n\sqrt{\log n}},$$

and completes the proof. □

Proof. Proof of Lemma 3.5.4.
Recall that the family is defined by

$$\varphi_{m,a}(x) = \left(1 - \sum_{j=1}^{m-1} a(j)\bar{V}_{m,j}\right) \alpha_0 * \Phi_1(x) + \sum_{j=1}^{m-1} a(j) f_0(x) V_{m,j}(x).$$

According to (3.53),

$$|\bar{V}_{m,j}| \leq \frac{\theta_m}{m} \|f_0\|_\infty \|V\|_1. \tag{C.11}$$

which combined with $\|V_{m,j}\|_\infty \leq \theta_m \|V\|_\infty$, gives that $\varphi_{m,a}$ is a positive function for large enough m, satisfying moreover:

$$\int \varphi_{m,a}(x)dx = \int f_0(x)dx + \sum_{j=1}^{m-1} a(j) \left(\int f_0(x)V_{m,j}(x)dx - \bar{V}_{m,j}\right) = 1,$$

and hence $\varphi_{m,a}$ is a probability density. The main point of the proof lies in checking that there exists a family of probability densities $\{\alpha_{m,a}\}_{a\in\mathcal{A}}$ on $\mathbb{R}$ such that for all integer m,

$$\forall a \in \mathcal{A}\,, \quad \alpha_{m,a} * \Phi_1 = \varphi_{m,a}. \tag{C.12}$$

We start with the proof of the existence of a family of real valued functions $\{\beta_{m,j}\}_{1\leq j\leq m-1}$ such that:

$$\forall 1 \leq j \leq m-1\,, \quad f_0 V_{m,j} = \beta_{m,j} * \Phi_1.$$

The following proof follows the lines of Theorem 2.2.1 in Taupin (1998). Consider the function $a_{m,j}$ defined by

$$a_{m,j}(t) = e^{t^2/2}(f_0 V_{m,j})^*(t). \tag{C.13}$$

The Fourier transform of the product $f_0 V_{m,j}$ is the convolution product between $f_0^* = \alpha_0^* \Phi_1^*$ and $V_{m,j}^*(u) = \frac{\theta_m}{m} e^{-iuj/m} V^*(\frac{u}{m})$ all having a compact support. This implies in particular that the function $a_{m,j}$ has a compact support, included in the algebraic sum of the preceeding supports, that is to say included in $[-2-2m; 2+2m]$. In particular, the function $a_{m,j}$ belongs to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$, and we can define $\beta_{m,j}$ as its inverse Fourier transform (denoted by $\bar{*}$) defined by

$$\beta_{m,j}(x) = a_{m,j}^{\bar{*}}(x) = \frac{1}{2\pi} \int e^{-itx} e^{t^2/2} (f_0 V_{m,j})^*(t) dt.$$

Then we get the identity:

$$\beta_{m,j} * \Phi_1 = a_{m,j}^{\bar{*}} * \Phi_1 = [a_{m,j}^* \Phi_1^*]^{\bar{*}} = [(f_0 V_{m,j})^*]^{\bar{*}} = f_0 V_{m,j}.$$

By construction $f_0 V_{m,j}$ is a real valued function. Consequently, $\mathrm{Im}(\beta_{m,j} * \Phi_1) \equiv 0$ and therefore $\Phi_1 * \mathrm{Im}(\beta_{m,j}) \equiv 0$. This implies that $\Phi_1^*(\mathrm{Im}(\beta_{m,j}))^* \equiv 0$ and thus $\mathrm{Im}(\beta_{m,j}) \equiv 0$. We finally get that $\beta_{m,j}$ is a real valued function. Now we define $\alpha_{m,a}$ in the following way:

$$\alpha_{m,a} = \left(1 - \sum_{j=1}^{m-1} a(j)\bar{V}_{m,j}\right) \alpha_0 + \sum_{j=1}^{m-1} a(j)\beta_{m,j}. \tag{C.14}$$

The function $\alpha_{m,a}$ is a real valued function that satisfies the equalities:

$$\begin{aligned}\alpha_{m,a} * \Phi_1 &= \left(1 - \sum_{j=1}^{m-1} a(j)\bar{V}_{m,j}\right) \alpha_0 * \Phi_1 + \sum_{j=1}^{m-1} a(j)\beta_{m,j} * \Phi_1 \\ &= \left(1 - \sum_{j=1}^{m-1} a(j)\bar{V}_{m,j}\right) f_0 + \sum_{j=1}^{m-1} a(j) f_0 V_{m,j}\end{aligned}$$

and therefore

$$\alpha_{m,a} * \Phi_1 = \varphi_{m,a}.$$

The last thing is to prove that $\alpha_{m,a}$ is a probability density on $\mathbb{R}$. The main point lies in proving that $\alpha_{m,a}$ is a positive function. This, combined with the facts that $\varphi_{m,a}$ is a probability density, $\varphi_{m,a} = \alpha_{m,a} * \Phi_1$ and Fubini's Theorem, give us that it integrates to one. It remains thus to prove that, for m large enough, it is a positive function.
The first step is to note that

$$\|\alpha_{m,a} - \alpha_0\|_\infty = \left\| \sum_{j=1}^{m-1} a(j)\bar{V}_{m,j}\alpha_0 - \sum_{j=1}^{m-1} a(j)\beta_{m,j} \right\|_\infty \underset{\mathrm{m}\to\infty}{\longrightarrow} 0, \tag{C.15}$$

so that for each compact set K in $\mathbb{R}$, we can find an integer m_0 large enough such that, for all m greater or equal to m_0 and for all x in K, we have $\alpha_{m,a}(x) \geq 0$. Indeed, easy calculations provide that $a_{m,j}$ can be written

$$a_{m,j}(t) = e^{t^2/2} \frac{\theta_m}{m} \int \alpha_0^*(u-t) e^{-(u-t)^2/2} e^{-iuj/m} V_{m,j}^* \left(\frac{u}{m}\right) du,$$

and therefore

$$\begin{aligned}|\beta_{m,j}(x)| &\leq \frac{\theta_m}{2\pi m} \int\int e^{t^2/2} \mathbb{1}_{\{|t| \leq 2+2m\}} |\alpha_0^*(u-t)| e^{-(u-t)^2/2} \left| V^*\left(\frac{u}{m}\right)\right| dudt \\ &\leq \frac{\theta_m}{2\pi} e^{2m^2+4m+2} \|\alpha_0^*\|_1 \|V^*\|_1. \end{aligned} \tag{C.16}$$

