The Annals of Statistics
1982, Vol. 10, No. 3, 811–824

CONSISTENCY OF TWO NONPARAMETRIC MAXIMUM PENALIZED LIKELIHOOD ESTIMATORS OF THE PROBABILITY DENSITY FUNCTION[1]

BY V. K. KLONIAS

The Johns Hopkins University

We study the consistency properties of a nonparametric estimator f_n of a density function f on the real line, which is known as the "first MPLE of Good and Gaskins," and which is obtained by maximizing the likelihood functional multiplied by the roughness penality $\exp\{-\alpha \int (f'/f)^2 f\}$ with $\alpha > 0$. Under modest assumptions on the density function f, and letting $\alpha = \alpha_n \to \infty$ and $\alpha_n/n \to 0$ a.s. as $n \to \infty$ we demonstrate the a.s. convergence of f_n to f, with rates, in the Hellinger, L_1, L_2, $\sup_F$ and Sobolev norms, as well as in integrated mean absolute deviation. Finally, the corresponding estimator for f supported on the half-line, is derived and the computational feasibility as well as the consistency properties of the estimator are indicated.

1. Introduction. Let $X_1, X_2, \cdots, X_n$ be independent observations from a distribution function F with density function f, assumed to have finite Fisher information and such that $\int \{F(1-F)\}^\rho < +\infty$ for some $\rho > 1/2$, and denote the likelihood functional by $\ell(f) \equiv \prod_{i=1}^n f(X_i)$. If the maximization of $\ell(f)$ is carried out over all possible density functions f, then the supremum (infinity) is achieved at an average of Dirac delta functions, which is not an acceptable estimator of the density function. To avoid this "Dirac catastrophe," Good and Gaskins (1971) introduced the maximum penalized likelihood method of density estimation (MPLE), which consists of maximizing $\log \ell(f) - \Phi(f)$, where $\Phi(\cdot)$ is a certain "flamboyance functional." The proposed functionals $\Phi(f)$ were functions of f and its derivatives, e.g. $\Phi(f) = \int (f'/f)^2 f$, the Fisher information functional. The problem can be stated as follows:

PROBLEM P1. $\max\{\log \ell(f) - \alpha\Phi(f)\}$ subject to $\int f = 1, f \geq 0$, where $\alpha > 0$.

However, to avoid having to deal with a nonnegativity constraint on f, Good and Gaskins suggested an alternative formulation of the problem in terms of the square root of the density function, $v = f^{1/2}$, i.e.:

PROBLEM P2. $\max\{2 \log \ell(v) - \alpha\bar{\Phi}(v)\}$ subject to $\int v^2 = 1$, $v(X_i) \geq 0$, $i = 1, \cdots, n$, where Φ is a functional of the root-density function.

The solution of this problem is then squared to obtain a density estimate. De Montricher, Tapia and Thompson (1975) showed that Problems P1 and P2 are not always equivalent. This is the case, for example, if $\Phi(f) = \bar{\Phi}(v) = \int (v'')^2$; but for the problem of interest here, namely the case $\Phi(f) = \bar{\Phi}(v) = \int (v')^2$, they showed the two problems to be equivalent and established the existence and uniqueness of the solution. The solution,

Received April 1981; revised February 1982.

[1] The material of this paper is adapted primarily from Chapter 3 of the author's dissertation (1980) at the University of Rochester. The research on the material of Section 6 was supported in part by the Office of Naval Research Contract No. N00014-79-C-0801.

AMS 1970 *subject classification.* Primary 62G05, 62G10, 60F15, 41A25; secondary 60F25, 40A30, 62E10.

Key words and phrases. Nonparametric density estimation, maximum penalized likelihood method, strong consistency, rates of convergence, likelihood functional, Fisher information functional, nonparametric density estimator with positive support, exponential spline function.

denoted throughout this paper for $f_n = u_n^2$, is an exponential spline with knots at the sample points (see equation (2.2)) and is known as the "first estimator of Good and Gaskins."

A fundamental question is whether f_n is *consistent* for the estimation of f, i.e., whether $\|f_n - f\| \to 0$, $n \to \infty$, in some stochastic sense for some suitable norm $\|\cdot\|$. This question was first investigated by Good and Gaskins (1971) and they presented a heuristic proof of the pointwise convergence in probability of the cumulative distribution function corresponding to the solution of Problem P1, for a general penalty functional Φ (see Good and Gaskins, 1980). An estimator related to the MPLE and corresponding to a discretized version of Problem P1 (DMPLE) on a finite interval (a, b), was investigated by Scott, Tapia and Thompson (1980) and the DMPLE corresponding to a discretized version of a penalty functional of the form $\Phi(f) = \int_a^b \{f'(x)\}^2\, dx$ was shown to converge pointwise w.p.1.; see also Tapia and Thompson (1978). For some other types of MPLE's see de Montricher (1981) and Silverman (1982). However, the question of the consistency of f_n, the solution to the infinite dimensional Problems P1 and P2, for a penalty proportional to the Fisher information (or any other of the original MPLE problems suggested by Good and Gaskins, 1971) has not been treated, except for the investigation by Good and Gaskins mentioned above. In the present paper the consistency of f_n is established in the senses of L_1, L_2, $\sup_{\mathbb{R}}$ and Sobolev norms, and corresponding rates of convergence are provided.

Section 2 formulates the MPLE problem, with penalty proportional to the Fisher information, in the context of the general MPLE problem over a Sobolev space of order $m \geq 1$, following de Montricher et al. (1975), and discusses the consistency proofs that follow. In Section 3 we present a Glivenko-Cantelli type theorem for the rate of convergence of the empirical distribution with respect to a particular metric, a modification of a similar result of Wellner (1977). This result enables us (in Section 5) to derive the a.s. convergence of the log-likelihood functional evaluated at f_n, which in turn implies the a.s. convergence of f_n in the Hellinger distance. In Section 4, it is indicated that the a.s. convergence of f_n—the MPLE of the density function—to the true density f in the L_1, L_2, $\sup_{\mathbb{R}}$ and Sobolev norms (see Proposition 4.1), can be deduced, through the set of inequalities in Lemma 4.1, from the a.s. L_2, $\sup_{\mathbb{R}}$ and Sobolev norm convergence of u_n—the MPLE of the root density function—which are established in Section 5. The assumptions on f stated in the beginning of this section are enough to secure the convergence of f_n in L_1, L_2 and $\sup_{\mathbb{R}}$ norms. The convergence of f_n in the Sobolev norm is proved under the additional assumption of a square integrable second derivative of $f^{1/2}$. Also in Sections 4 and 5, as corollaries to the main propositions, the a.s. convergences of several functionals of f_n are derived and f_n is shown to also converge in integrated mean absolute deviation. Rates of convergence are provided for all types of convergences mentioned above.

In Section 6 we derive the "first MPLE of Good and Gaskins" in the case that f has support only the half line, essentially by reflecting f around zero and invoking the results for f having support $\mathbb{R}$. The numerical evaluation and the consistency properties of the estimator are deduced similarly.

2. Preliminaries. The natural setting for the solution of the MPLE problem in the form P2, was shown by de Montricher, Tapia and Thompson (1975) to be provided by the Sobolev subspaces of $L_2(\mathbb{R})$, defined next.

