The Annals of Statistics
1989, Vol. 17, No. 4, 1532–1549

THE GRENANDER ESTIMATOR: A NONASYMPTOTIC APPROACH

BY LUCIEN BIRGÉ

CNRS and Université Paris X, Nanterre

In this paper we shall investigate some nonasymptotic properties of the Grenander estimator of a decreasing density f. This estimator is defined as the slope of the smallest concave majorant of the empirical c.d.f. It will be proved that its risk, measured with $\mathbb{L}^1$-loss, is bounded by some functional depending on f and the number n of observations. For classes of uniformly bounded densities with a common compact support, upper bounds for the functional are shown to agree with older results about the minimax risk over these classes. The asymptotic behavior of the functional as n goes to infinity is also in accordance with the known asymptotic performances of the Grenander estimator.

1. Introduction. This paper is meant to illustrate some of the problems that occur in density estimation: choice of a proper estimator, evaluation of its risk. Although a lot of things are supposed to be known in density estimation, answering a question like "I have 1000 observations from a density, which is certainly very close to unimodal. What should I do?" is not so obvious.

There are many ways of estimating densities and many different types of results to obtain. They depend widely upon the class of densities you consider, the type of loss function, the kind of uniformity you are looking for and the asymptotic approach. This variety will appear immediately after looking at some papers or books like Bretagnolle and Huber (1979), Ibragimov and Has'minskii (1981), Prakasa-Rao (1983), Devroye and Györfi (1985), Birgé (1986) and Devroye (1987). Some familiarity with the subject rapidly convinces the reader that the different approaches are not easily comparable. To be somewhat more precise, let us briefly sketch our problem. Suppose we are given n i.i.d. observations from some unknown density f on the line. If $\hat{f}_n$ is an estimator, its risk will be defined through $\mathbb{L}^1$ distance by

$$R_n(f, \hat{f}_n) = \mathbb{E}_f\left[\int |f(x) - \hat{f}_n(x)|\, dx\right].$$

The choice of $\mathbb{L}^1$ is motivated here mainly by its invariance properties and further references to previous works. It is always possible to find a consistent sequence of estimators $\hat{f}_n$ such that

$$R_n(f, \hat{f}_n) \xrightarrow[n \to +\infty]{} 0, \tag{1.1}$$

but this is not a very interesting result since this convergence can be arbitrarily slow as shown in Birgé (1986). In order to get some uniformity in (1.1), it is generally necessary to restrict one's attention to a given compact (or at least

Received July 1986; revised December 1988.
AMS 1980 *subject classifications*. Primary 62G05; secondary 60E15.
Key words and phrases. Decreasing densities, Grenander estimator, local nonasymptotic risk.

σ-compact) class $\mathscr{F}$ of densities. In this case it is generally possible to show that

$$(1.2)\qquad C_1 r(n) \le \inf_{\hat{f}_n} \sup_{f \in \mathscr{F}} R_n(f, \hat{f}_n) \le C_2 r(n), \qquad r(n) \xrightarrow[n \to +\infty]{} 0.$$

The constants C_1, C_2 and the rate of convergence $r(n)$ (usually of the form $n^{-\delta}$) depend on $\mathscr{F}$. It is also often possible to find some good estimator $\hat{f}_n$ which satisfies

$$(1.3)\qquad \lim_{n \to +\infty} R_n(f, \hat{f}_n)/r(n) = R(f) < C_2,$$

where $R(f)$ is some functional defined on $\mathscr{F}$. (1.2) and (1.3) illustrate two different points of view regarding the same problem—global nonasymptotic and local asymptotic, respectively—and both of them have advantages and disadvantages. (1.2) is safer because it is valid for all n but puts emphasis on the worst f in $\mathscr{F}$ and does not take care of more regular cases which might be more likely to happen (incidentally, the author must apologize for having used this minimax approach in his previous papers). It is also difficult to choose between various estimators using minimax criteria. On the other hand, the asymptotic point of view illustrated by (1.3) is much more precise, but what is asymptotically justified can be far from reasonable with an ordinary (even large) amount of observations especially in nonparametric problems like density estimation where the rates of convergence are slower than $n^{-1/2}$. It is therefore generally difficult to judge the validity of (1.3) when $n = 1000$, which is already a lot in practice.

Our purpose in this paper is to show how these two points of view might be misleading in some circumstances, that important distortions can occur between them and, mainly, to present an alternative choice: the local nonasymptotic point of view. It will not lead to very precise results but it seems to be rather interesting. We do not intend to develop it in great generality, but in a specific case: estimation of a decreasing density on the line. This particular choice is motivated by two reasons: (i) It has been one of the author's concerns during the last years and (ii) it is possible to develop reasonably precise computations in this context.

In this case, the two competing points of view are illustrated by the papers of Birgé (1987a, 1987b) and Groeneboom (1985), respectively. If $\mathscr{F}(a, H, L)$ is the set of all decreasing densities bounded by H with support on $[a; a + L]$, then

$$(1.4)\qquad 0.195 S^{1/3} \le \inf_{\hat{f}_n} \sup_{f \in \mathscr{F}(a, H, L)} n^{1/3} R_n(f, \hat{f}_n) \le 1.95 S^{1/3}$$

with $S = \log(1 + HL)$ and $n \ge 39S$.

If f is any bounded decreasing density on $[a; +\infty)$, with compact support and continuous first derivative, then from Groeneboom (1985),

$$(1.5)\qquad \lim_{n \to +\infty} n^{1/3} R_n(f, \tilde{f}_n) = 0.82 \int_a^{+\infty} |f(x) f'(x)/2|^{1/3}\, dx.$$

Actually, Groeneboom proved a much stronger result using second derivatives but, according to him, (1.5) will be true in our context even if only one derivative is assumed. $\tilde{f}_n$ denotes a specific estimator, known as the Grenander estimator and defined as the slope of the smallest concave majorant of the empirical c.d.f. This estimator was introduced by Grenander (1956) and is described and studied in Barlow, Bartholomew, Bremner and Brunk (1972) and Grenander (1980). Let us denote by $I(f)$ the integral in the right-hand side of (1.5). Then it can be derived from Birgé (1987a) that when $HL \geq e - 1$,

$$\sup_{f \in \mathscr{F}(a, H, L)} I(f) = C(S) S^{1/3}, \qquad S = \log(HL + 1), \qquad 0.65 < C(S) < 0.75,$$

but this is rather exceptional since this supremum is obtained for a particular shape of f. In some other cases, $I(f)$ is much smaller than this upper bound and (1.5) is asymptotically much better than (1.4). It is also noted that $I(f)$ is not continuous with respect to $\mathbb{L}^1$ norm, which means that convergence in (1.5) is certainly not uniform. Therefore we do not know what confidence to give to (1.5) when n is only moderately large.

