C. R. Acad. Sci. Paris, Ser. I 334 (2002) 509–513

Statistique/*Statistics*
(Probabilités/*Probability Theory*)

Least squares cross-validation for the kernel deconvolution density estimator

Élie Youndjé [a], **Martin T. Wells** [b]

[a] UMR CNRS 6085, Université de Rouen, 76821 Mont Saint Aignan, France

[b] Department of Social Statistics, Cornell University, 364 Ives Hall, Ithaca, NY 14851-0952, USA

Received 29 June 2001; accepted after revision 28 January 2002

Note presented by Paul Deheuvels.

Abstract Assume we have i.i.d. replications from the corrupted random variable $Y = X + \varepsilon$, where X and ε are independent. We propose a data-driven bandwidth based on cross-validation ideas, for the kernel deconvolution estimator of the density of X. The proposed method is shown to be asymptotically optimal. ***To cite this article: É. Youndjé, M.T. Wells, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 509–513.*** © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

Validation croisée pour l'estimateur à noyau de la déconvolution d'une densité

Résumé En présence d'un échantillon i.i.d. d'une variable aléatoire corrumpue $Y = X + \varepsilon$, avec X et ε indépendants. Nous proposons une méthode basée sur la validation-croisée, pour choisir la largeur de la fenêtre de l'estimateur à noyau de la densité de X. L'optimalité asymptotique de la méthode proposée est établie. ***Pour citer cet article : É. Youndjé, M.T. Wells, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 509–513.*** © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

Version française abrégée

Soient $Y_1, \ldots, Y_n$ un n-échantillon i.i.d. de la variable

$$Y = X + \varepsilon,$$

où X et ε sont indépendantes. Nous supposons que g, f et h sont les densités de Y, X et ε respectivement et que h est connue. Nous considérons le problème d'estimer f à l'aide de $Y_1, \ldots, Y_n$, notre but principal étant de bâtir une règle de sélection de la fenêtre pour l'estimateur à noyau de f. Pour une fonction τ soit Φ_τ sa transformée de Fourier c'est-à-dire

$$\Phi_\tau(t) \triangleq \int e^{ixt} \tau(x)\, dx.$$

E-mail addresses: Elie.youndje@univ-rouen.fr (É. Youndjé); mtw1@cornell.edu (M.T. Wells).

S1631-073X(02)02291-4/FLA

L'estimateur à noyau de f est défini par

$$f_b(x) \triangleq \frac{1}{nb}\sum_{j=1}^{n} g_n\left(\frac{x-Y_i}{b}\right)$$

avec

$$g_n(x) \triangleq \frac{1}{2\pi}\int e^{-ixt}\frac{\Phi_K(t)}{\Phi_h(t/b_n)}\,dt,$$

K le noyau, $b = b_n$ la largeur de fenêtre. Le noyau g_n est connu dans la littérature comme noyau de déconvolution. L'estimateur $f_b(x)$ a été très étudié par plusieurs auteurs, nous pouvons citer entre autres Fan [2] et [3], Stefanski et Carroll [12] et Masry [8–10] pour les propriétés de convergence. Dans cette Note sous l'hypothèse que $\Phi_h(t) = O(t^{-\beta})$ quand $t \to \infty$, nous allons proposer une règle de sélection de la fenêtre et prouver son optimalité. Comme presque toujours dans ce genre de problème nous voulons choisir b minimisant l'erreur quadratique intégrée

$$ISE(b) = \int \left|f_b(x) - f(x)\right|^2 dx,$$

mais comme cette quantité n'est pas calculable dans les situations pratiques, nous proposons le critère de sélection (dont la motivation est donnée dans la partie principale de la Note)

$$CV(b) \triangleq \int |f_b|^2(x)\,dx - 2\widehat{ST}(b)$$

avec

$$\widehat{ST}(b) \triangleq \frac{1}{n(n-1)}\sum_{j\neq k}\mathrm{Re}\,\frac{1}{2\pi}\int\frac{\Phi_K(tb)}{|\Phi_h(t)|^2}e^{-it(Y_k-Y_j)}\,dt.$$

Le Théorème 0.1 ci-dessus montre que minimiser $CV(b)$ est équivalent à minimiser $ISE(b)$.

