

Estimation de la fonction de répartition : revue bibliographique

Rémi Servien¹

Title

Distribution function estimation : a review

Résumé

L'estimation de la fonction de répartition d'une variable aléatoire est un volet important de l'estimation non paramétrique. De nombreuses méthodes ont été proposées et étudiées afin de modifier efficacement l'outil brut qu'est la fonction de répartition empirique. Dans cet article, nous effectuons un point bibliographique sur les différentes méthodes envisagées dans le cas de variables aléatoires réelles.

Mots-clés : Fonction de répartition, Estimation non paramétrique, Efficacité des estimateurs.

Abstract

The estimation of the distribution function of a real random variable is an important topic in non parametric estimation. A number of methods have been proposed and studied to improve the efficiency of the raw empirical distribution function in a broad variety of context. The present paper aims at giving an overview of these methods.

Keywords : Distribution function, Non parametric estimation, Efficiency of estimators

Mathematics Subject Classification: (62G05)

¹Institut de Mathématiques et de Modélisation de Montpellier, UMR CNRS 5149, Equipe de Probabilités et Statistique, Université Montpellier II, Cc 051 – Place Eugène Bataillon, 34095 Montpellier Cedex 5 France
rservien@math.univ-montp2.fr

1 Introduction

Un problème récurrent en statistique est celui de l'estimation d'une densité f ou d'une fonction de répartition F à partir d'un échantillon de variables aléatoires réelles X_1, X_2, \dots, X_n indépendantes et de même loi inconnue. Les fonctions f et F , tout comme la fonction caractéristique, décrivent complètement la loi de probabilité des observations et en connaître une estimation convenable permet de résoudre nombre de problèmes statistiques. Cette estimation tient donc naturellement une place importante dans l'étude de nombreux phénomènes de nature aléatoire. Elle peut être menée, sous des hypothèses restrictives, à l'aide de techniques paramétriques comme la méthode des moments ou celle du maximum de vraisemblance. Les approches non paramétriques que nous privilégions ici sont plus flexibles et constituent toujours un complément utile, même lorsque certains modèles paramétriques semblent s'imposer.

Même si les fonctions de répartition et de densité caractérisent toutes les deux la loi de probabilité d'une variable, la densité a un net avantage sur le plan visuel. Elle permet d'avoir un aperçu très rapide des principales caractéristiques de la distribution (pics, creux, asymétries, ...), ce qui explique le volume important de littérature qui lui est consacré. La fonction de répartition contient bien sûr cette information mais de manière moins visible. Néanmoins c'est en terme de comportement local de la fonction de répartition que s'explique le plus facilement le comportement des estimateurs fonctionnels (vitesse de convergence, normalité asymptotique) et c'est finalement par un estimateur de la fonction de répartition que l'on passe pour estimer des probabilités d'ensembles : la probabilité qu'une variable se cantonne dans un intervalle donné ou qu'une observation au moins d'un nouvel échantillon dépasse un seuil fixé. Lorsqu'on veut donner une borne inférieure pour la probabilité qu'un paramètre θ inconnu appartienne à un intervalle de la forme $[\theta_n - \varepsilon, \theta_n + \varepsilon]$, où θ_n est un estimateur de θ , on a en fait besoin d'un estimateur de la fonction de répartition de θ_n .

Il est vrai que l'on peut souvent passer d'un estimateur de f à un estimateur de F par intégration et d'un estimateur de F à un estimateur de f par dérivation. Néanmoins une particularité est à souligner : c'est l'existence de la fonction de répartition empirique F_n , fonction de répartition de la loi empirique associée à l'échantillon. Il n'y a bien sûr pas d'équivalent pour la densité, ce qui différencie la nature de chacun des deux problèmes d'estimation. Il est important de noter que la fonction de répartition empirique ne fait appel à aucune structure algébrique ou topologique mais seulement à des notions ensemblistes et que les techniques d'estimation ne sont pas autre chose que des techniques de régularisation de cette référence empirique dont la donnée est équivalente à celle de l'échantillon.

Cette dernière remarque conduit donc à collecter en premier lieu les principaux résultats disponibles sur la fonction de répartition empirique. Ils constituent l'entrée en matière de la présente revue bibliographique. Puis nous passerons dans les sections suivantes à des estimateurs introduits plus spécifiquement pour la fonction de répartition. La section 3 sera consacrée au lissage local, dont un exemple bien connu est le lissage polynômial local. La méthode de lissage local décrite permet de donner un cadre général à l'estimation de fonctionnelles de la fonction de répartition et de leurs dérivées. Un cas particulier est la méthode du noyau dont l'application à la fonction de répartition est abordée dans la section 4. Les estimateurs à noyau sont, avec les estimateurs splines décrits en section 5, les méthodes de lissage les plus communément utilisées dans les logiciels statistiques. Les sections suivantes abordent des approches plus récentes : les Support Vector Machines (S.V.M.) dans la section 6, le level-crossing

dans la section 7 et les Systèmes de Fonctions Itérées (I.F.S.) dans la section 8. Nous mentionnons rapidement en section 9 quelques méthodes développées pour la densité et utilisables par intégration, comme celle des fonctions orthogonales, particulièrement celles des ondelettes. La section 10 traite du cas de la fonction de répartition d'une loi conditionnelle. Enfin, la prise en compte d'un biais éventuel sur les données est abordée en section 11.

2 Un estimateur naturel : la fonction de répartition empirique

La fonction de répartition empirique sera notée

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{]-\infty, x]}(X_i)$$

où

$$I_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon.} \end{cases}$$

Afin de passer en revue quelques résultats importants concernant cette fonction, nous notons $Z_n(x) = \sqrt{n}(F_n(x) - F(x))$ et définissons les statistiques suivantes

$$D_n^+ = \sup_{x \in \mathbb{R}} Z_n(x), \quad D_n^- = \sup_{x \in \mathbb{R}} (-Z_n(x)) \quad \text{et} \quad D_n = \sup_{x \in \mathbb{R}} |Z_n(x)|.$$

Kolmogorov [63] et Smirnov [88] introduisent et étudient ces trois statistiques et démontrent que leur distribution ne dépend pas de F . En outre, à l'aide des théorèmes de Donsker [32] et Doob [33, 34], il est prouvé que les statistiques D_n^+ et D_n^- ont la même loi et que nous avons les résultats asymptotiques suivants :

$$\lim_{n \rightarrow \infty} P(D_n^- > \lambda) = \exp(-2\lambda^2),$$

$$\lim_{n \rightarrow \infty} P(D_n > \lambda) = 2 \sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2\lambda^2).$$

Par la suite, Dvoretzky, Kiefer et Wolfowitz [35] déterminent une borne de la forme

$$P(D_n^- > \lambda) \leq C \exp(-2\lambda^2),$$

où C est une constante indéterminée. L'ensemble de cette démarche et de ces résultats sont analysés dans Hennequin et Tortrat [56]. Pour des propriétés liées aux statistiques d'ordre et de rang on pourra consulter le livre de Caperaa et Van Cutsem [20].

De nombreux auteurs essayent ensuite de trouver la meilleure constante C dans l'inégalité précédente. Devroye et Wise [30], Shorack et Wellner [86] ou Hu [58] font peu à peu diminuer cette constante. Finalement, Massart [68] démontre qu'il est possible de prendre $C = 1$ pourvu que

$$\lim_{n \rightarrow +\infty} P(D_n^- > \lambda) = \exp(-2\lambda^2) \leq 1/2.$$

Il montre aussi que, quel que soit λ ,

$$P(D_n > \lambda) \leq 2 \exp(-2\lambda^2),$$

et que ces bornes ne peuvent plus être améliorées, ces inéquations étant valides que F soit continue ou pas.

D'autre part, le théorème de Glivenko-Cantelli nous donne la convergence uniforme presque sûre de F_n vers F c'est-à-dire

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0 \text{ p.s.}$$

Pour des variables dépendantes (plus précisément ϕ -mélangeantes), ce résultat est amélioré plus tard par Collomb, Hassani, Sarda et Vieu [24] en convergence presque complète sous certaines hypothèses, notamment l'uniforme continuité de f .