Combining (C.11) with (C.16) we get that

$$\begin{aligned}\left\| \sum_{j=1}^{m-1} a(j)\bar{V}_{m,j}\alpha_0 \right\|_\infty &\leq \|\alpha_0\|_\infty \|f_0\|_\infty \|V\|_1 \theta_m \\ \left\| \sum_{j=1}^{m-1} a(j)\beta_{m,j} \right\|_\infty &\leq \frac{e^2}{2\pi} \|\alpha_0^*\|_1 \|V^*\|_1 m\theta_m e^{2m^2+4m},\end{aligned}$$

and the uniform convergence of $\alpha_{m,a}$ to α_0 in (C.15) follows under Assumption C.1.
The second step deals with what happens for large $|x|$. For this write $\beta_{m,j}$ in the form

$$\begin{aligned}\beta_{m,j}(x) &= \frac{\theta_m}{2\pi m} \int\int e^{itx} e^{t^2/2} e^{-iuj/m} e^{-(u-t)^2/2} \alpha_0^*(u-t) V^*\left(\frac{u}{m}\right) dudt \\ &= \frac{\theta_m}{2\pi m} \int\int e^{itx - iuj/m} e^{-u^2/2+ut} \alpha_0^*(u-t) V^*\left(\frac{u}{m}\right) dudt \\ &= \frac{\theta_m}{2\pi m} \int e^{ixu-iuj/m} \left(\int e^{-v(ix+u)} \alpha_0^*(v) dv \right) e^{u^2/2} V^*\left(\frac{u}{m}\right) du\end{aligned}$$

Simple calculations show that

$$\int e^{-v(ix+u)}\alpha_0^*(v)dv = \frac{-2}{(u+ix)^2} + \frac{e^{2(u+ix)} - e^{-2(u+ix)}}{2(u+ix)^3},$$

and then we finally obtain

$$\beta_{m,j}(x) = \frac{\theta_m}{2\pi m}\int e^{ixu-iuj/m}\left(\frac{-2}{(u+ix)^2} + \frac{e^{2(u+ix)} - e^{-2(u+ix)}}{2(u+ix)^3}\right)e^{u^2/2}V^*\left(\frac{u}{m}\right)du, \qquad \text{(C.17)}$$

and the bound:

$$\left|\sum_{j=1}^{m-1} a(j)\beta_{m,j}(x)\right| \;\leq\; \frac{Cm\theta_m e^{2m^2+4m}}{x^2}, \qquad \text{(C.18)}$$

where C is a positive constant and $m\theta_m e^{2m^2+4m} \underset{\mathrm{m}\to\infty}{\longrightarrow} 0$. Arguing as in Taupin (1998), we get that large enough $|x|$, we have

$$\left(1 - \sum_{j=1}^{m-1} a(j)\bar{V}_{m,j}\right)\alpha_0(x) \;\geq\; \frac{\delta}{x^2}, \qquad \text{(C.19)}$$

$$\text{(C.20)}$$

where δ is a positive constant. This lower bound is based on the following inequalities: for all $x > 0$ write $x = k\pi + t$ where $k > 0$, and $0 < t < \pi$. Then we get

$$\begin{aligned}\alpha_0(x) \;&=\; \int \frac{\sin^2(t-u)\sin^2(u)}{[u(k\pi+t-u)]^2}du \\ &\geq \int_{|u|\leq\pi/2} \sin^2(t-u)\sin^2(u)/[u(k\pi+t-u)]^2 du.\end{aligned}$$

But $(k-1/2)\pi < k\pi + t - u < (k+1)\pi - \pi/2$ leads to

$$\alpha_0(x) \;\geq\; [(k+3/2)\pi]^{-2}\int_{|u|\leq\pi/2} \sin^2(t-u)\sin^2(u)/u^2 du,$$

and thus

$$\begin{aligned}\alpha_0(x) \;&\geq\; C\,[(k+3/2)\pi]^{-2}\int_{|u|\leq\pi/2}\sin^2(t-u)du \\ &\geq\; K\,[(k+3/2)\pi]^{-2}.\end{aligned}$$

Then we use that

$$x^2 = k^2\pi^2 + 2tk\pi + t^2 > k^2\pi^2.$$

Moreover, there exists a constant C'' such that for k large enough, $(k+3/2)^2\pi^2 \leq C''k^2\pi^2 \leq C''x^2$, and then

$$(k+3/2)^{-2}\pi^{-2} \geq [C''x^2]^{-1},$$

which gives the result.

Combining (C.18) and (C.19), we get that for large enough $|x|$, the quantity $\alpha_{m,a}(x)$ is positive.
Note that Equation (C.18) follows from (C.17), by using that

$$\begin{aligned}\left|\sum_{j=1}^{m-1} a(j)\beta_{m,j}(x)\right| &\leq (m-1)\frac{\theta_m}{2\pi m}\int\left(\frac{2}{u^2+x^2}+\frac{e^{2|u|}}{(u^2+x^2)^{3/2}}\right)e^{u^2/2}1\!\!1_{|\frac{u}{m}|\leq 2}du\\ &\leq \frac{\theta_m}{\pi x^2}\int e^{u^2/2+2|u|}1\!\!1_{|\frac{u}{m}|\leq 2}du\\ &\leq \frac{4}{\pi x^2}m\theta_m e^{2m^2+4m}\end{aligned}$$

which converges to 0 under Assumption C.1. □

Proof. Proof of lemma 3.5.6.
We will consider the different cases $\ell = 0$ (concerning the estimation of the density of the observations) and $\ell \geq 1$ (concerning the derivatives of this density).
<u>First case:</u> $\ell = 0$.
Fix two functions a and a' in $\mathcal{A}_2$. By definition, the distance $\|\varphi_{m,a} - \varphi_{m,a'}\|_p$ is given by

$$\|\varphi_{m,a} - \varphi_{m,a'}\|_p = \left\|\sum_{j=1}^{m-1}(a(j)-a'(j))f_0(V_{m,j}-\bar{V}_{m,j})\right\|_p .$$

