The Annals of Statistics
1997, Vol. 25, No. 3, 970–981

ESTIMATION OF UNIMODAL DENSITIES WITHOUT SMOOTHNESS ASSUMPTIONS

BY LUCIEN BIRGÉ

Université Paris VI and URA CNRS 1321

The Grenander estimator of a decreasing density, which is defined as the derivative of the concave envelope of the empirical c.d.f., is known to be a very good estimator of an unknown decreasing density on the half-line $\mathbb{R}^+$ when this density is not assumed to be smooth. It is indeed the maximum likelihood estimator and one can get precise upper bounds for its risk when the loss is measured by the $\mathbb{L}^1$-distance between densities. Moreover, if one restricts oneself to the compact subsets of decreasing densities bounded by H with support on $[0, L]$ the risk of this estimator is within a fixed factor of the minimax risk. The same is true if one deals with the maximum likelihood estimator for unimodal densities with known mode. When the mode is unknown, the maximum likelihood estimator does not exist any more. We shall provide a general purpose estimator (together with a computational algorithm) for estimating nonsmooth unimodal densities. Its risk is the same as the risk of the Grenander estimator based on the knowledge of the true mode plus some lower order term. It can also cope with small departures from unimodality.

1. Introduction. Nonparametric density estimation has mainly been devoted, for a long time, to estimation of smooth densities (Hölderian or so, say) using linear methods like kernel estimators with fixed bandwidth or projection estimators (truncated series expansions with estimated coefficients). Typical examples would be (among many others) the papers by Bretagnolle and Huber (1979) or Efroimovich and Pinsker (1982). More recently, considerable attention has been given to estimation of functions with inhomogeneous smoothness, which requires new nonlinear methods such as variable bandwidth kernels or threshold estimators. Some illustrations can be found, for instance, in Kerkyacharian and Picard (1992) or Donoho and Johnstone (1994). These new methods, known under the generic term of "spatially adaptive," try to adjust automatically the smoothness of the estimator to the local smoothness of the unknown underlying density f. A very simple example is as follows: even if we know that f is decreasing on $[0, 1]$ and bounded by H, it can very well be steep at some places and flat elsewhere. Consequently, the best histogram for estimating f need not be based on a regular partition as can be seen from the results of Birgé (1987a, b) or Kogure (1987). In this particular situation (estimation of a decreasing density on $\mathbb{R}^+$) a special estimator, which takes the form of a variable binwidth histogram, has been known for a long time. It has been introduced by Grenander (1956) as the derivative of the least con-

Received June 1987; revised September 1996.
AMS 1991 *subject classifications*. Primary 62G05, 62G07; secondary 41A25.
Key words and phrases. Curve estimation, unimodal densities, Grenander estimator, spatial adaptation.

cave majorant of the empirical distribution function F_n which is merely the nonparametric maximum likelihood estimator (m.l.e. for short) restricted to decreasing densities on $\mathbb{R}^+$ [for a proof see, for instance, Grenander (1981) or Barlow, Bartholomew, Bremner and Brunk (1972)]. The so-called "Grenander estimator" has been studied by Robertson (1967), Prakasa Rao (1969), Kiefer and Wolfowitz (1976), Grenander (1981), Groeneboom (1985), Lo (1986) and Birgé (1989); the papers by Groeneboom and Birgé provide precise results about its asymptotic and nonasymptotic risk, respectively, when the loss function is taken to be the $\mathbb{L}^1$-distance between the densities. It can be derived from Birgé (1989) that the Grenander estimator approximately achieves the minimax risk over the class of decreasing densities on $[0, 1]$ bounded by H. This risk is of magnitude $C(H)n^{-1/3}$, as in the case of densities with variation bounded by H but with a different form of the constant $C(H)$ [see Birgé (1987a) for details]. More surprisingly, the nonasymptotic evaluations of Birgé (1989) show that the Grenander estimator (which is a histogram) generates a partition which is approximately the best one for the estimation of the unknown f with the value of n at hand. It is therefore an excellent example of a spatially adaptive estimator.