THÉORÈME 1.1. – *Sous les hypothèses* (A.1)–(A.7) *ci-dessous on a* :

$$\frac{ISE(\hat{b})}{ISE(b_0)} \longrightarrow 1, \quad p.s.,$$

où $b_0 \triangleq \operatorname{argmin}_{b\in B_n'} ISE(b)$ *et* $\hat{b} \triangleq \operatorname{argmin}_{b\in B_n'} CV(b)$.

---

1. Introduction

Let $Y_1,\ldots,Y_n$ be i.i.d. replicates from the corrupted random variable

$$Y = X + \varepsilon, \tag{1}$$

where X and ε are independent. Assume that g, f, h are the densities of Y, X, and ε, respectively, and suppose that h is known. We are going to consider the problem of estimating f using the sample $Y_1,\ldots,Y_n$, our main focus is to build a data-driven bandwidth and show that it is asymptotically optimal. For a function τ let Φ_τ denotes its Fourier transform that is

$$\Phi_\tau(t) \triangleq \int e^{ixt}\tau(x)\,dx. \tag{2}$$

The kernel deconvolution estimator of f is defined by

$$f_b(x) \triangleq \frac{1}{nb}\sum_{j=1}^{n} g_n\left(\frac{x-Y_i}{b}\right), \tag{3}$$

where

$$g_n(x) \triangleq \frac{1}{2\pi} \int \mathrm{e}^{-\mathrm{i}xt} \frac{\Phi_K(t)}{\Phi_h(t/b_n)} \,\mathrm{d}t, \tag{4}$$

K the kernel function, $b = b_n$ is the bandwidth. The kernel g_n is known as the deconvolution kernel in the literature. The estimator (3) has been abundantly studied, by many authors, we can cite among others Fan [1,2], Stefanski and Carroll [12], and Masry [8–10]. In most of these papers, depending on how $\Phi_h(t)$ goes to zero as t goes to infinity, the authors give the rate of convergence of the kernel estimator. In particular Fan [2] gives the optimal rates of convergence, Fan [3] gives conditions under which the estimator is asymptotically normal. Masry [10] provides pointwise and uniform strong consistency rates. In this paper, under the restriction that $\Phi_h(t) = \mathrm{O}(t^{-\beta})$ as $t \to \infty$, we will go a bit further and propose a bandwidth selection rule. In this paper we develop the cross-validation ideas as those in Rudemo [11], Bowman [1], and Hall [4] for the ordinary kernel estimator. Our rule is motivated by the integrated squared error representation:

$$\begin{aligned} ISE(b) &= \int \left| f_b(x)\right|^2 \mathrm{d}x - 2\,\mathrm{Re}\left(\int f_b(x) f(x)\,\mathrm{d}x\right) + \int f^2(x)\,\mathrm{d}x \\ &\triangleq FT(b) - 2ST(b) + TT, \end{aligned}$$

where for any complex number z, $\mathrm{Re}(z)$ denotes its real part. Since the last summand is independent of b, the goal if minimizing $ISE(b)$ is equivalent to that of minimizing $FT(b) - 2ST(b)$. On the other hand, using Parseval equality we have

$$\begin{aligned} ST(b) &= \mathrm{Re}\left(\int f_b(x) f(x)\,\mathrm{d}x\right) \\ &= \mathrm{Re}\left(\frac{1}{2\pi}\int \Phi_{f_b}(t)\overline{\Phi}_f(t)\,\mathrm{d}t\right) \\ &= \mathrm{Re}\left(\frac{1}{2\pi}\int \frac{\Phi_K(bt)\Phi_n(t)}{\Phi_h(t)}\overline{\Phi}_f(t)\,\mathrm{d}t\right) \\ &= \mathrm{Re}\left(\frac{1}{2\pi}\int \frac{\Phi_K(bt)\Phi_n(t)}{|\Phi_h(t)|^2}\overline{\Phi}_g(t)\,\mathrm{d}t\right). \end{aligned}$$