Now, we restrict our attention to the interval $[1/(2m);(2m-1)/(2m)]$ and lower bound the integral $\|\varphi_{m,a} - \varphi_{m,a'}\|_p^p$ by the restricted integral over this subset. This interval is the disjoint union of the intervals $[(2k-1)/(2m);(2k+1)/2m]$, when $1 \leq k \leq m-1$, so that we get:

$$\|\varphi_{m,a} - \varphi_{m,a'}\|_p^p \geq \sum_{k=1}^{m-1}\int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}}\left|\sum_{j=1}^{m-1}(a(j)-a'(j))f_0(x)(V_{m,j}(x)-\bar{V}_{m,j})\right|^p dx.$$

We use the following inequality on real numbers, valid for all $p \geq 1$,

$$|a|^p \geq 2^{-p}|a-b|^p - |b|^p , \tag{C.21}$$

to obtain that

$$\begin{aligned}\|\varphi_{m,a} - \varphi_{m,a'}\|_p^p \geq \sum_{k=1}^{m-1}\int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}} \Big[&2^{-p}|a(k)-a'(k)|^p f_0^p(x)|V_{m,k}(x)-\bar{V}_{m,k}|^p\\ &-\Big|\sum_{\substack{1\leq j\leq m-1\\ j\neq k}}(a(j)-a'(j))f_0(x)(V_{m,j}(x)-\bar{V}_{m,j})\Big|^p\Big]dx.\end{aligned}$$

Note that the functions a and a' take values in $\{0;1\}$, so that the quantity $|a(k)-a'(k)|^p$ equals $|a(k)-a'(k)|$ and is bounded by 1, for all $1 \leq k \leq m-1$. Use also the triangular inequality to obtain

$$\|\varphi_{m,a}-\varphi_{m,a'}\|_p^p \geq \sum_{k=1}^{m-1}\int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}} \Big[2^{-p}|a(k)-a'(k)|f_0^p(x)|V_{m,k}(x)-\bar{V}_{m,k}|^p - \Big(\sum_{\substack{1\leq j\leq m-1\\ j\neq k}} f_0(x)|V_{m,j}(x)-\bar{V}_{m,j}|\Big)^p\Big]dx.$$

The interval $[1/(2m);(2m-1)/(2m)]$ is included in $[0;1]$, and the density f_0 is positive on the compact interval $[0;1]$, so that we can write:

$$\|\varphi_{m,a}-\varphi_{m,a'}\|_p^p \geq 2^{-p}\sum_{k=1}^{m-1}|a(k)-a'(k)|\left(\inf_{[0;1]} f_0^p\right)\int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}}|V_{m,k}(x)-\bar{V}_{m,k}|^p dx - \sum_{k=1}^{m-1}\int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}}\Big(\sum_{\substack{1\leq j\leq m-1\\ j\neq k}} f_0(x)|V_{m,j}(x)-\bar{V}_{m,j}|\Big)^p dx. \quad \text{(C.22)}$$

Consider the first term, and use the Inequality (C.21) to get that

$$\int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}}|V_{m,k}(x)-\bar{V}_{m,k}|^p dx \geq 2^{-p}\int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}}|V_{m,k}(x)|^p dx - \frac{|\bar{V}_{m,k}|^p}{m}$$

with

$$2^{-p}\int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}}|V_{m,k}(x)|^p dx - \frac{|\bar{V}_{m,k}|^p}{m} = 2^{-p}\frac{\theta_m^p}{m}\int_{-1/2}^{1/2}|V|^p - \frac{|\bar{V}_{m,k}|^p}{m}$$

and therefore by using (C.11)

$$\int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}}|V_{m,k}(x)-\bar{V}_{m,k}|^p dx \geq 2^{-p}\frac{\theta_m^p}{m}\int_{-1/2}^{1/2}|V|^p - \frac{\theta_m^p}{m^{p+1}}\|f_0\|_\infty^p\|V\|_1^p.$$

Returning to (C.22), we get:

$$\|\varphi_{m,a}-\varphi_{m,a'}\|_p^p \geq 2^{-p}\frac{\theta_m^p}{m}\left(\inf_{[0;1]} f_0^p\right)\sum_{k=1}^{m-1}|a(k)-a'(k)|\left(2^{-p}\int_{-1/2}^{1/2}|V|^p - \frac{\|f_0\|_\infty^p\|V\|_1^p}{m^p}\right) - \sum_{k=1}^{m-1}\int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}} f_0(x)^p\Big(\sum_{\substack{1\leq j\leq m-1\\ j\neq k}}|V_{m,j}(x)-\bar{V}_{m,j}|\Big)^p dx.$$

By definition of the set $\mathcal{A}_2$, we have:

$$\sum_{k=1}^{m-1}|a(k)-a'(k)| \geq \frac{m-1}{4},$$

and since $m \geq 2$, it leads to

$$\|\varphi_{m,a} - \varphi_{m,a'}\|_p^p \geq \frac{2^{-p}}{8}\theta_m^p \left(\inf_{[0;1]} f_0^p\right) \left(\int_{-1/2}^{1/2} |V|^p - \frac{\|f_0\|_\infty^p \|V\|_1^p}{m^p}\right) - \sum_{k=1}^{m-1} \int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}} f_0(x)^p \Big(\sum_{\substack{1\leq j\leq m-1 \\ j\neq k}} |V_{m,j}(x) - \bar{V}_{m,j}| \Big)^p dx. \quad \text{(C.23)}$$

