The Annals of Statistics
1975, Vol. 3, No. 6, 1329–1348

NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATION OF PROBABILITY DENSITIES BY PENALTY FUNCTION METHODS[1]

BY G. F. DE MONTRICHER, R. A. TAPIA, AND J. R. THOMPSON

Princeton University and Rice University

Tne maximum likelihood estimate of a probability density function based on a random sample does not exist in the nonparametric case. For this reason and others based on heuristic Bayesian considerations Good and Gaskins suggested adding a penalty term to the likelihood. They proposed two penalty terms; however they did not establish existence or uniqueness of their maximum penalized likelihood estimates. Good and Gaskins also suggested an alternate approach for calculating the maximum penalized likelihood estimate which avoids the nonnegativity constraint on the estimate. In the present work the existence and uniqueness of both of Good's and Gaskins' maximum penalized likelihood estimates are rigorously demonstrated. Moreover, it is shown that one of these estimates is a positive exponential spline with knots only at the sample points and that in this case the alternate approach leads to the correct estimate; however in the other case the alternate approach leads to the wrong estimate. Finally, it is shown that a well-known class of reproducing kernel Hilbert spaces leads very naturally to maximum penalized likelihood estimates which are polynomial splines with knots at the sample points.

1. Introduction. Let Ω denote the interval (a, b). In this study we consider the problem of estimating the (unknown) probability density function $f \in L^1(\Omega)$ which gave rise to the random sample $x_1, \cdots, x_N \in \Omega$. The set Ω may be bounded or unbounded.

As usual define $L(v)$, the *likelihood* that $v \in L^1(\Omega)$ gave rise to the random sample $x_1, \cdots, x_N$ by

$$L(v) = \prod_{i=1}^{N} v(x_i) . \tag{1.1}$$

Let $H(\Omega)$ be a manifold in $L^1(\Omega)$ and consider the following optimization problem:

$$\text{maximize} \quad L(v)\,; \quad \text{subject to} \quad v \in H(\Omega)\,, \qquad \int_\Omega v(t)\,dt = 1 \quad \text{and} \quad v(t) \geqq 0 \qquad \forall\, t \in \Omega\,. \tag{1.2}$$

Here dt denotes the Lebesgue measure on Ω. By a *maximum likelihood estimate* based on the random sample $x_1, \cdots, x_N$ (corresponding to the manifold $H(\Omega)$) we mean any solution of the constrained optimization problem (1.2). The

Received August 1974; revised March 1975.

[1] Invited paper presented to annual meeting of the Institute of Mathematical Statistics in Edmonton, Alberta, Canada on August 15, 1974. Research sponsored by the Office of Naval Research under Contract NR-042-283.

AMS 1970 *subject classifications.* Primary 62G05; Secondary 62E10.

Key words and phrases. Density estimation, maximum likelihood estimation, spline, monospline.

estimate is said to be *parametric* if $H(\Omega)$ is a finite dimensional manifold and *nonparametric* if $H(\Omega)$ is an infinite dimensional manifold.

In general a nonparametric maximum likelihood estimate does not exist. To see this observe that the nonexistent solution is idealized by a linear combination of Dirac delta spikes at the samples and gives a value of $+\infty$ to the likelihood functional. Hence, in any infinite dimensional manifold which has the property that it is possible to construct a sequence of functions which integrate to one, are nonnegative and converge pointwise to a Dirac delta spike, the likelihood will be unbounded and a maximum likelihood estimate will not exist. Morever most infinite dimensional manifolds in $L^1(\Omega)$ will have this property, e.g., the continuous functions, the differentiable functions, the infinitely differentiable functions and the polynomials. Of course we may consider pathological examples of infinite dimensional manifolds on which the likelihood is bounded. As an extreme case let $H(\Omega)$ be all continuous functions on Ω which vanish at the samples $x, \cdots, x_N$. Then the likelihood is identically zero on the infinite dimensional manifold $H(\Omega)$ and is therefore bounded. Furthermore any member of $H(\Omega)$ which is nonnegative and integrates to one is a maximum likelihood estimate for the random sample $x_1, \cdots, x_N$.

The fact that in general the nonparametric maximum likelihood estimate does not exist implies that, except in the extreme case where the parametric form of the unknown density function f is known a priori, the parametric maximum likelihood approach for large dimensional problems must necessarily lead to unsmooth estimates and a numerically ill-posed problem. This leaves the practitioner with the following dilemma: For small dimensional problems he has no flexibility and the solution will be greatly influenced by the choice of the manifold $H(\Omega)$; while for large dimensional problems the solution must necessarily approximate a linear combination of Dirac delta spikes, be unsmooth and create numerical problems.

The following example will illustrate many of these points. For a given positive integer n partition Ω into n half-open half-closed disjoint intervals $T_1, \cdots, T_n$ of equal length $h = (b - a)/n$. Let $I(T_i)$ denote the characteristic function of the interval T_i and let ν_i denote the number of samples in the interval T_i. The well-known histogram estimate for f based on the random sample $x_1, \cdots, x_N$ is given by

$$f^* = \sum_{i=1}^n \frac{\nu_i}{Nh} I(T_i) . \tag{1.3}$$

LEMMA 1.1. *The probability density estimate* (1.3) *is the maximum likelihood estimate for the random sample* $x_1, \cdots, x_N$ *corresponding to the n-dimensional manifold* $H(\Omega)$ *where* $H(\Omega)$ *is the linear span of the characteristic functions* $\{I(T_i) : i = 1, \cdots, n\}$.

PROOF. A typical member ω of $H(\Omega)$ has the form

$$\omega = \sum_{i=1}^n y_i I(T_i) .$$

The nonnegativity constraint has the form $y_i \geqq 0$, $i = 1, \cdots, n$ and will not be active at the solution. To see this observe that if $\nu_i > 0$, then the optimal solution of problem (1.2) will have $y_i > 0$. Since the log is a concave function we may work with the log likelihood instead of the likelihood. Hence we are interested in determining $y_1, \cdots, y_n$ which maximizes

$$G(y_1, \cdots, y_n) = \sum_{i=1}^n \nu_i \log (y_i)$$

subject to the integral constraint $h \sum_{i=1}^n y_i = 1$. From the theory of Lagrange multipliers, see e.g., Fiacco–McCormick (1969) Chapter 2, we must have

$$\nu_i + \lambda y_i = 0\,, \qquad i = 1, \cdots, n \tag{1.4}$$
$$h \sum_{i=1}^n y_i = 1$$

for some scalar λ. It is not difficult to see that $y_i = \nu_i/(hN)$ gives the unique solution of (1.4) and that this solution satisfies the sufficiency conditions for a maximizer. This proves the lemma.

Notice that for a fixed sample as $n \to \infty$ the estimate f^* given by (1.3) has the property that $f^*(x_i) \to +\infty$, $i = 1, \cdots, N$ while $f^*(x) \to 0$ if $x \notin \{x_1, \cdots, x_N\}$. Hence for large n our maximum likelihood estimate is very unsmooth and unsatisfactory. Whether or not we obtain a reasonable estimate is completely dependent on the delicate and tricky art of choosing n, properly. It is also of interest to note that the numerical properties of (1.4) are very poor for large n.

For the reasons stated above and others based on heuristic Bayesian considerations Good and Gaskins (1971) suggested adding a penalty term to the likelihood which would penalize rough (unsmooth) estimates. They suggested two specific penalty terms; however, they left the reader in the unfortunate situation of not knowing whether their maximum penalized likelihood estimates exist. Good and Gaskins also suggested an alternate approach for constructing the maximum penalized likelihood estimate which avoids the nonnegativity constraint. Again they do not demonstrate that the original approach and the alternate approach give the same estimate or that the estimate obtained from the alternate approach exists.

In Section 2 we establish a general existence and uniqueness theory for a large class of maximum penalized likelihood estimates. This general theory is used to show that a well-known class of reproducing kernel Hilbert spaces (Sobolev spaces) lead quite naturally to maximum penalized likelihood estimates which are polynomial splines (monosplines in the terminology of Schoenberg (1968)) with knots at the sample points.

