The Annals of Statistics
1987, Vol. 15, No. 3, 995–1012

ESTIMATING A DENSITY UNDER ORDER RESTRICTIONS: NONASYMPTOTIC MINIMAX RISK

BY LUCIEN BIRGÉ

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

Let us consider the class of all unimodal densities defined on some interval of length L and bounded by H; we shall study the minimax risk over this class, when we estimate using n i.i.d. observations, the loss being measured by the $\mathbb{L}^1$ distance between the estimator and the true density. We shall prove that if $S = \text{Log}(HL + 1)$, upper and lower bounds for the risk are of the form $C(S/n)^{1/3}$ and the ratio between those bounds is smaller than 44 when S/n is smaller than 220^{-1}.

1. Introduction. Let us consider some interval I of length L on the real line and the set Θ of all unimodal densities on I, bounded by H. More precisely, any element f of Θ should satisfy $\int f(x)\,dx = 1$, $f(x) = 0$ if $x \notin I$, $0 \leq f(x) \leq H$ if $x \in I$, and there exists some m in I such that $f(x)$ is nondecreasing on $]-\infty; m[$ and nonincreasing on $]m; +\infty[$. Besides that, we do not assume any smoothness and f could be discontinuous. Suppose that we want to estimate the density f using n i.i.d. observations. If $\hat{f}_n(x)$ is any estimator, the loss will be measured by the $\mathbb{L}^1$ distance between the two densities (this is also the variation distance between the corresponding measures) and the resulting risk of $\hat{f}_n(x)$ will then be defined by

$$R_n(f, \hat{f}_n) = \mathbb{E}_f\left[\int |\hat{f}_n(x) - f(x)|\,dx\right]. \tag{1.1}$$

The minimax risk on Θ is given by

$$R_n(\Theta) = \inf_{\hat{f}_n} \sup_{f \in \Theta} R_n(f, \hat{f}_n). \tag{1.2}$$

Our main purpose in this paper will be to show that under some mild restrictions on n and HL,

$$\tfrac{1}{4}(S/n)^{1/3} \leq R_n(\Theta) \leq 11(S/n)^{1/3}, \qquad S = \text{Log}(HL + 1). \tag{1.3}$$

This means that we get a nonasymptotic evaluation of the minimax risk for this situation and, although the ratio between upper and lower bounds is not terrific, it is the only result of this type that we know of. As an illustration, we get for $S = 1.3$ and $n = 300$ a lower bound of 0.04. Of course, in such problems the choice of the loss function is widely arbitrary. The particular choice of $\mathbb{L}^1$-distance rather than $\mathbb{L}^p$, with $p > 1$, is motivated by the fact that $\mathbb{L}^1$ is scale

---

Received May 1984; revised May 1986.

[1]Most of this work was carried out during the author's visit at the Mathematical Sciences Research Institute, Berkeley.

AMS 1980 *subject classifications*. Primary 62G05; secondary 41A46.

Key words and phrases. Unimodal densities, nonasymptotic minimax risk, rates of convergence of estimates.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

invariant (not $\mathbb{L}^p$) and intrinsic (it only depends on the two measures, not on the choice of the dominating one). As a consequence the minimax risk depends only on the product HL which is invariant by a change of scale. Those properties are also shared by Hellinger distance, but $\mathbb{L}^1$ is more tractable in this context.

The motivation for such a study is a paper by Kiefer (1982) who proved in various situations that the rate of convergence in estimation cannot be improved by order restrictions, but also suggested that the constants might change. The "rate" means the factor depending on the number of observations (here $n^{-1/3}$) and the "constant" is a factor depending on the shape of the unknown density (in our case it will be $S^{1/3}$, which is some sort of an index of how peaked the densities could be). In order to investigate this problem we shall consider some classical family of densities with bounded variation. Let us denote by Λ the class of all densities having their support on I and a total variation bounded by $V = 2H$. Then Θ is the subset of all unimodal densities in Λ and Kiefer's problem becomes, "Is estimation improved if we restrict the class of densities from Λ to Θ?" The treatment of Λ is rather classical and could follow the lines of Bretagnolle and Huber (1979). Lower bounds are easily found using small perturbations and upper bounds are given by kernel estimators or even simple histograms. Such computations are classical and lead to bounds of the type

$$C_1(HL/n)^{1/3} \leq R_n(\Lambda) \leq C_2(HL/n)^{1/3}. \tag{1.4}$$

Obviously, the upper bound is also valid for $R_n(\Theta)$, but an attempt to extend the lower bound would not be successful for the following reason: The monotonicity restriction implies the replacement of sinusoidal perturbations by monotonous ones which should also be equal according to Bretagnolle's technique. This and the fact that we should only consider unimodal densities imply that the density to be perturbed should be triangular. If we build a system of small perturbations around a triangular function we get a lower bound of the form

$$R_n(\Theta) \geq C_3 n^{-1/3}. \tag{1.5}$$

The explanation for the absence of HL in the formula is the fact that for a triangular density the product of the basis by the height is constant. This also means that there is a gap between this lower bound and the upper bound in (1.4). It is also possible to check that the upper bounds for the risk of kernel estimators, or histograms, as given in (1.4), are sharp and the only possible improvement could be in the constant C_2. This comes from the fact that the risk is the result of a balance between a bias term and an error term and it is impossible to make both of them simultaneously small. This means that in order to get upper and lower bounds of comparable magnitude, new methods have to be developed and this will be the purpose of this paper.

Little is known about estimation of unimodal densities with unknown mode [see Wegman (1970)]. When the mode is known, the problem is simpler because it becomes quite similar to estimation of decreasing densities. This was studied a long time ago [see, for example, Prakasa Rao (1969)]. Recently, estimation of decreasing densities received a beautiful treatment by Groeneboom (1985) who found very sharp asymptotic bounds, but of a completely different type, since they are not uniform with respect to the underlying density f.

The techniques used in the present paper involve a precise description of the metric structure of our parameter space. We shall first derive lower bounds in Section 2 using perturbation arguments, but in order to take into account the size of our parameter space we need to use unequal perturbations. This will be taken care of by the techniques developed in Birgé (1986). The idea is roughly as follows: The parameter space being compact, one can approximate it with a given error by a finite subset. This set has to be more dense around a very peaked density than around a flat one. This means that estimation is more difficult around peaked densities and that we should build our systems of perturbations around such densities. Also, because of the similarity of the two situations, we shall, at the same time, derive lower bounds for classes of decreasing densities.

