Déconvolution adaptative de densité sur $\mathbb{R}^+$

Gwennaëlle Mabon

CREST
3 avenue Pierre Larousse
92245 Malakoff, France
gwennaelle.mabon@ensae.fr
MAP5, Université Paris Descartes
45 rue des Saints-Pères
75006 Paris, France

Résumé. On considère le problème d'estimation adaptative de densité dans le modèle de convolution : $Z = X + Y$ où X et Y sont des variables indépendantes positives. Le but est d'estimer la densité de X à partir de n observations de Z, en supposant la loi de Y connue. Ce problème de déconvolution, classique en statistique non-paramétrique, a été traité en utilisant une approche Fourier. Cependant dans ce travail les variables aléatoires ont la particularité d'être distribuées sur $\mathbb{R}^+$. Sachant cela, nous proposons un nouvel angle d'attaque en construisant un estimateur par projection sur la base de Laguerre. Nous présentons une majoration du risque quadratique intégré de cet estimateur. Enfin nous décrivons aussi une stratégie d'estimation adaptative pour sélectionner un espace de projection pertinent.

Mots-clés. Estimation non-paramétrique de densité. Estimation adaptative. Base de Laguerre. Déconvolution. Risque quadratique intégré.

Abstract. We consider the problem of adaptive density estimation in an additive model defined by $Z = X + Y$ with X independent of Y, when both random variables are nonnegative. We want to recover the density of X through n observations of Z, assuming that the distribution of Y is known. This issue can be seen as the classical statistical problem of deconvolution which has been tackled in many cases using Fourier-type approaches. Nonetheless, in the present case the random variables have the particularity to be $\mathbb{R}^+$ supported. Knowing that, we propose a new angle of attack by building a projection estimator with an appropriate Laguerre basis. We present upper bounds on the mean squared integrated risk of the density function estimator. We then describe a nonparametric adaptive strategy for selecting a relevant projection space.

Keywords. Nonparametric density estimation. Adaptive estimation. Laguerre basis. Deconvolution. Mean squared risk.

1 Introduction

Considérons le modèle de déconvolution de densité

$$Z_i = X_i + Y_i, \quad i = 1, \dots, n, \tag{1}$$

les variables X_i, de densité f inconnue, et Y_i, de densité g connue, ont la particularité d'être distribuées sur $\mathbb{R}^+$. Nous supposons que les X_i et Y_i sont deux séquences i.i.d. mutuellement indépendantes. L'objectif est d'estimer f à partir des observations Z_i de densité h.

Le problème de déconvolution de densité a été largement étudié dans la littérature. Les vitesses de convergence et leur optimalité ont été étudiées, par exemple, par Caroll et Hall (1988), Stefanski et Carroll (1990), Fan (1991), Efromovich (1997) pour des estimateurs à noyaux, Butucea (2004), Butucea et Tsybakov (2008) pour l'optimalité au sens minimax ; étendu à bruit inconnu notamment par Comte et Lacour (2011), Johannes (2009), Kappus et Mabon (2014). Ces travaux supposent que les variables sont distribuées sur $\mathbb{R}$ et sont encore valides dans notre cadre. Cependant dans ce travail, le but est de proposer une solution spécifique pour des variables positives. Les variables aléatoires positives apparaissent notamment dans les modèles d'actuariat ou d'assurance.

Dans ce papier, nous définissons un estimateur par projection de f sur la base de Laguerre et présentons une majoration du risque intégré de cet estimateur. Ensuite nous établissons une méthode de type sélection de modèle pour choisir un espace de projection pertinent. Ce dernier résultat est nouveau dans la littérature. Ajoutons que la méthode proposée peut être étendue à l'estimation de fonction de survie (qui ne consiste pas à intégrer l'estimateur de la densité) ainsi qu'à l'estimation de densité dans le cas d'observation directe. La méthodologie complète est présentée dans Mabon (2014), de même que certaines vitesses de convergence et une illustration numérique des résultats.

2 Procédure d'estimation

Notations. Pour deux fonctions réelles φ et ψ appartenant à $\mathbb{L}^2(\mathbb{R})$, on note $\|\varphi\|$ la norme $\mathbb{L}^2$ de φ définie par $\|\varphi\|^2 = \int_{\mathbb{R}} |\varphi(x)|^2 dx$, $\langle \varphi, \psi \rangle$ le produit scalaire entre φ et ψ défini par $\langle \varphi, \psi \rangle = \int_{\mathbb{R}} \varphi(x)\psi(x)dx$. Soit d un entier, pour deux vecteurs $\vec{u}$ et $\vec{v}$ appartenant à $\mathbb{R}^d$, on note $\|\vec{u}\|_{2,d}$ la norme euclidienne définie par $\|\vec{u}\|_{2,d}^2 = {}^t\vec{u}\vec{u}$ où ${}^t\vec{u}$ est le vecteur transposé de $\vec{u}$. Le produit scalaire entre $\vec{u}$ et $\vec{v}$ est $\langle \vec{u}, \vec{v} \rangle_{2,d} = {}^t\vec{u}\vec{v} = {}^t\vec{v}\vec{u}$.

