APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY
Appl. Stochastic Models Bus. Ind., 2005; **21**:175–185
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asmb.535

Some linear programming methods for frontier estimation

G. Bouchard[1,*,†], S. Girard[2], A. Iouditski[2] and A. Nazin[3,‡]

[1] *INRIA, ZIRST, 655 av. Europe, Montbonnot, Saint-Ismier 38334, Cedex, France*
[2] *SMS/LMC, Université Grenoble I, BP 53, Grenoble 38041, Cedex 9, France*
[3] *Institute of Control Sciences, RAS, Profsoyuznaya str., 65, Moscow 117997, Russia*

SUMMARY

We propose new methods for estimating the frontier of a set of points. The estimates are defined as kernel functions covering all the points and whose associated support is of smallest surface. They are written as linear combinations of kernel functions applied to the points of the sample. The weights of the linear combination are then computed by solving a linear programming problem. In the general case, the solution of the optimization problem is sparse, that is, only a few coefficients are non-zero. The corresponding points play the role of support vectors in the statistical learning theory. In the case of uniform bivariate densities, the L_1 error between the estimated and the true frontiers is shown to be almost surely converging to zero, and the rate of convergence is provided. The behaviour of the estimates on one finite sample situation is illustrated on simulations. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: functional estimate; linear programming; L_1 error; frontier estimation

1. INTRODUCTION

Many proposals are given in the literature for estimating a set S given a finite random set of points drawn from the interior. This problem of frontier or support estimation arises in classification [1], clustering problems [2], discriminant analysis [3] and outliers detection. Applications are found in medical diagnosis [4] as well as in condition monitoring of machines [5]. In image analysis, the segmentation problem can be considered under the support estimation point of view, where the support is a convex bounded set in $\mathbb{R}^2$ [6]. We also point out some applications in econometrics (e.g. Reference [7]). In such cases, the unknown support can be written

$$S = \{(x, y) : 0 \leqslant x \leqslant 1;\ 0 \leqslant y \leqslant f(x)\} \tag{1}$$

*Correspondence to: G. Bouchard, INRIA, ZIRST, 655 av. Europe, Montbonnot, 38334 Saint-Ismier Cedex, France.
†E-mail: Guillaume.Bouchard@inrialpes.fr
‡The work of A. Nazin has been carried out during his stay in INRIA Rhône-Alpes as invited professor, November–December 2002.

Contract/grant sponsor: Federal Office for Scientific, Technical and Cultural Affairs; contract/grant number: P5/24

where $f : [0, 1] \to (0, +\infty)$ is an unknown function. Here, the problem reduces to estimating f, called the production frontier (see, for instance, Reference [8]). The data consist of pair (X, Y) where X represents the input (labour, energy or capital) used to produce an output Y in a given firm. In such a framework, the value $f(x)$ can be interpreted as the maximum level of output which is attainable for the level of input x. Korostelev *et al.* [9] suppose f to be increasing and concave, from economical considerations, which suggests an adapted estimator, called the DEA (Data Envelopment Analysis) estimator. It is the lowest concave monotone increasing function covering all the sample points. Therefore it is piecewise linear and, up to our knowledge, it is the first frontier estimate computed thanks to a linear programming technique [10]. Its asymptotic distribution is established by Gijbels *et al.* [11].

An early paper was written by Geffroy [12] for independent identically distributed observations from a density ϕ. The proposed estimator is a kind of histogram based on the extreme values of the sample. This work was extended in two main directions.

On the one hand, piecewise polynomials estimates were introduced. They are defined locally on a given slice as the lowest polynomial of fixed degree covering all the points in the considered slice. Their optimality in an asymptotic minimax sense is proved under weak assumptions on the rate of decrease α of the density ϕ towards 0 by Korostelev and Tsybakov [6] and by Härdle *et al.* [13]. Extreme values methods are then proposed by Hall *et al.* [14] and by Gijbels and Peng [15] to estimate the parameter α.

On the other hand, different propositions for smoothing Geffroy's estimate were made in the case of a Poisson point process. Girard and Jacob [16] introduced estimates based on kernel regressions and orthogonal series method [17, 18]. In the same spirit, Gardes [19] proposed a Faber–Shauder estimate. Girard and Menneteau [20] introduced a general framework for studying estimates of this type and generalized them to supports writing

$$S = \{(x, y) : x \in E;\ 0 \leqslant y \leqslant f(x)\}$$

where $f : E \to (0, +\infty)$ is an unknown function and E an arbitrary set. In each case, the limit distribution of the estimator is established.