In order to solve (at least partially) the dilemma, we shall give a nonasymptotic bound for the risk $R_n(f, \tilde{f}_n)$. This bound is of the form $\mathscr{L}(f, n^{-1/2})$ where $f \to \mathscr{L}(f, z)$, $z > 0$, is some $\mathbb{L}^1$-continuous functional closely related to the error of histogram estimators. It might be viewed as the nonasymptotic counterpart of $I(f)$. Actually, although its definition might look strange at first sight, the functional $I(f)$ [and $\mathscr{L}(f, z)$ as well] has a nice interpretation in terms of the optimal risk for histograms. Let us begin with some heuristics. Suppose one wants to estimate a density on $[0, 1]$ using n i.i.d. observations and the estimator $\hat{f}_n$ is the histogram based on some subdivision $x_0 = 0, x_1, \dots, x_m = 1$. Let us denote by N_i the number of observations falling in $J_i = [x_{i-1}; x_i)$ and by $\bar{f}_i$ the average of f on J_i,

$$\bar{f}_i = (x_i - x_{i-1})^{-1} \int_{J_i} f(t)\, dt.$$

On J_i, $\hat{f}_n$ takes the value $(x_i - x_{i-1})^{-1} N_i / n$ and its risk is bounded by the sum of two terms: a *bias term*

$$B = \sum_{i=1}^{m} \int_{J_i} |f(t) - \bar{f}_i|\, dt$$

and an *error term*, which is the analogue of the classical *variance term*

$$E = \sum_{i=1}^{m} \left[\int_{J_i} (f(t)/n)\, dt \right]^{1/2}.$$

Therefore an optimal subdivision $\{x_i\}_{i=0, m}$ will tend to minimize $B + E$ and will usually not lead to a uniform partition. If we take for granted that the J_i's should be small when n is large and assume that f' exists and is continuous, the following approximations are valid for large n:

$$\int_{J_i} |f(t) - \bar{f}_i|\, dt \approx \tfrac{1}{4}(x_i - x_{i-1})^2 |f'(x_i)|,$$

$$\left[\int_{J_i} (f(t)/n)\, dt \right]^{1/2} \approx \left[(x_i - x_{i-1}) f(x_i)/n \right]^{1/2}.$$

In order to minimize the error, the ideal choice of $x_i - x_{i-1}$ should be $[4f(x_i)/(nf'(x_i)^2)]^{1/3}$, leading to the approximation

$$E + B \approx 3.2^{-4/3} \sum_{i=1}^{m} (x_i - x_{i-1}) \left| \frac{f(x_i)f'(x_i)}{n} \right|^{1/3} \approx \frac{3}{2} \int \left| \frac{f(t)f'(t)}{2n} \right|^{1/3} dt,$$

which is proportional to $I(f)$. We also note that the optimal partition is generally far from being uniform.

The same heuristics, in a nonasymptotic framework, can motivate the introduction of the functional $\mathscr{L}(f, z)$. Let us begin with some notation. If f is a nonnegative nonincreasing function on the interval $J = [a; b)$, b being possibly infinite, we set

$$l(J) = b - a, \qquad f(J) = \int_J f(t)\,dt, \qquad \Delta f(J) = f(b) - f(a),$$

$$(1.6) \qquad \begin{aligned} bf(J) &= \int_J |f(J)/l(J) - f(t)|\,dt \quad \text{if } l(J) < +\infty, \\ bf(J) &= f(J) \quad \text{if } l(J) = +\infty. \end{aligned}$$

When we use several intervals $J_1, \dots, J_m$ and no confusion is possible, we put f_i for $f(J_i)$, Δf_i for $\Delta f(J_i)$ and so on. Any finite increasing sequence $\{x_i\}_{0 \le i \le m}$ with $x_0 = a$, $x_m = b$ generates a partition $\mathscr{J}$ of J into intervals $J_i = [x_{i-1}; x_i)$, $1 \le i \le m$, and a functional $L(\mathscr{J}, f, z)$ defined for positive z by

$$L(\mathscr{J}, f, z) = \sum_{i=1}^{m} \left[bf(J_i) + z(f(J_i))^{1/2} \right] = \sum_{i=1}^{m} \left[bf_i + z f_i^{1/2} \right].$$

It is easily seen that $L(\mathscr{J}, f, n^{-1/2})$ is exactly the sum $B + E$ for a histogram based on the partition $\mathscr{J}$. Looking for the best partition of J will lead to the functional

$$\mathscr{L}^J(f, z) = \inf L(\mathscr{J}, f, z), \qquad z > 0,$$

where the infimum is taken over all relevant partitions $\mathscr{J}$ of J. Our heuristics show that for large n and smooth f, $\mathscr{L}^J(f, n^{-1/2})$ is approximately $3/2n^{-1/3}I(f)$. The main purpose of this paper will be to prove the validity of this approximation and show that the Grenander estimator has a nonasymptotic risk which is bounded by a small multiple of $\mathscr{L}^J(f, n^{-1/2})$. This means that it automatically adapts to the shape of f in order to pick an approximately optimal subdivision $\{x_i\}_{0 \le i \le m}$ without requiring any arbitrary choice of a smoothing parameter. This is certainly an advantage for the Grenander estimator, but we must temper our optimistic view by two remarks:

1. By construction, the Grenander estimator is a decreasing density and therefore can only be consistent when the underlying density is itself decreasing as shown in Devroye (1987). Nevertheless, we could also show in Birgé (1986) that it still behaves very reasonably when f is close to decreasing.
2. The rate of convergence is not better than $n^{-1/3}$ (except for piecewise constant densities) even if the underlying density is very smooth. Actually,

with smooth densities, it is always possible to improve the bias term using approximation by smooth functions rather than the piecewise constant. The balance between bias and error is modified, resulting in better rates of convergence. Clearly, our heuristics, using first order approximation, are intended mainly for the case where nothing is known about second derivatives. If f were known to have a continuous second derivative and a compact support, the rate of convergence of well-tuned kernel estimators would be at least $n^{-2/5}$. The Grenander estimator would then be far from optimal and its use not recommended. Unfortunately, it is not always clear that the existence of f'' can be assumed since f is unknown. The "good" performance of the Grenander estimator is related to the fact that we assume no more than a decreasing density. In this respect, the Grenander estimator can be viewed as a tool for the pessimistic statistician. The optimistic one will assume a nicely bounded second derivative and use a kernel estimator. If he is able to choose the smoothing parameter correctly, he will do better asymptotically, but if n is moderate and $\sup_x |f''(x)|$ fairly large, it is not clear at all that the kernel will do better.