F_n est également l'estimateur non paramétrique du maximum de vraisemblance (Kiefer et Wolfowitz [62], Efron et Tibshirani [37]) et, en général, une transformée $t(F)$ a pour estimateur du maximum de vraisemblance $t(F_n)$. D'autre part, parmi les estimateurs sans biais de $F(x)$, $F_n(x)$ est également l'unique estimateur de variance minimale (Yamato [101], Lehmann [66]). Cette dernière vaut par ailleurs $F(x)(1 - F(x))/n$. Yamato [101] élargit les recherches à une famille de fonctions englobant F_n . Il étudie les fonctions du type

$$F_n^*(x) = \frac{1}{n} \sum_{j=1}^n W_n(x - X_j)$$

où W_n est une fonction de répartition connue. Il démontre que, en tout point de continuité x de F , $F_n^*(x)$ est asymptotiquement non biaisé et

$$P\left[\sup_{-\infty < x < \infty} |F_n^*(x) - F(x)| \rightarrow 0 \right] = 1$$

si et seulement si $W_n \rightarrow e_0$ avec

$$e_0(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

Il obtient de plus la convergence vers une loi normale $\mathcal{N}(0, 1)$ de la distribution de

$$\frac{\sqrt{n}[F_n(x) - EF_n(x)]}{\sqrt{F(x)[1 - F(x)]}}.$$

Pour évaluer l'écart entre F et son estimateur, plusieurs autres critères peuvent être utilisés. Ainsi, Devroye et Györfi [28] choisissent comme critère la variation totale V qui se définit comme suit entre deux mesures de probabilité μ et ν

$$V(\mu, \nu) = \sup_A |\mu(A) - \nu(A)|,$$

avec le supremum calculé sur tous les boréliens A , et l'entropie relative I telle que

$$I(\mu, \nu) = \sup_{\{A_j\}} \sum_j \mu(A_j) \log \frac{\mu(A_j)}{\nu(A_j)}$$

où le supremum est pris sur l'ensemble des partitions finies en boréliens mesurables $\{A_j\}$ de \mathbb{R} . Au sens de ces critères, F_n est un mauvais estimateur. En effet, ils obtiennent $V(\mu, \mu_n) = 1$ et $I(\mu, \mu_n) = \infty$ pour des mesures non-atomiques. Mais ils démontrent ensuite la non-existence de bons estimateurs F_n^* , de mesure associée μ_n^* , tels que

$$V(\mu, \mu_n^*) \rightarrow 0 \text{ ou } I(\mu, \mu_n^*) \rightarrow 0,$$

qui rend ces critères non adaptés.

Aggarwal [2] introduit le problème consistant à déterminer le meilleur estimateur invariant de F , au sens des transformations monotones, avec la fonction de perte

$$L(F, a) = \int \{F(t) - a(t)\}^2 h(F(t)) dF(t).$$

L'éventuelle admissibilité, c'est-à-dire la minimisation de L , et minimaxité de l'estimateur sont également des questions importantes. Dans le cas particulier de fonction de perte de Cramér-von Mises, où $h(t) = t^{-1}(1-t)^{-1}$, F_n est le meilleur estimateur invariant. Aggarwal [2], Brown [18] et Yu [102] démontrent la non admissibilité de F_n pour $h(t) = t^\alpha(1-t)^\beta$, $\alpha, \beta \geq -1$ alors qu'il est admissible pour

$$L(F, a) = \int \{F(t) - a(t)\}^2 F(t)^\alpha (1 - F(t))^\beta dW(t),$$

avec $-1 \leq \alpha, \beta \leq 1$ et W une mesure finie (Cohen et Kuo [23]). Dvoretzky, Kiefer et Wolfowitz [35] et Phadia [76] prouvent que F_n est asymptotiquement minimax pour une grande variété de fonctions de pertes. Néanmoins, pour celle de Kolmogorov-Smirnov, Friedman, Gelman et Phadia [49] déterminent le meilleur estimateur invariant de F . Celui-ci est une fonction en escalier différente de F_n .

D'autres fonctions en escalier ont également été étudiées. Ainsi, si F est continue sur $[0, 1]$, nous savons qu'il existe un estimateur \tilde{F}_n linéaire par morceaux et tel que la distribution de $\sqrt{n}[\tilde{F}_n(x) - F(x)]$ converge faiblement vers un pont brownien $B(x)$ connu (Billingsley [14]). Beran [7] s'appuie sur ce résultat pour définir un nouvel estimateur \hat{F}_n tel que

$$\hat{F}_n = \begin{cases} H(x) & \text{si } \sup_x |\tilde{F}_n(x) - H(x)| \leq n^{-1/4} \\ \tilde{F}_n(x) & \text{sinon,} \end{cases}$$

où H est une fonction de répartition quelconque sur $[0, 1]$. Il obtient alors que, hormis le cas trivial où la distribution de l'échantillon est H , $\sqrt{n}[\hat{F}_n(x) - F(x)]$ converge vers $B(x)$. De plus, il démontre que tout estimateur \hat{F}_n régulier peut être représenté comme une convolution d'un pont brownien avec une autre fonction de répartition de $C[0, 1]$ dépendant uniquement de la densité f .

La fonction de répartition empirique ne tient pas compte d'une éventuelle information que nous pouvons avoir sur la fonction à estimer. Modarres [69] utilise une possible symétrie pour

bâtir un nouvel estimateur à partir de F_n :

$$\hat{F}^s(x) = \frac{1}{2}(F_n(x) + 1 - F_n(-x)) \quad \text{pour } x < 0$$

et démontre que cet estimateur est l'estimateur du maximum de vraisemblance sous une hypothèse de symétrie. Il bâtit ensuite un nouvel estimateur \hat{F}^{aux} en utilisant de l'information amenée par une covariable Y . Un échantillon auxiliaire $(Y_i)_{1 \leq i \leq m}$ avec $m \gg n$ permet de connaître la fonction de répartition empirique de la variable Y . En collectant ensuite un échantillon $Z_i = (X_i, Y_i)_{1 \leq i \leq n}$ et en maximisant sa vraisemblance, il obtient l'estimateur \hat{F}^{aux} . Cet estimateur est relativement simple et ne fait intervenir qu'un comptage des X_i dans certaines zones. En croisant les 2 modèles, il obtient l'estimateur \hat{F}^h suivant

$$\hat{F}^h(x) = \frac{1}{2}(\hat{F}^{aux}(x) + 1 - \hat{F}^{aux}(-x)) \quad \text{pour } x < 0,$$

montre sa normalité asymptotique et étudie ses propriétés. Il faut noter que cet estimateur ne maximise pas la vraisemblance de ce modèle et ne s'avère pas robuste envers l'hypothèse de symétrie. Enfin, il montre sur des simulations une amélioration du biais et de l'efficacité relativement à la fonction de répartition empirique.

L'estimateur de Grenander [53] est également obtenu directement à partir de la fonction de répartition empirique. Il est défini comme le plus petit majorant concave de F_n . Pour une fonction F_n de support réel $[a, b]$, F_{Gr} est la plus petite fonction concave telle que

$$F_{Gr}(t) \geq F_n(t) \text{ et } F_{Gr}(a) = 0, F_{Gr}(b) = 1.$$

Mais cet estimateur, bien qu'étant défini à partir de la fonction de répartition et sans aucun paramètre supplémentaire, est beaucoup plus souvent dérivé pour estimer une densité f qu'utilisé directement pour déterminer F . Dans le cas où la densité f est monotone, f_{Gr} l'est également et maximise la vraisemblance. Pour une étude plus approfondie de f_{Gr} nous renvoyons le lecteur à Devroye [27].