We also refer to Abbar [21] and Jacob and Suquet [22] who used a similar smoothing approach, although their estimates are not based on the extreme values of the Poisson process.

The estimate proposed in this paper can be considered to belong to the intersect of these two directions. It is defined as a kernel estimate obtained by smoothing some selected points of the sample. These points are chosen automatically by solving a linear programming problem to obtain an estimate of the support covering all the points and with smallest surface. Its advantages are the following: it can be computed with standard optimization algorithms (see e.g. Reference [23, Chapter 4]), its smoothness is directly linked to the smoothness of the chosen kernel and it benefits from interesting theoretical properties. For instance, we prove that it is almost surely convergent for the L_1 norm. The estimate is defined in Section 2. Its theoretical properties are established in Section 3. The behaviour of the estimate is illustrated in Section 4 on one finite sample situation. It is compared to a similar proposition found in Reference [24]. We conclude by describing our further work in Section 5. Proofs can be found in Reference [25].

2. BOUNDARY ESTIMATES

2.1. A linear programming problem

Let all the random variables be defined on a probability space $(\Omega, \mathscr{F}, P)$.

The problem under consideration is to estimate an unknown positive function $f : [0, 1] \to (0, \infty)$ on the basis of observations $Z_N = (X_i, Y_i)_{i=1,\ldots,N}$. The former represents an i.i.d. sequence with pairs (X_i, Y_i) being uniformly distributed in the set S defined as in (1). For the sake of simplicity, we consider in the following the extension of f on all $\mathbb{R}$ by introducing $f(x) = 0$ for all $x \notin [0, 1]$. Letting

$$C_f \triangleq \int_0^1 f(u)\,\mathrm{d}u = \int_{\mathbb{R}} f(u)\,\mathrm{d}u$$

each variable X_i is distributed in $[0, 1]$ with p.d.f. $f(\cdot)/C_f$ while Y_i has the uniform conditional distribution with respect to X_i in the interval $[0, f(X_i)]$.

The considered estimate of the frontier is chosen from the family of functions:

$$\begin{aligned} &\hat{f}_N(x) = \textstyle\sum_{i=1}^N K_h(x - X_i)\alpha_i, \quad K_h(t) = h^{-1}K(t/h) \\ &\alpha_i \geqslant 0, \quad i = 1, \ldots, N \end{aligned} \tag{2}$$

where K is a kernel function $K : \mathbb{R} \to [0, \infty)$ integrating to one and with bandwidth $h > 0$. Each coefficient α_i represents the importance of the point (X_i, Y_i) in the estimation. In particular, if $\alpha_i \neq 0$, the corresponding point (X_i, Y_i) can be called a support vector by analogy with support vector machines (SVM). We refer to Reference [26] for a review on this topic and to Reference [27, Chapter 8], for examples of application of SVM to quantile estimation. The constraint $\alpha_i \geqslant 0$ for all $i = 1, \ldots, N$ ensures that $\hat{f}_N(x) \geqslant 0$ for all $x \in \mathbb{R}$ and prevents the estimator from being too irregular.

Let us remark that the surface of the estimated support is given by

$$\int_{\mathbb{R}} \hat{f}_N(x)\,\mathrm{d}x = \sum_{i=1}^N \alpha_i$$

This suggests to define the vector parameter $\alpha = (\alpha_1, \ldots, \alpha_N)^{\mathrm{T}}$ from a linear program as follows:

$$J_P^* \triangleq \min_{\alpha} \mathbf{1}^{\mathrm{T}}\alpha \tag{3}$$

subject to

$$A\alpha \geqslant Y \tag{4}$$

$$\alpha \geqslant 0 \tag{5}$$

The following notations have been introduced:

$$\mathbf{1} \triangleq (1, 1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^N$$

$$A \triangleq \{K_h(X_i - X_j)\}_{i,j=1,\ldots,N}$$

$$Y \triangleq (Y_1, \ldots, Y_N)^{\mathrm{T}}$$

Hence, $A_\alpha = (\hat{f}_N(X_1), \ldots, \hat{f}_N(X_N))^{\mathrm{T}}$, and the vector constraint (4) means that $\hat{f}_N(X_i) \geqslant Y_i$, $i = 1, \ldots, N$. In other words, $\hat{f}_N$ defines the kernel estimate of the support covering all the points and with smallest surface. In practice (see Section 4 for an illustration) the solution of the linear program is sparse in the sense that $n(\alpha) = \#\{\alpha_i \neq 0\}$ is small (for a properly selected value of h, see Section 2.2) and thus the resulting estimate is fast to compute even for large samples.