In Section 3 a rigorous demonstration, using the theory developed in Section 2, is given of the existence and uniqueness of one of Good's and Gaskins' maximum penalized likelihood estimates. It is also shown that this estimate is a positive exponential spline with knots only at the sample points and that the alternate approach suggested by Good and Gaskins gives the correct estimate.

In Section 4 using the theory developed in Section 2 a rigorous demonstration of the existence and uniqueness of Good's and Gaskins' other maximum penalized estimate is given. Moreover, it is also demonstrated that the estimate obtained from Good's and Gaskins' alternate approach is not the maximum penalized likelihood estimate.

Much of our analysis uses the notions of the Fréchet gradient, the Fréchet derivative and the second Fréchet derivative in an abstract Hilbert space. The reader not familiar with these notions is referred to Tapia (1971).

Many important statistical considerations, e.g., consistency, unbiasedness, convergence rate, choice of penalty term and numerical implementation are unanswered here. Instead, it is the purpose of the present work to set the maximum penalized likelihood approach on solid approximation theoretic ground. Hopefully, the insights gained from this study will lead to investigation of the classical statistical properties of the estimates considered.

Before moving on to Section 2 a few brief historical comments are in order. Rosenblatt (1956) performed the first analytical study of the theoretical properties of histograms. Parzen (1962) constructed a class of estimators which properly included the histogram estimators and examined the consistency properties of the estimators in this class. These results have been improved upon recently by Wahba (1971). Kimeldorf and Wahba (1970) introduced the application of spline techniques in contemporary statistics. Boneva, Kendall and Stefanov (1971) and Schoenberg (1972) examined the use of spline techniques for obtaining from histograms smooth estimates of a probability density function. It is of interest to us that essentially all previous authors seem either to ignore the nonnegativity constraint or to attempt handling it with the seemingly clever trick of working with a function whose square is to be used as the estimate of the probability density; however in many cases this approach tacitly ignores the nonnegativity constraint. More will be said about the use of this approach in Sections 3 and 4.

2. Maximum penalized likelihood estimators. Let $H(\Omega)$ be as in Section 1 and consider a functional $\Phi : H(\Omega) \to R$. Given the random sample $x_1, \cdots, x_N \in \Omega$ the *Φ-penalized likelihood* of $v \in H(\Omega)$ is defined by

$$\hat{L}(v) = \prod_{i=1}^{N} v(x_i) \exp(-\Phi(v)) . \tag{2.1}$$

Consider the constrained optimization problem:

$$\text{maximize} \quad \hat{L}(v) \,; \quad \text{subject to} \qquad v \in H(\Omega)\,, \quad \int v(t)\,dt = 1 \quad \text{and} \quad v(t) \geqq 0\,, \quad \forall\, t \in \Omega\,. \tag{2.2}$$

The form of the penalized likelihood (2.1) is due to Good and Gaskins (1971). Their specific suggestions are analyzed in Sections 3 and 4.

Any solution to problem (2.2) is said to be a *maximum penalized likelihood estimate* based on the random sample $x_1, \cdots, x_N$ corresponding to the manifold $H(\Omega)$ and the penalty function Φ. The terms parametric and nonparametric have the same meaning in this context as they did in the context of Section 1. In the case when $H(\Omega)$ is a Hilbert space a very natural penalty function to use is $\Phi(v) = ||v||^2$ where $||\cdot||$ denotes the norm on $H(\Omega)$. Consequently when $H(\Omega)$ is a Hilbert space and we refer to the penalized likelihood functional on $H(\Omega)$ or to the maximum penalized likelihood estimate corresponding to $H(\Omega)$ with no reference to the penalty functional Φ we are assuming that Φ is the square of the norm in $H(\Omega)$. The Hilbert space inner product will be denoted by $\langle \cdot, \cdot \rangle$ so that $\langle x, x\rangle = ||x||^2$. When $H(\Omega)$ is a Hilbert space it is said to be a reproducing kernel Hilbert space if point evaluation is a continuous operation, i.e., $v_n \to v$ in $H(\Omega)$ implies $v_n(x) \to v(x)\ \forall\, x \in \Omega$; see Goffman and Pedrick (1965).

For problem (2.2) to make sense we would like $H(\Omega)$ to have the property that for $x_1, \cdots, x_N \in \Omega$ there exists at least on $v \in H(\Omega)$ such that

$$\int_\Omega v(t)\,dt = 1\,, \quad v(t) \geqq 0 \quad \forall\, t \in \Omega \quad \text{and} \quad v(x_i) > 0 \quad i = 1, \cdots, N\,. \tag{2.3}$$

PROPOSITION 2.1. *Suppose that $H(\Omega)$ is a reproducing kernel Hilbert space and D is a closed convex subset of $\{v \in H(\Omega) : v(x_i) \geqq 0\}$ with the property that D contains at least one function which is positive at the samples $x_1, \cdots, x_N$. Then the penalized likelihood functional* (2.1) *has a unique maximizer in D.*

PROOF. Since $H(\Omega)$ is a reproducing kernel Hilbert space we have $|v(x_i)| \leqq K_i||v||$ for $i = 1, \cdots, N$. It follows that

$$|\hat{L}(v)| \leqq C_1||v||^N \exp(-||v||^2)\,. \tag{2.4}$$

The function $\theta(\lambda) = \lambda^N \exp(-\lambda)^2$ is bounded above by $(N/2)^{N/2}\exp(-N/2)$; hence $|\hat{L}(v)| \leqq C_2$. If $M = \sup\{\hat{L}(v) : v \in D\}$, then there exists $\{v_j\} \subset D$ such that $\hat{L}(v_j) \to M$. From our hypothesis $M > 0$. Notice that $\theta(\lambda) \to 0$ as $\lambda \to +\infty$. Hence from (2.4) $||v_j|| \leqq C_3\ \forall\, j$. The ball $\{v \in H(\Omega) : ||v|| \leqq C_3\}$ is weakly compact. Hence $\{v_j\}$ contains a weakly convergent subsequence which we also denote by $\{v_j\}$. Let v^* denote the weak limit of $\{v_j\}$. We have that $v_j(x_i) \to v^*(x_i)$ as $j \to \infty$ for each $i = 1, \cdots, N$. The norm is a continuous convex functional; hence weakly lower semicontinuous so that $\liminf ||v_j|| \geqq ||v^*||$. It follows that

$$\lim_j \prod_{i=1}^N v_j(x_i)\exp(-||v_j||^2) \leqq \prod_{i=1}^N v^*(x_i)\exp(-||v^*||^2)\,. \tag{2.5}$$

However the left-hand side of (2.5) is equal to M and the right-hand side is equal to $\hat{L}(v^*)$; so $M \leqq \hat{L}(v^*)$. Now since D is closed and convex it is weakly closed; hence $v^* \in D$. This establishes the existence of a maximizer.

Since $M > 0$, maximizing $\hat{L}$ over D is equivalent to maximizing $J = \log \hat{L}$ over D. A straightforward calculation gives the second Fréchet derivative of J as

$$J''(v)(\mu, \eta) = -\sum_{i=1}^N \frac{\mu(x_i)\eta(x_i)}{v(x_i)^2} - 2\langle \mu, \eta\rangle\,. \tag{2.6}$$

Now since $J''(v)$ is negative definite J is strictly concave and can therefore have at most one maximizer on a convex set.

THEOREM 2.1. *Suppose $H(\Omega)$ is a reproducing kernel Hilbert space, integration over Ω is a continuous functional and there exists at least one $v \in H(\Omega)$ satisfying* (2.3) *Then the maximum penalized likelihood estimate corresponding to $H(\Omega)$ exists and is unique.*

PROOF. The proof follows from Proposition 2.1 since the constraints in (2.2) give a closed convex subset of $\{v \in H(\Omega) : v(x_i) \geqq 0, i = 1, \cdots, N\}$.