3 Estimateurs par lissage local

Afin d'obtenir une estimation plus régulière de la fonction de répartition, Berlinet [8] puis Lejeune et Sarda [67] lissent la fonction de répartition empirique dans différents espaces. Lejeune et Sarda utilisent la régression polynômiale locale. La minimisation de la norme pondérée L^2 débouche sur des choix optimaux que l'on retrouve parmi les estimateurs à noyaux décrits en section 4. Ce résultat est ensuite élargi par Abdous, Berlinet et Hengartner [1] qui proposent d'estimer différentes fonctionnelles $\phi(x, F)$ de la fonction F au point x . Pour cela ils substituent $\phi(x, F_n)$ à $\phi(x, F)$ et, dans le cas où $\phi(x, F)$ a r dérivées continues, choisissent de minimiser le critère suivant

$$J(a_0, \dots, a_r; x) = \int \frac{1}{h} K\left(\frac{z-x}{h}\right) \left\{ \phi(z, F_n) - \sum_{k=0}^r \frac{a_k}{k!} (z-x)^k \right\}^2 dz.$$

Les dérivées successives de $\phi(x, F)$ sont estimées par les $\hat{a}_0(x), \hat{a}_1(x), \dots, \hat{a}_r(x)$ minimisant le critère $J(a_0, \dots, a_r; x)$ au point x . Une expression explicite est ensuite obtenue à l'aide d'une

fonction K , une densité sur $[-1,1]$, et $Q_0(z), \dots, Q_n(z)$ une base orthonormale de $L^2(K)$ de l'espace P_r des polynômes de degré au moins r . On définit

$$K^{[m,r]}(u) = \left(\sum_{k=0}^r Q_k(u) \frac{d^m}{dw^m} Q_k(w) \Big|_{w=0} \right) K(u)$$

et on obtient alors le minimiseur $\hat{a}_0(x), \hat{a}_1(x), \dots, \hat{a}_r(x)$ par

$$\hat{a}_m = \frac{1}{h^{m+1}} \int \phi_n(z) K^{[m,r]} \left(\frac{z-x}{h} \right) dz.$$

En considérant l'estimateur

$$\hat{\theta}_{n,h}^{(m)}(x) = \frac{1}{h^{m+1}} \int_{-\infty}^{\infty} \phi(z, F_n) K^{[m,r]} \left(\frac{z-x}{h} \right) dz$$

de $\theta^{(m)}(x) = \phi^{(m)}(x, F)$, ils démontrent ensuite sous certaines conditions que $\hat{\theta}_{n,h}^{(m)}(x)$ converge p.s. vers $\phi^{(m)}(x, F)$. Berlinet [9] et Berlinet et Thomas-Agnan [11] élargissent ce résultat à un espace V de Hilbert à noyau reproduisant à la place de P_r . L'estimation de la projection $\Pi_V(F)$ de F sur V est alors déterminée par les équations

$$F_x(hv) = \int \Pi_V(F(x+h.))(u) K(u, v) K_0(u) d\lambda(u)$$

et

$$h^m F_x^{(m)}(hv) = \int \Pi_V(F(x+h.))(u) \frac{d^m K(u, v)}{dv^m} K_0(u) d\lambda(u)$$

où K_0 est un noyau et K le noyau reproduisant de V .

4 Estimateur à noyau

L'estimateur à noyau de la densité

$$\tilde{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n k \left(\frac{x - X_i}{h} \right),$$

avec un noyau k intégrable et d'intégrale 1 et une fenêtre $h > 0$, est un estimateur non paramétrique bien connu de $f(x)$ introduit par Akaike [3], Rosenblatt [80] et Parzen [74]. La littérature qu'il a suscitée est considérable. Dans le présent paragraphe nous nous limitons à ses applications orientées vers l'estimation de la fonction de répartition. On définit l'estimateur \tilde{F}_n à noyau de F par

$$\tilde{F}_n(x) = \frac{1}{n} \sum_{i=1}^n K \left(\frac{x - X_i}{h} \right),$$

où

$$K(x) = \int_{-\infty}^x k(y) dy.$$

Ses propriétés sont connues depuis longtemps, par exemple sa convergence uniforme vers F avec f continue (Nadaraya [72], Winter [98], Yamato [101]) puis sans conditions sur f (Singh, Gasser et Prasad [87]) ou sa normalité asymptotique (Watson et Leadbetter [97]). Winter [99] démontre aussi qu'il vérifie la propriété de Chung-Smirnov, c'est-à-dire

$$\limsup_{n \rightarrow \infty} \left\{ \left(\frac{2n}{\log \log n} \right)^{(1/2)} \sup_{-\infty < t < \infty} |\tilde{F}_n(t) - F(t)| \right\} \leq 1$$

avec probabilité 1. Azzalini [4] trouve une expression asymptotique pour l'erreur quadratique moyenne ou M.S.E. $(E(\tilde{F}_n(x) - F(x))^2)$ et détermine la fenêtre asymptotiquement optimale permettant d'avoir un M.S.E. plus faible que pour F_n . Reiss [77] prouve que l'inefficacité relative asymptotique de F_n par rapport à \tilde{F}_n tend rapidement vers l'infini quand la taille de l'échantillon augmente avec un choix approprié de noyau, par exemple

$$k(x) = \frac{9}{8} \left(1 - \frac{5}{3}x^2 \right) I_{[-1,1]}(x),$$

et certaines conditions vérifiées notamment lorsque le support de k est borné et

$$\int_{-\infty}^{\infty} tk(t)K(t)dt > 0. \tag{1}$$

Falk [39] donne ensuite une solution complète à ce problème en établissant la représentation de l'inefficacité relative de F_n par rapport à \tilde{F}_n sous les conditions ci-dessus notamment lorsque le support de k est borné. Le nombre $\psi(k) = \int 2k(x)K(x)xdx$ est introduit par Falk [40] comme une mesure de la performance asymptotique du noyau k . Mais il démontre qu'aucun noyau de carré intégrable ne minimise ψ . Il utilise alors le nombre $\phi(k) = \int k(x)^2dx$ défini par Epanechnikov [38] comme une mesure de la performance du noyau en estimation de la densité. Au sens de ϕ , le noyau d'Epanechnikov suivant

$$k(x) = (3/4)(1 - x^2)I_{(|x| \leq 1)}.$$

est le meilleur mais les noyaux gaussiens ou uniformes ont des performances très proches. En utilisant le critère ψ le noyau d'Epanechnikov est alors de loin le meilleur des trois.

Falk [40] montre ensuite que cette inefficacité relative s'applique aussi aux estimateurs des quantiles q_n de F_n par rapport aux quantiles \tilde{q}_n de \tilde{F}_n . Enfin, Golubev et Levit [51] donnent les conditions permettant de trouver un estimateur minimax du second ordre pour la fonction de perte carrée $L(F, a) = \int (F(t) - a(t))^2 dF(t)$.

Au sens de l'erreur quadratique moyenne intégrée ou I.M.S.E., le meilleur noyau est le noyau uniforme bien que les performances d'autres noyaux (Epanechnikov, normal, triangulaire) ne soient, en pratique, que légèrement moins bonnes (Jones [61]). Il est intéressant de noter que ce ne sera pas le meilleur noyau dans le cadre d'estimation de la densité.

L'expression asymptotique de l'I.M.S.E. est également étudiée par Swanepoel [93]. Pour une fonction f continue, il prouve que le meilleur noyau est le noyau uniforme $k(x) =$

$(1/2\omega)I_{[-\omega,\omega]}(x)$ pour une constante arbitraire $\omega > 0$ (ce qui démontre que les critères de Falk pour définir un noyau optimal ne sont en fait pas adaptés à la fonction de répartition) alors que, pour f discontinue en un nombre fini de points, c'est le noyau exponentiel $k(x) = (c/2) \exp(-c|x|)$ pour une constante arbitraire $c > 0$. \tilde{F}_n est ici aussi plus efficace que F_n pour $h_n = o(n^{-1/2})$ sous la condition (1).