In Section 3 we shall compute upper bounds following the general method developed in Birgé (1983): Take a finite approximation of the parameter space and apply to this finite set a robust version of the maximum likelihood. Then you get two error terms: one coming from the discretization and one from the size of your finite set. The construction is optimal when the two terms are of the same order. The only problem is the construction of the approximating sets. We shall use step functions whose shape was suggested by the previous investigation concerning lower bounds. The idea is that step lengths should be small and their heights large when the density is large, and vice-versa when the density is small. This is not very different from a balance between a bias and a variance term.

Putting upper and lower bounds together, we confirm (1.3) and Kiefer's suggestions: The rate is still $n^{-1/3}$ [which was already clear from (1.4) and (1.5)] but the effect of the shape is very different since HL has been replaced by $\text{Log}(HL + 1)$ in the formulas.

We shall conclude with various remarks and the evocation of some desirable improvements and open problems.

All constructions and results are given in the core of the paper, but most of the proofs are very technical and have been collected in an Appendix at the end of the paper.

2. Evaluation of lower bounds. We shall treat simultaneously the cases of decreasing and unimodal densities, which are very similar, in order to avoid repetitions. We shall thus consider two classes of functions: $\Theta_1(H, L)$ will be the set of all decreasing densities on $[0; L]$ bounded by H and $\Theta_2(H, L)$, the set of unimodal densities, on $[-L/2; L/2]$ bounded by H. The particular choices of $[0; L]$ or $[-L/2; L/2]$ are obviously irrelevant because our problem is clearly translation invariant. It is also scale invariant, as already mentioned, and the risk will only depend on the product HL or on the more suitable quantity $S = \text{Log}(HL + 1)$. We shall therefore replace the notation $\Theta_j(H, L)$ by $\Theta_j(S)$ for $j = 1, 2$. The similarity of the two problems comes from the following transformation: Starting from a decreasing density on $[0; L]$, contract it into a decreasing function on $[0; L/2]$, the corresponding measure now having a total mass of

1/2. Then take its reflection, with respect to the vertical axis, and put both together. This gives a unimodal density on $[-L/2; L/2]$ with mode at zero. Once the problem of decreasing densities is solved, applying this transformation will give the solution for unimodal densities.

Since the construction is rather technical, we shall first try to give the main idea for the decreasing case. We want to build perturbations around some given density which will be a very peaked hyperbola (in order to get a maximum value of H and a support of length L). Each perturbation is a step from some point M of coordinates (x, y) to P: $(x + h, y - k)$. Two choices will be possible:

(a) jump from M to $(x, y - k/2)$, then go to $(x + h, y - k/2)$ and jump to P; or

(b) go from M to $(x + h/2, y)$, jump to $(x + h/2, y - k)$ and then go to P.

If we have p successive steps with such choices for each step, this will result in 2^p different functions. They will be densities because our steps follow a density, and although the steps are unequal, we shall design them in such a way that for each step both Hellinger distance and variation distance between solutions (a) and (b) are constant. This is the idea. In order to be more precise, let us just fix some notation: We shall denote by J_j the sets $J_1 = \{1; 2; \dots; p\}$ and $J_2 = J_1 \cup \{-1; -2; \dots; -p\}$ for some integer p which will be explicitly defined later. Recall that $S = \operatorname{Log}(HL + 1)$ and define

$$\varepsilon = \frac{4}{5}(5jS/n)^{1/3}, \qquad u = \frac{L/j}{(1+\varepsilon)^p - 1}, \qquad q = jp,$$

$$A = \left(1 + \frac{5\varepsilon^2}{64}\right)\left(1 + \frac{\varepsilon}{2}\right)^{-2}\left(1 - \frac{3\varepsilon}{4} + \frac{7\varepsilon^2}{12}\right)^{-2}, \tag{2.1}$$

$$\lambda = \frac{1+\varepsilon}{uq\varepsilon(1+\varepsilon/2)}, \qquad x_i = u\left[(1+\varepsilon)^i - 1\right], \qquad 0 \le i \le p.$$

For each i in J_1, the length l_i of $I_i = [x_{i-1}; x_i[$ is $u\varepsilon(1+\varepsilon)^{i-1}$. Let m_i be $(x_i + x_{i-1})/2$ and define two functions $f_i(x)$ and $g_i(x)$ on I_i by

$$f_i(x) = \lambda(1+\varepsilon)^{-i}(1+\varepsilon/2),$$

$$g_i(x) = \begin{cases} \lambda(1+\varepsilon)^{-i+1} & \text{for } x \le m_i, \\ \lambda(1+\varepsilon)^{-i} & \text{for } x > m_i, \end{cases}$$

f_i, g_i being zero on I_i^c. For negative i we define I_i, f_i and g_i analogously, by symmetrization,

$$x_{-i} = -x_i, \qquad I_{-i} = -I_i, \qquad l_{-i} = l_i,$$

$$f_{-i}(x) = f_i(-x), \qquad g_{-i}(x) = g_i(-x).$$

Using Lemma A.1 we get

$$\int_{I_i} f_i(x)\,dx = \int_{I_i} g_i(x)\,dx = \frac{u\lambda\varepsilon(1+\varepsilon/2)}{1+\varepsilon} = q^{-1},$$

$$\int_{I_i} |f_i(x) - g_i(x)|\,dx = \frac{\lambda u\varepsilon^2}{2(1+\varepsilon)} = \frac{\varepsilon}{q(2+\varepsilon)} = \alpha,$$

$$\frac{1}{2}\int_{I_i}\left(f_i^{1/2}(x) - g_i^{1/2}(x)\right)^2 dx \le \frac{\varepsilon^2}{8q(2+\varepsilon)^2}\left(1 + \frac{5\varepsilon^2}{64}\right) = \beta.$$

For $j = 1$ or 2 we may consider the sets of 2^q density functions defined by

$$F_j = \left\{ f = \sum_{i\in J_j} \lambda_i f_i + (1-\lambda_i) g_i \,\middle|\, \lambda_i = 0 \text{ or } 1, \forall i \right\},$$

and check that $F_j \in \Theta_j$ as soon as $\lambda \le H$. For such systems of perturbations as F_j, we can get lower bounds for the minimax risk using Assouad's lemma as explained in Assouad (1983) or Birgé (1986). Let $R_j(n)$ be the minimax risk on F_j when the loss function is the $\mathbb{L}^1$-distance and n is the number of i.i.d. observations at hand. Then Assouad's lemma implies that

$$\begin{aligned} R_j(n) &\ge \frac{q\alpha}{2}\left[1 - \left(1 - (1-\beta)^{2n}\right)^{1/2}\right] \\ &= \frac{\varepsilon}{4(1+\varepsilon/2)}\left[1 - \left(1 - (1-\beta)^{2n}\right)^{1/2}\right]. \end{aligned} \tag{2.2}$$

This also gives a lower bound for the minimax risk over $\Theta_j(S)$ as soon as F_j is included in Θ_j, i.e., when $\lambda \le H$. This inequality will be satisfied for a convenient choice of p.