Base de Laguerre. Il s'agit d'une base orthonormale de fonctions $(\varphi_k)_{k\in\mathbb{N}}$ de $\mathbb{L}^2(\mathbb{R}^+)$

$$\varphi_k(x) = \sqrt{2}e^{-x}L_k(2x), \qquad \text{avec} \quad L_k(x) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} \frac{x^j}{j!}, \quad k \in \mathbb{N}, \quad x \geq 0, \tag{2}$$

où les L_k sont les polynômes de Laguerre. La base de Laguerre vérifie

$$\int_0^x \varphi_k(u)\varphi_j(x-u)\,\mathrm{d}u = 2^{-1/2}\left(\varphi_{k+j}(x) - \varphi_{k+j+1}(x)\right) \tag{3}$$

(voir formule 22.13.14, Abramowitz et Stegun (1964)), de même que

$$\sup_{x\in\mathbb{R}^+} |\varphi_k(x)| = \|\varphi_k\|_\infty \leq \sqrt{2}. \tag{4}$$

Notons $\mathcal{S}_m = \mathrm{Span}\{\varphi_0, \ldots, \varphi_{m-1}\}$ ainsi que pour une fonction p de $\mathbb{L}^2(\mathbb{R}^+)$

$$p(x) = \sum_{k\geq 0} b_k(p)\varphi_k(x) \quad \text{avec} \quad b_k(p) = \int_{\mathbb{R}^+} p(u)\varphi_k(u)\,\mathrm{d}u.$$

Estimateur par projection de la densité. Considérons l'hypothèse suivante

(A1) f et $g \in \mathbb{L}^2(\mathbb{R}^+)$.

Comme X et Y sont des variables positives indépendantes, nous avons la relation de convolution suivante

$$h(x) = \int_0^x f(u)g(x-u)\,\mathrm{d}u \tag{5}$$

or d'après **(A1)**, f et g admettent une décomposition sur la base de Laguerre, d'où

$$h(x) = \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} b_k(f)b_j(g)\int_0^x \varphi_k(u)\varphi_j(x-u)\,\mathrm{d}u. \tag{6}$$

Ainsi en écrivant h comme $\sum_{k=0}^{\infty} b_k(h)\varphi_k(x)$ et en appliquant (3) à (6), nous obtenons la relation

$$\sum_{k=0}^{\infty} b_k(h)\varphi_k(x) = \sum_{k=0}^{\infty}\varphi_k(x)\left(2^{-1/2}b_k(f)b_0(g) + \sum_{l=0}^{k-1} 2^{-1/2}\left(b_{k-l}(g) - b_{k-l-1}(g)\right)b_l(f)\right).$$

Nous avons finalement un système d'équations linéaires triangulaire infini. Nous pouvons alors écrire pour tout entier m : $\vec{h}_m = \mathbf{G}_m\vec{f}_m$ avec $\vec{h}_m$ et $\vec{f}_m$ des vecteurs de taille m de coordonnées respectives $b_k(f)$ et $b_k(h)$, $k = 0, 1, \ldots, m-1$, et $\mathbf{G}_m$ une matrice de Toeplitz triangulaire inférieure avec les éléments suivants

$$[\mathbf{G}_m]_{ij} = \begin{cases} 2^{-1/2}b_0(g) & \text{si } i = j, \\ 2^{-1/2}\left(b_{i-j}(g) - b_{i-j-1}(g)\right) & \text{si } j < i, \\ 0 & \text{sinon,} \end{cases} \tag{7}$$

Le principe de l'estimation par projection est de réduire l'estimation de f à celle de f_m, la projection de f sur $\mathcal{S}_m$, définie par $f_m(x) = \sum_{k=0}^{m-1} b_k(f)\varphi_k(x)$. Les $b_k(f)$ peuvent être retrouvés si la matrice $\mathbf{G}_m$ est inversible. Puisque $\mathbf{G}_m$ est une matrice triangulaire, elle est inversible si et seulement si tous ses coefficients diagonaux sont non nuls, i.e. $b_0(g)$ est non nul. Or par définition

$$b_0(g) = \int_{\mathbb{R}^+} g(u)\varphi_0(u)\,\mathrm{d}u = \sqrt{2}\int_{\mathbb{R}^+} g(u)e^{-u}\,\mathrm{d}u = \sqrt{2}\mathbb{E}[e^{-Y}] > 0.$$