Let $\mathbb{R}$ denote the real line and $L_2 \equiv L_2(\mathbb{R})$ the space of all Lebesgue measurable square integrable functions $u: \mathbb{R} \to \mathbb{R}$. It is well known that $(L_2, \langle \cdot, \cdot \rangle_2)$ is a Hilbert space for the inner product $\langle u, v \rangle_2 \equiv \int_{\mathbb{R}} uv$ and the induced norm $\|u\|_2 \equiv (\int_{\mathbb{R}} u^2)^{1/2}$. By a Sobolev space of order m on $\mathbb{R}$, denoted by $H^m \equiv H^m(\mathbb{R})$, where m is positive integer, is meant the space of all functions. $u: \mathbb{R} \to \mathbb{R}$ such that $u, u^{(1)}, \cdots, u^{(m)} \in L_2(\mathbb{R})$, where $u^{(s)}$ denotes the distributional derivative of order s of u (all derivatives in this paper are distributional derivatives and the first two are also denoted by u', u''). Endowed with the inner product

$$\langle u, v \rangle_{H^m} \equiv \textstyle\sum_{s=0}^m w_s \langle u^{(s)}, v^{(s)} \rangle_2, \qquad w_s \geq 0 \quad \text{and} \quad w_0, w_n > 0,$$

and the induced norm

$$\| u \|_{H^m} \equiv (\langle u, u \rangle_{H^m})^{1/2},$$

H^m is a *reproducing kernel Hilbert space* (RKHS), i.e. there exists a unique real valued function $k(x, y)$ defined on $\mathbb{R} \times \mathbb{R}$, called the *kernel*, such that

$$k_y(\cdot) \equiv k(\cdot, y) \in H^m \quad \text{for all} \quad y \in \mathbb{R},$$

and

$$\langle u, k_y \rangle_{H^m} = u(y) \quad \text{for all} \quad u \in H^m.$$

For example, for $m = 1$, the kernel of $H^1(\mathbb{R})$ (a space of special interest for this paper) is given by

$$k(x, y) = (4w_0 w_1)^{-1/2} \exp\{-(w_0/w_1)^{1/2} |x - y|\}. \tag{2.1}$$

For a treatment of the RKHS and Sobolev space the reader is referred to Aronszajn (1950), Parzen (1967), Gel'fand and Shilov (1964) and Yosida (1974).

De Montricher et al. (1975) give the general form of the solution to the optimization Problem P2 for penalty terms of the form $\Phi(v) = \| v \|_H^2$, $v \in H^m$. The case of interest here corresponds to $m = 1$ and can be stated as follows:

PROBLEM P3. $\max\{2 \log \ell(v) - \alpha \| v' \|_2^2\}$, $v \in H^1$, subject to $\| v \|_2 = 1$, $v(X_i) \geq 0$, $i = 1, \cdots, n$, with $\alpha > 0$.

Notice that $4 \| v' \|_2^2 = \int (f'/f)^2 f$, the Fisher information functional. The solution—the MPLE of the square root of the density function—is an exponential spline with knots at the sample points, given implicitly by

$$u_n(x) = (4\lambda_n \alpha)^{-1/2} \sum_{i=1}^n u_n(X_i)^{-1} \exp\{-(\lambda_n/\alpha)^{1/2} |x - X_i|\}, \qquad x \in \mathbb{R}, \tag{2.2}$$

where λ_n is the Lagrange multiplier associated with the constraint on the L_2-norm of v and satisfies (de Montricher et al., 1975, Proposition 3.3)

$$n/2 \leq \lambda_n < n. \tag{2.3}$$

The associated MPLE of the density function is then given by

$$f_n = u_n^2. \tag{2.4}$$

Notice that the values of the estimator at the sample points $u_n(X_i)$, $i = 1, \cdots, n$, and λ_n are the solutions to a system of $n + 1$ nonlinear equations, consisting of the condition $\int u_n^2 = 1$ and the n relations obtained from (2.2) after setting $x = X_j$, $j = 1, \cdots, n$; see also Ghorai and Rubin (1979). Hence both random variables λ_n and each $u_n(X_i)$ are complicated functions of the sample, so that it does not seem promising to use equation (2.2) directly for the study of the statistical properties of u_n and f_n. Instead we investigate the behavior of f_n through $\ell(f_n)$, the likelihood functional evaluated at f_n. Then, through inequality (5.4), it is shown that the convergence of the log-likelihood functional evaluated at f_n implies the convergence of f_n with respect to the Hellinger distance, which is at the core of the other modes of convergence of f_n and u_n (see Section 5).

It can be shown (see inequality (5.2)) that the proof of the convergence of the log-likelihood functional is reduced to that of the convergence of the empirical distribution function with respect to a particular metric and with some rate, which we present in the following section.

3. Rates of convergence for the empirical process w.r.t. a particular $\| \cdot / q \|$-metric. We modify a theorem of Wellner (1975) on almost sure "nearly linear" bounds for the empirical d.f. to provide a rate for its convergence w.r.t. a metric defined below.

The rate is utilized in Section 5 for the consistency results (see Lemma 5.2 and Proposition 5.1(ii)).

The lemma below is similar to the lemma on page 69 of Billingsley (1968), and its proof is omitted.

LEMMA 3.1. *Suppose* $Z_1, Z_2, \cdots$ *are independent random elements in a metric space, with* $\|X\|$ *denoting the distance of* X *from the zero element, and let* $S_n \equiv Z_1 + \cdots + Z_n$. *For all* $\lambda > c$ *and positive integers* $n \ge m$ *for which* $0 < \delta_{mn} \equiv \min_{m \le i \le n} P(\|S_n - S_i\| < c)$, *we have*

$$P(\max_{m \le i \le n} \|S_i\| \ge \lambda) \le P(\|S_n\| > \lambda - c)/\delta_{mn}. \tag{3.1}$$

The following theorem gives a rate of convergence for the uniform empirical distribution using a particular $\|\cdot/q\|$-metric, defined by

$$\|(f-g)/q\| \equiv \sup_{t \in (0,1)} |f(t) - g(t)|/q(t),$$

with $q(t) \equiv [t(1-t)]^\rho$, $\rho \in (½, 1)$. It is an extension of Theorem 4 on page 29 of Wellner (1975) (where the "$s = 0$" case appears).

THEOREM 3.1. *Let* $\xi_1, \xi_2, \cdots$ *be* i.i.d. $U(0, 1)$ r.v.'s *and* $\Gamma_n(\cdot)$ *the empirical distribution function of* $\xi_1, \xi_2, \cdots, \xi_n$. *Then, for* s, ρ *such that* $s > 0$, $½ < \rho < 1$ *and* $s + \rho < 1$, *we have*

$$n^s \|(\Gamma_n - I)/[I(1-I)]^\rho\| \to 0 \quad \text{a.s.} \quad \textit{as} \quad n \to \infty, \tag{3.2}$$

where I *denotes the identity function.*

PROOF. Let $A_n \equiv \{n^s \sup_{0<t\le\theta} |\Gamma_n(t) - t|/t^\rho \ge \varepsilon\}$ for given $0 < \theta < 1$ and $\varepsilon > 0$. Also let $Y_i \equiv h_\theta(\xi_i) - \int_{(0,\xi_i)} (1-I)^{-1} h_\theta \, dI$, where $h_\theta(s) = s^{-\rho} I_{[0,\theta]}(s)$; $Y_1, Y_2, \cdots$ are i.i.d. with mean zero and $E|Y_1|^\tau < +\infty$, for any $\tau < 1/\rho$.