Noting that

$$\alpha_n(t) \triangleq \frac{1}{n(n-1)} \sum_{j\neq k} \mathrm{e}^{-\mathrm{i}t(Y_k - Y_j)}$$

is an unbiased estimator of $\Phi_g(t)\overline{\Phi}_g(t)$ we can approximate $ST(b)$ by

$$\widehat{ST}(b) \triangleq \frac{1}{n(n-1)} \sum_{j\neq k} \mathrm{Re}\,\frac{1}{2\pi}\int \frac{\Phi_K(tb)}{|\Phi_h(t)|^2} \mathrm{e}^{-\mathrm{i}t(Y_k - Y_j)}\,\mathrm{d}t.$$

It is clear that, $E(\widehat{ST}(b)) = E(ST(b))$. Therefore, the proposed cross-validation functional is given by

$$CV(b) \triangleq \int |f_b|^2(x)\,\mathrm{d}x - 2\widehat{ST}(b). \tag{5}$$

It seems reasonable to choose b minimizing $CV(b)$ as our proposed bandwidth for estimating the density f.

In the next section, we will state conditions under which the minimizer of (5) minimizes $ISE(b)$ asymptotically. The remainder of the paper is composed of the proofs of this result.

2. The optimality result

In order to prove the main result of this paper we will assume,

(A.1) $\int |x| h(x)\,\mathrm{d}x < \infty$ and $\int |x|^2 h(x)\,\mathrm{d}x < \infty$,

(A.2) $t^\beta \Phi_h(t) \to c$ as $t \to \infty$, for some $\beta > 1$, $c \neq 0$ and $\Phi_h(t) \neq 0$, $\forall t \in \mathbb{R}$,

(A.3) K is symmetric, $\int K(x)\,\mathrm{d}x = 1$, $\int |x| K(x) < \infty$ and $0 < \int |x|^2 K(x)\,\mathrm{d}x < \infty$,

(A.4) $\int |t|^{2\beta} |\Phi_K(t)|\,\mathrm{d}t < \infty$,

(A.5) $\int |t|^{\beta-1} |\Phi'_K(t)|\,\mathrm{d}t < \infty$ and $\int |t|^\beta |\Phi''_K(t)|\,\mathrm{d}t < \infty$,

(A.6) g, f, h are in L^2; f is twice continuously differentiable; and f, f'' and g are bounded,

(A.7) Let

$$B_n = \left[An^{-1/(2\beta+5)-\nu}, Bn^{-1/(2\beta+5)+\nu}\right], \quad 0 < A < B < \infty,\ 0 < \nu < \frac{1}{(2\beta+5)(2\beta+6)}$$

and let B'_n be a finite subset of B_n such that $\#B'_n = n^\rho$, for some $\rho > 0$.

Remark 1. – Conditions (A.1)–(A.7), are set of assumptions needed to obtain asymptotic equivalence between measures of accuracy in Youndjé and Wells [13]. The main restriction is to assume that $\beta > 1$. This automatically excludes the supersmooth case (*see* Fan [2]). But, this case is rather difficult to deal with, because the rates of convergence are drastically slow in the supersmooth case (Fan [2]).

Theorem 2.1 below is the main result of this paper. It shows that minimizing $CV(b)$ yields an optimal bandwidth.

THEOREM 2.1. – *Under assumptions* (A.1)–(A.7) *we have*

$$\frac{ISE(\hat{b})}{ISE(b_0)} \longrightarrow 1, \quad a.s.,$$

where $b_0 \triangleq \operatorname{argmin}_{b \in B'_n} ISE(b)$ *and* $\hat{b} \triangleq \operatorname{argmin}_{b \in B'_n} CV(b)$.

After completing this work, it was brought to our attention that a paper by Hesse [6] presents Theorem 2.1 in probability, and we would like to stress that this work was done independently.