Let us consider the second term appearing in this inequality. Combine the triangular inequality, (C.11), and then the definition of the function V to get

$$\begin{aligned} \sum_{\substack{1\leq j\leq m-1 \\ j\neq k}} |V_{m,j}(x) - \bar{V}_{m,j}| &\leq \sum_{\substack{1\leq j\leq m-1 \\ j\neq k}} \theta_m |V(mx-j)| + m\|f_0\|_\infty \|V\|_1 \frac{\theta_m}{m} \\ &\leq \theta_m \Big(\sum_{\substack{1\leq j\leq m-1 \\ j\neq k}} \frac{2}{\pi(mx-j)^2} + \|f_0\|_\infty \|V\|_1 \Big). \end{aligned}$$

But when x belongs to the interval $[(2k-1)/(2m);(2k+1)(2m)]$, the variable $mx-j$ ranges the interval $[k-j-1/2;k-j+1/2]$ and then we have the upper bound:

$$\begin{aligned} \sum_{\substack{1\leq j\leq m-1 \\ j\neq k}} \frac{2}{\pi(mx-j)^2} &= \sum_{1\leq j\leq k-1} \frac{2}{\pi(mx-j)^2} + \sum_{k+1\leq j\leq m-1} \frac{2}{\pi(mx-j)^2} \\ &\leq \sum_{1\leq j\leq k-1} \frac{2}{\pi(k-j-1/2)^2} + \sum_{k+1\leq j\leq m-1} \frac{2}{\pi(k-j+1/2)^2} \\ &\leq \frac{4}{\pi} \sum_{n\geq 1} \frac{1}{(n-1/2)^2} \,. \end{aligned} \quad \text{(C.24)}$$

The last quantity is a finite constant denoted by C_1. Then, when x belongs to the interval $[(2k-1)/(2m);(2k+1)/(2m)]$, we have

$$\sum_{\substack{1\leq j\leq m-1 \\ j\neq k}} |V_{m,j}(x) - \bar{V}_{m,j}| \leq \theta_m \left(C_1 + \|f_0\|_\infty \|V\|_1\right) \cdot$$

Now, returning to (C.23), we obtain

$$\|\varphi_{m,a} - \varphi_{m,a'}\|_p^p \geq \theta_m^p \Big[\frac{2^{-p}}{8} \left(\inf_{[0;1]} f_0^p\right) \int_{-1/2}^{1/2} |V|^p - (C_1 + \|f_0\|_\infty \|V\|_1)^p \|f_0\|_p^p - \frac{2^{-p}}{8} \left(\inf_{[0;1]} f_0^p\right) \frac{\|f_0\|_\infty^p \|V\|_1^p}{m^p} \Big].$$

Let us study the right hand side of this inequality. The third term is negligible in front of the two others, as m tends to infinity. The last thing to check is that the quantity

$$\frac{2^{-p}}{8} \left(\inf_{[0;1]} f_0^p\right) \int_{-1/2}^{1/2} |V|^p - (C_1 + \|f_0\|_\infty \|V\|_1)^p \|f_0\|_p^p \,, \quad \text{(C.25)}$$

is positive. Use the function $V_\lambda(\cdot) = V(\lambda\cdot)$ instead of V (there is no loss of generality as the only important thing we used is the compact support of the Fourier transform of V). Then we have

$$\int_{-1/2}^{1/2} |V_\lambda|^p = \frac{1}{\lambda}\int_{-\lambda/2}^{\lambda/2} |V|^p$$

and instead of the upper-bound (C.24), we get

$$\sum_{\substack{1\leq j\leq m-1\\ j\neq k}} |V_\lambda(mx-j)| \leq \frac{1}{\lambda^2}\sum_{\substack{1\leq j\leq m-1\\ j\neq k}} |V(mx-j)| \leq \frac{C_1}{\lambda^2}.$$

Moreover,

$$\|V_\lambda\|_1 = \frac{1}{\lambda}\|V\|_1.$$

Now, the quantity (C.25) becomes:

$$\frac{1}{\lambda}\left[\frac{2^{-p}}{8}\left(\inf_{[0;1]} f_0^p\right)\int_{-\lambda/2}^{\lambda/2}|V|^p - \frac{1}{\lambda^{p-1}}\left(\frac{C_1}{\lambda}+\|f_0\|_\infty\|V\|_1\right)^p\|f_0\|_p^p\right],$$

which is positive for some λ_0 large enough, as $p>1$. Using this remark, we conclude that there exists a positive constant C such that

$$\|\varphi_{m,a}-\varphi_{m,a'}\|_p^p \geq C\theta_m^p.$$

<u>Second case:</u> $\ell \geq 1$.
By definition $\|\varphi^{(\ell)}_{m,a}-\varphi^{(\ell)}_{m,a'}\|_p$ given by

$$\|\varphi^{(\ell)}_{m,a}-\varphi^{(\ell)}_{m,a'}\|_p = \left\|\sum_{j=1}^{m-1}(a(j)-a'(j))\left(f_0^{(\ell)}(V_{m,j}-\bar{V}_{m,j})+\sum_{k=0}^{\ell-1}\binom{\ell}{k}f_0^{(k)}V_{m,j}^{(\ell-k)}\right)\right\|_p$$

satisfies the inequality

$$\|\varphi^{(\ell)}_{m,a}-\varphi^{(\ell)}_{m,a'}\|_p \geq A_1 - A_2 - A_3 \tag{C.26}$$

with

$$A_1 = \left\|\sum_{j=1}^{m-1}(a(j)-a'(j))f_0V_{m,j}^{(\ell)}\right\|_p \quad ; \quad A_2 = \left\|\sum_{j=1}^{m-1}(a(j)-a'(j))f_0^{(\ell)}(V_{m,j}-\bar{V}_{m,j})\right\|_p$$

$$\text{and } A_3 = \left\|\sum_{j=1}^{m-1}(a(j)-a'(j))\sum_{k=1}^{\ell-1}\binom{\ell}{k}f_0^{(k)}V_{m,j}^{(\ell-k)}\right\|_p.$$