Using an inequality from Wellner (1977, page 484), we have

$$P(A_n) \le 2^{\tau+1} \varepsilon^{-\tau} E|Y_1|^\tau n^{-(\tau-1-s\tau)} \tag{3.3}$$

for any τ such that $1 < \tau < 1/\rho < 2$. If $s < 1 - \tau^{-1}$, this implies

$$\textstyle\sum_{n=1}^\infty P(A_{[n^\eta]}) < +\infty \quad \text{whenever} \quad \eta > 1/(\tau - 1 - s\tau) > \rho/(1 - s - \rho). \tag{3.4}$$

We hereafter assume $\eta > (\tau - 1 - s\tau)^{-1}$ and also $\eta \ge 1$.

Next we handle the probabilities for k such that $n^\eta < k \le (n+1)^\eta$. Write $S_k \equiv \sum_{i=1}^k Q_i$ where $Q_i(t) \equiv I_{[0,t]}(\xi_i) - t$, and define

$$\|z/q\|_0^\theta \equiv \sup_{0<t\le\theta} |z(t)|/q(t) \quad \text{with} \quad q(t) \equiv t^\rho.$$

Then

$$\max_k \|(\Gamma_k - I)/q\|_0^\theta = \max_k \|S_k/q\|_0^\theta/k < n^{-\eta} \max_k \|S_k/q\|_0^\theta.$$

Hence

$$\begin{aligned} &P(\max_k n^s \|(\Gamma_k - I)/q\|_0^\theta > 2\lambda) \\ &\qquad \le P(\max_k \|S_k/q\|_0^\theta \ge 2\lambda n^{\eta-s}) \le P(\|S_{[(n+1)^\eta]}/q\|_0^\theta \ge \lambda n^{\eta-s})/c_{\lambda,n}, \end{aligned}$$

by Lemma 3.1 with

$$c_{\lambda,n} \equiv \min_k P(\|(S_{[(n+1)^\eta]} - S_k)/q\|_0^\theta < \lambda n^{\eta-s}),$$

so that

$$\begin{aligned} P(\max_k n^s \|(\Gamma_k - I)/q\|_0^\theta \ge 2\lambda) \le P([(n+1)^\eta]^s \|(\Gamma_{[(n+1)^\eta]} - I)/q\|_0^\theta \\ \ge \lambda n^{\eta-s}/([(n+1)^\eta]^{1-s})/c_{\lambda,n} \end{aligned}$$

$$\leq P([(n+1)^\eta]^s \|(\Gamma_{[(n+1)^\eta]} - I)/q\|_0^\theta \geq \lambda 2^{-\eta(1-s)})/c_{\lambda,n}$$

$$= P(A_{[(n+1)^\eta]})/c_{\lambda,n},$$

for $\lambda = 2^{\eta(1-s)}\varepsilon$, since

$$\lambda n^{\eta-s}/[(n+1)^\eta]^{(1-s)} \geq \lambda n^{s(\eta-1)} \Big/ \left(1+\frac{1}{n}\right)^{\eta(1-s)} > \lambda 2^{-\eta(1-s)}$$

since $n \geq 1$. Summing the probabilities in the above inequality we have

$$\sum_n P(\max_k n^s \|(\Gamma_k - I)/q\|_0^\theta \geq 2\lambda) \leq \sum_n P(A_{[(n+1)^\eta]})/c_{\lambda,n}. \tag{3.5}$$

We next use inequality (3.3) again, to control the denominator in the RHS of (3.5), i.e. to find for what values of η $c_{\lambda,n}$ stays bounded away from 0:

$$\begin{aligned}
1 - c_{\lambda,n} &= \max_k P(\|(\Gamma_{([(n+1)^\eta]-k)} - I)/q\|_0^\theta \geq \varepsilon 2^{\eta(1-s)} n^{\eta-s}/([(n+1)^\eta]-k)) \\
&\leq \max_k 2^{\tau+1-\eta\tau(1-s)} \varepsilon^{-\tau} E|Y_1|^\tau n^{-\tau(\eta-s)}([(n+1)^\eta]-k) \\
&= 2^{\tau+1-\eta\tau(1-s)} \varepsilon^{-\tau} E|Y_1|^\tau n^{-\tau(\eta-s)}([(n+1)^\eta]-[n^\eta]-1) \\
&< 2^{\tau+1-\eta\tau(1-s)} \varepsilon^{-\tau} E|Y_1|^\tau((n+1)^\eta - n^\eta)/n^{\tau(\eta-s)} \\
&< 2^{\tau+1-\eta\tau(1-s)} \varepsilon^{-\tau} E|Y_1|^\tau \eta(n+1)^{\eta-1}/n^{\tau(\eta-s)}.
\end{aligned}$$

For $c_{\lambda,n}$ to be bounded away from zero, this last term must be strictly less than one for large n, which means $\tau(\eta - s) > \eta - 1$ or equivalently $\eta > (s\tau - 1)/(\tau - 1)$ which is true since $\eta \geq 1 > (s\tau - 1)/(\tau - 1)$.

We then have that for $\eta > 1 \vee (1/(\tau - 1 - s\tau))$, (3.4) is true and the series in (3.5) converges and hence

$$n^s \sup_{0<t\leq\theta} |\Gamma_n(t) - t|/t^\rho \to 0 \quad \text{a.s.} \qquad \text{as} \quad n \to \infty.$$

To handle the interval $[\theta, 1]$ we use Chung's (1947) version of the Glivenko-Cantelli theorem. □

A similar result (with a different proof) appeared in Mason (1981) while this paper was being refereed.

4. Consistency of the density estimator. It is convenient to summarize here our notation and label our assumptions. Let $X_1, X_2, \cdots$ be i.i.d. random variables with distribution function F which is absolutely continuous with density function f, $v = f^{1/2}$, and make the following assumptions.

ASSUMPTION A1. $v \in H \equiv H^1(\mathbb{R})$, i.e. $\|v'\|_2 < +\infty$,

ASSUMPTION A2. $E_f|X|^\tau < +\infty$ for some $\tau > 1$.

The empirical distribution function of the sample $\mathbf{X}_n \equiv (X_1, \cdots, X_n)$ is denoted by F_n and the MPLE of v is denoted by $u_n \in B_n \equiv B(\mathbf{X}_n)$. We set $f_n \equiv u_n^2$ and $\mathscr{F}_n(x) \equiv \int_{(-\infty,x]} f_n(z)\, dz$, for the MPLE of f and the corresponding estimator of F, respectively. The ordered sample is denoted by $X_{(1)}, \cdots, X_{(n)}$.

It should be noted that Assumption A1 is slightly less restrictive than assuming v to be absolutely continuous on $\mathbb{R}$. The purpose of Assumption A2 is to guarantee that $\int \{F(1 - F)\}^\rho < +\infty$ for some $\rho \in ((2 \wedge \tau)^{-1}, 1)$ (see Remark 5.1), which may be a less intuitive assumption to make. Also Assumptions A1 and A2 guarantee that the entropy, as well as $\int |\log f|^\gamma$ dF for every $\gamma > 1$ (needed in Proposition 3.1(i) to secure the rates), are finite (see Klonias, 1981a).