Néanmoins, \tilde{F}_n ne fournit pas toujours une meilleure estimation que F_n . En effet, dans le cas d'une fonction F uniformément lipschitzienne Fernholz [45] obtient que

$$\sqrt{n} \|\tilde{F}_n - F_n\|_\infty \rightarrow 0 \text{ p.s.}$$

et que $\sqrt{n} \|\tilde{F}_n - F\|_\infty$ et $\sqrt{n} \|F_n - F\|_\infty$ ont la même distribution asymptotique. De plus, Shirahata et Chu [85] démontrent que sous certaines hypothèses sur F , l'erreur quadratique intégrée (I.S.E.) $(\int_{-\infty}^{\infty} (\tilde{F}_n(x) - F(x))^2 dF(x))$ de \tilde{F}_n est presque sûrement supérieure à celle de F_n .

5 Estimateurs splines

Les méthodes splines connaissent un large spectre d'applications de par la simplicité de leur mise en oeuvre, de la régularité des courbes obtenues et de la multiplicité des conditions que l'on peut imposer aux solutions. Néanmoins, les fonctions à estimer doivent obéir à certaines conditions de régularité et le nombre d'observations doit être suffisamment grand pour éviter les phénomènes classiques de sur ou sous-lissage. Pour plus de précision sur le sujet on pourra se référer à Besse et Thomas-Agnan [13], Wahba [96] ou Green et Silverman [52].

Berlinet [8] utilise des splines cubiques pour lisser la fonction de répartition empirique et obtenir un estimateur uniformément asymptotiquement sans biais de la fonction de répartition. Les splines cubiques sont ici définies comme un polynôme de degré au plus 3 interpolant la fonction de répartition empirique sur l'ensemble des points de l'échantillon S_n avec certaines conditions aux limites.

Une approche différente est étudiée par Restle [78] qui définit l'estimateur F_{sp} de F sous forme d'une spline naturelle cubique en définissant les valeurs de F_{sp} aux points X_1, \dots, X_n comme

$$\mathbf{F}_{sp} = (\mathbf{I} + \alpha \mathbf{W}^{-1} \mathbf{K})^{-1} \mathbf{F}_n$$

où α est le paramètre de lissage, la matrice \mathbf{W} contient des poids et la matrice aléatoire \mathbf{K} ne dépend que des écarts $H_i = T_{i+1} - T_i$ de l'échantillon. A cause des conditions de continuité imposées aux splines naturelles cubiques, les valeurs aux points T_1, \dots, T_n caractérisent la fonction F_{sp} complètement. Cet estimateur possède une erreur quadratique intégrée de l'ordre de $O_p(n^{-1})$ et le supremum de la différence absolue de F et F_{sp} est de l'ordre de $O_p(n^{-1/4})$. De plus, la probabilité que cet estimateur soit monotone tend vers 1.

6 Les Support Vector Machines

L'idée originale des Support Vector Machines (S.V.M.) est publiée par Vapnik [94] puis reprise dans Burges [19], Schölkopf, Burges et Smola [83], Schölkopf et Smola [84] et plus récemment dans Vapnik et Kotz [95]. Ces dernières années ont vu une explosion du nombre de travaux exploitant la méthode des S.V.M. dont le but premier est de résoudre certains problèmes de classification. Elle est basée sur l'utilisation de fonctions dites noyaux qui permettent une séparation optimale (sans problème d'optimum local) des points observés X_1, \dots, X_n en différentes catégories. Afin de remédier au problème de l'absence de séparateur linéaire, l'idée des S.V.M. est de reconsidérer le problème dans un espace de dimension supérieure. Dans ce nouvel espace, il existe un séparateur linéaire qui permet de classer au mieux les points observés dans les groupes qui conviennent. On peut ensuite projeter le séparateur linéaire dans l'espace d'origine pour visualiser le résultat de la classification. Le changement d'espace se fait au moyen d'une fonction symétrique $k(\cdot, \cdot)$ répondant au critère de Mercer, c'est à dire telle que, pour $1 \leq i, j \leq n$, $(k(X_i, X_j))_{i,j}$ est une matrice définie positive. Ce critère autorise un changement "dans les deux sens" ce qui permet à partir de l'expression de l'hyperplan dans l'espace de dimension plus élevé de classer les éléments dans l'espace de description initial.

Afin d'utiliser les S.V.M., le problème d'estimation de la densité est vu comme le problème suivant : on choisit tout d'abord un ensemble de densités $f(x, \alpha)$ où α est l'ensemble des paramètres à déterminer. La résolution du problème

$$\int_{-\infty}^x f(t, \alpha) dt = F(x)$$

revient donc à résoudre le problème inverse

$$Af(\cdot, \alpha) = F(\cdot, \alpha)$$

où A est une application linéaire de l'espace de Hilbert des fonctions $f(\cdot, \alpha)$ dans celui des fonctions $F(\cdot, \alpha)$. L'estimation de la densité est donc maintenant traduite en un problème inverse qui se résout en utilisant F_n pour estimer F . On peut alors utiliser les S.V.M. pour estimer la densité de différentes manières (Schölkopf, Burges et Smola [83]). Mohamed et Farag [71] choisissent tout d'abord de coupler l'approche S.V.M. et la théorie de champ moyen (qui permet des approximations efficaces). Puis ils incorporent l'algorithme EM à l'approche précédente (Mohamed, El-Baz et Farag [70]). Les estimateurs de $f(x)$ et $F(x)$ sont alors

$$\bar{f}_n^*(x) = \sum_{i=1}^n \omega_i k(x, X_i) \text{ et } \bar{F}_n^*(x) = \sum_{i=1}^n \omega_i K(x, X_i)$$

où k est un noyau vérifiant certaines propriétés (Vapnik [94]) dérivé de K et ω_i obtenus dans Mohamed, El-Baz et Farag [70]. L'algorithme EM est ensuite utilisé pour trouver les paramètres optimaux pour le noyau, ici la matrice de covariance pour un noyau gaussien, et permet un gain substantiel (notamment en terme de distance de Levy) par rapport à la première approche où on utilise les paramètres ad-hoc.

7 Le level-crossing

Le level-crossing, passage à niveau en français, est défini dans Leadbetter [65]. Si nous notons $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ les statistiques d'ordre de l'échantillon X_1, X_2, \dots, X_n et prenons $x \in \mathbb{R}$ alors on dit que l'intervalle $[X_{(j)}, X_{(k)})$ "croise" le niveau x si et seulement si $x \in [X_{(j)}, X_{(k)})$, $1 \leq j \leq k < n$. Il est ensuite possible de définir une fonction $r_n(x)$ comptant le nombre de "croisements" se produisant au niveau x à partir d'un échantillon de départ $\{X_i, i = 1, \dots, n\}$.

L'espérance de cette fonction $r_n(x)$ s'exprime en fonction de $F(x)$. Huang et Brill [59] appliquent le level-crossing à l'estimation de la fonction de répartition. L'estimateur obtenu diffère de la fonction de répartition empirique qui donne un poids équivalent de $1/n$ à chaque valeur de l'échantillon. Ils essayent d'améliorer l'efficacité en faisant varier les poids des observations. Des poids moins importants sont attribués aux données situées dans les queues de l'échantillon. Le nouvel estimateur est alors défini comme

$$\hat{F}_n^*(x) = \sum_{i=1}^n I_{(-\infty, x]}(X_{(i)}) p_{n,i}, \quad n \geq 2,$$

avec

$$p_{n,i} = \begin{cases} \frac{1}{2} \left[1 - \frac{n-2}{\sqrt{n(n-1)}} \right], & \text{pour } i = 1, n, \\ \frac{1}{\sqrt{n(n-1)}}, & \text{pour } i = 2, \dots, n-1. \end{cases}$$

Nous remarquons que les valeurs $X_{(1)}$ et $X_{(n)}$ ont moins de poids que les valeurs centrales de l'échantillon. Ils obtiennent tout d'abord

$$P \left\{ \lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} |\hat{F}_n^*(x) - F(x)| = 0 \right\} = 1$$

et démontrent ensuite que pour toute fonction F continue, il existe un nombre réel $z_n \in (0, \frac{1}{2})$, tel que l'efficacité de leur estimateur par rapport à F_n est plus grande que 1 sur les intervalles ouverts $(0, z_n)$ et $(1 - z_n, 1)$. Ces résultats théoriques sont ensuite complétés par des simulations sur lesquelles le level-crossing donne de meilleures estimations que la fonction de répartition empirique.