Par conséquence $\mathbf{G}_m$ est inversible. Nous pouvons alors écrire que $\mathbf{G}_m^{-1}\vec{h}_m = \vec{f}_m$. Comme $b_k(h) = \mathbb{E}[\varphi_k(Z_1)]$, la projection de f sur $\mathcal{S}_m$ est estimée par

$$\hat{f}_m(x) = \sum_{k=0}^{m-1} \hat{b}_k(f)\varphi_k(x) \quad \text{avec} \quad \hat{\vec{f}}_m = \mathbf{G}_m^{-1}\hat{\vec{h}}_m \quad \text{et} \quad \hat{b}_k(h) = \frac{1}{n}\sum_{i=1}^{n}\varphi_k(Z_i) \tag{8}$$

avec $\hat{\vec{h}}_m = (\hat{b}_0(h), \ldots, \hat{b}_{m-1}(h))$ et $\hat{\vec{f}}_m = (\hat{b}_0(f), \ldots, \hat{b}_{m-1}(f))$.

Majoration du risque intégré.

Proposition 2.1 *Si **(A1)** est vraie, pour $\mathbf{G}_m$ définie par (7) et $\hat{f}_m$ défini par (8), alors*

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{2m}{n}\varrho^2(\mathbf{G}_m^{-1}). \tag{9}$$

où $\varrho^2(\mathbf{G}_m^{-1})$ est la plus grande valeur propre en valeur absolue de ${}^t\mathbf{G}_m^{-1}\mathbf{G}_m^{-1}$.

Les deux termes à droite de l'Équation (9) correspondent à un terme de biais qui diminue avec m et à un terme de variance qui augmente avec m. Une étude détaillée des vitesses de convergence résultantes est menée dans Mabon (2014). Des sous-espaces de $\mathbb{L}^2(\mathbb{R}^+)$ dénommés Laguerre-Sobolev (cf Bongioanni et Torrea (2009)) permettent de déterminer l'ordre du biais. Des conditions données par Comte *et al.* (2013) sur la densité g permettent d'évaluer l'ordre du rayon spectrale et par la même celui de la variance. Ainsi la vitesse de convergence d'un estimateur $\hat{f}_{m_n}$ sur $\mathcal{S}_{m_n}$ (pour m_n bien choisi) est obtenue.

3 Sélection de modèle

Ajoutons deux hypothèses supplémentaires :

(A2) $\mathcal{M}_n^{(1)} = \{1 \leq m \leq d_1, m\varrho^2\left(\mathbf{G}_m^{-1}\right) \leq n\}$, où $d_1 < n$ peut dépendre de n.

(A3) $\forall b > 0$, $\sum_{m\in\mathcal{M}_n^{(1)}} \varrho^2\left(\mathbf{G}_m^{-1}\right) e^{-bm} < C(b) < \infty$.

De plus pour $m \in \mathcal{M}_n^{(1)}$, définissons les sous-espaces associés $\mathcal{S}_{d_1}^m \subseteq \mathbb{R}^{d_1}$

$$\mathcal{S}_{d_1}^m = \left\{\vec{t}_{m,d_1} \in \mathbb{R}^{d_1} \,/\, \vec{t}_{m,d_1} = {}^t(b_0(t), b_1(t), \ldots, b_{m-1}(t), 0, \ldots, 0)\right\}.$$

Cet espace est introduit pour avoir des modèles emboîtés. Ainsi, lorsque la dimension passe de m à $m+1$, seul le $m+1$-*ème* coefficient doit être calculer. Nous posons alors le contraste suivant pour tout $\vec{t} \in \mathbb{R}^{d_1}$

$$\gamma_n(\vec{t}) = \|\vec{t}\|^2_{2,d_1} - 2\langle \vec{t}, \mathbf{G}_{d_1}^{-1}\widehat{\vec{h}}_{d_1}\rangle_{2,d_1}.$$

Remarquons que pour $\vec{t}_{m,d_1} \in \mathcal{S}^m_{d_1}$, grâce aux coordonnées nulles de $\vec{t}_{m,d_1}$ et à la forme triangulaire inférieure de $\mathbf{G}_{d_1}$ et $\mathbf{G}_m$, nous avons

$$\langle \vec{t}_{m,d_1}, \mathbf{G}_{d_1}^{-1}\widehat{\vec{h}}_{d_1}\rangle_{2,d_1} = \langle \vec{t}_m, \mathbf{G}_m^{-1}\widehat{\vec{h}}_m\rangle_{2,m} = \langle \vec{t}_m, \hat{\vec{f}}_m\rangle_{2,m}.$$

