The Annals of Statistics
1987, Vol. 15, No. 3, 1013–1022

ON THE RISK OF HISTOGRAMS FOR ESTIMATING DECREASING DENSITIES

BY LUCIEN BIRGÉ

Université Paris X and U.A. C.N.R.S. 743

Suppose we want to estimate an element f of the space Θ of all decreasing densities on the interval $[a; a+L]$ satisfying $f(a^+) \le H$ from n independent observations. We prove that a suitable histogram $\hat{f}_n$ with unequal bin widths will achieve the following risk:

$$\sup_{f\in\Theta} \mathbb{E}_f\left[\int |\hat{f}_n(x) - f(x)|\,dx\right] \le 1.89(S/n)^{1/3} + 0.20(S/n)^{2/3},$$

with $S = \operatorname{Log}(HL+1)$. If $n \ge 39S$, this is only ten times the lower bound given in Birgé (1987). An adaptive procedure is suggested when a, L, H are unknown. It is almost as good as the original one.

1. Introduction. The problem of estimating decreasing densities has been studied by Grenander (1956, 1980), Barlow, Bartholomew, Bremner and Brunk (1972), Prakasa Rao (1969) and others but essentially from a local asymptotic point of view. Only recently, the global error in estimation (using variation distance as loss function) was considered by Groeneboom (1985) and the present author. The results of Groeneboom are very precise asymptotics while Birgé (1987) gives nonasymptotic lower bounds for the minimax risk together with upper bounds for estimation of unimodal densities. These bounds are obviously valid for decreasing densities but they are of no practical use because the estimators which reach the bounds are not computable. Groeneboom deals with a very practical estimator, but his results are truly asymptotic. Classical estimators like kernel estimators or histograms could also be used in this context; they would not lead to the right bounds [Birgé (1987)]. Therefore, our purpose in this paper is to design a simple estimator and derive upper bounds for its risk that are within a factor ten of the lower bound of Birgé (1987).

To be more precise, denote by $\Theta(a, H, L)$ the set of all decreasing densities on $[a; a+L]$ which are bounded by H. Given such a density $f(x)$ and an estimator $\hat{f}_n(x)$ depending on n i.i.d. random variables, we define the risk of $\hat{f}_n$ at f by

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

By translation and scale invariance, the minimax risk over the class $\Theta(a, H, L)$ only depends on the product HL. This motivates the introduction of the parameter $S = \operatorname{Log}(HL+1)$ and the minimax risk

$$R_n(S) = \inf_{\hat{f}_n} \sup_{f\in\Theta(a,H,L)} R_n(\hat{f}_n, f).$$

Received May 1984; revised May 1986.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62C20.
Key words and phrases. Histograms, minimax risk, decreasing densities.

Birgé (1987) proved that the following universal lower bound holds for $S \geq 1.3$ and $n \geq 39S$:

$$R_n(S) > 0.195(S/n)^{1/3}. \tag{1.1}$$

We shall design an estimator $\hat{f}_n(x)$ and derive the bound

$$\sup_{f \in \Theta(a, H, L)} R_n(f, \hat{f}_n) \leq 1.89(S/n)^{1/3} + 0.20(S/n)^{2/3}, \tag{1.2}$$

which implies

$$R_n(S) < 1.95(S/n)^{1/3}, \quad \text{for } n \geq 39S. \tag{1.3}$$

The estimator $\hat{f}_n$ is a histogram with unequal bin widths and is therefore very easy to implement. The choice of the bin widths is motivated by the study of the metric structure and entropy properties of the classes $\Theta(a, H, L)$ as described in Birgé (1987).