2. Nonasymptotic risk of the Grenander estimator. Although our main concern here is the classical Grenander estimator, we shall take a somewhat more general point of view in order to deal with various situations like estimation of a monotone tail or a unimodal density with known mode and begin with the introduction of the relevant concepts.

If G is any right-continuous nondecreasing and bounded function on the real line and J some interval $[a; b)$ where a is finite and b possibly infinite, we shall denote by $\tilde{G}^J$ (omitting the superscript when it is clear) the smallest concave majorant of G on J. This is a nondecreasing function with derivative $\tilde{g}$ and it satisfies

$$\tilde{G}^J(a) = G(a), \qquad \tilde{G}^J(b) = G(b), \qquad \tilde{G}^J(x) \geq G(x) \quad \text{for } a < x < b$$

with an obvious meaning when b is infinite. It is well-known [see, for instance, Barlow et al. (1972)] that if the derivative g of G is a step function (histogram), $\tilde{g}$ can be derived from g by the technique known as "pooling adjacent violators."

When F is a concave distribution function on $[a; +\infty)$, $X_1, \dots, X_n$ are n i.i.d. variables from F and F_n is the corresponding empirical c.d.f., then the smallest concave majorant $\tilde{F}_n$ of F_n is the classical Grenander estimator as described in Grenander (1956, 1980), Barlow et al. (1972) and Groeneboom (1985) and its density $\tilde{f}_n$ is a nondecreasing step function looking like a histogram. If we restrict ourselves to some subinterval J of $[a; +\infty)$, $\tilde{F}_n^J$ will be called the (restricted) Grenander estimator on J and its derivative denoted by $\tilde{f}_n^J$. Let us notice that $\tilde{F}_n^J$ is not the restriction of $\tilde{F}_n$ to J: $\tilde{F}_n^J(x) \leq \tilde{F}_n(x)$ for x in J and the inequality is generally strict for some x's.

Our purpose in this section will be to bound the risk of the estimator $\tilde{f}_n^J$ in terms of $\mathscr{L}^J(f, z)$ when f is the true underlying density. Part of the argument is based on the following lemma.

LEMMA 1. *Let* $\mathscr{J} = \{J_i\}_{1 \le i \le m}$ *be some partition of* J, F *an absolutely continuous distribution function,* F_n *the corresponding empirical c.d.f. and* $\tilde{F}_n = \tilde{F}_n^J$; $\tilde{F}_n^i = \tilde{F}_n^{J_i}$ *the related Grenander estimators. Define*

$$\bar{F}_n(x) = \sum_{i=1}^{m} \tilde{F}_n^i(x) 1_{J_i}(x)$$

and f *and* $\bar{f}_n$ *to be the respective derivatives of* F *and* $\bar{F}_n$. *Then*

$$\mathbb{E}\left[\int_J |\tilde{f}_n(x) - f(x)|\, dx\right] \le \mathbb{E}\left[\int_J |\bar{f}_n(x) - f(x)|\, dx\right].$$

PROOF. The definition of the Grenander estimator implies that the smallest concave majorant of $\bar{F}_n$ is $\tilde{F}_n$. Therefore $\tilde{f}_n$ is deduced from $\bar{f}_n$ by "pooling adjacent violators." Since f is decreasing, the conclusion follows from Proposition 1 of Birgé (1987b). □

We are now in a position to prove

THEOREM 1. *If* F *is a distribution function which is concave on* J *and* $\tilde{F}_n^J$ *is its Grenander estimator on* J *with respective derivatives* f *and* $\tilde{f}_n$, *then*

$$\mathbb{E}_f\left[\int_J |f(x) - \tilde{f}_n(x)|\, dx\right] \le 2\mathscr{L}^J(f, Kn^{-1/2}), \qquad K = 1/2\left(C\sqrt{\pi/2} + 1\right), \tag{2.1}$$

where $C \le 1.18$. *According to a classical conjecture we can take* $C = 1$, *which leads to* $K \simeq 1.13$.

PROOF. The proof relies in an essential way on the classical inequality by Dvoretzky, Kiefer and Wolfowitz (1956), Lemma 2,

$$\mathbb{P}\left[\sup_x (F_n(x) - F(x)) > t\right] \le A \exp(-2nt^2).$$

It was conjectured in Birnbaum and McCarty (1958) that $A = 1$ and this conjecture was supported by numerical computations. Very recently, Massart (1988) showed that it is true at least when $2nt^2 \ge \log 2$, which implies by integration that

$$\mathbb{E}\left[\sup_x (F_n(x) - F(x))\right] \le C/2(\pi/(2n))^{1/2} \tag{2.2}$$

with $C = 1.18$ according to Massart and $C = 1$ if the conjecture is true as expected.

By Lemma 1 it is enough to prove that for any partition $\mathscr{J} = \{J_i\}_{1 \le i \le m}$ of J,

$$\mathbb{E}_f\left[\int_J |f(x) - \bar{f}_n(x)|\, dx\right] \le 2 \sum_{i=1}^{m} \left[bf_i + K(f_i/n)^{1/2}\right],$$

where $\bar{f}_n$ is the derivative of $\bar{F}_n$. This is certainly true if we have for any arbitrary subinterval I of J the inequality

$$\mathbb{E}_f\left[\int_I |f(x) - \tilde{f}_n^I(x)|\,dx\right] \le 2\left[bf(I) + K(f(I)/n)^{1/2}\right]. \tag{2.3}$$

In order to prove (2.3) we first assume that I is finite and that N observations fall in it, N having a binomial distribution $\mathscr{B}(n, f(I))$. Then with $\tilde{f}_n = \tilde{f}_n^I$,

$$\int_I |f(x) - \tilde{f}(x)|\,dx \le bf(I) + |f(I) - N/n| + b\tilde{f}_n(I).$$

The only difficulty comes from the last term. Define G to be the conditional c.d.f. of the observations that fall in I and g its derivative. Then

$$\mathbb{E}_f\left[b\tilde{f}_n(I)|N\right] = N/nE_g\left[b\tilde{g}_N(I)|N\right]$$

because the joint distribution of the N observations falling in I given N is the same as the distribution of N i.i.d. variables from G. If $U(x)$ is the uniform c.d.f. on I, then

$$\begin{aligned} 1/2b\tilde{g}(I) &= \sup_{x\in I}\left[\tilde{G}_N(x) - U(x)\right] = \sup_{x\in I}\left[G_N(x) - U(x)\right] \\ &\le \sup_{x\in I}\left[G_N(x) - G(x)\right] + \sup_{x\in I}\left[G(x) - U(x)\right] \\ &\le \sup_{x\in I}\left[G_N(x) - G(x)\right] + 1/2bg(I). \end{aligned}$$