8 Les Systèmes de Fonctions Itérées

Les systèmes de fonctions itérées (I.F.S.), introduits dans Barnsley et Demko [6], sont principalement utilisés pour la modélisation de fractales. Cette théorie est entièrement basée sur la propriété d'invariance par changement d'échelle. Ainsi une application à l'approximation de fonction est récemment proposée dans Forte et Vrscay [48]. Le principe est le suivant : on suppose que ω est une fonction contractante d'un espace métrique M dans lui-même. Le théorème du point fixe nous donne ensuite l'existence et l'unicité d'un point fixe x' dans M tel que $\omega(x') = x'$. De plus, pour tout point de départ x_0 , la suite de points (x_n) définie par $x_{n+1} = \omega(x_n)$ converge vers x' lorsque n tend vers l'infini. C'est pourquoi x' est appelé attracteur de l'I.F.S.

On peut maintenant supposer qu'il existe n fonctions contractantes ω_i , $i = 1, \dots, n$, dans M . Chacune de ces fonctions ω_i a son propre point fixe x'_i et si on applique une de ces fonctions, par exemple ω_l , alors la suite (x_n) converge vers le point fixe x'_l . Ce sont ces résultats que Iacus et La Torre [60] utilisent en interprétant l'estimation de F comme la recherche du point fixe d'une famille de fonctions contractantes sur un espace métrique complet. Le problème se définit alors à l'aide de N fonctions affines ω et d'un vecteur de paramètres p . Pour une famille donnée W , le résultat dépend uniquement du choix de p . Il faut donc maintenant déterminer p en résolvant un problème d'optimisation quadratique sous contraintes. Cette approche est appelée *approche inverse*. Ils démontrent qu'une fois la famille de fonctions choisie le problème inverse revient à résoudre un problème quadratique dépendant uniquement des moments non centrés de l'échantillon.

Le choix $p_i = 1/N$ nous permet d'obtenir une version lissée de la fonction de répartition empirique et, par conséquent, de bonnes propriétés asymptotiques et cela même si le support de la loi n'est pas compact. Différentes simulations sur des fonctions bêta sont testées et l'I.F.S. apparaît comme plus efficace que les fonctions de répartition empiriques ou à noyaux en terme d'erreur quadratique moyenne ou de norme infinie. Cette méthode est notamment très performante dans les cas de données manquantes ou de petits échantillons mais nécessite une fonction à support compact ou au moins une connaissance du support.

9 D'autres estimateurs

Comme il a été dit précédemment tout estimateur de la densité fournit par intégration un estimateur de la fonction de répartition. Nous évoquons donc dans le présent paragraphe quelques familles importantes de tels estimateurs qui n'ont pas été mentionnées auparavant. Nous resterons volontairement succincts, le nombre de références sur le sujet concernant l'estimation de la densité étant considérable. Les premiers travaux traitant de manière unifiée des estimateurs de la densité et des conditions de leur convergence sont ceux de Foldes et Revesz [47] et Bleuez et Bosq [15, 16]. Ces auteurs ont considéré la famille des estimateurs de la densité de la forme

$$f_n(t) = \frac{1}{n} \sum_{j=1}^n K_{r(n)}(X_j, t),$$

où $(K_r(x, t), r \in I)$ est une famille de fonctions réelles mesurables et I une partie non bornée de \mathbb{R}^+ . Cette famille contient les estimateurs à noyaux vus précédemment et ceux utilisant la méthode des fonctions orthogonales qui correspond au choix

$$K_r(x, t) = \sum_{i=0}^r e_i(x)e_i(t),$$

$r \in \mathbb{N}$ et où les e_i forment une base de hilbertienne de $L^2(\mathbb{R})$. Elle englobe également les histogrammes, de nombreux estimateurs basés sur des partitions (Bosq et Lecoutre [17], Berlinet et Biau [10]) et les estimateurs par ondelettes.

La théorie des ondelettes s'est développée au milieu des années 80 et est très utilisée dans de nombreux domaines (analyse harmonique, traitement du signal, compression d'images, statistique fonctionnelle...). Son succès est dû à son adaptativité aux données et à sa facilité d'implémentation. Une ondelette est une onde "localisée". Lorsqu'on décompose une fonction en séries de Fourier, on la décompose en fait en fréquence. La décomposition en ondelettes ajoute une dimension, la décomposition est double : fréquentielle et spatiale. On décompose en effet une fonction comme une somme d'oscillations de fréquences précises se produisant à un endroit précis. La construction d'une base d'ondelettes dans un espace convenable permet d'estimer une fonction si on sait estimer ses coordonnées dans la base choisie. On peut citer comme références Donoho, Johnstone, Kerkyacharian et Picard [31], Härdle, Kerkyacharian, Picard et Tsybakov [55] ou Herrick, Nason et Silverman [57]. Mais la plupart sont appliquées à l'estimation de la densité et il n'existe pas, à notre connaissance, d'application visant à obtenir des propriétés spécifiques d'estimateurs de la fonction de répartition.

Nous pouvons aussi évoquer brièvement une approche statistique différente : l'approche bayésienne. Dans la perspective bayésienne un état de connaissance initial est traduit par une loi de probabilité a priori sur les paramètres du modèle choisi. Le théorème de Bayes permet de passer ensuite à une loi dite a posteriori sur les paramètres conditionnellement aux données observées. Cette approche est notamment développée en détails dans Robert [79] et Bernardo et Smith [12]. Son application à l'estimation de la fonction de répartition, sous différents modèles, est évoquée dans Korwar et Hollander [64], Susarla et Van Ryzin [92] et Pantazopoulos, Pappis, Fifis, Costopoulos, Vaughan et Gasparini [73]. En statistique non paramétrique la construction des estimateurs bayésiens est souvent difficile. On pourra consulter Bosq et Lecoutre [17] au sujet de l'utilisation d'un processus de Dirichlet comme loi a priori. L'estimateur résultant a des propriétés asymptotiques analogues à celles de la fonction de répartition empirique.

Enfin, les méthodes combinatoires, développées par Devroye et Lugosi [29], apportent un regard nouveau sur le choix efficace d'estimateurs non paramétriques et sont applicables aux estimateurs de la fonction de répartition.

10 Fonction de répartition conditionnelle

Lorsque les données sont disponibles sous la forme de couples $(X_i, Y_i)_{1 \leq i \leq n}$, on peut chercher à estimer la fonction de répartition conditionnelle $\pi(y|x) = P(Y_i \leq y | X_i = x)$. Roussas [81] définit son estimateur basé sur l'estimateur à noyau comme

$$\pi_G(y|x) = \int_{-\infty}^y q_n(x, y') / p_n(x)$$

où

$$p_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \text{ et } q_n(x, y') = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h^{1/2}}\right) K\left(\frac{y' - Y_i}{h^{1/2}}\right).$$

Sous certaines hypothèses fortes, il démontre sa convergence en probabilité ainsi que celle de ses quantiles (voir aussi Samanta [82]). Stone [89] définit lui un estimateur de type

$$\hat{\pi}_n(A|X) = \sum_{i=1}^n W_{ni}(X)I_A(Y_i)$$

où W_{ni} est une fonction de poids propre à chaque X_i . Il obtient la consistance de l'estimateur sous certaines conditions sur les fonctions de poids. En s'appuyant sur ce travail, Cleveland [22] définit l'estimation polynômiale locale. L'idée sous-jacente est en fait un simple développement de Taylor : soit x et y fixés et supposons que la fonction qui à z associe $\pi(y|z)$ est suffisamment régulière dans un voisinage de x . Alors, par une approximation de Taylor,

$$\pi(y|z) \approx \pi(y|x) + (z - x)\pi^{(1)}(y|x) + \dots + (z - x)^p \frac{\pi^{(p)}(y|x)}{p!}$$

où $\pi^{(v)}(y|x)$ est la $v^{\text{ème}}$ dérivée de $\pi(y|z)$ par rapport à z , évaluée en x . Nous approximons donc localement $\pi(y|z)$ par un polynôme de degré p en $(z - x)$. Fan [41, 42] et Fan et Gijbels [43, 44] montrent que cet estimateur présente plusieurs avantages sur les autres méthodes de régression (comme celle du noyau de Gasser et Müller [50]) : elle contourne la principale difficulté des estimateurs à noyaux classiques en corrigeant automatiquement les effets de bord tout en conservant les propriétés d'optimalité théorique. Il est à noter que Ferrigno et Ducharme [46] construisent un test permettant de rejeter (ou de valider) un modèle en fonction de sa "distance" à l'estimateur polynômial local.