The construction of the Grenander estimator can easily be extended to unimodal densities with a known mode at M (which need not be unique). In a similar way, the m.l.e. restricted to unimodal densities with mode at M is defined as the derivative of the distribution function obtained by the union of the greatest convex minorant of F_n over $(-\infty, M]$ and its least concave majorant over $[M, +\infty)$. The preceding results carry over without any difficulty to this new Grenander estimator but it should be noted that this construction depends in a crucial way on an a priori knowledge of the position of the mode, which is clearly unrealistic for most practical problems. If the mode is unknown, the preceding method fails since the m.l.e. over the family of all unimodal densities does not exist any more because it tends to put an infinite density at one of the observations. Estimation of unimodal densities has been studied by Wegman in a series a papers [see Wegman (1968, 1969, 1970a, b)]. In order to deal with an unknown mode, he restricts the maximization of the likelihood to a smaller class of functions by assuming a modal interval of length greater than some positive ε which is a tuning parameter to be chosen by the statistician. The asymptotic properties of this constrained m.l.e. are the same as those of the Grenander estimator except for an interval of length ε around the mode and it is known to have a spike near the mode when ε is too small. Other estimators have been designed by Reiss (1973, 1976) and Prakasa Rao (1983) and further references and results can be found in Barlow, Bartholomew, Bremner and Brunk (1972) or Robertson, Wright and Dykstra (1988).

Since the Grenander estimator is a very good spatially adaptive estimator, even from a nonasymptotic point of view, it would be desirable, in the case of an unknown mode, to build an estimator which is close to the Grenander and to have a nonasymptotic evaluation of its performances as compared to those of the Grenander (which assumes a known mode). The following study was essentially motivated because we did not know of any result in this direction. Our main conclusion is that it is not essentially more difficult to estimate a nonsmooth unimodal density when the mode is unknown than if it were known. We shall actually provide in the next section a feasible method to solve this problem together with a precise upper bound on the additional $\mathbb{L}^1$-risk consequent to the fact that the mode is not known. As we shall see in (2.6) below (with $\eta = 1/n$), the increase of the risk (with n observations), as compared to the situation of a known mode, can be bounded by $6n^{-1/2}$ which is of smaller order than the overall risk (of magnitude $n^{-1/3}$ except for flat densities) that one incurs when estimating a nonsmooth unimodal density with a known mode.

Actually, the idea underlying the construction given in Definition 3 is extremely simple: consider all possible Grenander estimators, each one corresponding to a different position of the mode and choose the one which is closest to the empirical c.d.f. At first sight, this does not look like a very practical method, leading to an untractable optimization problem but we shall see in Section 3 that this problem can very well be solved approximately on a computer since it can be reduced to a simpler one. A practical implementation of this algorithm together with simulations and industrial applications have actually been successfully developed by Reboul.

The main features of this study, which make it quite different from related works on the subject are the following.

1. No assumption at all is made on the underlying density apart from the fact that it is unimodal. It can, in particular, have an unbounded support or be discontinuous.
2. Nothing is supposed concerning the behavior of the density near the mode which might not even be unique. In particular, the density might be very flat around the mode which would lead to a very slow rate of convergence of any estimator of the mode [as can easily be checked using methods similar to those of Has'minskii (1979)].
3. We provide an explicit construction of the estimator together with a computational algorithm and the main emphasis is put on a precise control of the risk which will be proved to be close and asymptotically equivalent to the risk of the Grenander estimator that one would build if the true mode were known.
4. The method is completely data driven and does not make use of any extra parameter to be chosen by the statistician (as Wegman's method does since it involves choosing the length of a modal interval) or any preliminary estimator of the mode.

It should be noticed that our estimator is not a restricted m.l.e. and since it is an "all purpose estimator" which is supposed to cope with possibly discontinuous, long-tailed or unbounded densities, it will not have a smooth appearance but merely be a histogram-looking estimator just like the Grenander estimator and it might be spiky near the mode. Its use should be limited to situations when the true density is not known to be smooth, since it is well known that histogram-type estimators are not optimal for estimating smooth densities. The estimation of smooth unimodal densities would clearly involve different methods as described in Mammen (1991), for instance.

As pointed out by Devroye (1987), the Grenander estimator (for a decreasing density) can only be consistent if the true density is decreasing and the same fact is true for our estimator which is by construction unimodal and therefore cannot estimate consistently a multimodal density. Nevertheless, this is an asymptotic point of view and, with a moderate number of observations, a unimodal estimator can perform rather well if the true density is reasonably close to unimodal. This is actually the philosophy underlying the construction of sieves estimators for nonparametric estimation [see, for instance, Birgé and Massart (1994)]. Actually our estimator will clearly be systematically biased for estimating a nonunimodal density but, apart from this bias, which will be moderate if the density is close to unimodal, the general behavior of the estimator will not be affected by this departure from unimodality. A similar issue was raised in Wegman (1968) although he was not concerned by precise nonasymptotic risk evaluations.

2. Construction and performance bounds for our estimator. Let us first give a precise definition of what we call a *unimodal* density.