Unfortunately, $\hat{f}_n$ depends on the values of a, H, L which are usually unknown. Therefore, we also give an adaptive version $\hat{h}_n$ of $\hat{f}_n$, using estimates of a, H and L instead of the true values. This estimator also makes sense when S is infinite. When S is finite, we shall be able to give upper bounds for the risk of $\hat{h}_n$ which are very similar to the right-hand side of (1.2) and asymptotically equivalent. The bounds are given in terms of some rough estimates S_n of S. When S is finite, S_n is bounded by S; otherwise S_n will grow with n. This is a kind of *a posteriori* justification to the introduction of classes like $\Theta(a, H, L)$. Actually the true justification is (1.1), which shows that no uniform convergence of estimators can be expected without some restrictions on the behavior of f in the tails and near a [Birgé (1987)]. The parameter S gives a very rough indicator of how "peaked" the density is: The larger S is, the more difficult it is to estimate f.

2. Performance of histograms with unequal bin widths. In this section a, H, L are assumed known and $S = \mathrm{Log}(HL + 1)$. Define the numbers t, γ and the integer p by

$$t = (2S)^{2/3}n^{1/3} + 2S/3, \qquad \gamma = \exp(S/t) - 1, \qquad p - 1 < t \leq p. \tag{2.1}$$

Then

$$(1 + \gamma)^p - 1 \geq HL. \tag{2.2}$$

The construction of our histogram starts with a partition of $[a; a + L]$ into p contiguous intervals $I_0; I_1; \ldots; I_{p-1}$ of increasing lengths such that if $I_j = [x_j; x_{j+1})$ ($x_0 = a$ and $x_p = a + L$)

$$l_j = \frac{L\gamma(1 + \gamma)^j}{(1 + \gamma)^p - 1} = x_{j+1} - x_j, \qquad 0 \leq j \leq p - 1, \tag{2.3}$$

and $\sum_{j=0}^{p-1} l_j = L$ as required. Given n observations, we denote by N_j the number of them which belong to I_j ($n = \sum_{j=0}^{p-1} N_j$). The histogram corresponding to this partition is given by

$$\hat{f}_n(x) = \sum_{j=0}^{p-1} (nl_j)^{-1} N_j 1_{I_j}(x)$$

and satisfies

THEOREM 1. *If p and γ satisfy* (2.1) *the risk of the estimator $\hat{f}_n(x)$ can be bounded as follows*:

$$\mathbb{E}_f \left[\int |\hat{f}_n(x) - f(x)| \, dx \right] \le 1.89(S/n)^{1/3} + 0.20(S/n)^{2/3}.$$

PROOF. Let us define

$$f_j = f(x_j), \qquad \bar{f}_j = l_j^{-1} \int_{x_j}^{x_{j+1}} f(x) \, dx.$$

By assumption, f is decreasing. Therefore,

$$f_j \ge \bar{f}_j \ge f_{j+1}, \qquad \sum_{j=1}^{p} f_j l_{j-1} \le \sum_{j=0}^{p-1} \bar{f}_j l_j = 1. \tag{2.4}$$

The estimation error, measured by the $\mathbb{L}^1$-distance between f and $\hat{f}_n$, is

$$\begin{aligned} \int |\hat{f}_n(x) - f(x)| \, dx &= \sum_{j=0}^{p-1} \int_{I_j} |\hat{f}_n(x) - f(x)| \, dx \\ &\le \sum_{j=0}^{p-1} \int_{I_j} |f(x) - \bar{f}_j| \, dx + \sum_{j=0}^{p-1} \int_{I_j} |\hat{f}_n(x) - \bar{f}_j| \, dx = B + R. \end{aligned}$$

The error splits into a bias term B and a random term R, and we shall bound each of these. The monotonicity of f implies that

$$\int_{I_j} |f(x) - \bar{f}_j| \, dx \le (f_j - f_{j+1}) l_j / 2,$$

which in turn implies the bound for the bias term,

$$\begin{aligned} B &\le \frac{1}{2} \sum_{j=0}^{p-1} l_j (f_j - f_{j+1}) \\ &\le \frac{1}{2} \left(f_0 l_0 + \gamma \sum_{j=1}^{p-1} f_j l_{j-1} \right) \le \frac{\gamma}{2} \left[1 + \frac{HL}{(1+\gamma)^p - 1} \right] \le \gamma, \end{aligned} \tag{2.5}$$

using (2.3), (2.4), the fact that $f_0 \le H$ and (2.2).