Let us note that the above-described estimator (2)–(5) might be derived as the maximum likelihood estimate related to the approximation family (2). Indeed, the joint probability density function for observations Z_N given parameter function $f(x)$ can be written

$$p(Z_N \,|f) = \prod_{i=1}^{N} \frac{f(X_i)}{C_f} \frac{1}{f(X_i)} \mathbf{1}\{0 \leqslant Y_i \leqslant f(X_i)\}$$

where $\mathbf{1}\{.\}$ is the indicator function. Moreover,

$$C_f|_{f=\hat{f}_N} = \sum_{i=1}^{N} \alpha_i$$

and therefore, the Log-Likelihood function is

$$L(\alpha) \triangleq \log p(Z_N \,|\hat{f}_N) = -N \log \sum_{i=1}^{N} \alpha_i + \sum_{i=1}^{N} \log \mathbf{1}\{Y_i \leqslant \hat{f}_N(X_i)\}$$

and its maximization over the set of non-negative parameters α is equivalent to problem (3)–(5).

2.2. Choice of the bandwidth

The bandwidth parameter h has to be properly selected. The asymptotic results presented in Section 3 (see Corollaries 1 and 2) only provide an order of magnitude and thus are useless in practical situations. Assuming that (X, Y) is uniformly distributed on S, we have two different estimations of $E[Y]$:

- $\hat{m}_1 = (1/N) \sum_{i=1}^{N} Y_i$.
- $\hat{m}_2 = (1/N) \sum_{i=1}^{N} \hat{f}_N(X_i; h)/2$

These estimations are expected to be equal whenever the estimated frontier $\hat{f}_N$ is close to f. Hence, we propose to select the value $\hat{h}_N$ minimizing the quantity

$$D(h) = \frac{1}{N} \left| \sum_{i=1}^{N} Y_i - \frac{1}{2} \sum_{i=1}^{N} \hat{f}_N(X_i; h) \right|$$

This criterion is tested on simulations in Section 4.

2.3. Comparison with other methods

Let us remark that other solution for estimating α in (2) have already been proposed. Girard and Menneteau [20] considered a partition $\{I_r : 1 \leqslant r \leqslant k\}$ of $[0, 1]$, with $k \to \infty$. For all $1 \leqslant r \leqslant k$, they introduce

$$D_r = \{(x, y) : x \in I_r, \;\; 0 \leqslant y \leqslant f(x)\}$$

the slice of S built on I_r, $Y_r^* = \max\{Y_i; (X_i, Y_i) \in D_r\}$, and the estimates

$$\hat{\alpha}_i = \begin{cases} \lambda(I_r) Y_r^* & \text{if } \exists r \in \{1, \ldots, k\};\ Y_i = Y_r^* \\ 0 & \text{otherwise} \end{cases}$$

where λ is the Lebesgue measure. They propose the following frontier estimate

$$\hat{f}_N(x) = \sum_{r=1}^{k} K_h(x - x_r)\lambda(I_r) Y_r^*$$

where x_r is the centre of I_r. This approach suffers from a practical difficulty: the choice of the partition and more precisely the choice of k. In our context, solving the linear problem (3)–(5) directly yields the support vectors, once h has been chosen with the method described in Section 2.2.

In this sense, the estimate proposed in Reference [24] is similar to $\hat{f}_N$. It is defined by the Fourier expansion:

$$\hat{g}_N(x) = c_0 + \sum_{k=1}^{M} a_k \cos(2\pi k x) + \sum_{k=1}^{M} b_k \sin(2\pi k x)$$

where the vector of parameters $\beta = (c_0, a_1, \ldots, a_M, b_1, \ldots, b_M)^{\mathrm{T}}$ is solution of the linear programming problem:

$$\min c_0 \left(= \int_0^1 \hat{g}_N(x)\, \mathrm{d}x \right) \tag{6}$$

under the constraints

$$\hat{g}_N(X_i) \geqslant Y_i, \quad i = 1, \ldots, N \tag{7}$$

$$\sum_{k=1}^{M} k(|a_k| + |b_k|) \leqslant L/(2\pi) \tag{8}$$

Therefore, $\hat{g}_N$ defines the Fourier estimate of the support covering all the points (Equation (7)), L-Lipschitzian (Equation (8)) and with smallest surface (Equation (6)). From the theoretical point of view, this estimate benefits from minimax optimality. It is compared to $\hat{f}_N$ on one numerical example in Section 4 for different choices of parameters M, L and h.