PROPOSITION 1. *If $S \ge 1 + 3\varepsilon/4$ and $\varepsilon \le 1.12$, then $\lambda \le H$ when p is given by*

$$SA/\varepsilon \le p < SA/\varepsilon + 1, \tag{2.3}$$

A being defined in (2.1).

We now just need a convenient evaluation of the right member of (2.2). One possibility is

PROPOSITION 2. *Assume that $\varepsilon \le 7/4$, $n \ge 3$ and p is given by* (2.3). *Then*

$$R_j(n) \ge 0.207(jS/n)^{1/3} - 0.137jS/n.$$

The proofs of these propositions rely on elementary calculus and will be sketched in the Appendix. Now let us denote by $R_n^{(j)}(S)$ the minimax risk for estimating a density of $\Theta_j(S)$ using n i.i.d. observations, i.e., for $j = 1, 2$,

$$R_n^{(j)}(S) = \inf_{\hat{f}_n(x)} \sup_{f \in \Theta_j(S)} \mathbb{E}_f\left[\int |f(x) - \hat{f}_n(x)|\, dx\right].$$

A convenient rewriting of the assumptions of Propositions 1 and 2 leads to

THEOREM 1. *Suppose that S and n satisfy the inequalities*

$$1 + 0.6(5jS/n)^{1/3} \leq S \leq 0.55n/j. \tag{2.4}$$

Then

$$R_n^{(j)}(S) \geq 0.207(jS/n)^{1/3} - 0.137jS/n. \tag{2.5}$$

We can see that (2.4) is not a severe restriction and will hold, if, for instance,

$$1.6 \leq S \leq n/(5j).$$

3. Upper bounds for the risk. Our techniques are derived from the metric theory developed by Le Cam (1973) and Birgé (1983). This involves a discretization of the parameter space $\Theta_2(S)$, which will be replaced by an ε-net N and a robust version of the maximum likelihood on N. It works as follows: The set F of nonnegative bounded functions on $[-L/2; L/2]$ is a metric space with distance d given by

$$d(f, g) = \int |f(x) - g(x)|\, dx.$$

$\Theta_2(S)$ being a compact subset can be approximated by a finite subset N of $\mathscr{F}$, and N will be called an ε-net for $\Theta_2(S)$, if for any f in $\Theta_2(S)$, we can find g in N such that $d(f, g) \leq \varepsilon$. Suppose that we have found N; the construction of a robust maximum likelihood on N is described in Birgé (1983) and involves only the possibility of testing between two balls of radius ε. In the case of n i.i.d. observations and the distance d (which is variation distance), this possibility has been proved in studies about robust testing as can be seen in Huber (1965), Huber and Strassen (1973) or Birgé (1984a). The only result we need is this one: Given two balls of probability measures with respective centers f and g in $\mathscr{F}$ and radii ε, there exists a test between these balls using n i.i.d. observations, and the supremum α_n of both errors of this test satisfies

$$\alpha_n \leq \exp\left[-(n/8)(d(f, g) - 2\varepsilon)^2\right]. \tag{3.1}$$

Such results are explained in detail in Section 4 of Birgé (1983). We have only to find some N with a cardinal as small as possible. We shall first build nets for unimodal densities, when the mode is approximately known, and then put all nets corresponding to the various positions of the mode together. The construction is as follows: Assume some integer $p \geq S$ is given and fix

$$\begin{aligned} &\varepsilon = \exp(S/p) - 1, \\ &y_j = L^{-1}\left[(1 + \varepsilon)^j - 1\right], \quad 0 \leq j \leq p, \quad Y = \{y_0; \dots; y_p\}. \end{aligned} \tag{3.2}$$

Then we cover the interval $I = [-L/2; L/2]$ which serves as a support for the densities in $\Theta_2(S)$ by k intervals of length l_0 with

$$\frac{LH}{\varepsilon} \leq k < \frac{LH}{\varepsilon} + 1, \qquad l_0 = \varepsilon/H.$$

Take one of those k intervals for I_0 and starting with $I_0 = [x_{-1}; x_1]$ define

$$x_{i+1} = \min\left(x_i + l_0(1+\varepsilon)^i; L/2\right), \qquad I_i = [x_i; x_{i+1}], \quad i \geq 1,$$

$$x_{i-1} = \max\left(x_i - l_0(1+\varepsilon)^{-i}; -L/2\right), \qquad I_i = [x_{i-1}; x_i], \quad i \leq -1.$$

We get a new covering of I by the intervals I_i for $-p_2 \leq i \leq p_1$, p_1 and p_2 being the smallest integers such that

$$x_{p_1+1} = L/2 \quad \text{and} \quad x_{-p_2-1} = -L/2.$$

Then by construction if l_i is the length of I_i,

$$l_i \leq (1+\varepsilon)l_{i-1}, \quad 0 < i \leq p_1, \quad l_i \leq (1+\varepsilon)l_{i+1}, \quad -p_2 \leq i < 0. \tag{3.3}$$

Let us consider the family of all unimodal functions g on I satisfying for $-p_2 \leq i \leq p_1$,

$$\text{if } x \in I_i, \qquad g(x) = g_i \in Y,$$

$$g_{i+1} \leq g_i, \quad \text{for } i \geq 0, \quad \text{and} \quad g_{i-1} \leq g_i, \quad \text{for } i \leq 0,$$

i.e, functions taking their values in Y, constant on each interval I_i and having a mode in I_0. Obviously, this family depends on the initial choice of I_0 among k possibilities which imply that we can get k such families. The union $F(\varepsilon)$ of these k families of unimodal functions has the following properties:

PROPOSITION 3. *Suppose that $p \geq 13.5S$. Then if $3\eta = 7.51\varepsilon$, $F(\varepsilon)$ is an η-net for $\Theta_2(S)$ which means that for any f in $\Theta_2(S)$ we can find some g in $F(\varepsilon)$ such that*

$$\int |f(x) - g(x)|\,dx \leq \eta = 7.51\varepsilon/3. \tag{3.4}$$

Moreover, $p \geq 10$ and the cardinal of $F(\varepsilon)$ is bounded by

$$\operatorname{card} F(\varepsilon) \leq 0.336S^{-1}\exp(2pc^2), \qquad c = (\operatorname{Log} 4 + 1/27)^{1/2}. \tag{3.5}$$