In order to bound R we notice that N_j is a binomial random variable with parameters n and $\bar{f}_j l_j$. The definition of $\hat{f}_n(x)$ then implies that

$$\mathbb{E}_f \left[\int_{I_j} |\hat{f}_n(x) - \bar{f}_j| \, dx \right] \le \left[n^{-1} \bar{f}_j l_j (1 - \bar{f}_j l_j) \right]^{1/2}.$$

Finally,

$$
(2.6)\quad
\begin{aligned}
\mathbb{E}_f(R) &\le pn^{-1/2}\left[\frac{1}{p}\sum_{j=0}^{p-1}\left(\bar{f}_j l_j\left(1-\bar{f}_j l_j\right)\right)^{1/2}\right]\\
&\le pn^{-1/2}\left[p^{-1}\sum_{j=0}^{p-1}\bar{f}_j l_j\left(1-p^{-1}\sum_{j=0}^{p-1}\bar{f}_j l_j\right)\right]^{1/2}=\left(\frac{p-1}{n}\right)^{1/2},
\end{aligned}
$$

from the concavity of the function $x \mapsto (x(1-x))^{1/2}$.

From (2.5), (2.6) and (2.1) we get the following bound for the risk:

$$
(2.7)\quad
\begin{aligned}
\mathbb{E}_f\left[\int|\hat{f}_n(x)-f(x)|\,dx\right] &\le B+\mathbb{E}_f(R)\le \exp(S/t)-1+(t/n)^{1/2}\\
&\le 3/2(2S/n)^{1/3}+1/8(2S/n)^{2/3}.
\end{aligned}
$$

The last inequality is a consequence of our choice of t and Lemma A.2 of Birgé (1987). □

REMARKS.

(i) These computations are valid for all values of S and n but the bound is trivial when $n \le 1.2S$.

(ii) The choice of γ in (2.1) was made for simplicity. For large values of S, a larger value of γ could reduce the bias term. Numerical optimization using (2.5) and (2.7) would be convenient in this case and t should also be modified.

(iii) An unpleasant feature of $\hat{f}_n(x)$ is that it is supposed to estimate a decreasing density but is not likely to be itself decreasing. A decreasing density can be constructed using the pool adjacent violators algorithm [Barlow, Bartholomew, Bremner and Brunk (1972)]. If $\hat{f}_n$ is increasing on two consecutive bins, replace it by its mean value and pool the bins together, then repeat the procedure. Whatever the order of pooling, this leads to a unique decreasing step function $\tilde{f}_n(x)$ with $R_n(\tilde{f}_n(x), f) \le R_n(\hat{f}_n(x), f)$. This bound follows from the next proposition. The proof is easy and will be omitted.

PROPOSITION 1. *Suppose that we are given a decreasing function f and an increasing function $\hat{f}$ on some interval $[a; b]$. Fix $\bar{f} = (b-a)^{-1}\int_a^b \hat{f}(x)\,dx$. Then*

$$\int_a^b|\bar{f}-f(x)|\,dx \le \int_a^b|\hat{f}(x)-f(x)|\,dx.$$

3. Adaptive estimation. The preceding section was based on the assumption that we want to estimate an unknown density with support on $[a; a+L]$, bounded by H. In practical situations we have no *a priori* knowledge of a, L and H and we do not even know if they are finite. Without such knowledge, the construction of Section 2 is impossible. To obtain estimates we substitute estimators of the unknown parameters. As we shall see from the formulas, we do not lose much in doing so and our procedure will lead to computable bounds for the risk even when S is infinite.