Ainsi, $\hat{\vec{f}}_m = \underset{\vec{t}_{m,d_1}\in\mathcal{S}^m_{d_1}}{\operatorname{argmin}}\ \gamma_n(\vec{t}_{m,d_1})$. Nous définissons la pénalité

$$\operatorname{pen}_1(m) = \frac{2\kappa_1 m}{n}\varrho^2\left(\mathbf{G}_m^{-1}\right) \tag{10}$$

où κ_1 est une constante numérique (voir commentaires plus bas).

Théorème 3.1 *Supposons que* ***(A1)**-**(A3)*** *sont vraies. Soit* $\hat{f}_{\widehat{m}_1}$ *défini par (8) et*

$$\widehat{m}_1 = \underset{m\in\mathcal{M}_n^{(1)}}{\operatorname{argmin}} \left\{\gamma_n(\widehat{\vec{f}}_m) + \operatorname{pen}_1(m)\right\}$$

avec pen_1 *définie par (10), alors il existe une constante numérique positive* κ_1 *telle que*

$$\mathbb{E}\|f - \hat{f}_{\widehat{m}_1}\|^2 \leq 4 \inf_{m\in\mathcal{M}_n^{(1)}} \left\{\|f - f_m\|^2 + \operatorname{pen}_1(m)\right\} + \frac{C}{n}, \tag{11}$$

où C *dépend de* $\|f\|$ *et* $\|g\|$*.*

D'après la preuve (voir Mabon (2014)), $\kappa_1 = 32$ conviendrait. En pratique les valeurs théoriques sont généralement trop grandes et la constante est calibrée à partir de simulations préliminaires. Une fois choisie, elle reste fixée pour les simulations ultérieures.

L'inégalité oracle (11) établit une majoration oracle non asymptotique. Cela montre que le compromis biais variance est automatiquement réalisé à une constante multiplicative près. Ainsi les vitesses de convergence sont atteintes sans avoir à spécifier le cadre.

Il est courant dans la littérature de supposer que les densités de probabilité appartiennent à un certain modèle semi-paramétrique, ce qui n'est pas le cas ici. Dans le cadre du problème de déconvolution avec une approche Fourier, des articles comme Comte *et al.* (2006), supposent que les transformées de Fourier de f et g ont un certain comportement. Dans ce travail seules les conditions **(A2)** et **(A3)**, peu contraignantes, sont supposées sur le comportement de la norme spectrale de $\mathbf{G}_m^{-1}$. En effet, si l'hypothèse **(A3)** est relâchée alors une procédure adaptative peut toujours être obtenue, à une perte logarithmique près, avec la pénalité $\operatorname{pen}(m) = 2\kappa m\varrho^2\left(\mathbf{G}_m^{-1}\right)\log(n)/n$.

À notre connaissance, il s'agit d'un nouveau résultat, et les extensions à l'estimation de la survie que nous proposons sont également originales et intéressantes.

Bibliographie

[1] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York, ninth dover printing, tenth gpo printing edition.
[2] Bongioanni, B. and Torrea, J. L. (2009). What is a Sobolev space for the Laguerre function systems? *Studia Math.*, 192(2):147–172.
[3] Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *The Canadian Journal of Statistics*, 32(2):181–192.
[4] Butucea, C. and Tsybakov, A. (2008). Sharp optimality in density deconvolution with dominating bias I. *Theory Proba. Appl.*, 52(1):24–39.
[5] Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186.
[6] Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society: Series B*, 73:601–627.
[7] Comte, F., Cuenod, C.-A., Pensky, M., and Rozenholc, Y. (2013). Laplace deconvolution and its application to dynamic contrast enhanced imaging. MAP5 2012-17.
[8] Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics*, (34):431–452.
[9] Efromovich, S. (1997). Density estimation for the case of supersmooth measurement errors. *Journal of the American Statistical Association*, 92:526–535.
[10] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19(3):1257–1272.
[11] Johannes, J. (2009). Deconvolution with unknown error distribution. *The Annals of Statistics*, 37(5a):2301–2323.
[12] Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution. *Electronic Journal of Statistics*. 8(2) : 2879-2904
[13] Mabon, G. (2014). Adaptive deconvolution on the nonnegative real line. MAP5 2014-33.
[14] Stefanski, S. and Carroll, R. (1990). Deconvoluting kernel density estimators. *Statistics*, 21:169–184.