Since $bf(I) = f(I)bg(I)$, we get

$$\mathbb{E}_f\left[b\tilde{f}_n(I)|N\right] \le N/n\left[2\mathbb{E}_g\left[\sup_{x\in I}\left[G_N(x) - G(x)\right]|N\right] + bf(I)/f(I)\right]$$

and using (2.2),

$$E_f\left[b\tilde{f}_n(I)|N\right] \le N/n\left[\left[C/2(2\pi/N)^{1/2}\right] + bf(I)/f(I)\right].$$

A final integration with respect to N leads to (2.3) when I is finite since $E[N^{1/2}] \le [nf(I)]^{1/2}$. When I is infinite, the result is trivial from our definition of $bf(I)$. □

This theorem clearly applies to the classical Grenander estimator of a decreasing density on $[a; +\infty)$. But it is also valid for the estimation of the right-hand tail of a density (if this tail is known to be decreasing) by a restricted Grenander estimator. Clearly, by symmetry all these results are immediately translated in terms of increasing densities on $(-\infty; a]$ and the greatest convex minorant of the empirical c.d.f. Therefore it is also possible to estimate monotonous left-hand tails and, using both techniques simultaneously, unimodal densities when the mode is known. The problem is more delicate when the mode is unknown and will be treated elsewhere. Another interesting problem is to estimate a decreasing density f on $J = [a; +\infty)$ when a is unknown. A possible solution is as follows. Denote by $X_{(1)}$ the smallest observation and by $\tilde{F}_n^1$ the restricted Grenander estimator on $I = [X_{(1)}; +\infty)$ with $\tilde{F}_n^1(X_{(1)}) = 1/n$. We denote its derivative by

$\tilde{f}_n^1$, put $\tilde{f}_n^1(X_{(1)}) = H$ and take as final estimator of the unknown density f,

$$\tilde{f}_n(x) = H1_{[X_{(1)}-(Hn)^{-1};\, X_{(1)})}(x) + (n-1)/n\tilde{f}_n^1(x). \tag{2.4}$$

COROLLARY 1. *If $\tilde{f}_n(x)$ is defined by* (2.4), *then*

$$\mathbb{E}\left[\int_J |\tilde{f}_n(x) - f(x)|\, dn\right] \le (2 + n^{-1})\mathscr{L}^J(f, Kn^{-1/2}) + 3/n,$$

K being defined as in Theorem 1.

PROOF. $\int_{-\infty}^{X(1)}|\tilde{f}_n(x) - f(x)|\,dx$ has an expectation bounded by $2/n$. In order to control $\int_I|\tilde{f}_n^1(x) - f(x)|\,dx$, we take conditional expectations with respect to $X_{(1)}$ and apply Theorem 1 to the conditional distributions, obtaining

$$\begin{aligned}
f(I)\mathbb{E}_f&\left[\int_I\left|\frac{f(x)}{f(I)} - \frac{n}{(n-1)}\tilde{f}_n^1(x)\right|dx|X_{(1)}\right]\\
&\le 2f(I)\mathscr{L}^I\left(\frac{f}{f(I)}, K(n-1)^{-1/2}\right)\\
&\le 2\max\left[1, \left(\frac{f(I)n}{n-1}\right)^{1/2}\right]\mathscr{L}^I(f, Kn^{-1/2}).
\end{aligned}$$

Since $\mathscr{L}^I(f, z) \le \mathscr{L}^J(f, z)$, as will be checked in the next section, we can write

$$\begin{aligned}
\mathbb{E}_f\left[\int_I\left| f(x) - \tilde{f}_n^1(x)\right| dx\Big|X_{(1)}\right] &\le \mathbb{E}_f\left[\left|\frac{nf(I)}{n-1} - 1\right|\int_I\tilde{f}_n^1(x)\,dx\Big|X_{(1)}\right]\\
&\quad + 2\max\left[1, \left(\frac{nf(I)}{n-1}\right)^{1/2}\right]\mathscr{L}^J(f, Kn^{-1/2})\\
&\le \frac{n-1}{n}Y + (2 + Y)\mathscr{L}^J(f, Kn^{-1/2}),
\end{aligned}$$

$$Y = \left|\frac{nf(I)}{n-1} - 1\right|.$$

The conclusion follows from the fact that

$$\mathbb{E}[f(I)] = \frac{n}{n+1}, \qquad \mathrm{Var}(f(I)) = \frac{n}{(n+1)^2(n+2)}$$

and therefore $E(Y^2) \le n^{-2}$ for $n \ge 3$. □

3. Some properties of the functional. As we shall see, the constants 2 and K in (2.1) are probably far from optimal. Nevertheless, even if we were able to improve them, the bound would still be in terms of $\mathscr{L}^J(f, z)$ and it would be interesting to derive some properties of this functional and relate it to previous results about estimation of decreasing densities. At first sight $\mathscr{L}^J(f, n^{-1/2})$

appears as a rather complicated object which is not likely to be continuous with respect to f (only upper semicontinuous) and difficult to relate to n. But a closer look will show that this is not true. Let us first state a few useful facts. Since bf is a decreasing function, i.e., $bf(I) \le bf(J)$ when $I \subset J$, we get

$$\mathscr{L}^I(f, z) \le \mathscr{L}^J(f, z) \quad \text{for } I \subset J \tag{3.1}$$

and also

$$\mathscr{L}^J(f, z) \le \sum_{j=1}^{k} \mathscr{L}^{J_j}(f, z) \quad \text{if } J = \bigcup_{j=1}^{k} J_j. \tag{3.2}$$

Finally if f is a step function on J related to the partition $\mathscr{J} = \{J_1, \dots, J_m\}$, i.e., $f(x) = \sum_{i=1}^{m} f(J_i)/l(J_i)1_{J_i}(x)$, then

$$\mathscr{L}^J(f, z) \le z \sum_{i=1}^{m} [f(J_i)]^{1/2} \le z[mf(J)]^{1/2}. \tag{3.3}$$

We are now in a position to study some more delicate properties of $\mathscr{L}^J(f, z)$.