To begin the construction we fix some integer l with $4 \le l \le n/2$ and define $\lambda = ((l-1)/l)(n/(n-l-1))$. Let $X_{(1)}, \dots, X_{(n)}$ be the n ordered observations,

$$I = (X_{(l)}; X_{(n)}), \qquad L' = X_{(n)} - X_{(l)}, \qquad H' = \frac{\lambda l}{n}\left[X_{(l)} - X_{(1)}\right]^{-1}.$$

The σ-algebra generated by $X_{(1)}$, $X_{(l)}$ and $X_{(n)}$ is denoted by $\mathscr{F}$. Obviously, $n' = n - l - 1$ observations lie in I and they have a decreasing conditional density given $\mathscr{F}$. Therefore, the algorithm of the preceding section with $a' = X_{(l)}$, L', H' and n' replacing a, L, H and n, respectively, leads to a decreasing estimator $\hat{g}_{n'}$ on $I = (a'; a' + L')$. Formulas (2.1) with n' and $S' = \operatorname{Log}(H'L' + 1)$ in place of n and S define the numbers γ', p' used in the construction of $\hat{g}_{n'}$. Our adaptive estimator will then be

$$\hat{h}_n = \frac{n-l}{n+1}\hat{g}_{n'} + \frac{l+1}{n+1}\left[X_{(l)} - X_{(1)}\right]^{-1}1_{[X_{(1)}, X_{(l)}]},$$

which is a density. The estimator $\hat{h}_n$ is not necessarily decreasing but we could modify it as described earlier in Proposition 1.

THEOREM 2. *If*

$$\delta = \operatorname{Log}\left[\frac{l-1}{l-2}\frac{n}{n-l-1}\right] \quad \text{and} \quad S_n = \mathbb{E}\left[\operatorname{Log}\left[\frac{l-2}{n}\frac{X_{(n)} - X_{(1)}}{X_{(l)} - X_{(1)}} + 1\right]\right],$$

then

$$\begin{aligned} R_n(f, \hat{h}_n) \le \frac{1}{n+1}\Bigg[2(l+1) + \left[\frac{(n-l)(l+1)}{n+2}\right]^{1/2} \\ +(n-l)\left[\frac{3}{2}\left(\frac{2(S_n+\delta)}{n-l-1}\right)^{1/3} + \frac{1}{8}\left(\frac{2(S_n+\delta)}{n-l-1}\right)^{2/3}\right]\Bigg] \\ < \frac{2l + l^{1/2}}{n} + \frac{9}{4n} + \frac{3}{2}\left(\frac{2(S_n+\delta)}{n}\right)^{1/3} + \frac{1}{8}\left(\frac{2(S_n+\delta)}{n}\right)^{2/3}. \end{aligned} \tag{3.1}$$

REMARK. The expected values S_n are always well defined, but possibly infinite, and increasing in n, for fixed l. They are finite whenever $\int |\operatorname{Log} x| f(x)\,dx$ is finite.

PROOF. Define

$$M_i = \int_{-\infty}^{X_{(i)}} f(x)\,dx, \qquad 1 \le i \le n, \qquad H_0 = f(X_{(l)}^+)$$

and let us first recall some well-known facts which we shall use repeatedly: If $\{U_{(i)}\}_{1 \le i \le n}$ are the order statistics from n uniform random variables on $[0; 1]$, the joint law of the $\{M_i\}_{1 \le i \le n}$ is the same as the law of the $\{U_{(i)}\}_{1 \le i \le n}$. Moreover, if we put $U_{(0)} = 0$ and $U_{(n+1)} = 1$, for $0 \le k < l \le n+1$ the law of