Suppose (a, b) is a finite interval. For each integer $s \geqq 1$ we let $H_0^s(a, b)$ denote the Sobolev space of functions defined on $[a, b]$ whose first $s - 1$ derivatives are absolutely continuous and vanish at a and at b and whose sth derivative is in $L^2(a, b)$. The inner product in $H_0^s(a, b)$ is defined by

$$\langle \mu, v \rangle = \int_a^b \mu^{(s)}(t) v^{(s)}(t)\, dt\, . \tag{2.7}$$

It is well known that the space $H_0^s(a, b)$ is a Hilbert space with the inner product given by (2.7).

LEMMA 2.1. *The space $H_0^s(a, b)$ is a reproducing kernel Hilbert space and integration over (a, b) is a continuous operation.*

PROOF. Suppose $u, u_n \in H_0^s(a, b)$ and $u_n \to u$ in $H_0^s(a, b)$. By the Cauchy–Schwarz inequality in $L^2(a, b)$ we have for $x \in [a, b]$

$$|u(x) - u_n(x)| = |\int_a^x (u'(t) - u_n'(t)\, dt| \leqq (b - a)||u' - u_n'||_{L^2(a,b)}\, ; \tag{2.8}$$

hence point evaluation is a continuous operation. A straightforward integration by parts and the Cauchy–Schwarz inequality lead to

$$|\int_a^b [u(t) - u_n(t)]\, dt| \leqq \tfrac{1}{2}(b^2 - a^2)||u' - u_n'||_{L^2(a,b)}\, ; \tag{2.9}$$

hence integration over (a, b) is a continuous operation.

THEOREM 2.2. *The maximum penalized likelihood estimate corresponding to the Hilbert space $H_0^s(a, b)$ exists, is unique and is a polynomial spline (monospline) of degree $2s$. Moreover, if the estimate is positive in the interior of an interval, then in this interval it is a polynomial spline (monospline) of degree $2s$ and of continuity class $2s - 2$ with knots exactly at the sample points.*

PROOF. The existence and uniqueness are a consequence of Lemma 2.1 and Theorem 2.1, since clearly there exists at least one $v \in H_0^s(a, b)$ satisfying (2.3).

When no confusion can arise we will delete the variable of integration in definite integrals. Consider an interval $I_+ = [\alpha, \beta] \subset [a, b]$. Let $I = \{t \in [a, b]: t \notin [\alpha, \beta]\}$. Define the two functionals J_+ and J_- on $H_0^s(a, b)$ by

$$J_+(v) = \sum_i \log v(x_i) - \int_{I_+} [v^{(s)}]^2\, ,$$

and

$$J_-(v) = \sum_i \log v(x_i) - \int_{I_-} [v^{(s)}]^2\, ,$$

where the summation in the first formula is taken over all i such that $x_i \in I_+$

and the summation in the second formula is taken over all i such that $x_i \in I_-$. It should be clear that

$$J(v) = J_+(v) + J_-(v)$$

where as before $J(v) = \log \hat{L}(v)$ and $\hat{L}$ is the penalized likelihood in $H_0^s(a, b)$. Let v_* denote the maximum penalized likelihood estimate for the samples $x_1, \cdots, x_N$. Suppose v_* is positive on the interval I_+. We claim that v_* restricted to this interval solves the following constrained optimization problem:

$$\text{maximize} \quad J_+(v)\,; \quad \text{subject to}$$

$$(2.10) \qquad v \in H_0^s(\alpha, \beta)\,, \quad v^{(m)}(\alpha) = v_*^{(m)}(\alpha)\,, \quad v^{(m)}(\beta) = v_*^{(m)}(\beta)\,, \quad m = 0, \cdots, s-1\,,$$

$$\textstyle\int_{I_+} v = \int_{I_+} v_* \qquad \text{and} \qquad v(t) \geqq 0\,, \qquad t \in I_+\,.$$

To see this observe that if v_+ satisfies the constraints of problem (2.10) and $J_+(v_*) < J_+(v_+)$, then the function v^* defined by

$$v^*(t) = v_+(t)\,, \quad t \in I_+$$
$$\quad\;\; = v_*(t)\,, \quad t \in I_-$$

satisfies the constraints of problem (2.2) with $H_0^s(a, b)$ playing the role of $H(\Omega)$ and $J(v_*) = J_+(v_*) + J_-(v_*) < J_+(v_+) + J_-(v_+) = J(v^*)$, which in turn implies that $\hat{L}(v_*) < \hat{L}(v^*)$; however this contradicts the optimality of v^*. Now define the functional G on $H_0^s(\alpha, \beta)$ by

$$G(v) = J_+(v_* + v) \qquad \text{for} \quad v \in H_0^s(\alpha, \beta)\,.$$

Consider the constrained optimization problem

$$(2.11) \qquad \text{maximize} \quad G(v)\,; \quad \text{subject to}$$
$$v \in H_0^s(\alpha, \beta) \qquad \text{and} \qquad \textstyle\int_{I_+} v = 0\,.$$

If v satisfies the constraints of problem (2.11), then $v_* + tv$ satisfies the constraints of problem (2.10) for t sufficiently small, since v_* is positive in I_+. It follows that the zero function is the unique solution of problem (2.11). From the theory of Lagrange multipliers we therefore must have

$$(2.12) \qquad \nabla G(0) + \lambda v_0 = 0\,,$$

where λ is a real number, $\nabla G(0)$ is the Fréchet gradient of G at 0 and v_0 is the Fréchet gradient of the functional $v \to \int_{I_+} v$ in the space $H_0^s(\alpha, \beta)$. Clearly in this case v_0 is merely the Riesz representer of the functional $v \to \int_{I_+} v$. Specifically

$$\textstyle\int_{I_+} v_0^{(s)} v^{(s)} = \int_{I_+} v.$$

Integrating by parts in the distribution sense we see that $v_0^{(2s)} = 1$; hence v_0 is a polynomial of degree $2s$ in $[\alpha, \beta]$. A straightforward calculation shows that

$$(2.13) \qquad \nabla G(0) = \left(\sum_i \frac{v_i}{v_*(x_i)} - 2v_*\right)$$

where the summation is taken over i such that $x_i \in I_+$ and v_i is the Riesz representer of the functional $v \to v(x_i)$ in $H_0^s(\alpha, \beta)$, i.e.,

$$\int_{I_+} v_i^{(s)} v^{(s)} = v(x_i) .$$

As before integrating by parts in the distribution sense we see that $v_i^{(2s)} = \delta_i$ where δ_i is the Dirac mass at the point x_i. It follows that v_i is a polynomial spline of degree $2s - 1$ and of continuity class $2s - 2$ with a knot exactly at the sample point x_i. From (2.12) and (2.13) we have that v_* restricted to the interval $[\alpha, \beta]$ is a polynomial spline of degree $2s$ and of continuity class $2s - 2$ with knots exactly at the sample points in $[\alpha, \beta]$. A simple continuity argument takes care of the case when v_* is only positive on the interior of $[\alpha, \beta]$. Schoenberg (1968) defines a monospline to be the sum of a polynomial of degree $2s$ and a polynomial spline of degree $2s - 1$. This proves the theorem.

Before we analyze the Good and Gaskins estimates in Section 3 and Section 4 we must develop more background on Sobolev spaces. The reader desiring a more complete treatment is referred to Lions and Magenes (1968).

By the Sobolev space of order s on the real line we mean

(2.14) $$H^s(-\infty, \infty) = \{\mu \in S' : (1 + \omega^2)^{s/2} F[\mu](\omega) \in L^2(-\infty, \infty)\}$$

where S' is the space of distributions with polynomial decrease at infinity and $F[\mu]$ denotes the Fourier transform of μ. The norm of $\mu \in H^s(-\infty, \infty)$ is given by

(2.15) $$\|\mu\|_{H^s(-\infty,\infty)} = \|(1 + \omega^2)^{s/2} F[\mu](\omega)\|_{L^2(-\infty,\infty)} .$$

If s is an integer, then $\mu \in H^s(-\infty, \infty)$ if and only if $\mu, \mu^{(1)}, \cdots, \mu^{(s)} \in L^2(-\infty, \infty)$ and an equivalent norm is given by

(2.16) $$[\sum_{i=0}^{s} w_i \|\mu^{(i)}\|^2_{L_2(-\infty,\infty)}]^{\frac{1}{2}}$$

where $w_i \geqq 0$ and $w_0, w_s > 0$.