DEFINITION 1. A density f on the real line is called unimodal if there exists some number M (not necessarily unique) such that f is nondecreasing on $(-\infty, M)$ and nonincreasing on $(M, +\infty)$. Any such M is called a mode of the density. The density f is said to be *decreasing* if $f(x) = 0$ for $x < M$ and *increasing* if $f(x) = 0$ for $x > M$.

One should notice that with such a definition the exponential or the uniform densities are unimodal as well as the normal one.

We now have to introduce a canonical way of building unimodal densities from arbitrary distribution functions.

DEFINITION 2. Let F be a distribution function on the line and r a continuity point of F. The *unimodal regularization* $\tilde{F}^r$ of F with mode at r is the continuous distribution function defined by $\tilde{F}^r(r) = F(r)$, $\tilde{F}^r$ is the largest convex minorant (or the convex envelope) of F on $(-\infty, r]$ and $\tilde{F}^r$ is the smallest concave majorant (or the concave envelope) of F on $[r, +\infty)$.

It follows from this definition that $\tilde{F}^r$ has a unimodal density $\tilde{f}^r$ with mode at r which is constant on any interval $[a, b]$ such that $\inf_{a \le x \le b} |\tilde{F}^r(x) - F(x)| > 0$. When F has a derivative f, it can be proved [see Barlow, Bartholomew, Bremner and Brunk (1972), Chapter 7] that $\tilde{f}^r$ is the unimodal density with mode at r which is the closest to f with respect to $\mathbb{L}^2$-distance. It is also the conditional expectation of f with respect to a convenient σ-lattice as defined in Wegman (1968).

Let us now assume that we observe n i.i.d. random variables with common distribution F and unimodal density f on the real line. Without loss of generality, 0 will be taken as a mode of f. As a particular case, f can be a decreasing function supported by $\mathbb{R}^+$. We denote by F_n and $\tilde{F}_n$ respectively the empirical distribution of the n observations and its unimodal regularization at 0. When f is decreasing on $\mathbb{R}^+$, the derivative $\tilde{f}_n$ of $\tilde{F}_n$ (which is a piecewise constant decreasing density) is the Grenander estimator of f and it has been proved by Groeneboom (1985) [see (1.5) of Birgé (1989)] that, if f is bounded and compactly supported with a continuous derivative f',

$$n^{1/3}\,\mathbb{E}[\|\tilde{f}_n - f\|_1] \to_{n\to+\infty} 0.82\int_0^\infty \left|\frac{f(x)f'(x)}{2}\right|^{1/3} dx, \tag{2.1}$$

where $\|\cdot\|_p$ denotes the $\mathbb{L}^p$-norm for $1 \le p \le \infty$. A nonasymptotic analogue of (2.1) which is true without any smoothness assumption has been given by Birgé (1989), Theorem 1:

$$\mathbb{E}[\|\tilde{f}_n - f\|_1] \le 2\mathscr{L}(f, 1.24n^{-1/2}), \tag{2.2}$$

where the functional $\mathscr{L}(f, z)$ is defined in the following way for positive z. Let $\mathscr{I}$ be any partition of $\mathbb{R}^+$ generated by some increasing sequence $x_0 = 0 < x_1 < \cdots < x_m = +\infty$, then

$$\begin{aligned} &\mathscr{L}(f, z) = \inf_{\mathscr{I}} \sum_{i=1}^{m}\left[\int_{x_{i-1}}^{x_i} |f(x) - f_i|\,dx + z f_i^{1/2}\right] \quad \text{where} \\ &f_i = \int_{x_{i-1}}^{x_i} \frac{f(x)}{x_i - x_{i-1}}\,dx, \end{aligned} \tag{2.3}$$

with the obvious convention that $f_m = 0$. It comes from elementary computations that $\mathscr{L}(f, n^{-1/2})$ is an upper bound for the risk of the best histogram for estimating f. Moreover, it follows from Birgé (1989) that, if f is bounded with a compact support and a continuous derivative,

$$\lim_{z\to 0} z^{-2/3}\mathscr{L}(f, z) = \frac{3}{2}\int_0^\infty \left|\frac{f(x)f'(x)}{2}\right|^{1/3} dx, \tag{2.4}$$

and therefore that the asymptotic upper bound for the risk of the Grenander estimator which can be derived from (2.2) is only within a factor 17/4 of the true one given by (2.1). When f is a unimodal density (with mode at 0), these results remain true provided that all the partitions $\mathscr{I}$ which define $\mathscr{L}(f, z)$ are now generated by increasing sequences $x_0 = -\infty < x_1 < \cdots < x_m = +\infty$ with one of the x_i's equal to 0.