Continuity properties. We shall begin with an auxiliary result which could prove interesting by itself. The definition of $\mathscr{L}^J(f, z)$ involves arbitrary finite partitions but we shall show that, up to some approximation, we can restrict ourselves to consider partitions with a bounded number of elements. From now on, J denotes the interval of interest and we omit the corresponding superscript.

LEMMA 2. *If f is decreasing on J and $\varepsilon > 0$, there exists a partition $\mathscr{J}' = \{J_1', \dots, J_m'\}$ of J such that*

$$m \le 2f(J)/\varepsilon + 1, \qquad L(\mathscr{J}', f, z) \le \mathscr{L}(f, z)(1 + \varepsilon^{1/2}/z).$$

PROOF. Let us consider an arbitrary partition $\mathscr{J} = \{J_i\}_{1 \le i \le M}$ of J. After a convenient renumbering of the indices, we may assume that

$$f(J_i) \ge \varepsilon \quad \text{for } i \le p, \qquad f(J_i) < \varepsilon \quad \text{for } p+1 \le i \le M.$$

We derive a new partition $\mathscr{J}' = \{J_i'\}_{1 \le i \le m}$ of J by choosing $J_i' = J_i$ if $i \le p$ and we pool all remaining adjacent J_i's in order to get the J_i''s with indices larger than p. Then

$$m - p \le p + 1 \quad \text{and} \quad p\varepsilon \le f(J),$$

which gives the bound on m. Also

$$L(\mathscr{J}', f, z) \le L(\mathscr{J}, f, z) + \sum_{i=p+1}^{m} bf(J_i') < L(\mathscr{J}, f, z) + \sum_{i=p+1}^{M} f(J_i).$$

The conclusion follows from the fact that

$$\sum_{i=p+1}^{M} f(J_i) \le \varepsilon^{1/2} \sum_{i=p+1}^{M} (f(J_i))^{1/2} \le \varepsilon^{1/2} z^{-1} L(\mathscr{J}, f, z). \qquad \square$$

We are now in a position to prove that $\mathscr{L}(f, z)$ is $\mathbb{L}^1$-continuous, which is satisfactory since it serves as a bound for a risk function which is itself $\mathbb{L}^1$-continuous.

PROPOSITION 1. *Suppose that f and g are nonnegative decreasing functions defined on $(a; +\infty)$ and $(b; +\infty)$, respectively, and set them to zero outside of their intervals of definition. Then*

$$|\mathscr{L}(g, z) - \mathscr{L}(f, z)| \leq C_1 \left[\int |f(x) - g(x)|\, dx \right]^{1/4}$$

with a constant C_1 depending only (for $z \leq 1$) on $\max\{\int f(x)\, dx, \int g(x)\, dx\}$.

PROOF. We shall assume that $J = (b; +\infty)$, $a \geq b$, and set $\eta = \int |f(x) - g(x)|\, dx$. Then applying Lemma 2 with $\varepsilon = z^2 \eta^{1/2}$ we can find a partition $\mathscr{J} = \{J_i\}_{0 \leq i \leq m}$ of $(b; +\infty)$ with $J_0 = (b; a]$ and such that

$$\sum_{i=1}^{m} \left(z f_i^{1/2} + b f_i \right) \leq \mathscr{L}(f, z)(1 + \eta^{1/4}),$$

$$m + 1 \leq 2\left(f(J) \eta^{-1/2} / z^2 + 1 \right), \qquad \mathscr{L}(g, z) \leq \sum_{i=0}^{m} \left(z g_i^{1/2} + b g_i \right).$$

Using these inequalities and

$$\left| \sum_{i=1}^{m} (b f_i - b g_i) \right| + b g_0 \leq \eta, \qquad \sum_{i=1}^{m} \left| f_i^{1/2} - g_i^{1/2} \right| \leq \sum_{i=1}^{m} | f_i - g_i |^{1/2},$$

we deduce

$$\mathscr{L}(g, z) \leq \mathscr{L}(f, z)(1 + \eta^{1/4}) + \eta + z\left[2\eta\left(f(J) \eta^{-1/2} z^{-2} + 1 \right) \right]^{1/2},$$

and since $\mathscr{L}(f, z) \leq z + f(J)$,

$$\mathscr{L}(g, z) - \mathscr{L}(f, z) \leq \eta^{1/4}\left[z + f(J) + (2 f(J))^{1/2} \right] + \eta + (2\eta)^{1/2} z.$$

The reverse inequality is proved analogously. □

Some upper bounds for $\mathscr{L}(f, z)$. We shall see here that some mild restrictions on f imply that $z^{-2/3}\mathscr{L}(f, z)$ is bounded, which implies uniform rates of convergence for $\hat{f}_n$ over large classes of decreasing densities and will also prove useful in our asymptotic derivations. We begin with a case which has been studied in Birgé (1987b).

PROPOSITION 2. *Assume that f is a decreasing density on $J = [a; a + L)$ with $\Delta f(J) \leq H$ [Δf being defined in (1.6)] and set*

$$S = \log(HL + 1), \qquad t = (2S/z)^{2/3} + 2S/3.$$

Then $t \geq 1.24z^{-2/3} + 0.46$ *and*

$$\mathscr{L}(f, z) \leq (1 + t^{-1})^{1/2}\left[3/2(2Sz^2)^{1/3} + 1/8(2Sz^2)^{2/3}\right]. \tag{3.4}$$

PROOF. Generate a partition $\mathscr{J}$ of J with an increasing sequence $\{x_i\}_{0 \leq i \leq p}$ such that $x_0 = a$, $x_p = a + L$ and $p - 1 < t \leq p$. Then

$$L(\mathscr{J}, f, z) \leq \sum_{i=1}^{p} \left[z f_i^{1/2} + 1/2(x_i - x_{i-1})(f(x_i) - f(x_{i-1}))\right].$$

Some computations, perfectly similar to those in the proof of Theorem 1 in Birgé (1987b) with the same values of the x_i's, lead to

$$L(\mathscr{J}, f, z) \leq zp^{1/2} + \exp(S/t) - 1 \leq \left[(t+1)/t\right]^{1/2}\left[zt^{1/2} + \exp(S/t) - 1\right].$$

The result follows from Lemma A2 of Birgé (1987a). □

The proposition can be extended to decreasing functions which are not necessarily densities by the following corollary, which will prove useful in the sequel.