LEMMA 2.2. *The Sobolev space* $H^s(-\infty, \infty)$ *is a reproducing kernel Hilbert space if and only if* $s > \frac{1}{2}$.

PROOF. The dual of H^s is H^{-s}. A reproducing kernel Hilbert space is a space such that the Dirac distributions are in the dual; hence we want

$$(1 + \omega^2)^{-s/2} F(\delta_x) \in L^2(-\infty, \infty)$$

where δ_x is the Dirac distribution at the point x. Since the Fourier image of a Dirac mass if a constant we must have

$$(1 + \omega^2)^{-s/2} \in L^2(-\infty, \infty) .$$

This proves the lemma.

3. The first maximum penalized likelihood estimator of Good and Gaskins. Motivated by information theoretic considerations Good and Gaskins (1971) consider the maximum penalized likelihood estimate corresponding to the penalty function

$$\Phi_1(v) = \alpha \int_{-\infty}^{\infty} \frac{v'(t)^2}{v(t)}\, dt \quad (\alpha > 0) .$$

They do not define the manifold $H(\Omega)$; but it is obvious from the constraints that must be satisfied and the fact that

$$\tfrac{1}{4}\Phi_1(v) = \alpha \int_{-\infty}^{\infty} \left(\frac{dv^{\frac{1}{2}}}{dt}\right)^2 dt$$

what the underlying manifold $H(\Omega)$ should be, namely $v^{\frac{1}{2}} \in H^1(-\infty, \infty)$. This leads us to analyzing the following constrained optimization problem:

$$\text{(3.1)} \quad \begin{gathered} \text{maximize} \quad L_1(v) = \prod_{i=1}^{N} v(x_i) \exp(-\Phi_1(v)) \,; \quad \text{subject to} \\ v^{\frac{1}{2}} \in H^1(-\infty, \infty) \,, \quad \int_{-\infty}^{\infty} v(t)\, dt = 1 \quad \text{and} \quad v(t) \geqq 0 \\ \forall\, t \in (-\infty, \infty) . \end{gathered}$$

In an effort to avoid the nonnegativity constraint in problem (3.1) Good and Gaskins considered working with the $v^{\frac{1}{2}}$ instead of v. Specifically if we let $u = v^{\frac{1}{2}}$, then restating problem (3.1) in terms of u we obtain

$$\text{(3.2)} \quad \begin{gathered} \text{maximize} \quad \prod_{i=1}^{N} u(x_i)^2 \exp(-4\alpha \int_{-\infty}^{\infty} u'(t)^2\, dt) \,; \quad \text{subject to} \\ u \in H^1(-\infty, \infty) \quad \text{and} \quad \int_{-\infty}^{\infty} u(t)^2\, dt = 1 . \end{gathered}$$

Problem (3.2) is solved for u^* and then $v^* = (u^*)^2$ is accepted as the solution to problem (3.1). This seemingly clever trick is somewhat standard in the literature. The following lemma tells us when this trick can be used.

LEMMA 3.1. *Let H be a subset of $L^2(\Omega)$ and J a functional defined on H. Consider Problem* I

$$\begin{gathered} \textit{maximize} \quad J(v^{\frac{1}{2}}) \,; \quad \textit{subject to} \\ v^{\frac{1}{2}} \in H \,, \quad \int_{\Omega} v(t)\, dt = 1 \quad \textit{and} \quad v(t) \geqq 0 \qquad \forall\, t \in \Omega \end{gathered}$$

and Problem II

$$\begin{gathered} \textit{maximize} \quad J(u) \,; \quad \textit{subject to} \\ u \in H \quad \textit{and} \quad \int_{\Omega} u(t)^2\, dt = 1 . \end{gathered}$$

Let u^* be a solution of Problem II. Then $v^* = (u^*)^2$ solves Problem I if and only if $|u^*| \in H$ and $J(u^*) = J(|u^*|)$.

PROOF. Observe that $v^* = |u^*|$ hence if $|u^*| \notin H$, then $J(v^*)^{\frac{1}{2}}$ is not defined and meaningless. While if $|u^*| \in H$ and $J(u^*) = J(|u^*|)$, then for any $v \geqq 0$ such that $v^{\frac{1}{2}} \in H$ we have

$$J(v^{\frac{1}{2}}) \leqq J(u^*) = J(|u^*|) = J(v^*)^{\frac{1}{2}} .$$

Now noticing that the proper constraints are always satisfied we have the lemma.

Two points are immediately of interest. The first being that the conditions of the lemma clearly hold when u^*, the solution to Problem II, is nonnegative. The second being that the space $H^1(-\infty, \infty)$ and the function $\hat{L}$ in problems (3.1) and (3.2) satisfy the conditions of the lemma. Hence Good's and

Gaskins' alternate approach gives the correct estimate in this case; however in their other case (analyzed in Section 4) this is unfortunately not true.

Problem (3.2) cannot possibly have a unique solution. To see this notice that if u^* is a solution, then so is $-u^*$. Adding the nonnegativity constraint to problem (3.2) and restating in the form obtained by taking the square root of the objective functional (since it is nonnegative) we arrive at the following constrained optimization problem:

$$\text{(3.3)} \qquad \text{maximum} \quad \hat{L}(v) = \prod_{i=1}^{N} v(x_i) \exp(-\Phi(v)); \quad \text{subject to}$$
$$v \in H^1(-\infty, \infty), \qquad \int_{-\infty}^{\infty} v(t)^2\, dt = 1 \quad \text{and} \quad v(t) \geqq 0, \quad \forall\, t \in (-\infty, \infty)$$

where

$$\Phi(v) = 2\alpha \int_{-\infty}^{\infty} v'(t)^2\, dt$$

and α is given in problem (3.1).

PROPOSITION 3.1.

(i) *If v solves problem* (3.1), *then $v^{\frac{1}{2}}$ solves problem* (3.2) *and problem* (3.3).

(ii) *If u solves problem* (3.2), *then $|u|$ solves problem* (3.3) *and u^2 solves problem* (3.1).

(iii) *If v solves problem* (3.3), *then v solves problem* (3.2) *and v^2 solves problem* (3.1).

PROOF. The proof follows from Lemma 3.1 and the fact that if $v \geqq 0$, then

$$\Phi(v^{\frac{1}{2}}) = \tfrac{1}{2}\Phi_1(v)$$

and

$$\hat{L}_1(v) = \hat{L}(v^{\frac{1}{2}})^2 .$$

COROLLARY 3.1. *If problem* (3.3) *has a unique solution, then problem* (3.1) *has a unique solution; and although problem* (3.2) *cannot have a unique solution, it will have solutions and the square of any of these solutions will give the unique solution of problem* (3.1).

The remainder of this section is dedicated to demonstrating that problem (3.3) has a unique solution which is a positive exponential spline with knots only at the sample points. The same will then be true of Good's and Gaskins' first maximum penalized estimate.

Along with problem (3.3) we will consider the constrained optimization problem obtained by only requiring nonnegativity at the sample points:

$$\text{maximize} \quad \hat{L}(v); \quad \text{subject to}$$
$$\text{(3.4)} \qquad v \in H^1(-\infty, \infty), \qquad \int_{-\infty}^{\infty} v(t)^2\, dt = 1 \quad \text{and} \quad v(x_i) \geqq 0, \quad i = 1, \cdots, N.$$

Given $\lambda > 0$ and α in problem (3.3) we may also consider the constrained optimization problem:

$$\text{(3.5)}\qquad \begin{aligned} &\text{maximum}\quad \hat{L}_\lambda(v) = \textstyle\prod_{i=1}^N v(x_i)\exp(-\Phi_\lambda(v))\,; \quad \text{subject to} \\ &v \in H^1(-\infty, \infty)\,, \qquad \textstyle\int_{-\infty}^{\infty} v(t)^2\,dt = 1 \quad \text{and} \quad v(x_i) \geqq 0\,, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1, \cdots, N \end{aligned}$$

where

$$\Phi_\lambda(v) = 2\alpha \textstyle\int_{-\infty}^{\infty} v'(t)^2\,dt + \lambda \int_{-\infty}^{\infty} v(t)^2\,dt\,.$$

Our study of problem (3.5) will begin with the study of the following constrained optimization problem:

$$\text{(3.6)}\qquad \begin{aligned} &\text{maximize}\quad \hat{L}_\lambda(v)\,; \quad \text{subject to} \\ &v \in H^1(-\infty, \infty) \qquad \text{and} \qquad v(x_i) \geqq 0\,, \qquad i = 1, \cdots, N \end{aligned}$$

where $\hat{L}_\lambda$ is given by problem (3.5). Let $L^2 = L^2(-\infty, \infty)$.