We can now define our estimator. Starting from the empirical distribution function F_n, we can consider all the possible unimodal regularizations $\tilde{F}_n^t$ of F_n corresponding to the values t which are not equal to the observations and choose as an estimator of the unknown f the derivative of one of these $\tilde{F}_n^t$ which is closest to F_n. More precisely, we give the following.

DEFINITION 3. Let a nonnegative number η be given and the real number r be chosen to satisfy

$$\|\tilde{F}_n^r - F_n\|_\infty \le \inf_t \|\tilde{F}_n^t - F_n\|_\infty + \eta.$$

The estimator $\hat{f}_n$ is then defined as the derivative of $\tilde{F}_n^r$.

Let us immediately observe that our estimator need not be unique but that it always exists as soon as $\eta > 0$. If several values of r are possible, any of them will do and lead to an estimator with the following property.

THEOREM 1. *Let the true density f be unimodal with mode at* 0 *and $\tilde{f}_n$ be the corresponding Grenander estimator* (*based on the true mode* 0) *built from n i.i.d. observations with density f. Let F_n, F and $\tilde{F}_n$ be, respectively, the empirical c.d.f. and the distributions with densities f and $\tilde{f}_n$; then*

$$\frac{1}{2}\|\tilde{f}_n - \hat{f}_n\|_1 \le \eta + \|\tilde{F}_n - F_n\|_\infty \le \eta + 2\|F - F_n\|_\infty. \tag{2.5}$$

It is now easy to derive the performances of $\hat{f}_n$.

COROLLARY 1. *The following bound on the risk of $\hat{f}_n$ holds for all n*:

$$\begin{aligned}\mathbb{E}[\|f - \hat{f}_n\|_1 &\le \mathbb{E}[\|f - \tilde{f}_n\|_1] + 2\left[\left(\frac{2\pi}{n}\right)^{1/2} + \eta\right] \\ &\le 2\left[\mathscr{L}(f, 1.24n^{-1/2}) + \left(\frac{2\pi}{n}\right)^{1/2} + \eta\right].\end{aligned} \tag{2.6}$$

PROOF. It has been proved by Massart (1990) that for positive t, $\mathbb{P}[\|F - F_n\|_\infty > t] \le 2\exp(-2nt^2)$ and therefore that $\mathbb{E}[\|F - F_n\|_\infty] \le [\pi/(2n)]^{1/2}$. Our conclusion then follows from (2.2) and (2.5). □

Choosing $\eta = 1/n$ we see that the additional risk due to the fact that the mode is unknown is not larger than $(8\pi/n)^{1/2} + 2/n \le 6n^{-1/2}$ as soon as $n \ge 5$. One should also note that the $\mathbb{L}^1$-consistency of $\hat{f}_n$ implies the convergence of its mode to 0 when this is the unique mode of f. Our estimator therefore also provides a consistent estimator of the mode.

From the asymptotic point of view we get the following, which derives immediately from Corollary 1 and Theorem 2 of Birgé (1989).

COROLLARY 2. *Let f be a unimodal density with mode at zero such that f' is defined and continuous except possibly at a finite number of points. Assume moreover that $\int_{|x|>1} f^{1/2}(x)\,dx < \infty$, $\int_{|x|<1} f^{2+\delta}(x)\,dx < \infty$ for some positive δ and that $\eta = o(n^{-1/3})$, then*

$$\limsup_n n^{1/3}\mathbb{E}[\|f - \hat{f}_n\|_1] < 3.5\int_0^\infty \left|\frac{f(x)f'(x)}{2}\right|^{1/3} dx.$$

REMARK. A slight strengthening of the assumptions allowing the use of (2.1) implies the stronger result:

$$\limsup_n n^{1/3}\mathbb{E}[\|f-\hat{f}_n\|_1] = 0.82\int_0^{\infty}\left|\frac{f(x)f'(x)}{2}\right|^{1/3} dx.$$

It now remains to prove Theorem 1. The proof is based on the following elementary lemma.