COROLLARY 2. *If* f *is a nonnegative decreasing function on* J *and* $S' = \log(1 + \Delta f(J) l(J)/f(J))$,

$$\mathscr{L}^J(f, z) \leq (1 + z^{2/3}/2)\Big[3/2\big(2S'(zf(J))^2\big)^{1/3} + 1/8(2S'z^2)^{2/3}(f(J))^{1/3}\Big], \tag{3.5}$$

$$\mathscr{L}^J(f, z) \leq (z/2)^{2/3}(1 + z^{2/3}/2)\Big[3(\Delta f(J) l(J) f(J))^{1/3} + 2^{-5/3}(z^2 \Delta f(J) l(J))^{1/3}\Big]. \tag{3.6}$$

PROOF. $g(x) = f(x)/f(J)$ is a density on J satisfying $\Delta g(J) = \Delta f(J)/f(J)$ and

$$\mathscr{L}^J(f, z) = f(J)\mathscr{L}^J\big(g, z(f(J))^{-1/2}\big).$$

(3.5) follows if we apply (3.4) to g with $(1 + t^{-1})^{1/2} < 1 + 1/2z^{2/3}$ and (3.6) since $\log(1 + x) \leq \min(x, x^{1/2})$. □

We may also want to deal with long-tailed or unbounded functions. In this case the following propositions are helpful.

PROPOSITION 3. *If* f *is a decreasing function on* $J = [a; +\infty)$ *and* $\int_J f^{1/2}(t)\,dt = M < +\infty$, *then*

$$z^{-2/3}\mathscr{L}^J(f, z) \leq \left[3/2(f(a)M^2)^{1/3} + z^{2/3}(f(a)M^2)^{1/6}\right].$$

PROOF. We consider the partition $\mathscr{J}=\{J_i\}_{1\le i\le m+1}$ induced by the numbers $x_i=a+il$, $0\le i\le m$, and $J_{m+1}=(x_m;+\infty)$. Then

$$L(\mathscr{J},f,z)\le\sum_{i=1}^{m}\left[zf_i^{1/2}+1/2(f(x_{i-1})-f(x_i))\right]+zf_{m+1}^{1/2}+f_{m+1}$$

$$\le zMl^{-1/2}+z(lf(a))^{1/2}+lf(a)/2+zf_{m+1}^{1/2}+f_{m+1}.$$

The result follows if we choose $l=(zM/f(a))^{2/3}$ and let m go to infinity. □

PROPOSITION 4. *If f is a decreasing function on $\mathscr{J}=[0;a]$ such that $\int_0^a f^p(t)\,dt=M<+\infty$ for some $p>2$ and $H=\lim_{x\to a-}f(x)>0$, then*

$$z^{-2/3}\mathscr{L}^J(f,z)\le 3/2(H/(H-h))^{p-1}(aMH^{2-p}/(p-2))^{1/3},$$
$$h^3=z^2a^{-2}MH^{2-p}/(p-2)$$

provided that $Mz^2<(p-2)a^2H^{p+1}$.

PROOF. From our assumption $H>h$. For j such that $jh\le H<(j+1)h$, define the partition $\mathscr{J}=\{J_i)_{j\le i\le m}$ of J by $J_m=\{x|f(x)\ge m\}$, $J_j=\{x|f(x)<(j+1)h\}$ and $J_i=\{x|ih\le f(x)<(i+1)h\}$ for $m>i>j$. If $l_i=l(J_i)$, we have for $i<m$,

$$f_i\le(i+1)l_ih,\qquad bf_i\le l_ih/2,\qquad \int_{J_i}f^p(t)\,dt\ge l_ih^pi^p$$

and

$$L(\mathscr{J},f,z)\le ah/2+zh^{1/2}\sum_{i=j}^{m-1}\left[(i+1)l_i\right]^{1/2}+\int_{J_m}f(t)\,dt+z\left[\int_{J_m}f(t)\,dt\right]^{1/2}$$

and letting m go to infinity,

$$\mathscr{L}^J(f,z)\le ah/2+zh^{1/2}\sum_{i\ge j}\left[(i+1)l_i\right]^{1/2}.$$

Using the Cauchy–Schwarz inequality and our definition of j we get

$$\left[\sum_{i\ge j}\left[(i+1)l_i\right]^{1/2}\right]^2\le\left[\sum_{i\ge j}(i+1)^pl_i\right]\left[\sum_{i\ge j}(i+1)^{1-p}\right]$$

$$\le\left(\frac{j+1}{j}\right)^ph^{-p}\int_J f^p(t)\,dt\int_j^{+\infty}t^{1-p}\,dt$$

$$=\frac{Mh^{-p}}{p-2}\left(1+\frac{1}{j}\right)^pj^{2-p}$$

$$\le\frac{Mh^{-p}}{p-2}\left(\frac{H}{H-h}\right)^p\left(\frac{H-h}{h}\right)^{2-p}$$

$$=\left(\frac{H}{H-h}\right)^{2p-2}\frac{Mh^{-2}H^{2-p}}{p-2}.$$

Then from our choice of h we derive

$$\mathscr{L}^J(f, z) \le \left(\frac{H}{H-h}\right)^{p-1}\left[ah/2 + z\left(\frac{Mh^{-1}H^{2-p}}{p-2}\right)^{1/2}\right]$$

$$= \frac{3}{2}\left(\frac{H}{H-h}\right)^{p-1}\left(\frac{z^2aMH^{2-p}}{p-2}\right)^{1/3}. \qquad \square$$

REMARK. These propositions could be used to extend the results of Birgé (1987b) (which were restricted to densities with compact support) to densities belonging to $\mathbb{L}^{1/2} \cap \mathbb{L}^p$, $p > 2$. Proposition 4 also shows the superiority of the Grenander estimator over ordinary histograms or kernel estimators when the underlying density is not bounded. Under the assumptions of this proposition, the rate of convergence of the Grenander estimator will be $n^{-1/3}$ which is impossible for histograms as shown by Theorem 5 in Devroye and Györfi [(1985), page 98] and similarly for kernel estimators with fixed bandwidth.