Stute [90, 91] définit lui un estimateur de type plus-proche-voisins

$$\pi_n(y|x) = \frac{1}{nh} \sum_{i=1}^n Y_i K \left(\frac{F_n(x) - F_n(X_i)}{h} \right)$$

où K est un noyau. Sous certaines conditions, il obtient sa normalité asymptotique et celle de ses quantiles.

Hall, Wolff et Yao [54] proposent deux méthodes différentes pour estimer cette fonction. Ils définissent ainsi un estimateur à noyau adapté (qui sera équivalent à un estimateur localement linéaire) et un estimateur faisant intervenir un modèle logistique. L'estimateur à noyau se définit comme

$$\tilde{\pi}(y|x) = \frac{\sum_{i=1}^n I_{(Y_i \leq y)} p_i(x) K \left(\frac{X_i - x}{h} \right)}{\sum_{i=1}^n p_i(x) K \left(\frac{X_i - x}{h} \right)},$$

où K est un noyau de fenêtre h et p_i , pour $1 \leq i \leq n$, des poids uniquement définis par plusieurs équations.

Le modèle logistique général pour $P(x) = \pi(y|x)$ admettant $r - 1$ dérivées continues est de la forme

$$L(x, \mathbf{b}) = A(x, \mathbf{b}) / \{1 + A(x, \mathbf{b})\},$$

où $A(\cdot, \mathbf{b})$ est une fonction non-négative dépendant du vecteur de paramètres $\mathbf{b} = (b_1, \dots, b_r)$ représentant $P(x)$ et ses $r - 1$ dérivées. On peut noter qu'il est possible de trouver des fonctions A relativement simples. Ce modèle amène à l'estimateur $\hat{\pi}(y|x) = L(0, \hat{\mathbf{b}}_{xy})$ où $\hat{\mathbf{b}}_{xy}$ minimise la fonction

$$R(\mathbf{b}; x, y) = \sum_{i=1}^n \{I_{(Y_i \leq y)} - L(X_i - x, \mathbf{b})\}^2 K \left(\frac{X_i - x}{h} \right).$$

Ces deux estimateurs font intervenir un noyau K . Ils seront consistants selon le choix toujours délicat de la fenêtre h .

Il est aussi possible d'estimer F en utilisant F_n et les techniques classiques de régression dans le modèle

$$F_n(X_i) = F(X_i) + \varepsilon_i, \quad i = 1, \dots, n.$$

En utilisant un estimateur localement linéaire de même type que celui décrit ci-dessus, Cheng et Peng [21]) déterminent un estimateur ayant une moyenne d'erreur quadratique intégrée plus petite que l'estimateur à noyau classique.

11 Données biaisées

En pratique, il peut arriver qu'observer directement la variable X de densité f soit impossible mais qu'on puisse, par contre, observer une variable Y à travers un échantillon Y_1, \dots, Y_n et possédant la densité

$$g(y) = w(y)f(y)/\mu(f)$$

où $w(x)$ va être la fonction de biais et $\mu(f) = E_f\{w(X)\} = 1/E_g\{w^{-1}(Y)\}$. On rencontre également cette situation dans le cas de données manquantes (Berlinet et Thomas-Agnan [11]). Dans ce qui suit on suppose que la fonction $w(x)$ est connue, intégrable et que

$$\forall x, \quad 0 < c_1 < w(x) < c_2 < \infty.$$

Patil, Rao et Zelen [75] référencent les différentes propriétés de ces fonctions biaisées pour laquelle Cox [25] propose la distribution suivante :

$$\tilde{F}_{co}(x) = \hat{\mu} n^{-1} \sum_{l=1}^n w^{-1}(Y_l) I_{(Y_l \leq x)},$$

avec

$$\hat{\mu} = \frac{1}{n^{-1} \sum_{l=1}^n w^{-1}(Y_l)}.$$

Cet estimateur est celui du maximum de vraisemblance dans le cadre non-paramétrique, il est asymptotiquement efficace et est un estimateur minimax du premier ordre. Efromovitch [36] modifie l'estimateur de Cox et propose le suivant :

$$\tilde{F}_e(x) = x + \sqrt{2} \sum_{j=1}^J \hat{\theta}_j (\pi j)^{-1} \sin(\pi j x),$$

où

$$\hat{\theta}_j = n^{-1} \sqrt{2} \hat{\mu} \sum_{l=1}^n w^{-1}(Y_l) \cos(\pi j Y_l),$$

ce qui est l'estimateur moyen de Cox pour θ_j , et J une constante calculable, fonction de n . Il obtient

$$|E\{\tilde{F}_e(x)\} - F(x)| \leq \frac{o(1)}{\sqrt{n} \log(n)}$$

et la normalité asymptotique de son estimateur qui est de plus un estimateur minimax du second ordre en x .

12 Conclusion

L'estimateur le plus simple, la fonction de répartition empirique a de bonnes propriétés de convergence mais possède certains inconvénients comme celui de ne pas prendre en compte une éventuelle information supplémentaire ou bien le fait d'être une fonction en escalier. Dès que l'on restreint quelque peu le modèle envisagé pour les données il existe des estimateurs qui sont préférables à la fonction de répartition empirique. Néanmoins l'existence de ce premier estimateur, au contraire de ce qui se passe pour la densité, donne des facilités quant à l'utilisation de méthodes de lissage. Les méthodes d'estimation que nous avons passées en revue fournissent en général une fonction qui possède les propriétés caractérisant une fonction de répartition sauf parfois en ce qui concerne la masse totale ($\lim_{x \rightarrow +\infty} F(x) = 1$), ce qui est facilement corrigé, ou bien en ce qui concerne la croissance. On peut alors penser à appliquer certaines méthodes d'estimation de fonctions monotones (Delecroix, Simioni et Thomas-Agnan [26]), la régression ou les splines isotoniques (Barlow, Bartholomew, Brewner et Brunk [5] et Wegman et Wright [100]) afin de garantir la croissance de l'estimateur.

Remerciements. Cet article a été rédigé dans le cadre d'une thèse co-financée par le CNRS et la région Languedoc-Roussillon. Je remercie mon directeur de thèse, Alain Berline, pour son implication dans la rédaction de ce travail ainsi que Christine Thomas-Agnan qui m'a communiqué quelques références pertinentes.

Qu'il me soit également permis de remercier les deux rapporteurs anonymes : leurs commentaires, leurs suggestions et leurs critiques constructives m'ont permis d'améliorer substantiellement la qualité de ce travail.

Références

- [1] B. ABDOUS, A. BERLINET et N. HENGARTNER – « A general theory for kernel estimation of smooth functionals of the distribution function and their derivatives », *Revue Roumaine de Mathématiques Pures et Appliquées* **48** (2003), p. 217–232.
- [2] O. AGGARWAL – « Some minimax invariant procedures for estimating a cumulative distribution function », *The Annals of Mathematical Statistics* **26** (1955), p. 450–462.
- [3] H. AKAIKE – « An approximation to the density function », *Annals of the Institute of Statistical Mathematics* **6** (1954), p. 127–132.
- [4] A. AZZALINI – « A note on the estimation of a distribution function and quantiles by a kernel method », *Biometrika* **68** (1981), p. 326–328.
- [5] R. BARLOW, D. BARTHOLOMEW, J. BREWNER et H. BRUNK – *Statistical inferences under order restrictions*, Wiley, New-York, 1972.
- [6] M. BARNSLEY et S. DEMKO – « Iterated function systems and the global construction of fractals », *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **399** (1985), p. 243–275.
- [7] R. BERAN – « Estimating a distribution function », *The Annals of Statistics* **5** (1977), p. 400–404.