The purpose is to lower bound A_1, and then upper bound A_2 and A_3 by quantities that are negligible in front of A_1. We begin with the first term A_1, which will be treated in the same way as the case $\ell=0$. We restrict our attention to the interval $[1/(2m);(2m-1)/(2m)]$, which is the disjoint union of the intervals $[(2k-1)/(2m);(2k+1)/2m]$, when $1 \leq k \leq m-1$, and use the Convexity Inequality (C.21) to write

$$\begin{aligned} A_1^p &= \int \Big| \sum_{j=1}^{m-1} (a(j)-a'(j)) f_0(x) m^\ell \theta_m V^{(\ell)}(mx-j) \Big|^p dx \\ &\geq m^{\ell p}\theta_m^p \sum_{k=1}^{m-1} \int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}} \Bigg[2^{-p}|a(k)-a'(k)| f_0^p(x) |V^{(\ell)}(mx-k)|^p \\ &\qquad - \Bigg| \sum_{\substack{1\leq j\leq m-1\\ j\neq k}} (a(j)-a'(j)) f_0(x) V^{(\ell)}(mx-j) \Bigg|^p \Bigg] dx. \end{aligned}$$

Same computations as in the case $\ell = 0$ lead to

$$\begin{aligned} A_1^p &\geq m^{\ell p}\theta_m^p \Big[\frac{2^{-p}}{8} \left(\inf_{[0;1]} f_0^p \right) \int_{-1/2}^{1/2} |V^{(\ell)}|^p \\ &\quad - \sum_{k=1}^{m-1} \int_{\frac{2k-1}{2m}}^{\frac{2k+1}{2m}} f_0^p(x) \Big(\sum_{1\leq j\leq m-1; j\neq k} |V^{(\ell)}(mx-j)| \Big)^p dx \Big], \end{aligned} \tag{C.27}$$

and the point is to upper bound the quantity

$$\sum_{1\leq j\leq m-1; j\neq k} |V^{(\ell)}(mx-j)|$$

by a finite constant. Remember the definition of the function V:

$$V(u) = \frac{\cos(u)-\cos(2u)}{\pi u^2} = \frac{1}{\pi} \frac{\mathrm{P}(\cos(u))}{\mathrm{Q}(u)},$$

where P is a polynomial function, and $Q(u) = u^2$. Now, compute the derivative with respect to the variable u,

$$V^{(\ell)}(u) = \frac{1}{\pi} \sum_{k=0}^{\ell} \binom{\ell}{k} (\mathrm{P} \circ \cos)^{(\ell-k)}(u) \left(\frac{1}{\mathrm{Q}} \right)^{(k)}(u).$$

Easy calculations give

$$\left(\frac{1}{\mathrm{Q}} \right)^{(k)}(u) = \frac{(-1)^k (k+1)!}{u^{2+k}},$$

and by obvious upper-bounds on trigonometric functions, there exist a constant M_ℓ such that, for all integer $0 \leq k \leq \ell$,

$$|(\mathrm{P}\circ\cos)^{(\ell-k)}(u)| \leq M_\ell.$$

Therefore the derivative $V^{(\ell)}$ satisfies

$$|V^{(\ell)}(u)| \leq \frac{M_\ell}{\pi} \sum_{k=0}^{\ell} \binom{\ell}{k} \frac{(k+1)!}{|u|^{2+k}} \leq \begin{cases} \kappa_\ell/|u|^2 & \text{if } |u| \geq 1 \\ \kappa_\ell/|u|^{2+\ell} & \text{if } |u| \leq 1 \end{cases} \tag{C.28}$$

where κ_ℓ is a positive constant. For all x in the interval $[(2k-1)/(2m);(2k+1)(2m)]$, the variable $mx-j$ ranges the interval $[k-j-1/2;k-j+1/2]$ and then we have

$$\sum_{1\leq j\leq m-1;j\neq k}|V^{(\ell)}(mx-j)| = \sum_{1\leq j\leq k-1}|V^{(\ell)}(mx-j)| + \sum_{k+1\leq j\leq m-1}|V^{(\ell)}(mx-j)|$$

which is bounded in the following way

$$\sum_{1\leq j\leq m-1;j\neq k}|V^{(\ell)}(mx-j)| \;\leq\; |V^{(\ell)}(mx-k+1)| + |V^{(\ell)}(mx-k-1)| + 2\kappa_\ell\sum_{n\geq 2}\frac{1}{(n-1/2)^2}.$$

It follows that $\sum_{1\leq j\leq m-1;j\neq k}|V^{(\ell)}(mx-j)|$ is bounded by

$$\begin{aligned}\kappa_\ell(2^{2+\ell}1\!\!1_{[1/2;1]}(mx-k+1) + 1\!\!1_{[1;3/2]}(mx-k+1))\\ + \kappa_\ell(2^{2+\ell}1\!\!1_{[-1;-1/2]}(mx-k-1) + 1\!\!1_{[-3/2;-1]}(mx-k-1))\\ + 2\kappa_\ell\sum_{n\geq 2}\frac{1}{(n-1/2)^2},\end{aligned}$$

and finally

$$\sum_{1\leq j\leq m-1;j\neq k}|V^{(\ell)}(mx-j)| \leq 2\kappa_\ell(1+2^{2+\ell}) + 2\kappa_\ell\sum_{n\geq 2}\frac{1}{(n-1/2)^2}\,. \tag{C.29}$$

The last quantity is a finite constant denoted by C_ℓ. Combining (C.29) with (C.27), we get

$$A_1^p \geq m^{\ell p}\theta_m^p\left[\frac{2^{-p}}{8}\left(\inf_{[0;1]} f_0^p\right)\int_{-1/2}^{1/2}|V^{(\ell)}|^p - C_\ell^p\|f_0\|_p^p\right].$$