Before the development of the consistency results, the following remark concerning the weight of the penalty term of Problem P3 is appropriate.

REMARK 4.1. According to our method of proving the consistency of the MPLE u_n (and then f_n), it is necessary to let the weight α of the penalty term depend on the sample size, i.e., $\alpha = \alpha_n$, in such a way that $\alpha_n \to \infty$ and $\alpha_n/n \to 0$ as $n \to \infty$ (see Proposition 5.1(ii) and e.g. Theorem 5.1, respectively). We then choose $\alpha_n = c_n n^t$ with $t \in (0, 1)$ and $c_n > 0$, such that $c_n \to c$—some positive constant—as $n \to \infty$. In practice we may want to let c_n depend on the sample, in which case we note that our proofs remain valid as long as $c_n \to c$ a.s.

We summarize the results on the consistency of f_n in the following theorem.

THEOREM 4.1. *Under* A1 *and* A2, *we have*

(i) $n^d \| f_n - f \|_p \to 0$ a.s. *as* $n \to \infty$, *where* $p = 1, 2$, *for* $d < ((1-t)/2) \wedge (t/4 - (2(2 \wedge \tau)^{-1} - 1)/4)$,

(ii) $n^d \| f_n - f \|_\infty \to 0$ a.s. *as* $n \to \infty$, *for* $d < ((1-t)/4) \wedge (t/2 - (2 \wedge \tau)^{-1}/2)$. *Also under the additional assumption* $\| v'' \|_2 < +\infty$, *and for* $d < ((1-t)/4) \wedge (3t/4 - (1 + (2 \wedge \tau)^{-1})/4)$, *we have*

(iii) $n^d \| f_n - f \|_H \to 0$ a.s. *as* $n \to \infty$.

With the aid of Lemma 4.1 below, the theorem is seen to follow from Theorems 5.1 through 5.3 in the next section. The lemma reduces the study of the consistency properties of f_n—the MPLE of the density function—to that of the convergence properties of u_n—the MPLE of the root density function—in L_2, $\sup_R$ and Sobolev norms and with the same rates, pending proper behavior of $\| u_n' \|_2$ (see Proposition 5.2 and Corollary 5.1). Corresponding convergences of u_n are established in Theorems 5.1–5.3.

LEMMA 4.1.

(i) $\| f_n - f \|_1 \le 2 \| u_n - v \|_2$,

(ii) $\| f_n - f \|_2 \le (\| u_n' \|_2^{1/2} + \| v' \|_2^{1/2}) \| u_n - v \|_2$,

(iii) $\| f_n - f \|_\infty \le 2 \| v' \|_2^{1/2} \| u_n - v \|_\infty + \| u_n - v \|_\infty^2$,

(iv) $\| f_n' - f' \|_2 \le 2[\| u_n' \|_2^{1/2} \| u_n' - v' \|_2 + \| v' \|_2 \| u_n - v \|_\infty]$.

PROOF. (i) For a proof, see e.g. LeCam (1970, page 803), or Pitman (1979, page 7).
(ii) Using, at the third stage below, the inequality

$$\| v \|_\infty^2 \le \| v \|_2 \| v' \|_2 \tag{4.1}$$

(see Klonias, 1981b), we have

$$\| f_n - f \|_2^2 = \int (u_n + v)^2 (u_n - v)^2 \le \| u_n + v \|_\infty^2 \| u_n - v \|_2^2$$
$$\le (\| u_n \|_\infty + \| v \|_\infty)^2 \| u_n - v \|_2^2 \le (\| u_n' \|_2^{1/2} + \| v' \|_2^{1/2})^2 \| u_n - v \|_2^2.$$

(iii) Using (4.1) at the last stage below, we have

$$| f_n - f | = | u_n - v | | 2v + (u_n - v) | \le 2 | v | | u_n - v | + | u_n - v |^2$$
$$\le 2 \| v' \|_2^{1/2} | u_n - v | + | u_n - v |^2.$$

(iv) Using (4.1) at the last stage below, we have

$$\| f_n' - f' \|_2 = 2 \| u_n u_n' - v v' \|_2 \le 2[\| u_n (u_n' - v') \|_2 + \| v' (u_n - v) \|_2]$$
$$\le 2[\| u_n \|_\infty \| u_n' - v' \|_2 + \| v' \|_2 \| u_n - v \|_\infty]$$
$$\le 2[\| u_n' \|_2^{1/2} \| u_n' - v' \|_2 + \| v' \|_2 \| u_n - v \|_\infty]. \qquad \square$$

As corollaries to Theorem 4.1 we can derive the integrated mean absolute deviation convergence of f_n as well as the convergence of some functionals of interest evaluated at f_n.

COROLLARY 4.1. *Under* A1 *and* A2 *and for the same values of d as in Theorem* 4.1(i), *we have*
(i) $n^d E\|f_n - f\|_1 \to 0$ *as* $n \to \infty$,
(ii) $n^d \|\mathscr{F}_n - F\|_\infty \to 0$ a.s. *as* $n \to \infty$,
(iii) $n^d |\|f_n\|_2 - \|f\|_2| \to 0$ a.s. *as* $n \to \infty$.

PROOF. (i) Notice that $\|f_n - f\|_1 \le \|f_n\|_1 + \|f\|_1 = 2$; the result now follows from Theorem 4.1(i) with $p = 1$, and the dominated convergence theorem.
(ii) The result is a consequence of Theorem 4.1(i) with $p = 1$, and the inequality: $|\mathscr{F}_n(x) - F(x)| \le \|f_n - f\|_1$.
(iii) The result follows from Theorem 4.1(i) with $p = 2$. □

COROLLARY 4.2. *Under* A1, A2 *and the additional assumption* $\|v''\|_2 < +\infty$, *and for the same value of d as in Theorem* 4.1(iii), *we have*

$$n^d |\|f_n'\|_2 - \|f'\|_2| \to 0 \quad \text{a.s.} \qquad as \quad n \to \infty.$$

In the following section we show that u_n converges a.s. to v in L_2, $\sup_{\mathbb{R}}$ and Sobolev norms, thus proving Theorem 4.1, in view of Lemma 4.1.

5. Consistency of the root density estimator. In this section we study the consistency properties of the MPLE u_n—the root density estimator. As indicated in the previous section, the convergence of u_n implies that of f_n in each mode included in Theorem 4.1. Since the converse is not generally true, one could also view the convergence properties of u_n as additional consistency properties of f_n.

Basic to the MPLE problem is inequality (5.1), to be derived next, which along with inequality (5.3) reduces the problem of the L_2-norm convergence of u_n (i.e. Hellinger distance convergence of f_n) to that of the convergence of the log-likelihood which in turn is shown (see inequality (5.2)) to be implied by the convergence of the empirical distribution, studied in Section 3.