LEMMA 1. *Let $r < t$ and $\tilde{F}^r$, $\tilde{F}^t$ be the corresponding unimodal regularizations of some distribution function F. Then*

$$\|\tilde{f}^r - \tilde{f}^t\|_1 = 2\max\left\{\sup_{r\le x\le t}[\tilde{F}^r(x)-F(x)];\ \sup_{r\le x\le t}[F(x)-\tilde{F}^t(x)]\right\}.$$

PROOF. It follows from the definition of unimodal regularizations that

$$\begin{cases} \tilde{F}^t(x)\le \tilde{F}^r(x)\le F(x), & \text{for } x\in(-\infty,r],\\ \tilde{F}^t(x)\le F(x)\le \tilde{F}^r(x), & \text{for } x\in[r,t],\\ F(x)\le \tilde{F}^t(x)\le \tilde{F}^r(x), & \text{for } x\in[t,+\infty). \end{cases} \tag{2.7}$$

Moreover if we define q and u by

$$q=\sup\{x<r\,|\,F(x)=\tilde{F}^t(x)\}\quad\text{and}\quad u=\inf\{x>t\,|\,F(x)=\tilde{F}^r(x)\},$$

$\tilde{F}^r(x)=\tilde{F}^t(x)$ and therefore $\tilde{f}^r(x)=\tilde{f}^t(x)$ for both $x<q$ and $x>u$. It also follows that $\tilde{F}^r$ is linear on the interval $[t,u]$ and $\tilde{F}^t$ linear on $[q,r]$. Therefore $\tilde{f}^r$ is constant on (t,u) and since it is unimodal with mode at r, it is nondecreasing on (q,r) and nonincreasing on (r,u). Similarly, $\tilde{f}^t$ is nondecreasing on (q,t), nonincreasing on (t,u) and constant on (q,r). It follows from this analysis and (2.7) that $\tilde{f}^r\ge\tilde{f}^t$ on (q,r) and that $\tilde{f}^r\le\tilde{f}^t$ on (t,u). Therefore one can find some point s in $[r,t]$ such that

$$\tilde{f}^r(x)\ge\tilde{f}^t(x)\quad\text{for } x\in(q,s)\quad\text{and}\quad \tilde{f}^r(x)\le\tilde{f}^t(x)\quad\text{for } x\in(s,u)$$

and since $\tilde{f}^r$ and $\tilde{f}^t$ are densities,

$$\tfrac{1}{2}\|\tilde{f}^r-\tilde{f}^t\|_1=\int_q^s[\tilde{f}^r(x)-\tilde{f}^t(x)]\,dx=[\tilde{F}^r(s)-\tilde{F}^t(s)]=\sup_{r\le x\le t}[\tilde{F}^r(x)-\tilde{F}^t(x)].$$

Since $\tilde{F}^t\le\tilde{F}^r$ on $[r,t]$, the distribution functions $\tilde{F}^r$ and $\tilde{F}^t$ are, respectively, convex and concave and both are piecewise linear, the maximum distance between them necessarily obtains at a point where the slope of one of them changes, which is a point of contact with F. This completes the proof of the lemma. □

PROOF OF THEOREM 1. Since $\hat{f}_n=\tilde{f}_n^r$ and $\tilde{f}_n=\tilde{f}_n^0$, it follows from the preceding lemma that

$$\|\hat{f}_n-\tilde{f}_n\|_1\le 2\max\{\|\tilde{F}_n^r-F_n\|_\infty;\|F_n-\tilde{F}_n\|_\infty\}\le 2[\|F_n-\tilde{F}_n\|_\infty+\eta]$$

from the definition of $\tilde{F}_n^r$. The second inequality in (2.5) follows from Marshall's Lemma [see Barlow, Bartholomew, Bremner and Brunk (1972), pages 70 and 227] which says that $\|F_n - \tilde{F}_n\|_\infty \le 2\|F_n - F\|_\infty$. $\square$

3. A computational algorithm for the estimator. We assume here that $n \ge 4$, $\eta > 0$ and denote by $X_{(1)}, \dots, X_{(n)}$ the order statistic, keeping in mind that our aim is to minimize over the possible values r the quantity $d(r) = \|\tilde{F}_n^r - F_n\|_\infty$. In order to locate the minimum of d it will be useful to define

$$d^-(r) = \sup_{x \le r}[F_n(x) - \tilde{F}_n^r(x)] \quad \text{and} \quad d^+(r) = \sup_{x \ge r}[\tilde{F}_n^r(x) - F_n(x)].$$

Then $d = \max(d^-, d^+)$ and we can derive from the definition of $\tilde{F}_n^r$ the following properties of $d^-(r)$ when r increases from $-\infty$ to $+\infty$:

1. it is a nondecreasing function on $\mathbb{R}$;
2. $d^-(x) = 0$ for $x < X_{(1)}$ and $d^-(x) = 1/n$ for $X_{(1)} < x < X_{(2)}$;
3. on each interval $(X_{(i)}, X_{(i+1)})$, d^- is a continuous function;
4. for each i, there exists some positive number ε_i such that the function $\tilde{F}_n^r \mathbb{1}_{(-\infty, X_{(i)}]}$ is constant with respect to r for $r \in (X_{(i)}, X_{(i)} + \varepsilon_i)$ and therefore on this interval, $d^-(r)$ is constant and equal to $d^-(X_{(i)}^+) = d^-(X_{(i)}^-) \vee 1/n$.