Asymptotic behaviour for small z. The preceding results show that $\mathscr{L}(f, z)$ goes to zero at a rate which is at least $z^{2/3}$ for a large class of functions. This is in accordance with the rate of convergence of the risk of the Grenander estimator $n^{-1/3}$ and suggests looking at the asymptotic behaviour of $z^{-2/3}\mathscr{L}(f, z)$ and its possible relation with the asymptotic functional

$$I(f) = \int |f(x)f'(x)/2|^{1/3}\,dx,$$

which describes the asymptotic risk of the Grenander estimator [see Groeneboom (1985)]. It is clear that no general result is to be found since $I(f)$ need not even be defined. Nevertheless $I(f)$ will exist under proper assumptions. For the sake of simplicity we shall restrict ourselves to the following:

A1. f is decreasing on $[0; +\infty)$ and for some $a > 0$, $\lambda > 0$,

$$\int_0^a f^{2+\lambda}(t)\,dt < +\infty, \qquad \int_a^{+\infty} f^{1/2}(t)\,dt < +\infty.$$

A2. f' is defined and continuous on the complement of some set whose closure is negligible.

These assumptions and Hölder inequality imply that $I(f)$ is well defined and we can prove

THEOREM 2. *If f satisfies* A1 *and* A2, *then*

$$\lim_{z\to 0} z^{-2/3}\mathscr{L}(f, z) = \tfrac{3}{2}I(f). \tag{3.7}$$

The original proof has been substantially shortened by the use of the following lemma due to Piet Groeneboom.

LEMMA 3. *Let f be a decreasing function with a continuous derivative on the compact interval J, such that f and f' are bounded away from zero on J. Then uniformly over all closed intervals $[a; b] \subset J$,*

$$\lim_{z \to 0} z^{-2/3} \mathscr{L}^{[a;\, b]}(f, z) = \tfrac{3}{2} \int_a^b |f(x) f'(x)/2|^{1/3}\, dx.$$

PROOF. For any interval $I \subset J$ we have

$$bf(I) > \tfrac{1}{4} l^2(I) \inf_{x \in J} |f'(x)|.$$

Since $z^{-2/3}\mathscr{L}[a; b](f; z)$ is uniformly bounded by Proposition 2, we may restrict ourselves to consider partitions $\mathscr{I}_z$ into intervals of length smaller than $Az^{1/3}$ if A is a large constant. From our assumptions we deduce that if x_I is the left-hand point of the interval I, we have uniformly on I,

$$bf(I) = \tfrac{1}{4}\, |f'(x_I)|\, l^2(I)(1 + o(l(I))), \qquad f(I) = f(x_I) l(I)(l + o(l(I))).$$

Then for all relevant partitions $\mathscr{I}_z$ with intervals of length smaller than $Az^{1/3}$,

$$L(\mathscr{I}_z, f, z) = \sum_{I \in \mathscr{I}_z} l(I)\left[l(I)|f'(x_I)|/4 + z[f(x_I)/l(I)]^{1/2}\right](1 + o(z)).$$

The function $g(t) = \alpha t + \beta t^{-1/2}$ is minimal for $t = (\beta/(2\alpha))^{2/3}$ with value $3(\beta/2)^{2/3}\alpha^{1/3}$. Therefore the choice

$$l(I) = \left[2zf^{1/2}(x_I)/|f'(x_I)|\right]^{2/3} \tag{3.8}$$

is optimal, leading to the lower bound

$$\tfrac{3}{2} z^{2/3} \sum_{I \in \mathscr{I}_z} l(I)[f(x_I)|f'(x_I)|/2]^{1/3}(1 + o(z)).$$

Since we can always define our partition by

$$x_0 = a, \qquad x_{i+1} = \inf(b, x_i + l_i), \qquad l_i = \left[2zf^{1/2}(x_i)/|f'(x_i)|\right]^{2/3},$$

we see that we can almost get an optimal partition with intervals I satisfying (3.8), apart from the last one, which will be smaller. Since its influence is $O(z^{4/3})$, we get

$$\inf_{\mathscr{I}_z} L(\mathscr{I}_z, f, z) = \tfrac{3}{2} z^{2/3} \sum_i l_i [f(x_i)|f'(x_i)|/2]^{1/3}(1 + o(z)) + O(z^{4/3})$$

and therefore

$$z^{-2/3}\mathscr{L}^{[a;\, b]}(f, z) = \tfrac{3}{2} \sum_i l_i [f(x_i)|f'(x_i)|/2]^{1/3}(1 + o(z)) + O(z^{2/3}).$$

Since $\sup_i l_i = O(z^{2/3})$, this sum can be approximated uniformly with respect to a and b by the integral $\int_a^b |f(x) f'(x)/2|^{1/3}\, dx$, which completes the proof. □

Proof of Theorem 2. Let us begin with the upper bound part. We split the support of f into $[0; a)$, $[a; a^{-1}]$ and $[a^{-1}; +\infty)$ for some small, conveniently chosen a and then split $[a; a^{-1}]$ into a finite number of intervals with the following properties: Intervals of the first type are those for which the assumptions of Lemma 3 are satisfied; intervals of type 2 are those on which f is constant; intervals of type 3 are the others, the total length of which is smaller than ε. Now, from (3.2), $z^{-2/3}\mathscr{L}(f, z)$ is smaller than the sum of the corresponding functionals on the various intervals. Using Propositions 3 and 4 we see that the contribution of the outer intervals is negligible if a is small enough. The contribution of intervals of type 2 is asymptotically zero since on those $\mathscr{L}(f, z) = O(z)$. The contribution of intervals of type 3, which we denote by $J_1, \dots, J_p$, is bounded using Corollary 2, which implies that

$$\limsup_{z \to 0} z^{-2/3} \sum_{i=1}^{p} \mathscr{L}^{J_i}(f, z) \le 3 \sum_{i=1}^{p} \left[\Delta f(J_i) l(J_i) f(J_i)\right]^{1/3}$$

and by Hölder's inequality, this is smaller than

$$3 \sum_{i=1}^{p} \left[\Delta f(J_i)\right]^{1/3} \sum_{i=1}^{p} \left[l(J_i)\right]^{1/3} \sum_{i=1}^{p} \left[f(J_i)\right]^{1/3} \le 3\left[f(a)\varepsilon\right]^{1/3}$$

and this is small for a good choice of ε. Finally, Lemma 3 implies that the total contribution of intervals of type 1 is smaller than $\frac{3}{2}I(f)$.