LEMMA 5.1. *Under assumption* A1, *we have*

$$\begin{aligned}(5.1)\qquad &(\alpha_n/n)(\|u_n'\|_2^2 - \|v'\|_2^2) + 2\int \log v \, d(F_n - F)\\ &\le 2\int \log u_n \, dF_n - 2\int \log v \, dF\\ &\le 2\int \log u_n \, d(F_n - F).\end{aligned}$$

PROOF.

$$\begin{aligned}2\int \log v\, dF - (\alpha_n/n)\|v'\|_2^2 + 2\int \log v \, d(F_n - F) &= \Big(2\textstyle\sum_{i=1}^n \log v(X_i) - \alpha_n\|v'\|_2^2\Big)/n\\ &\le \Big(2\textstyle\sum_{i=1}^n \log u_n(X_i) - \alpha_n\|u_n'\|_2^2\Big)/n\\ &= 2\int \log u_n \, dF + 2\int \log u_n \, d(F_n - F) - (\alpha_n/n)\|u_n'\|_2^2\\ &\le 2\int \log v \, dF + 2\int \log u_n \, d(F_n - F) - (\alpha_n/n)\|u_n'\|_2^2.\end{aligned}$$

The first inequality above is true since u_n is the MPLE, and the last one (which is also a consequence of (5.4)) involves a well known inequality concerning entropy; see, for example, (le6.6) in Rao (1973). We then derive (5.1) by subtracting $2\int \log v\, dF - (\alpha_n/n)\|u_n'\|_2^2$ from all sides. □

The two most commonly occurring functionals in the consistency proofs that follow are $\int \log v\, d(F_n - F)$ and $\int \log u_n\, d(F_n - F)$ which we prove, in Proposition 5.1 below, to converge a.s. to zero; but first we give a bound on the latter.

LEMMA 5.2. *If* $E_f|X| < +\infty$, *then*

$$\left|\int \log u_n\, d(F_n - F)\right| \le (\lambda_n/\alpha_n)^{1/2}\|(\Gamma_n - I)/h\| \int h\circ F, \tag{5.2}$$

where h is a positive measurable function on (0, 1) *and Γ_n is the uniform empirical distribution function associated with the random variables* $F(X_1), \cdots, F(X_n)$.

PROOF. Using equation (2.2) we have for $x > x_{(n)}$,

$$\log u_n(x) = \log\,[(4\alpha_n\lambda_n)^{-1/2}\textstyle\sum_{i=1}^n (\exp\{(\lambda_n/\alpha_n)^{1/2}x_i\}/u_n(x_i))] - (\lambda_n/\alpha_n)^{1/2}x$$

and $F_n(x) = 1$, and similarly for $x < x_{(1)}$,

$$\log u_n(x) = \log[(4\alpha_n\lambda_n)^{1/2}\textstyle\sum_{i=1}^n (\exp\{-(\lambda_n/\alpha_n)^{1/2}x_i\}/u_n(x_i))] + (\lambda_n/\alpha_n)^{1/2}x$$

and $F_n(x) = 0$, i.e. $\log u_n(x)$ is linear in x, outside the range of observations. Then under our assumption $\{F_n(x) - F(x)\}\log u_n(x)$ vanishes at $\pm\infty$ (since $F_n(x)$ is respectively 0 or 1 as $x < x_{(1)}$ or $x > x_{(n)}$), and hence

$$\begin{aligned}\left|\int \log u_n\, d(F_n - F)\right| &= \left|-\int (F_n - F)(u_n'/u_n)\right| \le \int |F_n - F|\,|u_n'/u_n| \\ &\le (\lambda_n/\alpha_n)^{1/2}\int |F_n - F| \\ &= (\lambda_n/\alpha_n)^{1/2}\int (|\Gamma_n\circ F - F|/h\circ f)\,h\circ F \\ &\le (\lambda_n/\alpha_n)^{1/2}\|(\Gamma_n - I)/h\|\int_0^1 [h(t)/(f\circ F^{-1})(t)]\,dt,\end{aligned}$$

where the second inequality above is a consequence of the inequality

$$(u_n')^2 \le (\lambda_n/\alpha_n)\,u_n^2, \tag{5.3}$$

appearing in de Montricher et al. (1975), page 1340. □

REMARK 5.1. In what follows we use $h(t) = \{t(1-t)\}^\rho$, $t \in (0, 1)$ with $\rho \in (½, 1)$ and assume $\int \{F(1-F)\}^\rho < +\infty$ or rather its sufficient condition $E|X|^\tau < +\infty$ for some $\tau > 1/\rho$ (this is Assumption A2). To see the sufficiency of the moment condition, notice that it implies $|x|^\tau F(x)\{1 - F(x)\} \to 0$ as $|x| \to \infty$ and hence there exist positive constants c, m such that $|x|^\tau F(x)\{1 - F(x)\} < c$ for all $x \in M \equiv \{x \in \mathbb{R}: |x| > m\}$, which implies

$$\int \{F(1-F)\}^\rho \le \int_{M^c} \{F(1-F)\}^\rho + c^\rho \int_M |x|^{-\tau\rho}\,dx < +\infty.$$

In view of Remark 5.1, the assumptions of Theorem 3.1 on s and ρ and Assumption A2 on τ, we must have $s < 1 - (2\wedge\tau)^{-1}$, whenever we invoke Theorem 3.1. It is this bound on s that determines the rate of the types of convergence of the estimate that follow.

PROPOSITION 5.1. *Under Assumptions* A1 *and* A2, *we have*

(i) $$n^d \int \log v \, d(F_n - F) \to 0 \quad \text{a.s.} \qquad \text{as} \quad n \to \infty,$$

for any d such that $0 \le d < \frac{1}{2}$, *and*

(ii) $$n^d \int \log u_n \, d(F_n - F) \to 0 \quad \text{a.s.} \qquad \text{as} \quad n \to \infty$$

for any d such that $d < (t/2) - (2(\tau \wedge 2)^{-1} - 1)/2$. (*Here t is as in Remark* 4.1 *and* τ *as in* A2; *actually* $\tau > 0$ *is sufficient for* (*i*).)

PROOF. (i) Notice that

$$2n^d \int \log v \, d(F_n - F) = n^{-(1-d)} \sum_{i=1}^{n} [\log f(X_i) - \int \log f \, dF],$$

and under Assumption A2 $\int f |\log f|^\gamma < +\infty$ for any $\gamma > 0$ (see Klonias, 1981a). Taking $\gamma = (1 - d)^{-1} \in (1, 2)$, the result then follows from the Marcinkiewicz version of the S.L.L.N. (see, e.g., Loève, 1963, section 16.4, page 243).
(ii) From inequality (5.2) of Lemma 5.2, and inequality (2.3), we have

$$n^d \left| \int \log u_n \, d(F_n - F) \right| < n^d (n/\alpha_n)^{1/2} \| (\Gamma_n - I)/\{I(1 - I)\}^\rho \| \int \{F(1 - F)\}^\rho.$$

Notice that under Assumption A2, $\int \{F(1 - F)\}^\rho < +\infty$ (see Remark 5.1); then the convergence follows as a result of Theorem 3.1, with $s = d + (1 - t)/2$. □

We derive next an inequality, to be used in the proof of the second part of Theorem 5.1, involving the Kullback-Liebler information functionals and the Hellinger distance of two probability density functions.

