Journal of the Korean
Statistical Society
Vol. 26, No. 1, 1997

Maximum Penalized Likelihood Estimate in a Sobolev Space†

Young J. Park [1] and Young H. Lee [2]

ABSTRACT

We show that the Maximum Penalized Likelihood Estimate uniquely exists in a Sobolev space which consists of bivariate density functions. The Maximum Penalized Likelihood Estimate is represented as the square of the sum of the solutions of the Modified Helmholtz's equation on the compact subset of R^2.

Key Words : Maximum penalized likelihood estimate; Sobolev space; Modified Helmholtz's equation.

1. INTRODUCTION

We are interested in the problem of estimating the unknown probability density function in the nonparametric respects. There are several methods introduced by many authors for nonparametric density estimation. Izenman

†This research was supported by Ewha Womans University Research Fund.

[1]Research Institute of Statistics, Seoul National University, Seoul, 151-742, Korea.
[2]Department of Mathematic Education, Ewha Womans University, Seoul, 120-750, Korea.

(1991) reviewed recent developments in nonparametric density estimation. In particular, Wegman and Wrigh (1983) worked on splines as a nonparametric function estimating technique. Good and Gaskin (1971) and Tapia and Thompson (1978) considered the optimization problem to maximize of the likelihood function based on the random samples $x_1, x_2, ..., x_n$ and the manifold H in $L^1(a,b)$. If the manifold H is infinite dimensional, then the MLE may not exist. This is mainly because the class H is too large containing unsmooth functions with sharp spikes. To overcome this problem, it must be better to subtract a penalty for roughness from the log likelihood function before maximizing it. Tapia and Thompson(1978) considered the following constrained optimization problem :

$$L(f) = \prod_{i=1}^{n} f(x_i) \exp(-\phi(f)),$$

subject to

$$f \in H, \int_a^b f(t)dt = 1, \text{ and } f(t) \geq 0, \text{ for all } t \in (a,b),$$

where H is an infinite dimensional vector space in $L^1(a,b)$ and the roughness penalty ϕ is a functional satisfying $\phi(f) \geq 0$ such that $\phi(f_0) < \infty$ for the true density f_0.

When problem (1.1) is solved, the probability density function f is said to be a maximum penalized likelihood estimate (MPLE).

The purpose of this paper is to extend the vector space $L^1(R)$ to the vector space $L^2(R^2)$. In particular, we consider that H is a Sobolev space which is a subspace of $L^2(R^2)$ with bivariate density functions and the penalty function is

$$\phi(f) = 4\alpha \int_{R^2} \mid \text{grad } f^{1/2} \mid^2 dt, \ \alpha > 0.$$

The plan of paper is as follows: Section 2 lists some definitions and properties. Section 3 discusses the existence of MPLE on the Sobolev space and obtains the form of MPLE. Finally, Section 4 contains the conclusions.

2. PENALIZED LIKELIHOOD FUNCTIONALS ON THE SOBOLEV SPACE

We list some definitions and review the properties for the vector space $L^2(R^2)$.

Definition 2.1. The vector space

$$L^2(R^2) = \{f : R^2 \to R, \ \int_{R^2} | f |^2 < \infty\}$$

with inner product

$$< f, g >_{L^2(R^2)} = \int_{R^2} f(t) \cdot g(t) dt$$

is an infinite dimensional Hilbert space. We write that $L^2 = L^2(R^2)$, for convenience.

Define the notaions of 'grad f' and '|grad f|' as for $x = (x_1, x_2)$

$$\text{grad} f = \nabla f = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}) \text{ and } |\nabla f|^2 = (\frac{\partial f}{\partial x_1})^2 + (\frac{\partial f}{\partial x_2})^2.$$

Definition 2.2. The vector space

$$H(R^2) = \{f : \ f \in L^2, \ \int_{R^2} |\nabla f|^2 < \infty\}$$

with inner product

$$< f, g > = \int_{R^2} f \cdot g + \int_{R^2} (\nabla f, \nabla g)$$

is an infinite dimensional Hilbert space and it is a Sobolev space.

Here $(\nabla f, \nabla g)$ means the Euclidean inner product of ∇f and ∇g.

Note. $|| \cdot ||$ denotes the norm on $H(R^2)$ with $||f||^2 = < f, f >$, where $< f, f > = \int_{R^2} f^2 + \int_{R^2} |\nabla f|^2$.

Definition 2.3 (Tapia and Thompson, 1978). A Hilbert space $H(T)$ of functionals on a set T is a Reproducing Kernel Hilbert Space (RKHS) if there exists a reproducing kernel functional $K(\cdot, \cdot)$ defined by on T with the two properties :

(i) $K(\cdot, t) \in H(T)$, for all $t \in T$

(ii) $f(t) = < f, \ K(\cdot, t) >$, for all $t \in T, f \in H(T)$.

Proposition 1(Tapia and Thompson,1978).