Similar properties hold for $d^+(r)$ (with obvious modifications) when r decreases from $+\infty$ to $-\infty$. One can then conclude that both d^- and d^+ can be extended to continuous functions on $[X_{(1)}, X_{(n)}]$ and define the monotone sequences $(a_i)_{1 \le i \le n}$ and $(b_i)_{1 \le i \le n}$ by $a_i = d^-(X_{(i)})$ and $b_i = d^+(X_{(i)})$ with $a_1 = b_n = 1/n$. Since there exists an index k such that $a_k \le b_k$ and $a_{k+1} \ge b_{k+1}$, the minimum value of d is obtained at a point r such that $d^-(r) = d^+(r)$ which belongs to the interval $(X_{(k)}, X_{(k+1)})$. Therefore the first step of the algorithm is to find the value of k by evaluating the sequences (a_i) and (b_i).

The algorithm. Let us first recall that one can easily compute $\tilde{F}_n^r$ and its derivative $\tilde{f}_n^r$ from the empirical c.d.f. F_n using the classical "pool adjacent violators algorithm" described in Barlow, Bartholomew, Bremner and Brunk [(1972), page 13] which transforms an arbitrary histogram into a decreasing or unimodal one. Since F_n is not absolutely continuous, it is necessary to replace it first by a linearized version, which, by derivation, will provide the suitable histogram. For any number r different from the $X_{(i)}$'s (and therefore a continuity point of F_n), we define the *linear regularization* G_n^r of F_n at r by

$$G_n^r(x) = 0 \quad \text{for } x \le X_{(1)}; \qquad G_n^r(x) = 1 \quad \text{for } x \ge X_{(n)};$$

$$G_n^r(X_{(i)}) = \frac{i-1}{n} \quad \text{for } X_{(i)} < r; \qquad G_n^r(r) = F_n(r); \quad \text{and}$$

$$G_n^r(X_{(i)}) = \frac{i}{n} \quad \text{for } X_{(i)} > r.$$

Moreover G_n^r is linear between these values. Then its derivative g_n^r is a histogram and the derivative $\tilde{g}_n^r$ of the unimodal regularization $\tilde{G}_n^r$ of G_n^r at r can be constructed from g_n^r by the "pool adjacent violators algorithm." Since it is easily seen that $\tilde{G}_n^r = \tilde{F}_n^r$ and therefore $\tilde{g}_n^r = \tilde{f}_n^r$, this method provides a simple construction of $\tilde{f}_n^r$.

We now want to compute the sequence (a_i). Let us define G^- as the restriction to $(-\infty, X_{(n)}+1)$ of $G_n^{X_{(n)}+1}$ (where we choose 1 here for simplicity, any other positive constant would do). If g^- is the derivative of G^-, one can get the derivative h_i^- of the convex envelope H_i^- of G^- restricted to the interval $(-\infty, X_{(i)}]$ by a "pool adjacent violators algorithm." Simultaneously one computes $a_i = \sup_{x\le X_{(i)}} G^-(x) - H_i^-(x)$. In order to minimize the amount of computation required for the evaluation of h_i^- and a_i one should do this recursively, starting from h_2^- and a_2 since $a_1 = 1/n$. The same method applies to the sequence (b_i) starting from $b_n = 1/n$.

The next step is to observe that $a_k \vee b_{k+1} \le d(x) \le a_{k+1} \vee b_k$ for $x \in (X_{(k)}, X_{(k+1)})$. Therefore, if the difference $(a_{k+1}\vee b_k)-(a_k\vee b_{k+1})$ is not larger than η, one can choose for $\hat{f}_n$ the derivative of any function $\tilde{F}_n^x$ with $x \in (X_{(k)}, X_{(k+1)})$ and we are done. If this is not the case, one can use a classical dichotomy argument on the interval $(X_{(k)}, X_{(k+1)})$, computing first $\tilde{f}_n^y$ with $y = (X_{(k)}+X_{(k+1)})/2$, then $d^-(y)$, $d^+(y)$ (using the above method based on the "pool adjacent violators" algorithm) in order to determine whether r belongs to $(X_{(k)}, y]$ or $[y, X_{(k+1)})$. Assuming that the former case obtains, one computes the difference $d^-(y)\vee d^+(X_{(k)})-d^-(X_{(k)})\vee d^+(y)$. If it is not larger than η then we are done, otherwise we iterate the procedure which will stop after a finite number of steps since d^- and d^+ are continuous functions on $[X_{(k)}, X_{(k+1)}]$ and $\eta > 0$.