The next thing to check is that the quantity

$$\frac{2^{-p}}{8}\left(\inf_{[0;1]} f_0^p\right)\int_{-1/2}^{1/2}|V^{(\ell)}|^p - C_\ell^p\|f_0\|_p^p \tag{C.30}$$

is positive. But the same argument as in the case $\ell=0$ concludes the proof. Use $V_\lambda(\cdot)=V(\lambda\cdot)$ instead of V (whose Fourier transform has a compact support). Then we have

$$\int_{-1/2}^{1/2}|V_\lambda^{(\ell)}|^p = \frac{1}{\lambda}\int_{-\lambda/2}^{\lambda/2}\lambda^{\ell p}|V^{(\ell)}|^p = \lambda^{\ell p-1}\int_{-\lambda/2}^{\lambda/2}|V^{(\ell)}|^p.$$

Using (C.28), we have

$$|V_\lambda^{(\ell)}(u)| = \lambda^\ell|V^{(\ell)}(\lambda u)| \leq \begin{cases}(\lambda^{\ell-2}\kappa_\ell)/|u|^2 & \text{if } |u|\geq 1\\ (\lambda^{-2}\kappa_\ell)/|u|^{2+\ell} & \text{if } |u|\leq 1\end{cases}$$

and the upper-bound (C.29) becomes

$$\sum_{1\leq j\leq m-1;j\neq k}|V_\lambda^{(\ell)}(mx-j)| \leq \lambda^{\ell-2}C'_\ell$$

where C'_ℓ is a positive constant. The quantity (C.30) becomes then

$$\lambda^{\ell p-1}\left(\frac{2^{-p}}{8}\left(\inf_{[0;1]} f_0^p\right)\int_{-\lambda/2}^{\lambda/2}|V^{(\ell)}|^p - \frac{(C'_\ell)^p}{\lambda^{2p-1}}\|f_0\|_p^p\right).$$

For $p>1/2$, the second term converges to zero as λ tends to infinity, while the first one converges to $2^{-p}/8(\inf_{[0;1]} f_0^p)\|V^{(\ell)}\|_p^p$. So we obtain that there exist some λ_0 large enough for which this quantity is positive. Using this remark, we conclude that there exists a positive constant C_1 such that

$$A_1 = \left\|\sum_{j=1}^{m-1}(a(j)-a'(j))f_0 V_{m,j}^{(\ell)}\right\|_p \geq C_1 m^\ell \theta_m. \tag{C.31}$$

We now come to upper bound the second term A_2. By definition and applying the triangular inequality we obtain

$$\begin{aligned} A_2 &= \left\|\sum_{j=1}^{m-1}(a(j)-a'(j))f_0^{(\ell)}(V_{m,j}-\bar{V}_{m,j})\right\|_p \\ &\leq \sum_{j=1}^{m-1}\|f_0^{(\ell)}(V_{m,j}-\bar{V}_{m,j})\|_p \\ &\leq \sum_{j=1}^{m-1}\left(\|f_0^{(\ell)}V_{m,j}\|_p + |\bar{V}_{m,j}|\ \|f_0^{(\ell)}\|_p\right) \\ &\leq \sum_{j=1}^{m-1}\left(\|f_0^{(\ell)}\|_\infty\|V_{m,j}\|_p + |\bar{V}_{m,j}|\ \|f_0^{(\ell)}\|_p\right). \end{aligned}$$

Again using the definition of $V_{m,j}$, we get

$$\|V_{m,j}\|_p = \frac{\theta_m}{m^{1/p}}\|V\|_p\ ,$$

which combined with Inequality (C.11) provide that for $p\geq 1$, there exists a positive constant C_2 such that

$$\begin{aligned} A_2 &\leq (m-1)\left(\frac{\theta_m}{m^{1/p}}\|V\|_p\|f_0^{(\ell)}\|_\infty + \frac{\theta_m}{m}\|f_0\|_\infty\|V\|_1\|f_0^{(\ell)}\|_p\right) \\ &\leq C_2 m^{1-1/p}\theta_m, \end{aligned} \tag{C.32}$$

and this quantity is negligible in front of $m^\ell\theta_m$ when $\ell\geq 1$. Finally, the third term A_3 verifies

$$\begin{aligned} A_3 &= \left\|\sum_{j=1}^{m-1}(a(j)-a'(j))\sum_{k=1}^{\ell-1}\binom{\ell}{k}f_0^{(k)}V_{m,j}^{(\ell-k)}\right\|_p \\ &\leq \sum_{j=1}^{m-1}\sum_{k=1}^{\ell-1}\binom{\ell}{k}\|f_0^{(k)}\|_\infty\|V_{m,j}^{(\ell-k)}\|_p\ . \end{aligned}$$

Note that

$$\|V_{m,j}^{(\ell-k)}\|_p = \theta_m \frac{m^{\ell-k}}{m^{1/p}} \|V^{(\ell-k)}\|_p \leq m^{\ell-1-1/p} \theta_m \max_{1\leq r\leq \ell-1} \|V^{(r)}\|_p.$$

And then

$$A_3 \leq 2^\ell m^{\ell-1/p} \theta_m \max_{1\leq r\leq \ell-1} \|f_0^{(r)}\|_\infty \max_{1\leq r\leq \ell-1} \|V^{(r)}\|_p \tag{C.33}$$

which is also negligible in front of $m^\ell \theta_m$. Combining (C.26), (C.31), (C.32) and (C.33), we conclude that for all $\ell \geq 1$, there exists a positive constant C such that

$$\|\varphi_{m,a}^{(\ell)} - \varphi_{m,a'}^{(\ell)}\|_p \geq C m^\ell \theta_m$$

and the proof is complete. □

Proof. Proof of Lemma 3.5.8.
The computations start exactly in the same way as for the proof of Lemma 3.5.4 until the discussion about the behaviour of $\beta_{m,j,\ell}(x)$ (which is the analogue of $\beta_{m,j}$ when f_0 is replaced by $f_{0,\ell}$) for large enough $|x|$. In fact, the expression of α_0 plays a role in establishing equation (C.17). Here, the function $\alpha^*_{0,\ell}$ is equal to the square of C_ℓ times the convolution product of $2\ell+2$ times the function $S^* = 1\!\!1_{[-1;1]}$. More precisely:

$$\alpha^*_{0,\ell} = C_\ell^2 (S^* * \ldots * S^*)^2,$$

where the convolution product appears $2\ell+2$ times. This function is an even piecewise polynomial, with support equal to $[-2\ell-2; 2\ell+2]$, and is 2ℓ times continuously differentiable. The behaviour of $\beta_{m,j,\ell}(x)$ is related to the quantity:

$$\int e^{-v(ix+u)} \alpha^*_{0,\ell}(v) dv \quad = \quad I(ix+u) + I(-ix-u),$$

where

$$I(c) = \int_0^{2\ell+2} e^{cv} \alpha^*_{0,\ell}(v) dv.$$

Integrating by parts, we get:

$$\begin{aligned} I(c) = & \left[\frac{e^{cv}}{c} \alpha^*_{0,\ell}(v)\right]_0^{2\ell+2} - \left[\frac{e^{cv}}{c^2} \left(\alpha^*_{0,\ell}\right)'(v)\right]_0^{2\ell+2} + \ldots \\ & - \left[\frac{e^{cv}}{c^{2\ell}} \left(\alpha^*_{0,\ell}\right)^{(2\ell-1)}(v)\right]_0^{2\ell+2} + \left[\frac{e^{cv}}{c^{2\ell+1}} \left(\alpha^*_{0,\ell}\right)^{(2\ell)}(v)\right]_0^{2\ell+2} - \int_0^{2\ell+2} \frac{e^{cv}}{c^{2\ell+1}} \left(\alpha^*_{0,\ell}\right)^{(2\ell+1)}(v) dv. \end{aligned}$$

Since $\alpha^*_{0,\ell}$ is 2ℓ times continuously differentiable and equal to zero outside $[-2\ell-2; 2\ell+2]$, its derivatives up to the order 2ℓ are equal to zero at the point $2\ell+2$. Moreover, $\alpha_{0,\ell}$ is an even function, so that $(\alpha^*_{0,\ell})^{(2k-1)}(0) = 0$ for all $k \leq \ell$. Now adding $I(c)$ with $I(-c)$, we obtain that:

$$I(c) + I(-c) = -\int_0^{2\ell+2} \frac{e^{cv} - e^{-cv}}{c^{2\ell+1}} \left(\alpha^*_{0,\ell}\right)^{(2\ell+1)}(v) dv,$$

and integrating by parts we get:

$$I(c) + I(-c) = -\left[\frac{e^{cv} - e^{-cv}}{c^{2\ell+2}} \left(\alpha^*_{0,\ell}\right)^{(2\ell+1)}(v)\right]_0^{2\ell+2} + \int_0^{2\ell+2} \frac{e^{cv} - e^{-cv}}{c^{2\ell+2}} \left(\alpha^*_{0,\ell}\right)^{(2\ell+2)}(v) dv.$$

This gives that

$$\int e^{-v(ix+u)}\alpha^*_{0,\ell}(v)dv = -\left[\frac{e^{(ix+u)v} - e^{-(ix+u)v}}{(ix+u)^{2\ell+2}}\left(\alpha^*_{0,\ell}\right)^{(2\ell+1)}(v)\right]_0^{2\ell+2} + \int_0^{2\ell+2}\frac{e^{(ix+u)v} - e^{-(ix+u)v}}{(ix+u)^{2\ell+2}}\left(\alpha^*_{0,\ell}\right)^{(2\ell+2)}(v)dv,$$

and is bounded in the following way:

$$|\int e^{-v(ix+u)}\alpha^*_{0,\ell}(v)dv| \leq \gamma_\ell \frac{e^{2(\ell+1)|u|}}{(u^2+x^2)^{\ell+1}}.$$

We use this expression to conclude that we have:

$$\left|\sum_{j=1}^{m-1} a(j)\beta_{m,j,\ell}(x)\right| \leq \frac{Cm\theta_m e^{2m^2+4(\ell+1)m}}{x^{2\ell+2}},$$

with, by assumption $m\theta_m e^{2m^2+4(\ell+1)m} \longrightarrow_{m\to\infty} 0$. The analogue of the lower bound (C.19) is established in the same way and gives that for large enough $|x|$ there exists a positive constant δ such that

$$\left(1 - \sum_{j=1}^{m-1} a(j)\bar{V}_{m,j}\right)\alpha_{0,\ell}(x) \geq \frac{\delta}{x^{2\ell+2}},$$

which ends up the proof of Lemma 3.5.8 in the same way as the proof of Lemma 3.5.4. □

Références

Arapostathis, A. and Marcus, S. (1990). Analysis of an identification algorithm arising in the adaptive estimation of Markov chains. *Mathematics of Control, Signals, and Systems 3*, 1–29.

Bakry, D., Milhaud, X., and Vandekerkhove, P. (1997). Statistique de chaînes de Markov cachées à espaces d'états fini. Le cas non stationnaire. *note C.R. Acad. Sci. Paris 325-I*, 203–206.

Baum, L. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist. 37*, 1554–1563.

Baum, L., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat. 41*, 164–171.

Bickel, P., Ritov, Y., and Rydén, T. (1998). Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist. 26*, 1614–1635.

Bickel, P. J. and Ritov, Y. (1996). Inference in hidden Markov models I : LAN in the stationary case. *Bernoulli 2*, 199–228.

Billingsley, P. (1965). *Ergodic theory and information*. New York : John Wiley & Sons Inc.

Birgé, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete 65*(2), 181–237.

Birgé, L. (2001). A new look at an old result : Fano's lemma. Technical report, Prépublications PMA.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc. 83*(404), 1184–1186.

Celeux, G., Diebolt, J., and Robert, C. (1993). Bayesian estimation of hidden Markov chains : a stochastic implementation. *Stat. Probab. lett. 16*, 77–83.

Chan, K. and Ledolter, J. (1995). Monte Carlo EM estimation in time series models involving counts. *J. Am. Statist. Assoc. 90*, 242–252.

Churchill, G. (1989). Stochastic models for heterogeneous DNA sequences. *Bull. Math. Biol. 51*, 79–94.