The proof can be found in the Appendix. Actually $F(\varepsilon)$ is too big for what we really need because there are functions g in $F(\varepsilon)$ such that $\int g(x)\,dx$ is far away from 1 and cannot be used for approximating densities. If we restrict ourselves to the subset N of $F(\varepsilon)$ defined by

$$N = \left\{ g \in F(\varepsilon) \middle| \left| \int g(x)\,dx - 1 \right| \leq \eta \right\},$$

N will also be an η-net for $\Theta_2(S)$ and, obviously, satisfies (3.5). We shall then base the construction of the estimator on this net N as indicated in Birgé (1983). Starting with n i.i.d. observations, this construction provides us with some estimator $\hat{g}_n$, with values in N, which means that the estimator is not necessarily stochastic but only satisfies

$$1 - \eta \leq \int \hat{g}_n(x)\,dx \leq 1 + \eta.$$

Bounds for the risk of such estimators are easily computed as soon as you have inequalities like (3.1) on the errors of the tests and (3.5) on the cardinality of N. The technique of Birgé (1983), Theorem 2.4, applied to our particular case leads to the following:

THEOREM 2. *Assume that $n \geq 6^3 S$ and choose p satisfying*

$$p - 1 < (7.51S/(2c))^{2/3} n^{1/3} + 2S/3 \leq p,$$
$$c = (\operatorname{Log} 4 + 1/27)^{1/2} \simeq 1.19. \tag{3.6}$$

Then

$$\sup_{f \in \Theta_2(S)} \mathbb{E}_f\left[\int |f(x) - \hat{g}_n(x)|\,dx\right]$$
$$< 10.5(S/n)^{1/3} + 0.82(S/n)^{2/3} + 2.4S^{-4/3}n^{-2/3}. \tag{3.7}$$

REMARK. The restriction $n \geq 6^3 S$ does not matter because (3.7) becomes uninteresting for smaller values of n.

PROOF. We first notice that (3.6) and the requirement on S/n imply that $p \geq 13.5S$. Then Proposition 3 is valid and $p \geq 10$. Consider the sequence $\{t_i\}_{i \geq 1}$ given by

$$t_i = 2 + t + a(i-1), \qquad i \geq 1, \quad a = \frac{4 \operatorname{Log} 2}{n\eta^2 t}, \quad t = \frac{4c}{\eta}[(p-1)/n]^{1/2}.$$

Now for some g in N such that $d(f, g) \leq \eta$ denote by N_i the number of points g' in N such that

$$t_i \eta \leq d(g; g') < t_{i+1}\eta;$$

then using (3.1) we get from Birgé (1983), Lemma 2.1,

$$P_f[d(f; \hat{g}_n) \geq (t_i + 1)\eta] \leq \sum_{j \geq i} N_j \exp\left[-\frac{n}{8}(t_j - 2)^2 \eta^2\right],$$

which implies

$$\mathbb{E}_f[d(f; \hat{g}_n)] \leq (t_1 + 1)\eta + \eta \sum_{i \geq 1} (t_{i+1} - t_i) P_f[d(f; \hat{g}_n) \geq (t_i + 1)\eta]$$
$$\leq (t_1 + 1)\eta + a\eta \sum_{j \geq 1} j N_j \exp\left[-\frac{n}{8}(t_j - 2)^2 \eta^2\right].$$

Our choice of a implies that the sequence $j\exp[-(n/8)(t_j - 2)^2\eta^2]$ is decreasing and we finally deduce, using the bound (3.5) for $\Sigma_j N_j$, that

$$\mathbb{E}_f\left[d(f; \hat{g}_n)\right] \le \eta(t+3) + \frac{4\operatorname{Log}2}{n\eta t}\exp\left[-\frac{n}{8}t^2\eta^2\right]\left(\frac{0.336}{S}\exp(2pc^2)\right).$$

Replacing η and t by their values and noticing that $p \ge 10$, we get

$$\begin{aligned}\mathbb{E}_f\left[d(f; \hat{g}_n)\right] &\le 4c\left(\frac{p-1}{n}\right)^{1/2} + 7.51\varepsilon + \frac{0.336\operatorname{Log}2\exp(2c^2)}{cS(n(p-1))^{1/2}}\\ &\le 4c\left(\frac{p-1}{n}\right)^{1/2} + 7.51\left[\exp\left(\frac{S}{p}\right) - 1\right] + 3.55\frac{(np)^{-1/2}}{S}.\end{aligned}$$

The choice of p and Lemma A.2 imply

$$\begin{aligned}\mathbb{E}_f\left[d(f; \hat{g}_n)\right] &\le \frac{3}{2}\left(32 \times 7.51c^2\frac{S}{n}\right)^{1/3}\\ &\quad + \frac{1}{8}\left(\frac{32c^2}{7.51^{1/2}}\frac{S}{n}\right)^{2/3} + 3.55\left(\frac{2c}{7.51}\right)^{1/3} S^{-4/3}n^{-2/3}\end{aligned}$$

and then the conclusion. □

4. Final remarks.

REMARK 1. The first thing to notice is the fact that our estimator $\hat{g}_n$ is theoretically computable but not practical because its computation involves more than 16^p tests with p behaving like $n^{1/3}$. It is a mere technical tool to get bounds for the minimax risk. Moreover, as we noticed in Section 3, it is not even a stochastic estimator in the sense that generally $\int_{-L/2}^{L/2}\hat{g}_n(x)\,dx \ne 1$. It would be possible to remedy that: N being an η-net for $\Theta_2(S)$, every point in N can be replaced by a point in $\Theta_2(S)$ which is at a distance smaller than η and we would then get a 2η-net for $\Theta_2(S)$ having the same cardinality. A similar construction would lead to a true stochastic estimator with slightly bigger constants in (3.7). Increasing the constants is certainly not desirable here and since we know that these estimators are not practically useful it is not a serious drawback if they are not stochastic.

REMARK 2. We restricted our proofs for upper bounds to the case of unimodal densities but similar results hold for decreasing ones. A previous version of this paper led to

$$R_n^{(1)}(S) < 7.66(S/n)^{1/3} + 0.55(S/n)^{2/3} + S^{-2/3}n^{-5/6}.$$

These bounds are actually superseded by the performance of a more practical estimator, a histogram with unequal bin widths. These results will be developed in a forthcoming paper.