REMARK. Various extensions of the principle of construction of our estimator to other models than density estimation have been developed by Reboul (1996) together with a practical implementation of the algorithm for simulations and industrial applications.

4. Approximate unimodality. In order to keep the presentation short and elementary, we shall restrict ourselves to the case of the Grenander estimator $\tilde{f}_n$ with a mode at 0, although the results could be extended with minor modifications but some additional technicalities to the general case of an unknown mode as explained in Birgé (1987c). Therefore, from now on, *unimodal* will mean unimodal with a mode at zero.

We want to understand the behavior of the Grenander estimator $\tilde{f}_n$ when the true density f is not exactly unimodal. Intuitively, since F_n converges to F and therefore $\tilde{F}_n$ to the unimodal regularization $\tilde{F}^0$ of F at 0, $\tilde{f}_n$ should converge to the derivative $\tilde{f}^0$ of $\tilde{F}^0$ which is known to be the best unimodal approximation of f in $\mathbb{L}^2$ [see Barlow, Bartholomew, Bremner and Brunk (1972), Chapter 7, for details]. Of course, $\tilde{f}^0$ need not be the best unimodal approximation of f in $\mathbb{L}^1$ but it follows from Proposition 1 of Birgé (1987a) that if g is a unimodal density, $\|\tilde{f}^0 - g\|_1 \leq \|f - g\|_1$ which implies that $\|\tilde{f}^0 - f\|_1 \leq 2\|f - g\|_1$ for any unimodal density g and therefore, up to a factor 2, $\tilde{f}^0$ is also the best unimodal approximation of f in $\mathbb{L}^1$. Of course, the problem of the behavior of $\tilde{f}_n$ as an estimator of f only makes sense when f is close to unimodal; that is, when $\|\tilde{f}^0 - f\|_1$ is small.

Let us first notice that the set J of those x's such that $\tilde{F}^0(x) \neq F(x)$ is a union of open intervals J_k such that $\tilde{F}^0$ and F coincide at the end points of J_k and $\tilde{F}^0$ is linear (therefore $\tilde{f}^0$ is constant) on J_k. This implies that if $f_k = \int_{J_k} f(x)\,dx / \int_{J_k} dx$, then

$$\|\tilde{f}^0 - f\|_1 = \sum_k \int_{J_k} |\tilde{f}^0(x) - f(x)|\,dx = \sum_k \int_{J_k} |f_k - f(x)|\,dx.$$

PROPOSITION 1. *The Grenander estimator $\tilde{f}_n$ with a mode at* 0 *satisfies*

$$\mathbb{E}[\|\tilde{f}_n - f\|_1] \leq 2\mathscr{L}'(\tilde{f}^0, 1.24n^{-1/2}) + \|\tilde{f}^0 - f\|_1,$$

where $\mathscr{L}'$ is defined as $\mathscr{L}$ by (2.3) *with the additional restriction that the minimization is now over all partitions $\mathscr{I}$ determined by sequences (x_j) including* 0 *and such that for any j, $F(x_j) = \tilde{F}^0(x_j)$.*

PROOF. Assuming, for the sake of simplicity, that f is a decreasing density on $\mathbb{R}^+$ we just follow the lines of the proof of Theorem 1 of Birgé (1989) with obvious modifications due to the fact that by our definition of $\mathscr{L}'$, if I is an element of $\mathscr{I}$, $\int_I(\tilde{f}^0 - f) = 0$ and $\tilde{F}^0 \geq F$. $\square$