Cover, Thomas M. and Thomas, Joy A. (1991). *Elements of information theory*. New York : John Wiley & Sons Inc.

Davis, K. B. (1975). Mean square error properties of density estimates. *Ann. Statist. 3*, 1025–1030.

Davis, K. B. (1977). Mean integrated square error properties of density estimates. *Ann. Statist. 5*, 530–535.

DeJong, P. and Shephard, N. (1995). The simulation smoother for time series models. *Biometrika 82*, 339–350.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. Discussion. *J. Royal Statist. Soc. 39-B*, 1–38.

Devroye, L. (1989). Consistent deconvolution in density estimation. *Canad. J. Statist. 17*(2), 235–239.

Douc, R., Moulines, E., and Rydèn, T. (2001). Asymptotic properties of the maximum likelihood estimator in autoregressive models with markov regime. Technical report. submitted.

Durbin, J. and Koopman, S. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika 84*, 669–684.

Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A 53*(1), 97–110.

Fan, J. (1991b). Global behavior of deconvolution kernel estimates. *Statist. Sinica 1*(2), 541–551.

Fan, J. (1991c). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist. 19*(3), 1257–1272.

Fan, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist. 21*(2), 600–610.

Fano, R. M. (1952). *Class notes for Transmission of Information.*

Feller, W. (1971). *An introduction to probability theory and its applications. Vol. II.* New York : John Wiley & Sons Inc.

Fredkin, D. and Rice, J. (1987). Correlation functions of a function of a finite-state Markov process with application to channel kinetics. *Math. Biosci. 87*, 161–172.

Gill, R. D. and Levit, B. Y. (1995). Applications of the Van Trees inequality : a Bayesian Cramér-Rao bound. *Bernoulli 1*(1-2), 59–79.

Golubev, G. K. and Levit, B. Y. (1996). Asymptotically efficient estimation for analytic distributions. *Math. Methods Statist. 5(3)*, 357–368.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application.* New York : Academic Press Inc. [Harcourt Brace Jovanovich Publishers]. Probability and Mathematical Statistics.

Ibragimov, I. A. and Has′minskiĭ, R. Z. (1981). *Statistical estimation.* New York : Springer-Verlag. Asymptotic theory, Translated from the Russian by Samuel Kotz.

Ibragimov, I. A. and Hasminskii, R. Z. (1983). Estimation of distribution density. *J. Soviet. Math. 25*, 40–57.

Ishwaran, H. (1999). Information in semiparametric mixtures of exponential families. *Ann. Statist. 27*(1), 159–177.

Jensen, J. L. and Petersen, N. V. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Statist. 27*(2), 514–535.

Juang, B. and Rabiner, L. (1991). Hidden Markov models for speech recognition. *Technometrics 33*, 251–272.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist. 27*, 887–906.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility : Likelihood inference and comparison with ARCH models. *Review of Economic studies 65*, 361–394.

Le Gland, F. and Mevel, L. (2000a). Basic properties of the projective product with application to products of column-allowable nonnegative matrices. *Math. Control Signals Systems 13*(1), 41–62.

Le Gland, F. and Mevel, L. (2000b). Exponential forgetting and geometric ergodicity in hidden Markov models. *Math. Control Signals Systems 13*(1), 63–93.

Leroux, B. (1992). Maximum-likelihood estimation for hidden Markov models. *Stochastic processes and their applications 40*, 127–143.

Lindsay, B. G. (1986). Exponential family mixture models (with least-squares estimators). *Ann. Statist. 14*(1), 124–137.

Liu, M. C. and Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist. 17*(4), 427–438.

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series*. London : Chapman & Hall.

Matias, C. and Taupin, M.-L. (2001). Estimation of some linear functionals in deconvolution models. Manuscript, Université Paris Sud.

Medgyessy, P. (1973). Decomposition of superposition of density functions on discrete distributions. II. *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl. 21*, 261–382.

Mevel, L. (1997). *Statistique asymptotique pour les modèles de Markov cachés*. PhD Thesis, Université de Rennes I.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov chains and stochastic stability*. London : Springer-Verlag London Ltd.

Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist. 7*(4), 307–330.

Nikol′skii, S. M. (1969). *Priblizhenie funktsii mnogikh peremennykh i teoremy vlozheniya*. Izdat. "Nauka", Moscow.

Pagano, M. (1974). Estimation of models of autoregressive signal plus white noise. *Ann. Statist. 2*, 99–108.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist. 27*(6), 2033–2053.

Pinelis, I. (1995). Optimum bounds on moment of sum of independent random vectors. *Sib. Adv . Math. 5*, 141–150.

Prohorov, Y. V. (1956). Convergence of random processes and limit theorems in probability theory. *Teor. Veroyatnost. i Primenen. 1*, 177–238.

Quinn, B. G. (1996). Statistical problems in the analysis of underwater sound. In *Athens Conference on Applied Probability and Time Series Analysis, Vol. II (1995)*, pp. 324–338. New York : Springer.

Rosenthal, H. P. (1970). On the subspaces of l^p, $(p > 2)$, spanned by sequences of independent random variables. *Israel J. Math. 8*, 273–303.

Seneta, E. (1981). *Non-negative matrices and Markov chains.* New York : Springer-Verlag.

Stefanski, L. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics 21*(2), 169–184.

Stefanski, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett. 9*(3), 229–235.

Sunyach, C. (1975). Une classe de chanes rcurrentes sur un espace mtrique complet. *Annales de l'Institut Poincar XI*, 325–343.

Taupin, M. L. (1998). *Semi-parametric estimation in the nonlinear errors-in-variables model.* Ph. D. thesis, Université Paris-Sud, 91405 Orsay France.

Taupin, M. L. (2001). Semi-parametric estimation in the non-linear errors-in-variables model. *Ann. Statist. 29*(1), 66–93.

van der Vaart, A. W. (1998). *Asymptotic statistics.* Cambridge : Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes.* New York : Springer-Verlag. With applications to statistics.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist. 20*, 595–602.

Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist. 18*(2), 806–831.