3. THEORETICAL RESULTS

In this section, we establish that $\hat{f}_N$ is almost surely convergent for the L_1 norm on $[0, 1]$. To this end, the basic assumptions on the unknown boundary function are introduced:

A1. $0 < f_{\min} \leqslant f(x) < f_{\max} < \infty$, for all $x \in [0, 1]$,
A2. $|f(x) - f(y)| \leqslant L_f |x - y|$, for all $x, y \in [0, 1]$, $L_f < \infty$.

The following assumptions on the kernel function are considered:

B1. $K(t) = K(-t) \geqslant 0$,
B2. $\int K(t)\, \mathrm{d}t = 1$,

B3. $|K(s) - K(t)| \leqslant L_K|s - t|$, $L_K < \infty$,
B4. $\int K^2(t)\,\mathrm{d}t < \infty$ and $\int t^2 K(t)\,\mathrm{d}t < \infty$.

In the sequel, we note $\|\hat{f}_N - f\|_1 = \int |\hat{f}_N(x) - f(x)|\,\mathrm{d}x$.

Theorem 1

Let $h \to 0$ and $\log N/(Nh^2) \to 0$ as $N \to \infty$. Let the above-mentioned assumptions A and B hold true. Then, estimator (2)–(5) has the following asymptotic properties:

$$\limsup_{N\to\infty} \varepsilon_1^{-1}(N)\|\hat{f}_N - f\|_1 \leqslant C < \infty \quad \text{a.s.}$$

with $\varepsilon_1(N) \triangleq \max\{h, \sqrt{\log N/(Nh^2)}\}$.

Corollary 1

The maximum rate of convergence which is guaranteed by Theorem 1

$$\|\hat{f}_N - f\|_1 = O_p((\log N/N)^{1/4})$$

is attained for $h \asymp (\log N/N)^{1/4}$.

This rate of convergence can be ameliorated at the price of a slight modification of the estimate. In the following, an additional constraint is considered in order to impose to each coefficient α_i to be of order $1/N$. The counterpart of this modification is that the new estimate $\tilde{f}_N$ will usually rely on more support vectors than $\hat{f}_N$.

Let us modify estimator (2)–(5) as follows:

$$\tilde{f}_N(x) = \sum_{i=1}^{N} K_h(x - X_i)\alpha_i \tag{9}$$

where vector $\alpha = (\alpha_1, \ldots, \alpha_N)^{\mathrm{T}}$ is defined from the modified linear program

$$J^*_{MP} \triangleq \min_{\alpha} \mathbf{1}^{\mathrm{T}}\alpha \tag{10}$$

subject to

$$A\alpha \geqslant Y \tag{11}$$

$$0 \leqslant \alpha \leqslant C_\alpha/N \tag{12}$$

with a constant

$$C_\alpha > f_{\max} \tag{13}$$

Remark

In fact, we need to ensure $C_\alpha > C_f$ which is implied by (13).

The modified estimator (9)–(13) differs from that of (2)–(5) by additionally bounding each α_i from above, see constraints (12).

Theorem 2

Let $h \to 0$ and $\log N/(Nh) \to 0$ as $N \to \infty$. Let kernel function $K(\cdot)$ has a finite support, that is $K(t) = 0\ \forall |t| \geqslant 1$, and the assumptions A and B hold true. Then, estimator (9)–(13) has the following asymptotic properties:

$$\limsup_{N\to\infty} \varepsilon_2^{-1}(N)\|\tilde{f}_N - f\|_1 \leqslant C < \infty \quad \text{a.s.}$$

with $\varepsilon_2(N) \triangleq \max\{h, \sqrt{\log N/(Nh)}\}$.

Remark

The support of $K(\cdot)$ is fixed to be the interval $[-1, 1]$ without loss of generality.