The meaning of Proposition 1 is actually easier to understand from an asymptotic point of view. Let us recall that under mild assumptions on $\tilde{f}^0$, it follows from (2.4) that

$$\mathscr{L}(\tilde{f}^0, z) = z^{2/3}\left[\frac{3}{2}\int\left|\frac{\tilde{f}^0(x)\tilde{f}^{0\,\prime}(x)}{2}\right|^{1/3} dx + o(1)\right]$$

when $z \to 0$. When J is the union of a finite number of intervals J_k, the result extends in a straightforward way to $\mathscr{L}'$ and becomes

$$\mathscr{L}'(\tilde{f}^0, z) = z^{2/3}\left[\frac{3}{2}\int_{J^c}\left|\frac{f(x)f'(x)}{2}\right|^{1/3} dx + o(1)\right]$$

since $\tilde{f}^0 = f$ on J^c and $\tilde{f}^0$ is piecewise constant (therefore with a derivative equal to 0) on J. This means that the upper bound for the risk given in Proposition 1 is the sum of a term which is essentially the same as if f where truly unimodal (and actually smaller since integration of $|ff'|^{1/3}$ is now restricted to J^c) and an additional bias term which is bounded by twice the distance from f to the set of unimodal densities which is clearly unavoidable since our estimator is by nature unimodal.

REFERENCES

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions*. Wiley, New York.

BIRGÉ, L. (1987a). Estimating a density under order restrictions: nonasymptotic minimax risk. *Ann. Statist.* **15** 995–1012.

BIRGÉ, L. (1987b). On the risk of histograms for estimating decreasing densities. *Ann. Statist.* **15** 1013–1022.

BIRGÉ, L. (1987c). Robust estimation of unimodal densities. Technical report, Univ. Paris X–Nanterre.

BIRGÉ, L. (1989). The Grenander estimator: a nonasymptotic approach. *Ann. Statist.* **17** 1532–1549.

BIRGÉ, L. and MASSART, P. (1994). Minimum contrast estimators on sieves. Technical Report 94.34., Univ. Paris–Sud.

BRETAGNOLLE, J. and HUBER, C. (1979). Estimation des densités: risque minimax. *Z. Wahrsch. Verw. Gebiete* **47** 119–137.

DEVROYE, L. (1987). *A Course in Density Estimation*. Birkhäuser, Boston.

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455.

EFROIMOVICH, S. YU. and PINSKER, M. S. (1982). Estimation of square-integrable probability density of a random variable. *Problems Inform. Transmission* **18** 175–189.

GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153.

GRENANDER, U. (1981). *Abstract Inference*. Wiley, New York.

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **2** 539–555. Wadsworth, Monterey, CA.

HAS'MINSKII, R. Z. (1979). Lower bound for the risks of nonparametric estimates of the mode. In *Contribution to Statistics*: *Hajek Memorial Volume* (J. Jureckova, ed.) 91–97. Academia, Prague.

KERKYACHARIAN, G. and PICARD, D. (1992). Density estimation in Besov spaces. *Statist. Probab. Lett.* **13** 15–24.

KIEFER, J. and WOLFOWITZ, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrsch. Verw. Gebiete* **34** 73–85.

KOGURE, A. (1987). Asymptotically optimal cells for a histogram. *Ann. Statist.* **15** 1023–1030.

LO, S. H. (1986). Estimation of a unimodal distribution function. *Ann. Statist.* **14** 1132–1138.

MAMMEN, E. (1991). Estimating a smooth monotone regression function. *Ann. Statist.* **19** 724–740.

MASSART, P. (1990). The tight constant in the D. K. W. inequality. *Ann. Probab.* **18** 1269–1283.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36.

PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation*. Academic Press, Orlando, FL.

REBOUL, L. (1996). Estimation of a function under shape restrictions, application to reliability. Technical report, Univ. Paris–Sud.

REISS, R. D. (1973). On the measurability and consistency of maximum likelihood estimates for unimodal densities. *Ann. Statist.* **1** 888–901.

REISS, R. D. (1976). On minimum distance estimators for unimodal densities. *Metrika* **23** 7–14.

ROBERTSON, T. (1967). On estimating a density which is measurable with respect to an *s*-lattice. *Ann. Math. Statist.* **38** 482–493.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

WEGMAN, E. J. (1968). On estimating a unimodal density. Ph.D. dissertation, Univ. Iowa.

WEGMAN, E. J. (1969). A note on estimating a unimodal density. *Ann. Math. Statist.* **40** 1661–1667.

WEGMAN, E. J. (1970a). Maximum likelihood estimation of a unimodal density function. *Ann. Math. Statist.* **41** 457–471.

WEGMAN, E. J. (1970b). Maximum likelihood estimation of a unimodal density II. *Ann. Math. Statist.* **41** 2169–2174.

I.S.U.P., BOÎTE 157
UNIVERSITÉ PARIS VI
4 PLACE JUSSIEU
F-75252 PARIS CEDEX 05
FRANCE
E-MAIL: lb@ccr.jussieu.fr