$U_{(l)} - U_{(k)}$ is $\beta(l-k, n+k-l+1)$. Recall that

$$\mathbb{E}(Y) = \frac{l}{n+1}, \qquad \mathrm{Var}(Y) = \frac{l(n-l+1)}{(n+1)^2(n+2)}, \qquad \mathbb{E}(1/Y) = \frac{n}{l-1},$$

when Y has $\beta(l, n-l+1)$ distribution. The conditional distribution of $U_{(n)} - U_{(l)}$ given $U_{(1)}, U_{(n)}$ is the distribution of $[U_{(n)} - U_{(1)}] \times Y$, where Y has a $\beta(n-l, l-1)$ distribution. Let g be the conditional density of observations falling in I. Then

$$g = 1_I \frac{f}{M_n - M_l} \quad \text{and} \quad g\left(X_{(l)}^+\right) = \frac{H_0}{M_n - M_l}.$$

The computations of Section 2 applied to g and its estimator $\hat{g}_{n'}$ show that

$$\mathbb{E}\left[\int_I |\hat{g}_{n'}(x) - g(x)|\,dx|\mathscr{F}\right]$$
$$\leq \gamma' + \left(\frac{p'-1}{n'}\right)^{1/2} + \frac{\gamma'}{2}\left[\frac{H_0 L'}{(M_n - M_l)\left[(1+\gamma')^{p'} - 1\right]} - 1\right],$$

the prime indicating that everything was constructed using H', L' $[H_0/(M_n - M_l)$ being the true upper bound of g instead of $H']$. As we have already seen, a good choice of p' allows us to bound the first two terms by

$$3/2(2S'/n')^{1/3} + 1/8(2S'/n')^{2/3}.$$

Let us now consider the last term. Since $(1+\gamma')^{p'} - 1 \geq H'L'$,

$$\text{(3.2)} \qquad \begin{aligned} \frac{H_0 L'}{(M_n - M_l)\left[(1+\gamma')^{p'} - 1\right]} &\leq \frac{H_0}{H'(M_n - M_l)} = \frac{H_0 n}{\lambda l}\frac{X_{(l)} - X_{(1)}}{M_n - M_l} \\ &\leq \frac{n}{\lambda l}\frac{M_l - M_1}{M_n - M_l}, \end{aligned}$$

because f is decreasing. From (2.1) we can check that γ' is an increasing function of S' and consequently of $H'L' = (\lambda l/n)((X_{(n)} - X_{(l)})/(X_{(l)} - X_{(1)}))$, which implies that γ' is decreasing with $X_{(l)}$ for fixed $X_{(1)}$ and $X_{(n)}$. On the contrary, $(n/\lambda l)((M_l - M_1)/(M_n - M_l)) - 1$ increases with M_l or $X_{(l)}$. Because γ' is positive, if

$$\text{(3.3)} \qquad \mathbb{E}\left[\frac{n}{\lambda l}\frac{M_l - M_1}{M_n - M_l} - 1 \middle| X_{(1)}, X_{(n)}\right] \leq 0,$$

then

$$\mathbb{E}\left[\frac{\gamma'}{2}\left[\frac{n}{\lambda l}\frac{M_l - M_1}{M_n - M_l} - 1\right] \middle| X_{(1)}, X_{(n)}\right] \leq 0$$

and, consequently, by (3.2)

$$\mathbb{E}\left[\frac{\gamma'}{2}\left[\frac{H_0 L'}{(M_n - M_l)(1+\gamma')^{p'} - 1} - 1\right]\right] \leq 0.$$

This means that (3.3) implies

$$\mathbb{E}\left[\int_I |\hat{g}_{n'}(x) - g(x)|\,dx\right] \le \tfrac{3}{2}\mathbb{E}\left[(2S'/n')^{1/3}\right] + \tfrac{1}{8}\mathbb{E}\left[(2S'/n')^{2/3}\right]. \tag{3.4}$$

In order to check (3.3), we notice that conditionally on $X_{(1)}$ and $X_{(n)}$, $((M_n - M_l)/(M_n - M_1))$ has a $\beta(n-l, l-1)$ distribution. Therefore,

$$\mathbb{E}\left[\frac{M_n - M_1}{M_n - M_l}\middle| X_{(1)}, X_{(n)}\right] = \frac{n-2}{n-l-1},$$

$$\mathbb{E}\left[\frac{n}{\lambda l}\frac{M_l - M_1}{M_n - M_l} - 1\middle| X_{(1)}, X_{(n)}\right] = \frac{n}{\lambda l}\left(\frac{n-2}{n-l-1} - 1\right) - 1 = 0.$$