PROPOSITION 3.2. *Problem* (3.6) *has a unique solution. Moreover if* v_λ *denotes this solution, then*

(i) v_λ *is a exponential spline with knots at the sample points* $x_1, \cdots, x_N$;
(ii) $v_\lambda(t) > 0$, $\forall\, t \in (-\infty, \infty)$; *and*
(iii) $||v_\lambda||_{L^2} \geqq (N/(4\lambda))^{\frac{1}{2}}$.

PROOF. From Lemma 2.2 $H^1(-\infty, \infty)$ is a reproducing kernel space. Also $||v||_\lambda^2 = \Phi_\lambda(v)$ gives a norm equivalent to the original norm on $H^1(-\infty, \infty)$. The existence of v_λ now follows from Proposition 2.1 with $D = \{v \in H^1(-\infty, \infty)\colon v(x_i) \geqq 0, i = 1, \cdots, N\}$. We will denote the Φ_λ inner product by $\langle\ ,\ \rangle_\lambda$. Let v_i be the representer in the Φ_λ inner product of the continuous linear functional given by point evaluation at the point x_i, $i = 1, \cdots, N$, i.e.

$$\langle v_i, \eta\rangle_\lambda = \eta(x_i)\,, \qquad \forall\, \eta \in H^1(-\infty, \infty)\,.$$

Equivalently

$$2\alpha \textstyle\int_{-\infty}^{\infty} v_i'(t)\eta'(t)\,dt + \lambda \int_{-\infty}^{\infty} v_i(t)\eta(t)\,dt = \eta(x_i)\,, \qquad \forall\, \eta \in H^1(-\infty, \infty)\,.$$

Integrating by parts in the distribution sense gives

$$\textstyle\int_{-\infty}^{\infty} [-2\alpha v_i''(t) + \lambda v_i(t)]\eta(t)\,dt = \eta(x_i)\,, \qquad \forall\, \eta \in H^1(-\infty, \infty)\,;$$

hence

$$\text{(3.7)}\qquad -2\alpha v_i'' + \lambda v_i = \delta_i\,, \qquad i = 1, \cdots, N$$

where $\delta_i(t) = \delta_0(t - x_i)$ and δ_0 denotes the Dirac distribution, i.e., $\int_{-\infty}^{\infty} \delta_0(t)\eta(t)\,dt = \eta(0)$. If we let v_0 be the solution of (3.7) for $i = 0$, then

$$\begin{aligned} v_0(t) &= \frac{1}{2(2\alpha\lambda)^{\frac{1}{2}}} \exp((\lambda(2\alpha))^{\frac{1}{2}} t)\,, \quad t < 0 \\ &= \frac{1}{2(2\alpha\lambda)^{\frac{1}{2}}} \exp(-(\lambda/(2\alpha))^{\frac{1}{2}} t)\,, \quad t > 0 \end{aligned}$$

and $v_i(t) = v_0(t - x_i)$ for $i = 1, \cdots, N$. Since v_λ is the maximizer we have that $v_\lambda(x_i) > 0$, $i = 1, \cdots, N$ we necessarily have that the Fréchet derivative of $\hat{L}_\lambda$

at v_λ must be the zero functional; equivalently the gradient of $\hat{L}_\lambda$ or for that matter the gradient of $\log \hat{L}_\lambda$ must vanish at v_λ since $\hat{L}_\lambda$ and $\log \hat{L}_\lambda$ have the same maxima. A calculation similar to that used in the proof of Proposition 2.1 gives

$$\nabla_\lambda \log \hat{L}_\lambda(v) = 2v - \sum_{i=1}^N \frac{v_i}{v(x_i)} \tag{3.8}$$

where ∇_λ denotes the gradient. It follows from (3.8) that

$$v_\lambda = \tfrac{1}{2} \sum_{i=1}^N \frac{v_i}{v_\lambda(x_i)} . \tag{3.9}$$

Properties (i) and (ii) are now immediate. Since $\langle v_i, v_\lambda \rangle_\lambda = v_\lambda(x_i)$ from (3.9) we have

$$||v_\lambda||_\lambda^2 = N/2 . \tag{3.10}$$

A straightforward calculation shows that

$$v_i'(t)v_j'(t) \leqq \frac{\lambda}{2\alpha} v_i(t)v_j(t) , \qquad \text{for } i, j = 1, \cdots, N .$$

So

$$\begin{aligned} v_\lambda'(t)^2 &= \frac{1}{4}\left[\sum_i \left(\frac{v_i'(t)}{v_\lambda(x_i)}\right)^2 + \sum_{i,j} \frac{v_i'(t)v_j'(t)}{v_\lambda(x_i)v_\lambda(x_j)}\right] \\ &\leqq \frac{\lambda}{8\alpha}\left[\sum_i \left(\frac{v_i(t)}{v_\lambda(x_i)}\right)^2 + \sum_{i,j} \frac{v_i(t)v_j(t)}{v_\lambda(x_i)v_\lambda(x_j)}\right] = \frac{\lambda}{2\alpha} v_\lambda(t)^2 . \end{aligned}$$

Integrating in t gives

$$2\alpha ||v_\lambda'||_{L^2(-\infty,\infty)}^2 \leqq \lambda ||v_\lambda||_{L^2(-\infty,\infty)}^2 .$$

By definition of the Φ_λ-norm and (3.10) we have property (iii). This proves the proposition.

PROPOSITION 3.3. *Problem* (3.4) *has a unique solution.*

PROOF. Let $B = \{v \in H^1(-\infty, \infty) : \int_{-\infty}^\infty v(t)^2\, dt \leqq 1 \text{ and } v(x_i) \geqq 0, i = 1, \cdots, N\}$. Clearly B is closed and convex. If $\hat{L}_\lambda$ is given by (3.5), then by Proposition 2.1 the functional has a unique maximizer in B; say u_λ. Now by property (iii) of Proposition 3.2 if we choose $0 < \lambda < \frac{1}{4}$, then v_λ the unique solution of problem (3.6) will be such that $||v_\lambda||_{L^2(-\infty,\infty)} > 1$. We will show that for this range of λ, $||u_\lambda||_{L^2(-\infty,\infty)} = 1$. Consider $v_\theta = \theta v_\lambda + (1 - \theta)u_\lambda$. We know that $\log \hat{L}_\lambda$ is a strictly concave functional (see the proof of Proposition 2.1). Moreover $\log \hat{L}_\lambda(v_\lambda) \geqq \log \hat{L}_\lambda(u_\lambda)$; hence $\log \hat{L}_\lambda(v_\theta) \geqq \log \hat{L}_\lambda(u_\lambda)$ for $0 < \theta < 1$. Now suppose $||u_\lambda||_{L^2(-\infty,\infty)} < 1$ and consider

$$g(\theta) = ||v_\theta||_{L^2(-\infty,\infty)} .$$

We have $g(0) < 1$ and $g(1) > 1$. So for some $0 < \theta_0 < 1$, $g(\theta_0) = 1$ and $\log \hat{L}_\lambda(u_\lambda) \leqq \log \hat{L}_\lambda(v_{\theta_0})$. This is a contradiction since u_λ is the unique maximizer of $\hat{L}_\lambda$ in B; hence $||u_\lambda||_{L^2(-\infty,\infty)} = 1$. This shows that u_λ is the unique solution of problem (3.5) for $0 < \lambda < \frac{1}{4}$. However, the term $\lambda \int_{-\infty}^{\infty} v(t)^2 \, dt$ is constant over the constraint set in problems (3.4) and (3.5); hence problems (3.4) and (3.5) have the same solutions for any $\lambda > 0$. This proves the proposition since we have demonstrated that problem (3.3) has a unique solution for at least one λ.