Corollary 2

The maximum rate of convergence which is guaranteed by Theorem 2

$$\|\tilde{f}_N - f\|_1 = O_p((\log N/N)^{1/3})$$

is attained for $h \asymp (\log N/N)^{1/3}$.

Note that the rates of convergence of the DEA and FDH estimators are $O_p(N^{-2/3})$ (see Reference [28]) and $O_p(N^{-1/2})$ (see Reference [29]), but stronger assumptions are required (monotonicity or concavity of the frontier function f).

4. NUMERICAL EXPERIMENTS

The simulations presented here illustrate the behaviour of the kernel estimator $\hat{f}_N$ compared to the estimator based on Fourier expansions $\hat{g}_N$ proposed in Reference [24]. Since the Fourier estimate $\hat{g}_N$ requires the unknown function to be periodic, we choose f such that $f(0) = f(1)$. Besides, to avoid boundary effects on the input domain, we consider functions that are nearly zero when x is close to 0 or 1. In more general situations, boundary corrections should be implemented (see Reference [30]). The chosen function

$$\begin{aligned} f(x) = 0.1 &+ 5(x - 0.1)\mathbf{1}_{\{x>0.1\}} \\ &- 5(x - 0.2)\mathbf{1}_{\{x>0.2\}} \\ &+ 1(x - 0.5)\mathbf{1}_{\{x>0.5\}} \\ &- 9(x - 0.8)\mathbf{1}_{\{x>0.8\}} \\ &+ 8(x - 0.9)\mathbf{1}_{\{x>0.9\}} \end{aligned}$$

is piecewise linear and locally Lipschitzian with a Lipschitz constant $L_f = 8$. Its graph is given on the right part of Figure 1 (dotted line). Here, $N = 50$ points are uniformly generated on the domain S upper bounded by f (left part of Figure 1). The smoothing parameter h is chosen according to Section 2.2. On the right, the kernel estimate (continuous line) is superimposed to the unknown function f. The squares represent the points for which $\hat{f}_N(X_i) = Y_i$. The triangles

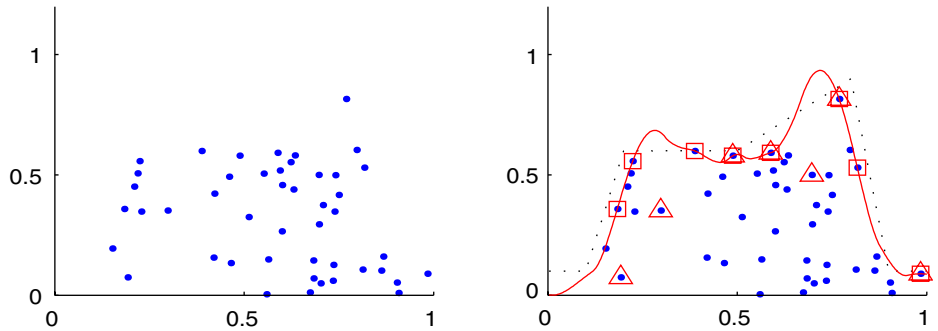

Figure 1. Illustration of the kernel estimation on a simulation.

Table I. Results for 1000 simulations with $N = 25$ points.

Estimate	h	L	M	Δ_N	np	$D(h)$
Kernel	0.120			0.113 (0.033)	4.476 (0.840)	43.692
	0.150			**0.111 (0.032)**	**3.613 (0.649)**	**38.232**
	0.180			0.122 (0.027)	3.093 (0.623)	34.362
	0.210			0.137 (0.023)	2.702 (0.590)	32.258
	0.240			0.153 (0.023)	2.434 (0.557)	33.793
	0.270			0.163 (0.028)	2.184 (0.403)	37.642
Fourier		3	4	0.142 (0.036)	4.534 (0.741)	
		5	**4**	**0.116 (0.043)**	**5.532 (1.008)**	
		7	4	0.124 (0.040)	6.636 (1.193)	
		9	4	0.135 (0.040)	7.312 (1.235)	
		11	4	0.143 (0.041)	7.690 (1.219)	
		13	4	0.150 (0.041)	7.870 (1.172)	
Fourier		3	8	0.143 (0.036)	4.545 (0.757)	
		5	**8**	**0.117 (0.043)**	**5.574 (1.055)**	
		7	8	0.126 (0.041)	6.771 (1.321)	
		9	8	0.139 (0.042)	7.691 (1.584)	
		11	8	0.150 (0.042)	8.376 (1.756)	
		13	8	0.159 (0.042)	8.877 (1.900)	

The mean value of Δ_N and np is given with the standard deviation in parentheses.

represent the support vectors (i.e. the points for which $\alpha_i > 0$). On this example, the estimate $\hat{g}_N$ is not visually different from $\hat{f}_N$.