REMARK 3. Let us denote the minimax risk for unimodal density estimation by $R_n(S)$ $[= R_n^{(2)}(S)]$. Assuming that $n \geq 220S$, Theorem 3 holds as well as Theorem 1 if $S \geq 5/4$ in order to satisfy (2.4). Then (2.5) and (3.7) hold together, which leads to the following inequalities for $R_n(S)$:

$$(2S/n)^{1/3}\left[0.207 - 0.137(2S/n)^{2/3}\right]$$
$$\leq R_n(S) \leq (S/n)^{1/3}\left[10.5 + (S/n)^{1/3}(0.82 + 2.4S^{-2})\right]$$

and, finally,

$$0.25(S/n)^{1/3} \leq R_n(S) \leq 10.9(S/n)^{1/3},$$

which implies (1.3).

REMARK 4. The restriction to classes like $\Theta_2(S)$ may look strange at first sight. Why not take the class of all unimodal densities? The answer is simple: There is no uniform rate of convergence over such a big class whatever the choice of the estimator. This is actually a consequence of our lower bounds (2.5). Since the class of all unimodal densities contains all classes $\Theta_2(S)$ for all values of S, we can fix S arbitrarily in (2.5) as soon as (2.4) is satisfied. The choice $n \geq 8$, $S = n/4$, leads to the lower bound 0.095. The minimax risk for estimating an arbitrary unimodal density is then always larger than 0.095 no matter how large the number of observations is. This result still holds if we restrict the class to uniformly bounded densities with unbounded support or unbounded densities with a fixed compact support. This means that if we want to get uniform rates of convergence, we have to restrict the size of the class of densities at hand by putting some restrictions on the tails and the growth of the densities. The same holds for decreasing densities. The choice of classes $\Theta_2(H, L)$ is not the only possible one and was made for the sake of simplicity but also because it could be adapted to treat more general cases as we shall see. A natural extension of the classes $\Theta_2(H, L)$ is a class of unbounded functions, with unbounded support, but some specified restrictions on the tails and the growth near the mode. The problem is what type of restriction to ask for. Different restrictions will lead to different rates. A possible way of dealing with this is as follows: For given H, L and n we know how to get a suitable estimator $\hat{g}_n$. It is proved in Birgé (1984b) that those estimators could be robustified. Let us call $\tilde{g}_n$ the robustified version. The robustness property can be expressed by the fact that if f is the true density [not necessarily belonging to $\Theta_2(H, L)$] and $d(f, \Theta_2) \leq k\varepsilon$ [ε given by (3.2)], then

$$R_n(\tilde{g}_n, f) \leq Ck(S/n)^{1/3}$$

and C is a universal constant. If we consider two nondecreasing sequences $\{H_n\}, \{L_n\}$, our computations will produce corresponding sequences S_n, p_n, ε_n and, as soon as $d(f, \Theta_2(H_n, L_n)) \leq k\varepsilon_n$, for suitable $\tilde{g}_n$

$$R_n(\tilde{g}_n, f) \leq Ck(S_n/n)^{1/3}.$$

Any pair of such sequences $\{H_n\}, \{L_n\}$ defines a class of functions f, not necessarily bounded or with compact support, satisfying

$$d(f, \Theta_2(H_n, L_n)) \leq k\varepsilon_n, \qquad n \geq 1,$$

and the property of belonging to such a class amounts to restrictions on the tails and growth at the mode of the functions in the class.

REMARK 5. A reasonable question is why did we choose $S = \mathrm{Log}(HL + 1)$ rather than $S = \mathrm{Log}(HL)$ which seems more natural. The main reason is technical: Our bounds are intended to be used for large values of HL. In this case it does not make a big difference but our proofs for upper bounds naturally lead to $HL + 1$, rather than HL, and this choice makes results simpler. A serious drawback of this choice is the restriction (2.4) on lower bounds which does not allow for small values of HL. This comes from Proposition 1. It would obviously be possible to get similar lower bounds with $S = \mathrm{Log}(HL)$ and the restrictions would be weakened and allow for smaller values of HL, but it does not seem that our proof for upper bounds could be extended to this case: hence our choice for S. Nevertheless, when HL is small, the densities involved are close to uniform and some asymptotic results of Groeneboom suggest that, in this case, the risk should be smaller. This leads to the idea that a good choice should be $S = \mathrm{Log}(HL)$. It would not change much for large HL but would improve the bounds in the case of small HL. We suspect that different proofs would be needed in this case and the upper bounds do not seem to be easy to get.

REMARK 6. Roughly speaking, the minimax risk for decreasing or unimodal densities behaves like $(S/n)^{1/3}$ and, although our results are intended to deal with finite n, it is interesting to compare them to asymptotic results for this problem. The only other theorems dealing with integrated risk and decreasing densities are from Groeneboom (1985). These concern the Grenander estimator [see Grenander (1980)]. This estimator which is given by the slope of the smallest concave majorant of the empirical c.d.f. is especially designed for estimating decreasing densities, but unfortunately, we do not know anything about its risk for finite n; the only result concerns its asymptotic risk at a given point. Let us denote this estimator by $\hat{F}_n$. Groeneboom proved that for smooth f it satisfies

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

and in a private communication he explained that if it is restricted to $\Theta_1(H, L)$, the largest values of the functional $\int |f(x) f'(x)|^{1/3}\, dx$ occur when the density is hyperbolic on $[0; L]$:

$$f(x) = \frac{HL}{tx + L} \quad \text{with } HL = \frac{t}{\mathrm{Log}(t + 1)}.$$

The function $t/\mathrm{Log}(t + 1)$ being increasing, $t \geq HL$ as soon as $HL \geq e - 1$ and it can be checked that

$$\int_0^L |f(x)f'(x)|^{1/3}\,dx = [\mathrm{Log}(t+1)]^{1/3} \geq [\mathrm{Log}(HL+1)]^{1/3},$$

which implies that if $HL \geq e - 1$,

$$\lim_n n^{1/3} R_n(\hat{R}_n, f) \geq 0.65 S^{1/3}.$$

This appears to be in accordance with our preceding results. In fact, the situation is slightly more complicated because for $HL = 10$, $\mathrm{Log}(t+1) \simeq \frac{3}{2}\mathrm{Log}(HL+1)$ while for $HL = 1 + x$, with small x, $t \simeq 2x$ and $\mathrm{Log}(t+1) \simeq 2\,\mathrm{Log}(HL)$. This suggests that our difficulties in the choice of S, as explained in Remark 5, are due to the very nature of our problem.