PROPOSITION 3.4. *Problem* (3.3) *has a unique solution which is positive and an exponential spline with knots at the points* $x_1, \cdots, x_N$.

PROOF. If we can demonstrate that $\tilde{v}$ the unique solution of problem (3.4) has these properties we will be through. Let $G(v) = \log \hat{L}(v)$ where $\hat{L}$ is given in problem (3.3) and let

$$g(v) = \int_{-\infty}^{\infty} v(t)^2 \, dt$$

for $v \in H^1(-\infty, \infty)$. Clearly $\tilde{v}(x_i) > 0$ for $i = 1, \cdots, N$; hance from the theory of Lagrange multipliers there exist λ such that $\tilde{v}$ satisfies the equations

$$G'(v) - \lambda g'(v) = 0 \quad \text{and} \quad g(v) = 1 \,. \tag{3.11}$$

Using $L^2(-\infty, \infty)$ gradients in the sense of distributions (3.11) is equivalent to

$$-4\alpha v'' + 2\lambda v = \sum_{i=1}^{N} \frac{\delta_i}{v(x_i)} \quad \text{and} \quad g(v) = 1 \tag{3.12}$$

where δ_i is the distribution such that $\int_{-\infty}^{\infty} v(t)\delta_i(t) \, dt = v(x_i)$, $i = 1, \cdots, N$. Since we have already established that problem (3.4) has a unique solution if follows that (3.12) must have a unique solution in $H^1(-\infty, \infty)$; namely $\tilde{v}$. If $\lambda \leqq 0$, then any solution of the first equation in (3.12) would be a sum of trigonometric functions and could not possibly satisfy the constraint $g(v) = 1$, i.e., cannot be contained in $L^2(-\infty, \infty)$. It follows that $\lambda > 0$. Now observe that

$$G - \lambda g = \log \hat{L}_\lambda$$

where $\hat{L}_\lambda$ is given by problem (3.5); hence if $\tilde{v}$ satisfies (3.11) (from the first equation alone) it must also be a solution of problem (3.6) for this λ and therefore has the desired properties according to Proposition 3.2. This proves the proposition.

PROPOSITION 3.5. *The first nonparametric maximum penalized likelihood estimate of Good and Gaskins exists and is unique*; *specifically the maximum penalized likelihood estimate corresponding to the penalty function*

$$\Phi(v) = \alpha \int_{-\infty}^{\infty} \frac{v'(t)^2}{v(t)} \, dt \quad (\alpha > 0)$$

and the manifold

$$H(\Omega) = \{v : v \geqq 0 \ \ \textit{and} \ \ v^{\frac{1}{2}} \in H^1(-\infty, \infty)\}$$

exists and is unique. Moreover the estimate is positive and an exponential spline with knots only at the sample points.

PROOF. The proof follows from Proposition 3.1 and Proposition 3.5.

4. The second maximum penalized likelihood estimator of Good and Gaskins. Consider the functional $\Phi: H^2(-\infty, \infty) \to R$ defined by

$$\Phi(v) = \alpha \int_{-\infty}^{\infty} v'(t)^2 \, dt + \beta \int_{-\infty}^{\infty} v''(t)^2 \, dt \tag{4.1}$$

for some $\alpha \geqq 0$ and $\beta > 0$. By a second maximum penalized likelihood estimate of Good and Gaskins we mean any solution of the following constrained optimization problem:

$$\begin{gathered} \text{maximize} \quad \hat{L}_1(v) = \prod_{i=1}^{N} v(x_i) \exp(-\Phi(v^{\frac{1}{2}})) ; \quad \text{subject to} \\ v^{\frac{1}{2}} \in H^2(-\infty, \infty), \quad \int_{-\infty}^{\infty} v(t)\, dt = 1 \quad \text{and} \quad v(t) \geqq 0 \quad \forall t \in (-\infty, \infty) . \end{gathered} \tag{4.2}$$

As in the first case (described in the previous section) Good and Gaskins suggest avoiding the nonnegativity constraint by calculating the solution of problem (4.2) from the following constrained optimization problem:

$$\begin{gathered} \text{maximize} \quad \prod_{i=1}^{N} v(x_i)^2 \exp(-\Phi(v)) ; \quad \text{subject to} \\ v \in H^2(-\infty, \infty) \quad \text{and} \quad \int_{-\infty}^{\infty} v(t)^2 \, dt = 1 \end{gathered} \tag{4.3}$$

where Φ is given by (4.1).

Notice that if $u \in H^2(-\infty, \infty)$ is positive at some points and negative at other points, then $|u|$ is in general not a member of $H^2(-\infty, \infty)$, since in general its derivative is not continuous. It follows from Lemma 3.1 that the only way we can obtain solutions to problem (4.2) from solutions of problem (4.3) is for problem (4.3) to have solutions which are completely of one sign. However we will presently show that this is not the case.

Along with problem (4.3) we consider the constrained optimization problem:

$$\begin{gathered} \text{maximize} \quad \hat{L}(v) = \prod_{i=1}^{N} v(x_i) \exp(-\tfrac{1}{2}\Phi(v)) ; \quad \text{subject to} \\ v \in H^2(-\infty, \infty), \quad \int_{-\infty}^{\infty} v(t)^2 \, dt = 1 \quad \text{and} \quad v(x_i) \geqq 0 , \\ i = 1, \cdots, N . \end{gathered} \tag{4.4}$$

Problem (4.4) was obtained from problem (4.3) by taking the square root of the functional to be maximized (since it is nonnegative) and requiring nonnegativity at the sample points; hence the two problems only differ by the nonnegativity constraints at the sample points. This simple difference will allow us to establish uniqueness of the solution of problem (4.4); whereas problem (4.3) cannot have a unique solution. By Lemma 3.2 we can obtain the solution of problem (4.2) from the solution of problem (4.4) if and only if these solutions are nonnegative. Moreover we will presently demonstrate that the solutions of problem (4.4) are not necessarily nonnegative. It will then follow that we cannot obtain the second estimate by considering problem (4.4). If we naively use v_*^2, where v_* solves problem (4.4), as an estimate for the probability density function giving rise to the random sample $x_1, \cdots, x_N$, then clearly v_*^2 will be nonnegative and integrate to one and is therefore a probability density; however the estimate obtained in this manner will not in the strict sense of our definition be a maximum penalized likelihood estimate. For this reason we will refer to this latter estimate as the *pseudo maximum penalized likelihood estimate* of Good and Gaskins.

The next six propositions are needed to show that the second maximum penalized likelihood estimate and the pseudo maximum penalized likelihood estimate of Good and Gaskins exist, are unique and are distinct, and that Good's and Gaskins' alternate approach cannot be used to obtain their second maximum penalized likelihood estimate.

PROPOSITION 4.1. *The second maximum penalized likelihood estimate and the pseudo maximum likelihood estimate of Good and Gaskins are distinct.*

PROOF. We will show that it is possible for problem (4.4) to have solutions which are not nonnegative. Toward this end let $N = 1$, $x_1 = 0$, $\alpha = 0$, and $\beta = 2$. Let $G(v) = \log \hat{L}(v)$, i.e.,

$$G(v) = \log v(0) - \int_{-\infty}^{\infty} v''(t)^2\, dt$$

and let

$$g(v) = \int_{-\infty}^{\infty} v(t)^2\, dt\,.$$

As in the proof of Proposition 3.4 using the theory of distributions and the theory of Lagrange multipliers we see that the solutions of problem (4.4) in this case are exactly the solutions of

$$v^{(iv)} + \lambda v = \frac{\delta_1}{2v(0)} \qquad \text{and} \qquad g(v) = 1 \tag{4.5}$$

where δ_1 is defined in the proof of Proposition 3.4. If we let $\tilde{v}$ denote the Fourier transform of v, then taking the Fourier transform of the first expression in (4.5) gives

$$\tilde{v}(\omega) = [2v(0)(\lambda + 16\Pi^4\omega^4)]^{-1}\,.$$

Since $||\tilde{v}||_{L^2(-\infty,\infty)} = ||v||_{L^2(-\infty,\infty)} = 1$ we must have

$$\int_{-\infty}^{\infty} \frac{d\omega}{(\lambda + 16\Pi^4\omega^4)^2} = 4v(0)^2\,. \tag{4.6}$$