3. Sketch of the proof of Theorem 1

Let us set

$$T \triangleq \frac{2}{n} \sum_{k=1}^{n} f(X_k) - \int f^2(x)\,\mathrm{d}x,$$

since T is independent of b, according to Härdle and Marron [5] (proof of Theorem 1), to prove Theorem 2.1, we are reduced to proving that,

$$\sup_{b \in B'_n} \left| \frac{ISE(b) - CV(b) - T}{ISE(b)} \right| \longrightarrow 0 \quad \text{a.s.} \tag{6}$$

It can be shown that for n large enough, there is a positive constant C such that

$$ISE(b) \geqslant \frac{C}{nb^{2\beta+1}} \quad \text{a.s.}$$

thus we are reduced to proving that

$$\sup_{b \in B'_n} nb^{2\beta+1} \left| ISE(b) - CV(b) - T \right| \longrightarrow 0 \quad \text{a.s.} \tag{7}$$

Let us set

$$U_{jk} \triangleq \frac{1}{2\pi} \operatorname{Re} \int \frac{\Phi_K(bt)}{|\Phi_h(t)|^2} \mathrm{e}^{-\mathrm{i}t(Y_k - Y_j)} \,\mathrm{d}t - \frac{1}{2\pi} \operatorname{Re} \int \frac{\Phi_K(bt)\mathrm{e}^{\mathrm{i}tY_j}\overline{\Phi}_g(t)\,\mathrm{d}t}{|\Phi_h(t)|^2} - f(X_k) + \int f^2(x)\,\mathrm{d}x$$

we have that

$$ISE(b) - CV(b) - T = \frac{2}{n(n-1)} \sum_{j \neq k} U_{jk}.$$

Therefore using the same steps as in the proof of Lemma 2 in Marron [7] it can be shown that

$$\sup_{b \in B_n'} nb^{2\beta+1} \left| \frac{1}{n(n-1)} \sum_{j \neq k} U_{jk} \right| \longrightarrow 0 \quad \text{a.s.,} \tag{8}$$

which completes the proof of our theorem. A comprehensive proof can be obtained from the authors on request.

References

[1] A.W. Bowman, An alternative method of cross-validation for the smoothing of density estimates, Biometrika 72 (1984) 353–360.
[2] J. Fan, On the optimal rates of convergence for nonparametric deconvolution problems, Ann. Statist. 19 (1991) 1257–1272.
[3] J. Fan, Asymptotic normality for deconvolution kernel density estimators, Sankhya Ser. A 53 (1991) 97–110.
[4] P. Hall, Large sample optimality of least squares cross-validation in density estimation, Ann. Statist. 11 (1983) 1156–1174.
[5] W. Härdle, J.S. Marron, Optimal bandwith selection in nonparametric regression estimation, Ann. Statist. 13 (1985) 1465–1481.
[6] C.H. Hesse, Data-driven deconvolution, Nonparametric Statist. 10 (1999) 343–373.
[7] J.S. Marron, A comparison of cross-validation techniques in density estimation, Ann. Statist. 15 (1987) 152–162.
[8] E. Masry, Multivariate probability density deconvolution for stationary random processes, IEEE Trans. Inform. Theory IT-37 (1991) 1105–1115.
[9] E. Masry, Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes, J. Multivariate Anal. 44 (1993) 47–68.
[10] E. Masry, Strong consistency and rates for deconvolution of multivariate densities of stationary processes, Stochastic Process. Appl. 47 (1993) 53–74.
[11] M. Rudemo, Empirical choice of histograms and kernel density estimators, Scand. J. Statist. 9 (1982) 65–78.
[12] L.A. Stefanski, R.J. Carroll, Deconvoluting kernel density estimators, Statistics 21 (1990) 169–184.
[13] E. Youndjé, M. Wells, Asymptotic equivalence of quadratic measure of errors for kernel deconvolution density, Technical report, Cornell Department of Statistical Sciences, 2000.