- [8] A. BERLINET – « Convergence des estimateurs splines de la densité », *Publications de l'Institut de Statistique de l'Université de Paris* **26** (1981), p. 1–16.
- [9] — , « Reproducing kernels and finite order kernels », *Nonparametric Functional Estimation and Related Topics* (G. Roussas, éd.), Kluwer Academic Publishers, Dordrecht, 1991.
- [10] A. BERLINET et G. BIAU – « Estimation de densité et prise de décision », *Décision et Reconnaissance de Formes en Signal* (R. Lengellé, éd.), Hermes, Paris, 2002.
- [11] A. BERLINET et C. THOMAS-AGNAN – *Reproducing kernel in Hilbert spaces in probability and statistics*, Kluwer, Boston, 2004.
- [12] J. BERNARDO et A. SMITH – *Bayesian theory*, Wiley, New-York, 1994.
- [13] P. BESSE et C. THOMAS-AGNAN – « Le lissage par fonctions splines en statistique : revue bibliographique », *Statistique et Analyse des données* **14** (1989), p. 55–83.
- [14] P. BILLINGSLEY – *Convergence of probability measures*, Wiley, New-York, 1968.
- [15] J. BLEUEZ et D. BOSQ – « Conditions nécessaires et suffisantes de convergence pour une classe d'estimateurs de la densité », *Comptes rendus de l'Académie des Sciences de Paris Série A* **282** (1976), p. 63–66.
- [16] — , « Conditions nécessaires et suffisantes de convergence pour une classe d'estimateurs de la densité par la méthode des fonctions orthogonales », *Comptes rendus de l'Académie des Sciences de Paris Série A* **282** (1976), p. 1023–1026.
- [17] D. BOSQ et J.-P. LECOUTRE – *Théorie de l'estimation fonctionnelle*, Economica, 1987.
- [18] L. BROWN – « Admissibility in discrete and continuous invariant non-parametric estimation problems, and in their multinomial analogs », *The Annals of Statistics* **16** (1988), p. 1567–1593.
- [19] C. BURGESS – « A tutorial on support vector machines for pattern recognition », *Data mining and knowledge discovery* **2** (1998), p. 1–47.
- [20] P. CAPERAA et B. VAN CUTSEM – *Méthodes et modèles en statistique non paramétrique*, Bordas, Paris, 1988.
- [21] M.-Y. CHENG et L. PENG – « Regression modelling for nonparametric estimation of distribution and quantiles functions », *Statistica Sinica* **12** (2002), p. 1043–1060.
- [22] W. CLEVELAND – « Robust locally weighted regression and smoothing scatterplots », *Journal of the American Statistical Association* **74** (1979), p. 829–836.
- [23] M. COHEN et L. KUO – « The admissibility of the empirical distribution function », *The Annals of Statistics* **13** (1985), p. 262–271.
- [24] G. COLLOMB, S. HASSANI, P. SARDA et P. VIEU – « Estimation non paramétrique de la fonction de hasard pour des observations dépendantes », *Statistique et Analyse des Données* **10** (1985), p. 42–49.
- [25] D. COX – « Some sampling problems in technology », *New Developments in Survey Sampling* (N. L. Johnson et H. Smith, éd.), Wiley, New-York, 1969.
- [26] M. DELECROIX, M. SIMIONI et C. THOMAS-AGNAN – « Functional estimation under shape constraints », *Journal of Nonparametrics Statistics* **6** (1996), p. 69–89.
- [27] L. DEVROYE – *A course in density estimation*, Birkhäuser, Boston, 1987.

- [28] L. DEVROYE et L. GYÖRFI – « Distribution and density estimation », *CISM Courses and Lectures* **434** (2002), p. 221–270.
- [29] L. DEVROYE et G. LUGOSI – *Combinatorial methods in density estimation*, Springer, New-York, 2001.
- [30] L. DEVROYE et G. WISE – « On the recovery of discrete probability densities from imperfect measurements », *Journal of the Franklin Institute* **307** (1979), p. 1–20.
- [31] D. DONOHO, I. JOHNSTONE, G. KERKYACHARIAN et D. PICARD – « Density estimation by wavelet thresholding », *The Annals of Statistics* **24** (1996), p. 508–539.
- [32] M. DONSKER – « Justification and extension of Doob’s heuristic approach to the Kolmogorov-Smirnov theorems », *The Annals of Mathematical Statistics* **23** (1952), p. 277–281.
- [33] J. DOOB – « Heuristic approach to the kolmogorov-smirnov theorem », *The Annals of Mathematical Statistics* **20** (1949), p. 393–403.
- [34] — , *Stochastic processes*, Wiley, New-York, 1953.
- [35] A. DVORETZKY, J. KIEFER et J. WOLFOWITZ – « Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator », *The Annals of Mathematical Statistics* **33** (1956), p. 642–669.
- [36] S. EFROMOVICH – « Distribution estimation for biased data », *Journal of Statistical Planning and Inference* **124** (2004), p. 1–43.
- [37] B. EFRON et R. TIBSHIRANI – *An introduction to the bootstrap*, Chapman & Hall, London, 1993.
- [38] A. EPANECHNIKOV – « Nonparametric estimation of a multivariate probability density », *Theory of Probability and its Applications* **14** (1969), p. 153–158.
- [39] M. FALK – « Relative efficiency and deficiency of kernel type estimators of smooth distribution functions », *Statistica Neerlandica* **37** (1983), p. 73–83.
- [40] — , « Relative deficiency of kernel type estimators of quantiles », *The Annals of Statistics* **12** (1984), p. 261–268.
- [41] J. FAN – « Design-adaptive nonparametric regression », *Journal of the American Statistical Association* **87** (1992), p. 998–1004.
- [42] — , « Local linear regression smoothers and their minimax efficiencies », *The Annals of Statistics* **21** (1993), p. 196–216.
- [43] J. FAN et I. GIJBELS – « Variable bandwidth and local linear regression smoothers », *The Annals of Statistics* **20** (1992), p. 2008–2036.
- [44] J. FAN et I. GIJBELS – *Local polynomial modelling and its applications*, Chapman & Hall, London, 1996.
- [45] L. FERNHOLZ – « Almost sure convergence of smoothed empirical distribution functions », *Scandinavian Journal of Statistics* **18** (1991), p. 255–262.
- [46] S. FERRIGNO et G. DUCHARME – « A global test of goodness-of-fit for the conditional distribution function », *Comptes Rendus Mathématiques de l’Académie des Sciences de Paris* **341** (2005), p. 313–316.