For the integral in (4.6) to exist we must have $\lambda > 0$. Now the inverse Fourier transform of $(\lambda + 16\Pi^4\omega^4)^{-1}$ is given by v where

$$\begin{aligned} v(t) &= \frac{e^{bt}}{8b^3}[\cos bt - \sin bt]\,, \quad t \leqq 0 \\ &= \frac{e^{-bt}}{8b^3}(\cos bt + \sin bt]\,, \quad t > 0 \end{aligned} \tag{4.7}$$

with $b = \lambda^{\frac{1}{4}}/2^{\frac{1}{2}}$. From (4.7) $v(0) = (8b^3)^{-1}$ and from (4.6) $v(0)^2 = \frac{1}{4}\lambda^{-\frac{3}{4}}K$ where $K = ||(1 + 16\Pi^4\omega^4)^{-1}||_{L^2(-\infty,\infty)}$. Hence $\lambda^{\frac{3}{4}} = 2K$ and $b = 2^{\frac{1}{2}}K$ which is clearly not nonnegative. This proves the proposition.

In the above proof we assumed the existence of a solution. At that point we did not know that this was so; however our proof is valid since in Proposition 4.4 we will show that problem (4.4) always has a unique solution.

COROLLARY 4.1. *Problem* (4.3) *has solutions which are not nonnegative.*

PROOF. The solution constructed in the proof of Proposition 4.1 also solves problem (4.3). To see this observe that if v satisfies the constraints of problem (4.3), then so does $-v$ and the functional values are the same. Also either v or $-v$ will satisfy the constraints of problem (4.4) since in the present case there is only one sample.

COROLLARY 4.2. *Good's and Gaskins' alternate approach cannot be used to obtain their second maximum penalized likelihood estimate.*

Given $\lambda > 0$ consider the constrained optimization problem:

$$\text{maximize} \quad \hat{L}_\lambda(v) = \prod_{i=1}^N v(x_i)\exp(-\Phi_\lambda(v))\,; \quad \text{subject to}$$
$$v \in H^2(-\infty, \infty)\,, \qquad \int_{-\infty}^{\infty} v(t)^2\,dt = 1 \quad \text{and} \quad v(x_i) \geqq 0\,, \tag{4.8}$$
$$i = 1, \cdots, N\,.$$

where

$$\Phi_\lambda(v) = \Phi(v) + \lambda \int_{-\infty}^{\infty} v(t)^2\,dt$$

with $\Phi(v)$ given by (4.1).

As before we also consider the constrained optimization problem obtained by dropping the integral constraint:

$$\text{maximize} \quad \hat{L}_\lambda(v)\,; \quad \text{subject to} \tag{4.9}$$
$$v \in H^2(-\infty, \infty) \qquad \text{and} \qquad v(x_i) \geqq 0\,, \qquad i = 1, \cdots, N\,.$$

PROPOSITION 4.2. *Problem* (4.9) *has a unique solution. Moreover if* v_λ *denotes this solution, then*

$$||v_\lambda||_{L^2(-\infty,\infty)} \to +\infty \qquad \textit{as} \quad \lambda \to 0\,.$$

PROOF. By Lemma 2.1 the Sobolev space $H^2(-\infty, \infty)$ is a reproducing kernel space. Moreover, if

$$||v||_\lambda^2 = \Phi_\lambda(v)\,,$$

then an integration by parts gives

$$\begin{aligned} ||v'||_{L^2}^2 &= |\langle v, v''\rangle_{L^2}| \leqq ||v||_{L^2}||v''||_{L^2} \\ &\leqq \tfrac{1}{2}[||v||_{L^2}^2 + ||v''||_{L^2}^2] \end{aligned} \tag{4.10}$$

where L^2 denotes $L^2(-\infty, \infty)$; hence $||\cdot||_\lambda$ is equivalent to the original norm on $H^2(-\infty, \infty)$. The existence and uniqueness of v_λ now follows from Proposition 2.1.

We must now show that $||v_\lambda||_{L^2} \to +\infty$ as $\lambda \to 0$. From the fundamental theorem of calculus we have

$$\begin{aligned} v(x)^2 &= \int_{-\infty}^{x} \frac{dv(t)^2}{dt}\,dt = 2\int_{-\infty}^{x} v(t)v'(t)\,dt \\ &\leqq 2||v||_{L^2}||v'||_{L^2}\,. \end{aligned} \tag{4.11}$$

Also, $||v''||_{L^2} \leqq ||v||_\lambda/\beta^{\frac{1}{2}}$ so that from (4.10) and (4.11)

(4.12) $$v(x)^2 \leqq 2||v||_{L^2}^{\frac{3}{2}}(||v||_\lambda/\beta^{\frac{1}{2}}) .$$

Evaluating (4.12) at x_i, taking logs (since $v(x_i) \geqq 0$) and summing over i gives

(4.13) $$\sum_{i=1}^N \log v(x_i) \leqq \frac{N}{4} \log \left(\frac{4}{\beta^{\frac{1}{2}}} ||v||_\lambda\right) + \frac{3N}{4} \log (||v||_{L^2}) .$$

Hence from (4.13) we see that

(4.14) $$\log \hat{L}_\lambda(v) \leqq \frac{3N}{4} \log (||v||_{L^2}) + \frac{N}{4} \log \left(\frac{4}{\beta^{\frac{1}{2}}} ||v||_\lambda\right) - ||v||_\lambda^2 .$$

In a manner exactly the same as that used to establish (3.10) we have that $||v_\lambda||_\lambda^2 = N/2$. Hence from (4.14) and the fact that $\log \hat{L}_\lambda(v) \leqq \log \hat{L}_\lambda(v_\lambda)$ we obtain

(4.15) $$\log \hat{L}_\lambda(v) \leqq \frac{3N}{4} \log (||v_\lambda||_{L^2}) + \frac{N}{8} \log (8N/\beta) - \frac{N}{2} ,$$

for any $v \in \{u \in H^2(-\infty, \infty) : u(x_i) \geqq 0, i = 1, \cdots, N\}$.

Let a and b be such that

$$a < \min_i (x_i) \qquad \text{and} \qquad \max_i (x_i) < b .$$

Given $\lambda > 0$ and ε and δ define the function θ_λ in the following piecewise fashion:

$$\begin{aligned} \theta_\lambda(t) &= \lambda^\varepsilon \exp(-(t-a)^2/2\sigma^2) && \text{for } t \in (-\infty, a) \\ &= \lambda^\varepsilon && \text{for } t \in [a, b] \\ &= \lambda^\varepsilon \exp(-(t-b)^2/2\sigma^2) && \text{for } t \in (b, +\infty) \end{aligned}$$

where $\sigma = \lambda^\delta$. Straightforward calculations can be used to show

$$\begin{aligned} \log (\textstyle\prod_{i=1}^N \theta_\lambda(x_i)) &= \varepsilon N \log (\lambda) , \\ ||\theta_\lambda||_{L^2}^2 &= (b-a)\lambda^{2\varepsilon} + ((\Pi\lambda)^{\frac{1}{2}})^{2\varepsilon+\delta} \\ ||\theta_\lambda'||_{L^2}^2 &= ((2\Pi\lambda)^{\frac{1}{2}})^{2\varepsilon-\delta} , \\ ||\theta_\lambda''||_{L^2}^2 &= 2((2\Pi\lambda)^{\frac{1}{2}})^{2\varepsilon-3\delta} , \end{aligned}$$

and

(4.16) $$||\theta_\lambda||_\lambda^2 = (b-a)\lambda^{2\varepsilon+1} + ((\Pi\lambda)^{\frac{1}{2}})^{2\varepsilon+\delta+1} + 4\alpha((2\Pi\lambda)^{\frac{1}{2}})^{2\varepsilon-\delta} + 2\beta((2\Pi\lambda)^{\frac{1}{2}})^{2\varepsilon-3\delta} .$$