Thus, a more precise comparison should be done. For each estimate, the L_1 error Δ_N as well as the number of effective parameters np (that is n_α and $n_\beta = \#\{\beta_i \neq 0\}$) are evaluated for $N = 25$ and 100. The average value and the standard deviation of these quantities are computed on 1000 replications. The estimation is carried out with different values of the parameters, namely h for the kernel estimate, and L and M for the Fourier estimate. For each value of h the criterion $D(h)$ presented in Section 2.2 is evaluated. The adaptive choice of L and M parameters is not implemented in this setting. The results are summarized in Tables I and II. The lowest

Table II. Results for 1000 simulations with $N = 100$ points.

Estimate	h	L	M	Δ_N	np	$D(h)$
Kernel	0.050			0.072 (0.015)	13.777 (1.351)	31.972
	0.070			0.058 (0.013)	9.976 (1.249)	23.200
	0.090			**0.058 (0.011)**	**7.558 (1.104)**	**17.611**
	0.110			0.061 (0.011)	5.696 (0.887)	15.210
	0.130			0.073 (0.011)	4.725 (0.714)	15.881
	0.150			0.082 (0.012)	3.964 (0.524)	17.337
Fourier		3	4	0.123 (0.021)	5.080 (0.716)	
		5	4	0.073 (0.019)	5.792 (0.781)	
		7	**4**	**0.060 (0.011)**	**7.850 (0.979)**	
		9	4	0.063 (0.012)	8.751 (0.579)	
		11	4	0.067 (0.014)	8.896 (0.362)	
		13	4	0.069 (0.015)	8.955 (0.235)	
Fourier		3	8	0.124 (0.021)	5.117 (0.752)	
		5	8	0.073 (0.020)	5.889 (0.863)	
		7	8	0.057 (0.012)	8.315 (1.470)	
		9	**8**	**0.057 (0.013)**	**10.620 (1.672)**	
		11	8	0.061 (0.014)	12.454 (1.858)	
		13	8	0.067 (0.014)	13.890 (1.879)	

The mean value of Δ_N and np is given with the standard deviation in parentheses.

error is emphasized for each estimate. It can be noted that the mean L_1 error of both estimates are very similar. In fact, the kernel estimate seems to give a slight lower error for small number of points and the Fourier estimate yields better results for large sample size situations, confirming its asymptotic optimality. Let us note that the standard deviation of the L_1 error is in general smaller for the kernel estimate. Regarding the number of parameters, the kernel estimate seems to be more parsimonious than the Fourier estimate. Finally, we can notice that our criterion $D(h)$ yields a reasonable choice of h, although a slight oversmoothing appears for small sample sizes.

5. FURTHER WORK

As noticed in Section 3, the rate of convergence of the L_1 error is not optimal. This problem can be overcomed by modifying the constraints of the linear programming problem in order to impose to the estimate the same Lipschitz properties as the unknown function f. We also plan to adapt our proofs to more general settings, namely non-uniform distributions with multi-dimensional support. The extension to the case where X is multidimensional is straightforward since it suffices to consider multivariate kernels. The non-uniform case requires assumptions on the behaviour of the distribution in the unknown boundary neighbourhood. From the practical point of view, the selection procedure of the smoothing parameter has to be improved for small sample sizes.

ACKNOWLEDGEMENTS

S. Girard was financially supported from the IAP research network nr P5/24 of the Belgian Government (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.