- [47] A. FOLDES et P. REVESZ – « A general method for density estimation », *Studia Scientiarum Mathematicarum* **9** (1974), p. 81–92.
- [48] B. FORTE et E. VRSCAY – « Solving the inverse problem for function/image approximation using iterated function systems », *Fractals* **2** (1995), p. 325–334.
- [49] Y. FRIEDMAN, A. GELMAN et E. PHADIA – « Best invariant estimation of a distribution function under the Kolmogorov-Smirnov loss function », *The Annals of Statistics* **16** (1988), p. 1254–1261.
- [50] T. GASSER et H. MÜLLER – « Estimating regression function and their derivatives by the kernel method », *Scandinavian Journal of Statistics* **3** (1984), p. 171–185.
- [51] G. GOLUBEV et B. LEVIT – « Distribution function estimation : adaptive smoothing », *Mathematical Methods of Statistic* **5** (1996), p. 383–403.
- [52] P. GREEN et B. SILVERMAN – *Nonparametric regression and generalized linear models. a roughness penalty approach*, Chapman & Hall, London, 1994.
- [53] U. GRENANDER – « On the theory of mortality measurement part II », *Skandinavisk Aktuarietidskrift* **39** (1956), p. 125–153.
- [54] P. HALL, R. WOLFF et Q. YAO – « Methods for estimating a conditional distribution function », *Journal of the American Statistical Association* **94** (1999), p. 154–163.
- [55] W. HÄRDLE, G. KERKYACHARIAN, D. PICARD et A. TSYBAKOV – « Wavelets, approximation and statistical applications », *Lecture Notes in Statistics* **129** (1999).
- [56] P. HENNEQUIN et A. TORTRAT – *Théorie des probabilités et quelques applications*, Masson, Paris, 1965.
- [57] D. HERRICK, G. NASON et B. SILVERMAN – « Some new methods for wavelet density estimation », *Sankhya* **A63** (2001), p. 394–411.
- [58] I. HU – « A uniform bound for the tail probability of Kolmogorov-Smirnov statistics », *The Annals of Statistics* **13** (1985), p. 811–826.
- [59] M. HUANG et P. BRILL – « A distribution estimation method based on level crossings », *Journal of Statistical Planning and Inference* **124** (2004), p. 45–62.
- [60] S. IACUS et D. LA TORRE – « A comparative simulation study on the IFS distribution function estimator », *Nonlinear Analysis : Real World Applications* **6** (2005), p. 858–873.
- [61] M. JONES – « The performance of kernel density functions in kernel distribution function estimation », *Statistics and Probability Letters* **9** (1990), p. 129–132.
- [62] J. KIEFER et J. WOLFOWITZ – « Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters », *The Annals of Mathematical Statistics* **27** (1956), p. 887–906.
- [63] A. KOLMOGOROV – « Sulla determinazione empirica di una legge de distribuzione », *Giornale dell’Istituto Italiano degli Attuari* **4** (1933), p. 83–91.
- [64] R. KORWAR et M. HOLLANDER – « Empirical Bayes estimation of a distribution function », *The Annals of Statistics* **4** (1976), p. 581–588.
- [65] M. LEADBETTER – « Point processes generated by level crossings », *Stochastic Point processes : Statistical Analysis, Theory and Applications* (P. Lewis, éd.), Wiley-Interscience, New-York, 1972.

- [66] E. LEHMANN – *Theory of point estimation*, Wiley, New-York, 1983.
- [67] M. LEJEUNE et P. SARDA – « Smooth estimators of distribution and density functions », *Computational Statistics and Data Analysis* **14** (1992), p. 457–471.
- [68] P. MASSART – « The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality », *The Annals of Probability* **18** (1990), p. 1269–1283.
- [69] R. MODARRES – « Efficient nonparametric estimation of a distribution function », *Computational Statistics and Data Analysis* **39** (2002), p. 75–95.
- [70] R. MOHAMED, A. EL-BAZ et A. FARAG – « Probability density estimation using advanced support vector machines and the EM algorithm », *International Journal of Signal Processing* **1** (2004), p. 260–264.
- [71] R. MOHAMED et A. FARAG – « Mean field theory for density estimation using support vector machines », *Seventh International Conference on Information Fusion, Stockholm* (2004), p. 495–501.
- [72] E. NADARAYA – « Some new estimates for distribution function », *Theory of Probability and its Application* **9** (1964), p. 497–500.
- [73] S. PANTAZOPOULOS, C. PAPPIS, T. FIFIS, C. COSTOPOULOS, J. VAUGHAN et M. GASPARINI – « Nonparametric Bayes estimation of a distribution function with truncated data », *Journal of Statistical Planning and Inference* **55** (1996), p. 361–369.
- [74] E. PARZEN – « On the estimation of a probability density and mode », *The Annals of Mathematical Statistics* **33** (1962), p. 1065–1076.
- [75] G. PATIL, C. RAO et M. ZELEN – « Weighted distribution », *Encyclopedia of Statistical Sciences* (N. L. Johnson et S. Kotz, éds.), Wiley, New-York, 1988.
- [76] E. PHADIA – « Minimax estimation of a cumulative distribution function », *The Annals of Statistics* **1** (1973), p. 1149–1157.
- [77] R.-D. REISS – « Nonparametric estimation of smooth distribution functions », *Scandinavian Journal of Statistics* **8** (1981), p. 116–119.
- [78] E. RESTLE – « Estimating cumulative distributions by spline smoothing », Thèse, Ecole Polytechnique Fédérale de Lausanne, 2001.
- [79] C. ROBERT – *L'analyse statistique bayésienne*, Economica, Paris, 1992.
- [80] M. ROSENBLATT – « Remarks on some non-parametric estimates of a density function », *The Annals of Mathematical Statistics* **27** (1956), p. 832–837.
- [81] G. ROUSSAS – « Nonparametric estimation of the transition distribution function of a Markov process », *The Annals of Mathematical Statistics* **40** (1969), p. 1386–1400.
- [82] M. SAMANTA – « Non-parametric estimation of conditional quantiles », *Statistics and Probability Letters* **7** (1989), p. 407–412.
- [83] B. SCHÖLKOPF, C. BURGESS et A. SMOLA – *Advances in kernel methods. support vector learning*, MIT Press, 1999.
- [84] B. SCHÖLKOPF et A. SMOLA – *Learning with kernels : Support vector machines, regularization, optimization and beyond*, MIT Press, 2002.
- [85] S. SHIRAHATA et I.-S. CHU – « Integrated squared error of kernel-type estimator of distribution function », *Annals of the Institute of Statistical Mathematics* **44** (1992), p. 579–591.

- [86] G. SHORACK et J. WELLNER – *Empirical processes with applications to statistics*, Wiley, New-York, 1986.
- [87] R. S. SINGH, T. GASSER et B. PRASAD – « Nonparametric estimates of distributions functions », *Communication in Statistics - Theory and Methods* **12** (1983), p. 2095–2108.
- [88] N. SMIRNOV – « Approximate laws of distribution of random variables from empirical data », *Uspekhi Matematicheskikh Nauk* **10** (1944), p. 179–206.
- [89] C. STONE – « Consistent nonparametric regression », *The Annals of Statistics* **5** (1977), p. 595–645.
- [90] W. STUTE – « Asymptotic normality of nearest neighbor regression function estimates », *The Annals of Statistics* **12** (1984), p. 917–926.
- [91] —, « Conditionnal empirical processes », *The Annals of Statistics* **14** (1986), p. 638–647.
- [92] V. SUSARLA et J. VAN RYZIN – « Empirical Bayes estimation of a distribution (survival) function from right censored observations », *The Annals of Statistics* **6** (1978), p. 740–754.
- [93] J. SWANEPOEL – « Mean integrated squared error properties and optimal kernels when estimating a distribution function », *Communication in Statistics - Theory and Methods* **17** (1988), p. 3785–379.
- [94] V. VAPNIK – *The nature of statistical learning theory*, Springer Verlag, New-York, 1995.
- [95] V. VAPNIK et S. KOTZ – *Estimation on dependences based on empirical data*, Springer Verlag, New-York, 2006.
- [96] G. WAHBA – *Spline models for observational data*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1990.
- [97] G. WATSON et M. LEADBETTER – « Hazard analysis II », *Sankhya* **26** (1964), p. 101–116.
- [98] B. WINTER – « Strong uniform consistency of integrals of density estimators », *The Canadian Journal of Statistics* **1** (1973), p. 247–253.
- [99] —, « Convergence rate of perturbed empirical distribution functions », *Journal of Applied Probability* **16** (1979), p. 163–173.
- [100] I. WRIGHT et E. WEGMAN – « Isotonic, convex and related splines », *The Annals of Statistics* **8** (1980), p. 1023–1035.
- [101] H. YAMATO – « Uniform convergence of an estimator of a distribution function », *Bulletin on Mathematical Statistics* **15** (1973), p. 69–78.
- [102] Q. YU – « Inadmissibility of the empirical distribution function in continuous invariant problems », *The Annals of Statistics* **17** (1989), p. 1347–1359.