If we want $||\theta_\lambda||_\lambda^2 \to 0$ as $\lambda \to 0$ it is sufficient to choose all exponents of λ in (4.16) positive. If we also want

$$\log (\textstyle\prod_{i=1}^N \theta_\lambda(x_i)) \to +\infty \qquad \text{as} \quad \lambda \to 0$$

we should choose $\varepsilon < 0$. This leads to the inequalities

(4.17) $$\begin{aligned} 2\varepsilon + 1 &> 0 \\ 2\varepsilon + \delta + 1 &> 0 \\ 2\varepsilon - \delta &> 0 \\ 2\varepsilon - 3\delta &> 0 \\ \varepsilon &< 0 . \end{aligned}$$

The system of inequalities (4.17) has solutions; specifically $\varepsilon = -\frac{1}{32}$ and $\delta = -\frac{1}{8}$ is one such solution. With this choice of ε and δ we see that $\log \hat{L}_\lambda(\theta_\lambda) \to +\infty$ as $\lambda \to 0$. It follows from (4.15) by choosing $v = \theta_\lambda$ that $||v_\lambda||_{L^2} \to +\infty$ as $\lambda \to 0$. This proves the proposition.

PROPOSITION 4.3. *Problem* (4.3) *has a unique solution.*

PROOF. By Proposition 4.2 there exists $\lambda > 0$ such that if v_λ is the unique solution of problem (4.9), then $||v_\lambda||_{L^2} > 1$. Now, if $B = \{v \in H^2(-\infty, \infty): \int_{-\infty}^{\infty} v(t)^2\, dt \leqq 1$ and $v(x_i) \geqq 0,\ i = 1, \cdots, N\}$, then B is closed and convex. The proof of the proposition is now exactly the same as the proof of Proposition 3.3.

PROPOSITION 4.4. *The pseudo maximum penalized likelihood estimate of Good and Gaskins exists and unique.*

PROOF. Since problems (4.4) and (4.8) have the same solutions the proposition follows from Proposition 4.3.

By the change of unknown function $v \to v^{\frac{1}{2}}$ we see that problem (4.2) is equivalent to the following constrained optimization problem:

$$\text{(4.18)}\qquad \begin{aligned} &\text{maximize}\quad \hat{L}(v) = \textstyle\prod_{i=1}^{N} v(x_i) \exp(-\frac{1}{2}\Phi(v))\,; \quad \text{subject to} \\ &v \in H^2(-\infty, \infty)\,, \quad \textstyle\int_{-\infty}^{\infty} v(t)^2\, dt = 1 \quad \text{and} \quad v(t) \geqq 0 \quad \forall\, t \in (-\infty, \infty) \end{aligned}$$

where $\Phi(v)$ is given by (4.1).

In turn for $\lambda > 0$ problem (4.18) is equivalent to

$$\text{(4.19)}\qquad \begin{aligned} &\text{maximize}\quad \hat{L}_\lambda(v)\,; \quad \text{subject to} \\ &v \in H^2(-\infty, \infty)\,, \quad \textstyle\int_{-\infty}^{\infty} v(t)^2\, dt = 1 \quad \text{and} \quad v(t) \geqq 0 \quad \forall\, t \in (-\infty, \infty) \end{aligned}$$

where $\hat{L}_\lambda$ is defined in problem (4.8).

As in the previous two cases we also consider the constrained optimization problem:

$$\text{(4.20)}\qquad \begin{aligned} &\text{maximize}\quad \hat{L}_\lambda(v)\,; \quad \text{subject to} \\ &v \in H^2(-\infty, \infty) \quad \text{and} \quad v(t) \geqq 0 \quad \forall\, t \in (-\infty, \infty) \end{aligned}$$

where $\hat{L}_\lambda(v)$ is defined in problem (4.8).

PROPOSITION 4.5. *Problem* (4.20) *has a unique solution. Moreover if v_λ^+ denotes this solution, then*

$$||v_\lambda^+||_{L^2} \to +\infty \qquad \textit{as} \quad \lambda \to 0\,.$$

PROOF. The existence of v_λ^+ follows from Proposition 2.1 as in the proof of Proposition 4.2. Let us first show that

$$\text{(4.21)}\qquad ||v_\lambda^+||_\lambda \leqq (N/2)^{\frac{1}{2}}\,.$$

From Lions (1968) we see that

$$\text{(4.22)}\qquad \hat{L}_\lambda'(v_\lambda^+)(\eta - v_\lambda^+) \leqq 0$$

for all nonnegative η in $H^2(-\infty, \infty)$. We have

$$\hat{L}_\lambda'(v)(\eta) = \sum_{i=1}^N \frac{\eta(x_i)}{v(x_i)} - 2\langle v, \eta\rangle_\lambda ;$$

hence

$$\hat{L}_\lambda'(v_\lambda^+)(v_\lambda^+) = N - 2\|v_\lambda\|_\lambda^2 . \tag{4.23}$$

Now choosing $\eta = 0$ in (4.22) and using (4.23) we arrive at (4.21). The functions θ_λ defined in the proof of Proposition 4.2 satisfy the constraints of this problem; hence

$$\log \hat{L}_\lambda(\theta_\lambda) \leqq \log \hat{L}_\lambda(v_\lambda^+) .$$

From (4.14) and (4.21) we have

$$\log \hat{L}_\lambda(\theta_\lambda) \leqq \frac{3N}{4} \log (\|v_\lambda^+\|_{L^2}) + \frac{N}{8} \log (8N/\beta) + \frac{N}{2} . \tag{4.24}$$

The proof now follows from (4.24) since $\log L_\lambda(\theta_\lambda) \to +\infty$ as $\lambda \to 0$.

PROPOSITION 4.6. *The second maximum penalized likelihood estimate of Good and Gaskins exists and is unique.*

PROOF. Using Proposition 4.5 the argument used to prove Proposition 4.3 shows that problem (4.19) has a unique solution which is also the unique solution of problem (4.18). This proves the proposition.

Acknowledgments. The authors would like to thank R. H. Byrd, B. F. Jones, P. E. Pfeiffer and W. A. Veech for helpful discussions. They also sincerely thank the referee and the editor for comments on the manuscript which led to a more readable paper.

REFERENCES

[1] BONEVA, L., KENDALL, D and STEFANOV, I. (1971). Spline transformations: Three new diagnostic aids for the statistical data-analyst. *J. Roy. Statist. Soc. Ser. B* **33** 1–77.

[2] FIACCO, A. V., and MCCORMICK, G. P. (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* Wiley, New York.

[3] GOFFMAN, C., and PEDRICK, G. (1965). *First Course in Functional Analysis.* Prentice-Hall, New Jersey.

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

[5] KIMELDORF, G. S. and WAHBA, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.* **41** 495–502.

[6] LIONS, J. S. (1968). *Contrôle Optimal de Systèmes Gouvernés par des Équations aux Dérivées Partielles.* Dunod, Paris.

[7] LIONS, J. L. and MAGENES, E. (1968). *Problèmes aux Limites Non Homogènes et Applications,* **1** Dunod, Paris.

[8] PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

[9] ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

[10] SCHOENBERG, I. J. (1968). Monosplines and quadrature formulae. *Theory and Application of Spline Functions* (T. N. E. Greville, ed.). Academic, New York.

[11] SCHOENBERG, I. J. (1972). Splines and histograms, with an appendix by Carl de Boor. Mathematics Research Center Report 1273, University of Wisconsin, Madison.

[12] TAPIA, R. A. (1971). The differentiation and integration of nonlinear operators. *Nonlinear Functional Analysis and Applications* (Louis B. Rall, ed.). Academic, New York.

[13] WAHBA, G. (1971). A polynomial algorithm for density estimation. *Ann. Math. Statist.* **42** 1870–1886.

G. F. DE MONTRICHER
54 RUE DU TOUC
3100 TOULOUSE
FRANCE

R. A. TAPIA AND J. R. THOMPSON
DEPARTMENT OF MATHEMATICAL SCIENCES
RICE UNIVERSITY
HOUSTON, TEXAS 77001