LEMMA 5.3. *Let f, g be probability density functions. Then*

(5.4) $$\| f^{1/2} - g^{1/2} \|_2^2 \le \int f \log(f/g).$$

PROOF. The inequality is trivial unless f vanishes whenever g vanishes, so assume the latter and let $v \equiv f^{1/2}$ and $w = g^{1/2}$. Since $\log(1 + z) < z$ for $z > -1$, we have

$$\begin{aligned} \int f \log(g/f) &= 2 \int_{[v>0]} f \log(w/v) \\ &= 2 \int_{[v>0]} f \log[1 + \{(w/v) - 1\}] \le 2 \int f((w/v) - 1) \\ &= -2\left(1 - \int vw\right) = -\| w - v \|_2^2. \quad \square \end{aligned}$$

We can now prove that the MPLE u_n of v, the root density, is L_2-strongly-consistent; we write $\| \cdot \|_p$, $p = 1, 2$, for the $L_p(\mathbb{R})$ - norm, and $\| \cdot \|_\infty$ for the $\sup_{\mathbb{R}}$-norm.

THEOREM 5.1. *Under* A1 *and* A2, *for* $0 \le d < (1 - t)/2$ *such that* $d < t/4 - (2(\tau \wedge 2)^{-1} - 1)/4$ *we have*

(i) $n^{2d}\{1 - (\lambda_n/n)\} \to 0$ a.s. *as* $n \to \infty$,
(ii) $n^d \| u_n - v \|_2 \to 0$ a.s. *as* $n \to \infty$.

PROOF. (i) It is easy to check that $\lambda_n \| u_n \|_2^2 + \alpha_n \| u_n' \|_2^2 = n$ (as in de Montricher et al., 1975, page 1340) λ_n is chosen so that $\| u_n \|_2 = 1$; hence $1 - (\lambda_n/n) = (\alpha_n/n) \| u_n' \|_2^2$. Then from the inequality between the left and right-most side of (5.1) we obtain

$$(5.5) \quad 0 < 1 - (\lambda_n/n) \le 2 \int \log u_n \, d(F_n - F) - 2 \int \log v \, d(F_n - F) + (\alpha_n/n) \| v' \|_2^2.$$

The result now follows from Proposition 5.1.

(ii) We have

$$\begin{aligned}
2 \int \log v \, dF &- 2 \int \log u_n \, dF \\
&= 2 \int \log u_n \, d(F_n - F) - \left\{ 2 \int \log u_n \, dF_n - 2 \int \log v \, dF \right\} \\
&\le 2 \int \log u_n \, d(F_n - F) - 2 \int \log v \, d(F_n - F) \\
&\quad - (\alpha_n/n)(\| u_n' \|_2^2 - \| v' \|_2^2) \\
&\le 2 \int \log u_n \, d(F_n - F) - 2 \int \log v \, d(F_n - F) \\
&\quad + (\alpha_n/n) \| v' \|_2^2,
\end{aligned}$$

where the first inequality above is a consequence of (5.1). This, along with (5.4) and Proposition 5.1, gives the result. □

REMARK 5.2. Notice that setting $r_n \equiv (\lambda_n/\alpha_n)^{1/2}$, the MPLE u_n given by (2.2), has the form

$$u_n(x) = \lambda_n^{-1} \sum_{i=1}^n u_n(x_i)^{-1} (r_n/2) e^{-r_n |x - x_i|}, \qquad x \in \mathbb{R},$$

and from Theorem 5.1(i) we have that λ_n is asymptotically equivalent to the sample size n and r_n to $c^{-1/2} n^{(1-t)/2}$. We could then think of u_n (at least asymptotically) as a kernel estimator of $f^{1/2}$, the kernel being the Laplace density and the bandwidth r_n^{-1}.

Notice also that Theorem 5.1(ii) and the fact that $\| u_n - v \|_2^2 \le 2$ imply, through the use of the dominated convergence theorem, the integrated mean square error convergence of u_n, i.e. $E \| u_n - v \|_2^2 \to 0$ as $n \to \infty$.

With the additional assumption $\| v'' \|_2 < +\infty$, we will show (see Corollary 5.2 at the end of this section) that $\| u_n' \|_2 \to \| v' \|_2$ a.s. (note that $4 \| u_n' \|_2^2$ is an estimate of the Fisher information). Under our present assumptions however, it is not known whether $\| u_n' \|_2$ converges. The weaker conclusions, though, of Proposition 5.2 below suffice for the result of Theorem 5.2.

PROPOSITION 5.2. *Under* A1 *and* A2, *we have*

(i) $$\limsup_n n^d (\| u_n' \|_2^2 - \| v' \|_2^2) = 0 \quad \text{a.s.},$$

for d such that $d < (3t - 1 - 2(\tau \wedge 2)^{-1})/2$. Also for any $t > (1 + 2(\tau \wedge 2)^{-1})/3$ and any sequence a_n such that $a_n \to 0$ (a.s.), we have

(ii) $$a_n \left| \| u_n' \|_2^2 - \| v' \|_2^2 \right| \to 0 \quad \text{a.s.}$$

PROOF. Using the left and right-most side of inequality (5.1) we have

$$(5.6)\quad -\|v'\|_2^2 < \|u_n'\|_2^2 - \|v'\|_2^2 \le (n/\alpha_n)\left\{2\int \log u_n\, d(F_n - F) - 2\int \log v\, d(F_n - F)\right\},$$

which readily gives both results as a consequence of Proposition 5.1. □

We now establish the strong consistency of u_n w.r.t. the supremum norm, including a rate of convergence.

THEOREM 5.2. *Under* A1 *and* A2 *and for* d *such that* $d < (1-t)/4$ *and* $d < t/2 - (\tau \wedge 2)^{-1}/2$, *we have*

$$n^d \|u_n - v\|_\infty \to 0 \quad \text{a.s.} \qquad as \quad n \to \infty.$$

PROOF. In Section 2 we presented the reproducing kernel

$$k_{\lambda_n,\alpha_n}(x, y) = (4\lambda_n\alpha_n)^{-1/2}\exp\{-(\lambda_n/\alpha_n)^{1/2}|x - y|\}$$

(see (2.1) of $H \equiv H^1(\mathbb{R})$ corresponding to the inner product $\langle\cdot,\cdot\rangle_{\lambda_n,\alpha_n}$, which induces the Sobolev norm

$$\|u\|_{\lambda_n,\alpha_n} = (\lambda_n\|u\|_2^2 + \alpha_n\|u'\|_2^2)^{1/2},\ u \in H.$$

Notice that $\|k_{\lambda_n,\alpha_n}(\cdot, y)\|_{\lambda_n,\alpha_n}^2 = (4\lambda_n\alpha_n)^{-1/2}$ for all $y \in \mathbb{R}$.