For the lower bound result,

$$\liminf_{z \to 0} z^{-2/3}\mathscr{L}(f, z) \ge \tfrac{3}{2}I(f), \tag{3.9}$$

we notice that $I(f)$ can be approximated as closely as we want by $\int_K |f(x)f'(x)/2|^{1/3}\,dx$, where K is finite union of intervals on which f satisfies the assumptions of Lemma 3. The arguments used in the proof of this lemma show that the trace of the optimal partitions on K is made of intervals of length $O(z^{2/3})$, which implies that restricting to K can only decrease the left-hand side of (3.9). The conclusion follows by an application of Lemma 3 to K. □

Remark. As a by-product of the proof we see, at least asymptotically, what the structure of an optimal partition is. Under the assumptions of Lemma 3, the size of the intervals should be of order $n^{-1/3}$ as expected for histograms. Actually, our heuristic developments (see the Introduction) concerning the optimal bin widths could be turned into a rigorous proof concerning the performance of histograms for estimating general densities using $\mathbb{L}^1$-loss under suitable assumptions (compact support and a continuous first derivative with a finite number of zeros, say, although these could certainly be weakened). The best histograms will clearly have a risk that is asymptotically bounded by $\frac{3}{2}n^{-1/3}\int |f(x)f'(x)/2|^{1/3}\,dx + o(n^{-1/3})$. A slightly more sophisticated analysis using the normal approximation would lead to the analogue for $\mathbb{L}^1$-norm of

Theorem 1 of Kogure (1987),

$$\limsup_n \left[\inf_{\hat{f}_n} n^{1/3} R_n\big(f, \hat{f}_n \big) \right] = c \int |f(x) f'(x)|^{1/3}\, dx,$$

where the infimum is taken over all possible histograms and c is a universal constant. Such a result is clearly not the subject of this paper and, anyway, does not lead, without suitable adaptation, to the effective construction of an optimal estimator since the optimal histograms depend in a crucial way on the unknown density f. This is a serious limit to its practical consequences. On the contrary, we would like to emphasize the adaptive nature of the Grenander estimator which automatically selects almost optimal partitions without knowledge (except for monotonicity) of the underlying density.

4. Conclusion. It is now possible to put all the pieces together. Our main result says that

$$R_n\big(f, \tilde{f}_n \big) \le 2\mathscr{L}\big(f, Kn^{-1/2} \big). \tag{4.1}$$

If we assume that the constant C in (2.2) is 1, then $K = 1.127$ and Theorem 2 leads to

$$\limsup_n n^{1/3} R_n\big(f, \tilde{f}_n \big) \le 3K^{2/3} I(f) < 3.25 I(f), \tag{4.2}$$

while Proposition 2 implies that for $S = \log(HL + 1)$ and $n \ge 40S$,

$$\sup_{f \in \mathscr{F}(a, H, L)} n^{1/3} R_n\big(f, \tilde{f}_n \big) \le 4.74 S^{1/3}. \tag{4.3}$$

This means that, up to multiplicative constants, our study leads simultaneously to the local asymptotic bound of Groeneboom (1985) and the nonasymptotic global result of Birgé (1987a). It also has the advantage of relating both points of view and providing a continuous upper bound for $R_n(f, \tilde{f}_n)$. Obviously, due to our crude methods the constants in (4.1) are likely to be too big, but the asymptotic constant in (4.2) is only four times the constant of Groeneboom. If we consider that (4.2) is only valid for smooth functions and that this smoothness induces some asymptotic bias reduction, it is likely that (4.1) is less than four times too large. Even if our computations are not very precise, it seems reasonable to consider $\mathscr{L}(f, n^{-1/2})$ as a good index of how difficult it is to estimate f using n observations.

Contrary to $\mathscr{L}(f, z)$, the asymptotic functional $I(f)$ is discontinuous, due to the fact that convergence in Theorem 2 is not uniform. This means that the asymptotic point of view could be very misleading when the convergence of $z^{-2/3}\mathscr{L}(f, z)$ to $I(f)$ occurs for very small values of z. It is easy to give simple examples of this fact: If f is a step function with m steps, by (3.3) R_n will converge to zero at a rate $n^{-1/2}$, which is not surprising since $I(f) = 0$. But such a phenomenon may be completely asymptotic if m is large, and it is perfectly possible that the right-hand side of (3.3) is much bigger than the bound given by (3.4) unless z is very small. The effect of a small value of $I(f)$ could be noticeable only with a very large number of observations. The opposite case could also occur: A bad behaviour of f in the tails could result in a very large value of $I(f)$ but will not affect $\mathscr{L}(f, n^{-1/2})$ for moderate n because the latter functional will be mainly determined for small n by the central part of the density.

The same types of arguments show that the minimax point of view could be just as misleading. The upper bound in (3.4), which is valid for the whole class $\mathscr{F}(a, H, L)$, is clearly much too big for most functions in the class and particularly for step functions with a small number of steps. This illustrates the well-known fact that there is no good choice between the two points of view, local asymptotic and global nonasymptotic. The intermediate point of view adopted here, local nonasymptotic, seems to be interesting because it can explain more precisely what happens and gives some idea of the reliability of the other points of view. The main drawbacks of those arguments are (i) in most cases, such local nonasymptotic evaluations are not available at the present time and (ii) when they are, as in our case, they are likely to be far from optimal, the constants being too large.

These considerations naturally lead to some interesting open questions. How do we improve on (4.1)? What about other estimators? Is it possible to improve on $\tilde{f}_n$ in a large number of situations? Is it possible to find optimality criteria in such a situation and how?

In any case, $\mathscr{L}(f, z)$ seems to be an interesting functional and is probably very closely related to the "difficulty in estimating f," if it is possible to give a reasonable meaning to this expression. It seems difficult to compute the exact value of $\mathscr{L}(f, z)$, but numerical methods would probably lead to reasonably precise upper bounds.

Acknowledgment. The author would like to thank Piet Groeneboom for his suggestions which greatly improved a former version of this paper and drastically shortened the original proof of Theorem 2.

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. (1986). On estimating a density using Hellinger distance and some other strange facts. *Probab. Theory Related Fields* **71** 271–291.

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.

Birnbaum, Z. W. and McCarty, R. C. (1958). A distribution-free upper confidence bound for $P(Y < X)$ based on independent samples of X and Y. *Ann. Math. Statist.* **29** 558–562.

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.

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View*. Wiley, New York.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

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

GRENANDER, U. (1980). *Abstact 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, Calif.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1983). Estimation of distribution density. *J. Soviet Math.* **21** 40–57.

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

MASSART, P. (1988). About the constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* To appear.

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

UNIVERSITÉ DE PARIS X
U.F.R. DE SCIENCES ECONOMIQUES
200 AVENUE DE LA RÉPUBLIQUE
92001 NANTERRE CEDEX
FRANCE