Let us come back to (3.4). For $0 < \alpha < 1$, Jensen's inequality implies that

$$\begin{aligned}
\mathbb{E}(S'^{\alpha}) \le [\mathbb{E}(S')]^{\alpha} &= \left[\mathbb{E}\left[\operatorname{Log}\left(\frac{\lambda l}{n}\frac{X_{(n)} - X_{(l)}}{X_{(l)} - X_{(1)}} + 1\right)\right]\right]^{\alpha} \\
&= \left[\mathbb{E}\left[\operatorname{Log}\left(\frac{l-1}{n-l-1}\frac{X_{(n)} - X_{(1)}}{X_{(l)} - X_{(1)}} + \frac{n-2l}{n-l-1}\right)\right]\right]^{\alpha} \\
&\le \left[\mathbb{E}\left[\delta + \operatorname{Log}\left(\frac{l-2}{n}\frac{X_{(n)} - X_{(1)}}{X_{(l)} - X_{(1)}} + 1\right)\right]\right]^{\alpha} = (\delta + S_n)^{\alpha}.
\end{aligned}$$

Finally,

$$\begin{aligned}
\int |f(x) - \hat{h}_n(x)|\,dx &= M_1 + \int_{X_{(1)}}^{X_{(l)}} |f(x) - \hat{h}_n(x)|\,dx \\
&\quad + \int_{X_{(l)}}^{X_{(n)}} \left|\frac{n-l}{n+1}\hat{g}_{n'}(x) - g(x)(M_n - M_l)\right| dx + (1 - M_n) \\
&\le 1 + M_1 - M_n + (M_l - M_1) + \frac{l+1}{n+1} \\
&\quad + \frac{n-l}{n+1}\int_I |g(x) - \hat{g}_{n'}(x)|\,dx + \left|M_n - M_l - \frac{n-l}{n+1}\right|.
\end{aligned}$$

Taking expectations and using the fact that

$$\mathbb{E}\left[\left|M_n - M_l - \frac{n-l}{n+1}\right|\right] \le \operatorname{Var}^{1/2}(M_n - M_l),$$

we get

$$\begin{aligned}
\mathbb{E}\left[\int |f(x) - \hat{h}_n(x)|\,dx\right] &\le \frac{l+1}{n+1} + \frac{l}{n+1} + \frac{1}{n+1} \\
&\quad + \frac{n-l}{n+1}E + \left(\frac{(n-l)(l+1)}{(n+1)^2(n+2)}\right)^{1/2},
\end{aligned}$$

where

$$E = \frac{3}{2}\left(\frac{2(S_n + \delta)}{n - l - 1}\right)^{1/3} + \frac{1}{8}\left(\frac{2(S_n + \delta)}{n - l - 1}\right)^{2/3}$$

The first inequality follows. The second inequality follows from our bounds on l. □

The last problem is the choice of l which would be easy if S_n were known, but S_n depends on f. We shall, therefore, give some heuristics for a reasonable choice of l, based on the following corollary. Define

$$k_n(S) = \tfrac{3}{2}(2S/n)^{1/3} + \tfrac{1}{8}(2S/n)^{2/3},$$

$$r_n(l, S) = \left[\frac{2l + l^{1/2}}{n} + \frac{9}{4n} + k_n(S + \delta)\right]k_n^{-1}(S).$$

Corollary 1.