5. Conclusion. The minimax risk for the classes considered in this paper is of order $[\mathrm{Log}(HL+1)/n]^{1/3}$, at least when HL is not close to one. But a few interesting questions still remain open:

— Is it possible to replace $\mathrm{Log}(HL+1)$ by $\mathrm{Log}(HL)$ when HL is close to one?

— How does one improve the ratio between the upper and lower bounds for the minimax risk, which is roughly 40 when S/n is small? This is clearly too big and some closer approximation to the minimax risk should be desirable, at least when n is not too small.

— The most important point is, How to find a reasonably practical estimator which achieves the bounds given by (3.7) and, if possible, an adaptive version, in order to cope with the fact that in practice S is unknown and has to be estimated. For the case of decreasing densities, we shall solve this problem in a forthcoming paper; but for unimodal densities we do not know of any practical solution apart from the usual histograms or kernel estimates as described in Bretagnolle and Huber (1979), which will not give the proper bounds (for large S) simply because they do not make use of the monotonicity properties of the densities and treat them as ordinary functions with bounded variation as mentioned in the Introduction.

APPENDIX

LEMMA A.1. *Suppose J and K are two nonintersecting sets having the same measure under μ: $\mu(J) = \mu(K) = l/2$ and f and g are positive functions satisfying*

$$g(x) = M(1+\varepsilon), \quad \text{for } x \in J, \qquad g(x) = M, \quad \text{for } x \in K,$$
$$f(x) = M(1+\varepsilon/2), \quad \text{for } x \in I = J \cup K.$$

Then

(i) $$\int_I |f(x) - g(x)|\,d\mu(x) = lM\varepsilon/2,$$

(ii) $$\frac{1}{2}\int_I \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 d\mu(x) \leq \frac{lM\varepsilon^2}{16(2+\varepsilon)}\left(1 + \frac{5\varepsilon^2}{64(1+\varepsilon)}\right).$$

PROOF. (i) is immediate. For (ii), we put $\delta = \varepsilon/(2+\varepsilon)$. Then

$$\frac{1}{2}\int_I \left(\sqrt{f} - \sqrt{g}\right)^2 d\mu = \frac{lM(1+\varepsilon/2)}{4}\left[(1-\sqrt{1+\delta})^2 + (1-\sqrt{1-\delta})^2\right]$$

$$= lM\left(1+\frac{1}{2}\varepsilon\right)\left[1 - \frac{1}{2}\left(\sqrt{1+\delta} + \sqrt{1-\delta}\right)\right].$$

Expanding the last bracketed term in power series we find

$$\frac{1}{2}\int_I \left(\sqrt{f} - \sqrt{g}\right)^2 d\mu = \frac{lM\varepsilon}{2}\sum_{i\geq 1}\delta^{2i-1}\frac{1\cdot 3\cdot \cdots \cdot (4i-3)}{2\cdot 4\cdot \cdots \cdot 4i}$$

$$= \frac{lM\varepsilon}{16}\frac{\varepsilon}{2+\varepsilon}\left[1 + \sum_{i\geq 1}\left(\frac{\varepsilon}{2+\varepsilon}\right)^{2i}\frac{1\cdot 3\cdot \cdots \cdot (4i+1)}{6\cdot 8\cdot \cdots \cdot (4i+4)}\right]$$

$$\leq \frac{lM\varepsilon^2}{16(2+\varepsilon)}\left[1 + \frac{15}{48}\frac{\varepsilon^2}{(2+\varepsilon)^2}\sum_{i\geq 0}\left(\frac{\varepsilon}{2+\varepsilon}\right)^{2i}\right]$$

$$= \frac{lM\varepsilon^2}{16(2+\varepsilon)}\left(1 + \frac{5\varepsilon^2}{64(1+\varepsilon)}\right).$$

□

PROOF OF PROPOSITION 1. It is easily checked by differentiation that

$$\text{(A.1)} \qquad \operatorname{Log}(1+x) < \frac{x(1+x/2)}{1+x}, \quad \text{for } x > 0.$$

If A is defined by (2.1), we can also check, using polynomial bounds for logarithms, that

$$\text{(A.2)} \qquad A \operatorname{Log}(1+\varepsilon) \leq \varepsilon, \quad \text{for } \varepsilon > 0.$$

From the definitions and (A.2) we deduce that

$$\lambda L = \frac{1+\varepsilon}{p\varepsilon(1+\varepsilon/2)}\left[(1+\varepsilon)^p - 1\right], \qquad p < \frac{SA}{\varepsilon} + 1 \leq \frac{S}{\operatorname{Log}(1+\varepsilon)} + 1.$$

Using (A.1) and the fact that $((1+\varepsilon)^x - 1)/x$ is increasing with x for $x > 0$ we get

$$\lambda L < \frac{(1+\varepsilon)\operatorname{Log}(1+\varepsilon)}{[S + \operatorname{Log}(1+\varepsilon)]\varepsilon(1+\varepsilon/2)}\left[(1+\varepsilon)e^S - 1\right]$$

$$< [S + \operatorname{Log}(1+\varepsilon)]^{-1}\left[(1+\varepsilon)e^S - 1\right].$$

To complete the proof we just have to check that

$$[S + \operatorname{Log}(1+\varepsilon)]^{-1}\left[(1+\varepsilon)e^S - 1\right] \leq HL = e^S - 1,$$

or, equivalently, if $S = 1 + a\varepsilon$, $a \geq 0$,

$$\varepsilon\left[a + (1 - a)e^{1+a\varepsilon}\right] \leq \left(e^{1+a\varepsilon} - 1\right)\operatorname{Log}(1 + \varepsilon).$$

Since $\operatorname{Log}(1 + \varepsilon) \geq \varepsilon - \varepsilon^2/2$ it is enough to check that

$$\left(e^{1+a\varepsilon} - 1\right)(\varepsilon/2 - a) + 1 \leq 0.$$

This is a decreasing function of a if $\varepsilon \leq 1.12$ and it is then enough to prove that

$$\left(e^{1+3\varepsilon/4} - 1\right)(2\varepsilon - 3) + 4 \leq 0, \quad \text{for } 0 \leq \varepsilon \leq 1.12.$$

We get a convex function of ε, which is negative for $\varepsilon = 0$, and then has a unique positive root. It suffices to check that this root is larger than 1.12. □

PROOF OF PROPOSITION 2. Our choice of p and ε implies that

$$n\beta \leq \frac{n\varepsilon^3}{8jSA(2 + \varepsilon)^2}\left(1 + \frac{5\varepsilon^2}{64}\right) = \frac{2}{25}\left(1 - \frac{3\varepsilon}{4} + \frac{7\varepsilon^2}{12}\right)^2 = B.$$