We then have

$$\begin{aligned}
|u_n(x) - v(x)|^2 &= |\langle k_{\lambda_n,\alpha_n}(\cdot, x), u_n - v\rangle_{\lambda_n,\alpha_n}|^2 \\
&\le \|k_{\lambda_n,\alpha_n}(\cdot, x)\|_{\lambda_n,\alpha_n}^2 \|u_n - v\|_{\lambda_n,\alpha_n}^2 \\
&= (4\lambda_n\alpha_n)^{-1/2}[\lambda_n\|u_n - v\|_2^2 + \alpha_n\|u_n' - v'\|_2^2] \\
&= (\lambda_n/4\alpha_n)^{1/2}[\|u_n - v\|_2^2 + (\alpha_n/\lambda_n)\|u_n' - v'\|_2^2] \\
&= (1/2)[(\lambda_n/\alpha_n)^{1/2}\|u_n - v\|_2^2 \\
&\quad + (\alpha_n/\lambda_n)^{1/2}(\|u_n'\|_2^2 + \|v'\|_2^2 - 2\langle u_n', v'\rangle_2)] \\
&\le (1/2)[(\lambda_n/\alpha_n)^{1/2}\|u_n - v\|_2^2 \\
&\quad + (\alpha_n/\lambda_n)^{1/2}(\|u_n'\|_2^2 - \|v'\|_2^2) + 2(\alpha_n/\lambda_n)^{1/2}\|v'\|_2^2 \\
&\quad + 2(\alpha_n/\lambda_n)^{1/4}[(\alpha_n/\lambda_n)^{1/2}(\|u_n'\|_2^2 - \|v'\|_2^2) + (\alpha_n/\lambda_n)^{1/2}\|v'\|_2^2]^{1/2}\|v'\|_2],
\end{aligned}$$

where both inequalities above are a consequence of the Cauchy-Schwarz inequality. The result now follows from Theorem 5.1 (ii) and Proposition 5.2 (ii). □

It is not known whether under the present assumptions u_n converges to v in the Sobolev norm (we can only prove a result of the type of Proposition 5.2 (ii)), but under the additional assumption: $v'' \in L_2$, we can obtain this type of convergence. Recall that $\|v\|_H^2 = \eta\|v\|_2^2 + \theta\|v'\|_2^2$, $v \in H$ and $\eta, \theta > 0$.

THEOREM 5.3. *Under* A1 *and* A2 *and the additional assumption* $\|v''\|_2 < +\infty$, *for* d *such that* $d < 3t/4 - (1 + 2(\tau \wedge 2)^{-1})/4$ *and* $d < (1-t)/4$, *we have*

$$n^d\|u_n - v\|_H \to 0 \quad \text{a.s.} \qquad \text{as} \quad n \to \infty.$$

PROOF. After an integration by parts and an application of the Cauchy-Schwarz inequality, we have from (5.6),

$$\| u' - v' \|_2^2 = \| u_n' \|_2^2 - \| v' \|_2^2 + 2 \int v'(v' - u_n')$$

$$= \| u_n' \|_2^2 - \| v' \|_2^2 + 2 \int v''(u_n - v)$$

$$\leq (n/\alpha_n) \left\{ 2 \int \log u_n \, d(F_n - F) - 2 \int \log v \, d(F_n - F) \right\}$$

$$+ 2 \| v'' \|_2 \| u_n - v \|_2,$$

and the result follows from Proposition 5.1 and Theorem 5.1 (ii).

As far as the rates are concerned, we must have $2d + 3(1 - t)/2 < 1 - (\tau \wedge 2)^{-1}$ for the first term on the right side of the inequality above to converge and for the second we need $4d + (1 - t)/2 < 1 - (\tau \wedge 2)^{-1}$ and $2d < (1 - t)/2$, and notice that the first and third of these relations imply the second. □

As an immediate corollary to Theorem 5.3, we obtain the convergence of the Fisher-information functional evaluated at u_n—i.e., $I(f_n)$ is a strongly consistent estimator of $I(f)$.

COROLLARY 5.1. *Under the assumptions of Theorem* 5.3 *and for the same values of d, we have*

$$n^d \,|\, \| u_n' \|_2 - \| v' \|_2 \,| \to 0 \quad \text{a.s.} \qquad as \quad n \to \infty.$$

In all of the consistency results so far, the rate t at which α_n tends to infinity, and the rate d of convergence of the estimator, are described through a set of inequalities relating t, d and τ (of the moment assumption, A2). In the remark that follows we give an idea of the possible values of t and of d.

REMARK 5.3. If we assume a finite second moment, i.e. $\tau = 2$, then:

(i) for any $t \in (0, 1)$, u_n converges a.s. w.r.t. the L_2-norm (Theorem 5.1 (ii)) and the highest rate of convergence ($< 1/6$; $2d < 1/3$ in Theorem 5.1) is achieved for $t = (2/3)^-$;

(ii) for any $t \in (1/2, 1)$, u_n converges a.s. w.r.t. the $\sup_{\mathbb{R}}$-norm (Theorem 5.2) and the highest rate of convergence ($<1/12$) is achieved for $t = 2/3$;

(iii) for any $t \in (2/3, 1)$, u_n converges a.s. w.r.t. the Sobolev norm (Theorem 5.3) and the highest rate of convergence ($<1/16$) is achieved for $t = 3/4$.

Hence under the assumption of a finite second moment, a choice of $t = (2/3)^+$ guarantees the convergence of u_n in all senses included in this section, and indeed the convergences in Theorems 5.1 and 5.2 attain their maximum rate (that our proofs allow).

6. A MPLE with positive support. Let f be the density function of a nonnegative random variable, such that $v = f^{1/2} \notin H^1(\mathbb{R})$, e.g., f is an exponential density; then the developments of the previous sections do not apply. However, we will show that if f, with support $\mathbb{R}_+ \equiv (0, +\infty)$, is such that $v \in H^1(\mathbb{R}_+)$, then the "first MPLE of Good and Gaskins" f_+, (we suppress the subscript n), exists, is unique, and is also an exponential spline with knots at the sample points, given by $f_+ = u_+^2$, where u_+—the MPLE of v—is given by (6.1) below. Finally, we indicate the computational feasibility and the consistency properties of the estimator.

Let $\| \cdot \|_2$, $\| \cdot \|_{2,+}$ denote the $L_2(\mathbb{R})$ and $L_2(\mathbb{R}_+)$ norms respectively, and consider the MPLE problem:

PROBLEM P4. $\max \prod_{i=1}^n u^2(X_i) \exp\{-\alpha \| u' \|_{2,+}^2\}$, $u \in H^1(\mathbb{R}_+)$ subject to $\| u \|_{2,+} = 1$ and $u(X_i) \geq 0$, $i = 1, 2, \cdots, n$.

PROPOSITION 6.1. *Problem* P4 *has a unique solution* u_+, *given implicitly by*

$$u_+(x) = (4\lambda\alpha)^{-1/2} \sum_{i=1}^{n} u_+(X_i)^{-1}[\exp\{-(\lambda/\alpha)^{1/2}|x - X_i|\} + \exp\{-(\lambda/\alpha)^{1/2}|x + X_i|\}], \quad x \in \mathbb{R}_+, \tag{6.1}$$

where $\lambda > 0$ *is the Lagrange multiplier corresponding to the constraint* $\| u \|_{2,+} = 1$.

PROOF. Let $\bar{u}(x) \equiv u(|x|)$ for all $x \in \mathbb{R}\backslash\{0\}$, $\bar{u}(0) \equiv \lim_{x\to 0+} u(x)$, and set $X_{-i} \equiv X_i$ for all $i = 1, \cdots, n$. Then Problem P4 is equivalent to

PROBLEM P5. $\max \prod_{|i|=1}^{n} \bar{u}^2(X_i)\exp\{-\alpha \| \bar{u}' \|_2^2\}$, $\bar{u} \in H_s$ subject to $\| \bar{u} \|_2^2 = 2$ and $\bar{u}(X_i) \geq 0$, $|i| = 1, \cdots, n$, where $H_s \equiv \{g \in H^1(\mathbb{R})\colon g(x) = g(-x) \text{ for all } x \in \mathbb{R}\}$.