(i) *For fixed n and l, $r_n(l, S)$ is a decreasing function of S.*
(ii) *If $l(n)$ is any sequence diverging to infinity with $l(n) = o(n^{2/3})$, then $r_n(l(n), S)$ converges to 1 uniformly for $S \geq \alpha > 0$.*
(iii) *If f belongs to some class $\Theta(a, H, L)$, then $S_n \leq S = \mathrm{Log}(HL + 1)$.*

Proof. Claims (i) and (ii) are easy and (iii) follows from

$$\mathbb{E}\left[\frac{X_{(n)} - X_{(1)}}{X_{(l)} - X_{(1)}}\right] \leq LH\mathbb{E}\left[(M_l - M_1)^{-1}\right] = \frac{nLH}{l - 2}$$

and Jensen's inequality. □

Claim (iii) means that in the case of f in $\Theta(a, H, L)$,

$$R_n(f, \hat{f}_n) \leq k_n(S), \qquad R_n(f, \hat{h}_n) \leq r_n(l, S)k_n(S).$$

From (ii), we see that the two bounds are asymptotically equivalent if l is conveniently chosen. The first result implies

$$R_n(f, \hat{h}_n) \leq r_n(l, S_0)k_n(S), \quad \text{for } S \geq S_0.$$

This suggests a heuristic choice for l: Minimize $r(l, S_0)$ for some value S_0 which is supposed to be smaller than the true S. Possibilities include $S_0 = \mathrm{Log}\, 2$ or 1 or a statistic very likely to underestimate S. One possible statistic is

$$\mathrm{Log}\left[\frac{m - 2}{n}\,\frac{X_{(n)} - X_{(1)}}{X_{(m)} - X_{(1)}} + 1\right],$$

with m between 10 and 20. Even if f or its support is unbounded this last statistic seems to be reasonable because of (i) and also because numerical investigations show that $r_n(l, S)$ varies rather slowly with S. Various computations performed with n between 200 and 1000 and $S \geq 1$ indicate that the optimal l is between 4 and 6 and in this range of values of n, $l = 5$ always appears as a very reasonable choice. In this case, $r_n(l, S)$ is smaller than 1.33 (smaller than 1.22 if $S \geq 2$). This indicates that even with small n and S, the ratio $r_n(l, S)$ is not much larger than one.

4. Concluding remarks.

— The main problem for adaptation comes from H. In the case of known a and L, it is natural to replace $X_{(1)}$ by a and $X_{(n)}$ by L but it leads to little improvement.

— Groeneboom (1985) studied some asymptotics for the Grenander estimator $\hat{k}_n$ [see Barlow, Bartholomew, Bremner and Brunk (1972)] and proved for smooth f that

$$\lim_n n^{1/3} R_n(\hat{k}_n, f) = 0.82 \int \left| \frac{f(x) f'(x)}{2} \right|^{1/3} dx. \tag{4.1}$$

Birgé (1987) points out that the supremum of the right-hand side over the class $\Theta(a, H, L)$ is larger than $0.65 S^{1/3}$. This is then comparable with $\lim_n n^{1/3} k_n(S)$ at least from a minimax point of view, apart from a better constant for the Grenander estimator. But this comparison is not really meaningful: The limit in (4.1) cannot be uniform since the right-hand side of (4.1) is a badly discontinuous function of f (with respect to the $\mathbb{L}^1$-norm) and the risk function is necessarily continuous. Secondly, (4.1) holds for smooth functions, whereas our results use no such assumptions. In particular, smoothness would asymptotically reduce our bias term and improve our bounds in the limit.

— Our estimator is constructed to get a low value of the minimax risk. The risk should depend on the whole shape of the density, not only on S. However, the risk of our estimator always is of the order $(S/n)^{1/3}$ even for nice densities (like piecewise linear with two pieces) while it is intuitively clear that it should be possible to do better.

Acknowledgment. I have benefitted from numerous suggestions of referees concerning my English grammar and style.

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

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

GRENANDER, U. (1980). *Abstract Inference.* Wiley, New York.

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proc. 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, Belmont, Calif.

KIEFER, J. (1982). Optimum rates for non-parametric density and regression estimates, under order restrictions. In *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.) 419–428. North-Holland, Amsterdam.

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

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