Also, some easy but tedious computations can be made to check that if $B \leq 1/5$ and $n \geq 3$,

$$1 - (1 - \beta)^{2n} \leq 2n\beta\left[1 - \tfrac{3}{8}n\beta\right]^2 \leq 2B\left[1 - \tfrac{3}{8}B\right]^2,$$

the requirement on B being satisfied for $\varepsilon \leq 7/4$. We finally get from (2.2)

$$\begin{aligned}
R_j(n) &\geq \frac{\varepsilon}{4(1 + \varepsilon/2)}\left[1 - \left(1 - \frac{3}{8}B\right)(2B)^{1/2}\right] \\
&= \frac{\varepsilon}{4(1 + \varepsilon/2)}\left[1 - \left(1 - \frac{3}{100}\left(1 - \frac{3\varepsilon}{4} + \frac{7\varepsilon^2}{12}\right)^2\right)\frac{2}{5}\left(1 - \frac{3\varepsilon}{4} + \frac{7\varepsilon^2}{12}\right)\right] \\
&= \frac{\varepsilon}{4(1 + \varepsilon/2)}\left[\frac{3}{5} + \frac{3\varepsilon}{10} - \frac{7\varepsilon^2}{30} + \frac{3}{250}\left(1 - \frac{3\varepsilon}{4} + \frac{7\varepsilon^2}{12}\right)^3\right] \\
&= \frac{3}{20}\varepsilon + \frac{\varepsilon}{40(1 + \varepsilon/2)}\left[-\frac{7\varepsilon^2}{3} + \frac{3}{25}\left(1 - \frac{3\varepsilon}{4} + \frac{7\varepsilon^2}{12}\right)^3\right].
\end{aligned}$$

The bracketed term is easily seen to have only one positive root $\varepsilon_0 < 7/4$ and $\varepsilon_0 \simeq 0.187$. Since $1 - 3\varepsilon/4 + 7\varepsilon^2/12$ has the minimum value $85/112$, we get

$$\begin{aligned}
R_j(n) &\geq \frac{3}{20}\varepsilon + \frac{\varepsilon}{40(1 + \varepsilon_0/2)}\left[-\frac{7\varepsilon^2}{3} + \frac{3}{25}\left(1 - \frac{3\varepsilon}{4} + \frac{7\varepsilon^2}{12}\right)^3\right] \\
&\geq \frac{3}{20}\varepsilon + \frac{\varepsilon}{40(1 + \varepsilon_0/2)}\left[-\frac{7\varepsilon^2}{3} + \frac{3}{25}\left(\frac{85}{112}\right)^3\right].
\end{aligned}$$

The conclusion follows if we replace ε and ε_0 by their values. □

PROOF OF PROPOSITION 3. In the preceding constructions we can always choose I_0 in such a way that a mode of f belongs to I_0. Then, if $f(x_i) = f_i$, l_i is the length of I_i and $f_{-1} \le f_1$ (the other case being symmetrical), we have

$$\sum_{i=-p_2}^{-1} l_i f_{i-1} + \sum_{i=1}^{p_1} l_i f_{i+1} + l_0 f_{-1} \le 1. \tag{A.3}$$

Define $\bar{f}_i = l_i^{-1} \int_{I_i} f(x)\,dx$. Since $y_0 = 0$ and $y_p = H$, for any i we can find some $j \ge 1$ with

$$\bar{f}_i = \lambda y_{j-1} + (1-\lambda) y_j, \qquad 0 \le \lambda \le 1.$$

We shall take g_i to be y_{j-1} if $\lambda > \frac{1}{2}$ or y_j if $\lambda \le \frac{1}{2}$. This defines an element g of $F(\varepsilon)$ (g being unimodal just as f) which will be used as an approximation of f. Let us check (3.4). Since

$$\bar{f}_i - y_{j-1} = (1-\lambda)\varepsilon L^{-1}(1+\varepsilon)^{j-1},$$

$$\bar{f}_i = L^{-1}\left[(1+\varepsilon)^{j-1}(1+(1-\lambda)\varepsilon) - 1\right]$$

and for $\lambda > \frac{1}{2}$

$$\bar{f}_i - g_i = \bar{f}_i - y_{j-1} = \frac{(L\bar{f}_i + 1)\varepsilon(1-\lambda)}{L(1+(1-\lambda)\varepsilon)} \le \frac{\varepsilon}{2+\varepsilon}\left(\bar{f}_i + L^{-1}\right).$$

An analogous bound holds for $\lambda \le \frac{1}{2}$ which implies

$$\int_{I_i} |\bar{f}_i - g(x)|\,dx \le \frac{l_i \varepsilon}{2+\varepsilon}\left(\bar{f}_i + L^{-1}\right).$$

Also using some classical bound on the bias we get

$$2\int_{I_i} |\bar{f}_i - f(x)|\,dx \le \begin{cases} l_i(f_i - f_{i+1}), & i \ge 1, \\ l_0(H - f_{-1}), & i = 0, \\ l_i(f_i - f_{i-1}), & i \le -1. \end{cases}$$

Putting everything together we find

$$\int |f(x) - g(x)|\,dx \le \frac{\varepsilon}{2+\varepsilon}\left[\sum_i l_i \bar{f}_i + L^{-1}\sum_i l_i\right] + \frac{1}{2}\left[\sum_{i=1}^{p_i} l_i(f_i - f_{i+1}) + \sum_{i=-p_2}^{-1} l_i(f_i - f_{i-1}) + Hl_0 - f_{-1}l_0\right],$$

which implies using $\sum_i \bar{f_i} l_i = 1$, (3.3) and then (A.3),

$$\int |f(x) - g(x)|\,dx \le \frac{2\varepsilon}{2+\varepsilon} + \frac{1}{2}\Bigg[(1+\varepsilon)l_0 f_1 + \varepsilon \sum_{i=1}^{p_1-1} l_i f_{i+1} + (1+\varepsilon) l_0 f_{-1} + \varepsilon \sum_{i=-p_2+1}^{-1} l_i f_{i-1} + H l_0 - f_{-1} l_0\Bigg]$$

$$\le \frac{2\varepsilon}{2+\varepsilon} + \frac{1}{2}\big[(1+\varepsilon)l_0 f_1 + (1+\varepsilon)l_0 f_{-1} + \varepsilon(1 - l_0 f_{-1}) + H l_0 - f_{-1} l_0\big]$$

$$\le \frac{2\varepsilon}{2+\varepsilon} + \frac{l_0}{2}\big[H + (1+\varepsilon) f_1\big] + \frac{\varepsilon}{2}$$

$$\le \frac{2\varepsilon}{2+\varepsilon} + \frac{\varepsilon(2+\varepsilon)}{2} + \frac{\varepsilon}{2} = \frac{\varepsilon}{2}\left[5 + \frac{\varepsilon^2}{2+\varepsilon}\right].$$

Equation (3.4) follows from the assumption that $\varepsilon \le 1/10$.