Notice that for $\bar{u} \in H^1(\mathbb{R})$, i.e., for $\bar{u}$ not necessarily symmetric, there exists a unique solution to Problem P5 given by

$$\bar{u}_0(x) = (4\lambda\,\alpha)^{-1/2} \sum_{|i|=1}^{n} \bar{u}_0(X_i)^{-1}\exp\{-(\lambda/\alpha)^{1/2}|x - X_i|\}, \qquad x \in \mathbb{R},$$

where λ is the Lagrange multiplier corresponding to the constraint $\| \bar{u} \|_2^2 = 2$. The arguments leading to this result are identical to those in de Montricher et al (1975) leading to (2.2). Hence to show that the spline function $\bar{u}_0$ is also the unique solution to Problem P5 and hence $u_+(x) \equiv \bar{u}(x)$ for $x \in \mathbb{R}_+$, the unique solution to Problem P4, we need only prove that $\bar{u}$ is in H_s—i.e., symmetric about zero. To this end notice that $\bar{u}_0$ is symmetric everywhere if it is symmetric at the knots, i.e., if $\bar{u}(X_i) = \bar{u}(-X_i)$ for $i = 1, \cdots, n$. But this is true since in system (6.2) below the variables $\bar{u}(X_i)$, $\bar{u}(-X_i)$, $i = 1, \cdots, n$ are interchangeable:

$$\begin{aligned} \bar{u}(X_j) &= (4\lambda\alpha)^{-1/2} \sum_{i=1}^{n} [\bar{u}(X_i)^{-1}\exp\{-(\lambda/\alpha)^{1/2}|X_j - X_i|\} \\ &\quad + \bar{u}(-X_i)^{-1}\exp\{-(\lambda/\alpha)^{1/2}|X_j + X_i|\}], \\ \bar{u}(-X_j) &= (4\lambda\alpha)^{-1/2} \sum_{i=1}^{n} [\bar{u}(X_i)^{-1}\exp\{-(\lambda/\alpha)^{1/2}|X_j + X_i|\} \\ &\quad + \bar{u}(-X_i)^{-1}\exp\{-(\lambda/\alpha)^{1/2}|X_j - X_i|\}], \\ j &= 1, \cdots, n. \end{aligned} \tag{6.2}$$

COROLLARY 6.1. *The "first* MPLE *of Good and Gaskins" when f has its support on* $\mathbb{R}_+$ *is given by* $f_+ = u_+^2$.

PROOF. This is a consequence of the nonnegativity of u_+ and Lemma 3.1 in de Montricher et al (1975).

REMARK 6.1. All the consistency results developed in the earlier sections for $f_n = u_n^2$, where u_n is given by (2.2), are also valid for f_+ and very little has to be changed in the way of proofs.

REMARK 6.2. Equation (6.1) gives u_+ only implicitly and the values of the estimate at the sample points have to be determined, i.e., system (6.2) has to be solved and λ to be chosen so that $\| \bar{u} \|_2^2 = 2$. In Chapter 4 of Klonias (1980) and in Hall and Klonias (1981), utilizing the particular structure of the "first MPLE of Good and Gaskins," an efficient method is presented for the resolution of the spline f_n, which can be easily adapted to determine the values of f_+ at the knots. The reader is also referred to Good and Gaskins (1971, 1980), Scott, Tapia and Thompson (1976), Tapia and Thompson (1978), and Ghorai and Rubin (1979), where methods for the numerical evolution of f_n are presented.

Acknowledgments. I would like to express my gratitude to Professor W. J. Hall, my thesis advisor, for many helpful discussions and for his advice in general during the preparation of my Ph.D. dissertation of which this is part. I would also like to thank

Professor Jon A. Wellner for his conjecture regarding Theorem 3.1. It is also a pleasure to acknowledge Professor R. J. Serfling for his encouragement, comments on earlier drafts, and suggestions that substantially improved the exposition.

REFERENCES

ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

CHUNG, K. L. (1947). Note on some strong laws of large numbers. *Amer. J. Math.* **69** 189–192.

DE MONTRICHER, G. F., TAPIA, R. A. and THOMPSON, J. R. (1975). Nonparametric maximum likelihood estimation of probability densities by penalty function methods. *Ann. Statist.* **3** 1329–1348.

DE MONTRICHER, G. F. (1981). On the consistency of maximum penalized likelihood estimators. Rice University, Texas.

GEL'FAND, I. M. and SHILOV, G. E. (1964). *Generalized Functions*, Vol. 1. Academic, New York and London.

GHORAI, J. and RUBIN, H. (1979). Computational procedures for maximum penalized likelihood estimate. *J. Statist. Comput. Simul.* **10** 65–78.

GOOD, I. J. and GASKINS, R. A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika* **58** 2, 255–277.

GOOD, I. J. and GASKINS, R. A. (1980). Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. (Invited paper), *J. Amer. Statist. Assoc.* **75** 42–73.

HALL, W. J. and KLONIAS, V. K. (1981). On the numerical evaluation of a maximum penalized likelihood estimator of the density function. *Tech. Report* No. 344, Department of Mathematical Sciences, The Johns Hopkins University.

KLONIAS, V. K. (1980). Nonparametric density estimation: Contributions to the maximum penalized likelihood method. Ph.D. dissertation, The University of Rochester, Rochester, N.Y.

KLONIAS, V. K. (1981a). Conditions for finite absolute moments of the logarithm of the probability density function. *Tech. Report No. 335, Dept. of Math. Sciences, The Johns Hopkins University.*

KLONIAS, V. K. (1981b). An upper bound to the probability density function in terms of the Fisher information. *Tech. Report No. 336 (ONR No. 81-1), Dept. of Math. Sciences, The Johns Hopkins University.*

LECAM, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Statist.* **41** 802–828.

LOÈVE, M. M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton, N.J.

MASON, DAVID M. (1981). Bounds for weighted empirical distribution functions. *Ann. Probability* **9** 881–884.

PARZEN, E. (1967). *Time Series Analysis Papers.* Holden-Day, San Francisco.

PITMAN, E. J. G. (1979). *Some Basic Theory for Statistical Inference.* Wiley, New York.

SCOTT, D. W., TAPIA, R. A. and THOMPSON, J. R. (1976). *Computer Science and Statistics: Ninth Annual Symposium on the Interface.*

SILVERMAN, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist.* **10** 795–810.

TAPIA, R. A., and THOMPSON, J. R. (1978). *Nonparametric Probability Density Estimation.* The Johns Hopkins University Press, Baltimore and London.

WELLNER, J. A. (1975). Convergence of the sequential uniform empirical process with bounds for centered beta rv's and a log-log law. Ph.D. Dissertation, University of Washington, Seattle.

WELLNER, J. A. (1977). A law of the iterated logarithm for functions of order statistics. *Ann. Statist.* **5** 481–494.

YOSIDA, K. (1974). *Functional Analysis*, 4th ed. Springer-Verlag; Berlin.

MATHEMATICAL SCIENCES
THE JOHNS HOPKINS UNIVERSITY
BALTIMORE, MARYLAND 21218