REFERENCES

1. Hardy A, Rasson JP. Une nouvelle approche des problèmes de classification automatique. *Statistique et Analyse des données* 1982; **7**:41–56.
2. Hartigan JA. *Clustering Algorithms*. Wiley: Chichester, 1975.
3. Baufays P, Rasson JP. A new geometric discriminant rule. *Computational Statistics Quarterly* 1985; **2**:15–30.
4. Tarassenko L, Hayton P, Cerneaz N, Brady M. Novelty detection for the identification of masses in mammograms. In *Proceedings*, *Fourth IEE International Conference on Artificial Neural Networks*, Cambridge, 1995; 442–447.
5. Devroye LP, Wise GL. Detection of abnormal behavior via non parametric estimation of the support. *SIAM Journal of Applied Mathematics* 1980; **38**:448–480.
6. Korostelev AP, Tsybakov AB. Minimax theory of image reconstruction. In *Lecture Notes in Statistics*, vol. 82. Springer: New York, 1993.
7. Deprins D, Simar L, Tulkens H. Measuring labor efficiency in post offices. In *The Performance of Public Enterprises: Concepts and Measurements*, Marchand M, Pestieau P, Tulkens H (eds). North-Holland: Amsterdam, 1984.
8. Härdle W, Hall P, Simar L. Iterated bootstrap with application to frontier models. *Journal of Productivity Analysis* 1995; **6**:63–76.
9. Korostelev AP, Simar L, Tsybakov AB. Efficient estimation of monotone boundaries. *The Annals of Statistics* 1995; **23**:476–489.
10. Charnes A, Cooper WW, Rhodes E. Measuring the inefficiency of decision making units. *European Journal of Operational Research* 1978; **2**:429–444.
11. Gijbels I, Mammen E, Park BU, Simar L. On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association* 1999; **94**:220–228.
12. Geffroy J. Sur un problème d'estimation géométrique. *Publications de l'Institut de Statistique de l'Université de Paris* 1964; **XIII**:191–200.
13. Härdle W, Park BU, Tsybakov AB. Estimation of a non sharp support boundaries. *Journal of Multivariate Analysis* 1995; **43**:205–218.
14. Hall P, Nussbaum M, Stern SE. On the estimation of a support curve of indeterminate sharpness. *Journal of Multivariate Analysis* 1997; **62**:204–232.
15. Gijbels I, Peng L. Estimation of a support curve via order statistics. *Discussion Paper 9905*, Institut de Statistique, Université Catholique de Louvain, 1999.
16. Girard S, Jacob P. Extreme values and kernel estimates of point processes boundaries. *Technical Report ENSAM-INRA-UM2*, 01-02, 2001.
17. Girard S, Jacob P. Extreme values and Haar series estimates of point processes boundaries. *Scandinavian Journal of Statistics* 2003a; **30**:369–384.
18. Girard S, Jacob P. Projection estimates of point processes boundaries. *Journal of Statistical Planning and Inference* 2003b; **116**:1–15.
19. Gardes L. Estimating the support of a Poisson process via the Faber–Shauder basis and extreme values. *Publications de l'Institut de Statistique de l'Université de Paris* 2002; **XXXXVI**:43–72.
20. Girard S, Menneteau L. Limit theorems for smoothed extreme values estimates of point processes boundaries. *Journal of Statistical Planning and Inference* 2005; to appear.
21. Abbar H. Un estimateur spline du contour d'une répartition ponctuelle aléatoire. *Statistique et analyse des données* 1990; **15**(3):1–19.
22. Jacob P, Suquet P. Estimating the edge of a Poisson process by orthogonal series. *Journal of Statistical Planning and Inference* 1995; **46**:215–234.
23. Bonnans FJ, Gilbert JC, Lemaréchal C, Sagastizábal C. Optimisation numérique. Aspects théoriques et pratiques. In *Mathématiques & Applications*, vol. 27. Springer: Paris, 1997.
24. Barron AR, Birgé L, Massart P. Risk bounds for model selection via penalization. *Probability Theory and Related Fields* 1999; **113**:301–413.
25. Bouchard G, Girard S, Iouditski A, Nazin A. Linear programming problems for frontier estimation. *Technical Report INRIA*, RR-4717, 2003. `http://www.inria.fr/rrrt/rr-4717.html`
26. Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines*. Cambridge University Press: Cambridge, 2000.
27. Schölkopf B, Smola A. *Learning With Kernels*. MIT University Press: Cambridge, 2002.

28. Kneip A, Park BU, Simar L. A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory* 1998; **14**:783–793.
29. Park BU, Simar L, Wiener Ch. The FDH estimator for productivity efficiency scores: asymptotic properties. *Econometric Theory* 2000; **16**:855–877.
30. Cowling A, Hall P. On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society B* 1996; **58**:551–563.