To get (3.5) we start with a simple combinatorial problem: Given $q+1$ increasing numbers $y_0; \dots; y_q$, what are the different ways of choosing p of them $x_1; \dots; x_p$, possibly with repetition, but in such a way that the resulting sequence is nondecreasing? We can easily do it by choosing p numbers among $\{0; \dots; q\}$, ordering them to get $k_1 \le k_2 \le \cdots \le k_p$ and fixing $x_i = y_{k_i}$. The number of choices of the k_i's is the number of combinations with repetitions of p numbers among $q+1$ which is known to be $\binom{p+q}{p}$ [Riordan (1958)]. For a given I_0 and $g_0 = y_j$ the number of choices for g is then $\binom{p_1+j}{p_1}\binom{p_2+j}{p_2}$ which implies, taking all possible values of j and I_0,

$$\operatorname{card} F(\varepsilon) \le k \sup_{p_1, p_2} \sum_{j=1}^{p} \binom{p_1+j}{p_1}\binom{p_2+j}{p_2},$$

the supremum being over all relevant values of p_1, p_2. Now, since $x_{-p_2} > -L/2$ and $x_{p_1} < L/2$, we find

$$L > l_0\left[-1 + \sum_{i=0}^{p_1-1}(1+\varepsilon)^i + \sum_{i=0}^{p_2-1}(1+\varepsilon)^i\right]$$

$$= H^{-1}\big[(1+\varepsilon)^{p_1} + (1+\varepsilon)^{p_2} - 2 - \varepsilon\big]$$

(this also being valid if p_1 or p_2 is 0), or, equivalently,

$$LH = (1+\varepsilon)^p - 1 > (1+\varepsilon)^{p_1} + (1+\varepsilon)^{p_2} - 1 - (1+\varepsilon)$$

and this implies that $p_1 + p_2 \le 2p - 2$ as soon as $p \ge 2$. By assumption $p \ge 13.5S \ge 13.5 \operatorname{Log} 2 = 9.36$, then $p \ge 10$,

$$\binom{p_1+j}{p_1}\binom{p_2+j}{p_2} \le \binom{p-1+j}{p-1}^2 = \frac{1}{4}\binom{2p}{p}^2\left[\prod_{i=j}^{p-1}\frac{i+1}{p+i}\right]^2,$$

if we define $\prod_{i=p}^{p-1}$ to be 1. This implies

$$\sum_{j=1}^{p}\binom{p_1+j}{p_1}\binom{p_2+j}{p_2} \le \frac{1}{4}\frac{((2p)!)^2}{(p!)^4}\left[1+\frac{p}{2p-1}+\frac{p(p-1)}{(2p-1)(2p-2)}+\cdots\right]^2.$$

Using Stirling's expansion we find with $p \ge 10$,

$$\operatorname{card} F(\varepsilon) \le \frac{k}{4}\frac{4\pi p(2p)^{4p}}{(2\pi p)^2 p^{4p}}\left[1+\frac{2p}{2p-1}\right]^2 < 0.336\frac{k}{p}2^{4p}.$$

Also using $e^x - 1 > x$ and $p \ge 13.5S$

$$k < \frac{LH}{\varepsilon} + 1 = \frac{e^S - 1}{e^{S/p} - 1} + 1 < (e^S - 1)\frac{p}{S} + 1 \le \frac{p}{S}\exp(2p/27),$$

which achieves the proof. □

LEMMA A.2. *Suppose that A, B, S are positive constants and n, p positive integers satisfying*

$$p - 1 < (2BS/A)^{2/3}n^{1/3} + 2S/3 \le p.$$

Then

$$A\left(\frac{p-1}{n}\right)^{1/2} + B\left[\exp\left(\frac{S}{p}\right) - 1\right] < \frac{3}{2}\left(\frac{2SA^2B}{n}\right)^{1/3} + \frac{1}{8}\left(\frac{2SA^2}{n}\right)^{2/3}B^{-1/3}.$$

PROOF. Consider the function $f(x) = C(x/n)^{1/2} + \exp(S/x) - 1$ as a function of x, with $C = A/B$. If $t = (2SC^2/n)^{1/3}$ and $x_0 = (2S/C)^{2/3}n^{1/3} + 2S/3$, then

$$f(x_0) = t(1 + t/3)^{1/2} + \exp\left[t/2(1 + t/3)^{-1}\right] - 1.$$

Some expansions and numerical computations prove that for $t > 0$,

$$t(1 + t/3)^{1/2} + \exp\left[t/2(1 + t/3)^{-1}\right] - 1 < 3t/2 + t^2/8,$$

which gives a bound on $f(x_0)$. The choice of p then leads to the conclusion. □

Acknowledgments. I would like to thank Piet Groeneboom and Persi Diaconis for very useful discussions and a number of stylistic corrections. I would also like to thank an anonymous referee for improving my English grammar.

REFERENCES

ASSOUAD, P. (1983). Deux remarques sur l'estimation. *C. R. Acad. Sci. Paris Sér. I* **296** 1021–1024.

BIRGÉ, L. (1983). Approximation dans les espaces méutriques et théorie de l'estimation. *Z. Wahrsch. verw. Gebiete* **65** 181–237.

BIRGÉ, L. (1984a). Sur un théorème de minimax et son application aux tests. *Probab. Math. Statist.* **3** 259–282.

BIRGÉ, L. (1984b). Stabilité et instabilité du risque minimax pour des variables indépendantes équidistribuées. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **20** 201–223.

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

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

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.

HUBER, P. J. (1965). A robust version of the probability ratio test. *Ann. Math. Statist.* **36** 1753–1758.

HUBER, P. J. and STRASSEN, V. (1973). Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Statist.* **1** 251–263.

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.

LE CAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53.

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

RIORDAN, J. (1958). *An Introduction to Combinatorial Analysis*. Wiley, New York.

WEGMAN, E. J. (1970). Maximum likelihood estimation of a unimodal density. I and II. *Ann. Math. Statist.* **41** 457–471, 2169–2174.

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