(i) A Hilbert space of functions defined on a set T is a proper functional Hilbert space if and only if it is RKHS.

(ii) Suppose that H is a RKHS and D is a closed convex subset of $\{f \in H : f(x_i) \geq 0, \ i = 1, ..., n\}$ with the property that D contains at least one function which is positive at the samples $x_1, x_2, ..., x_n$. Then, if $\phi(f) = ||f||^2$, the penalized likelihood function defined in (1) has a unique maximizer in D.

We consider the problem of maximizing the following likelihood function:

$$L(v) = \prod_{i=1}^{n} v(x_i) \exp(-\phi(v)) \tag{2.1}$$

subject to

$$v^{1/2} \in H(R^2), \quad \int_{R^2} v(t)dt = 1, \ v(t) \geq 0, \ \text{for all } t \in R^2$$

where the roughness penality, for $\alpha > 0$,

$$\phi(v) = 4\alpha \int_{R^2} |\nabla v^{1/2} |^2 dt.$$

We need to modify the problem 2, since it is difficult to find the MPLE in problem (2.1) . Letting $u = v^{1/2}$, then we obtain the problem (2.2) about u:

$$L(u) = \prod_{i=1}^{n} u(x_i)^2 \exp(-\phi(u)) \tag{2.2}$$

subject to

$$u \in H(R^2), \quad \int_{R^2} u(t)^2 dt = 1, \ u(t) \geq 0, \ \text{for all } t \in R^2$$

with penality

$$\phi(u) = 4\alpha \int_{R^2} |\nabla u|^2 dt.$$

But this problem for u can not have a unique solution because if u is a solution then so is $-u$. By taking the square root in problem (2.2) , we restate the constrained optimization problem as follows:

$$L_1(u) = \prod_{i=1}^{n} u(x_i) \exp(-\phi_1(u)) \tag{2.3}$$

subject to

$$u \in H(R^2), \quad \int_{R^2} u(t)^2 dt = 1, \ u(t) \geq 0, \ \text{for all } t \in R^2,$$

with

$$\phi_1(u) = 2\alpha \int_{R^2} |\nabla u|^2 dt.$$

Then problems (2.1) , (2.2) and (2.3) have the following relationships between their solutions. If problem (2.3) has a unique solution, then problem (2.1) and (2.2) also have a unique solution. The form of solution of problem (2.1) is given by the square of solution of problem (2.3) . Hence it is enough to show that problem (2.3) has a unique MPLE.

In order to get a norm which is equivalent to the original norm defined on the Sobolev space $H(R^2)$ for given $\lambda > 0$ we consider the constrained optimization problem

$$L_\lambda(v) = \prod_{i=1}^{n} v(x_i) \exp(-\phi_\lambda(v)), \tag{2.4}$$

subject to

$$v \in H(R^2), \ \int_{R^2} v(t)^2 dt = 1, \ v(t) \geq 0, \ \text{for all } t \in R^2,$$

with

$$\phi_\lambda(v) = 2\alpha \int_{R^2} |\nabla v(t)|^2 dt + \lambda \int_{R^2} v(t)^2 dt.$$

We note that the problem (2.3) and (2.4) have the same solution.

3. EXISTENCE AND FORM OF MAXIMUM PENALIZED LIKELIHOOD ESTIMATOR

In this section, we find the solution of problem (2.4) . By squaring it we have the solution for the primary problem (2.1) .

Theorem 1. Problem (2.3) has a unique solution.

The proof of Theorem 1 is given in Appendix.

Theorem 2. Let v_λ denote the solution problem (2.4) and $\mathcal{D}$ be a compact set of R^2 . Then its form is a sum of the solutions of modified Helmholtz's equation in $H(\mathcal{D})$. Also $v_\lambda(t) > 0$, for all $t \in \mathcal{D}$.

Proof. Let $< \cdot, \cdot >_\lambda$ denote the inner product associated with ϕ_λ , i.e.,

$$< v_1, v_2 >_\lambda = 2\alpha \int_{R^2} (\nabla v_1, \nabla v_2) + \lambda \int_{R^2} v_1 v_2.$$

Let v_i be the Rietz representer in the ϕ_λ inner product of the continuous linear functional given at the point x_i, $i = 1, ..., n$. That is

$$< v_i, \tau >_\lambda = \tau(x_i), \text{ for all } \tau \in H(\mathcal{D}). \tag{3.1}$$

By definition of $< \cdot, \cdot >_\lambda$ and by definition 2.2, formula (3.1) can be written as

$$2\alpha \int_{\mathcal{D}} (\nabla v_i(t), \nabla \tau(t))dt + \lambda \int_{\mathcal{D}} v_i(t) \cdot \tau(t)dt = \tau(x_i).$$

By integrating by parts the first term of the left hand side becomes

$$2\alpha \int_{\mathcal{D}} (\nabla v_i(t), \nabla \tau(t))dt = 2\alpha \int_{\mathcal{D}} \tau(t)\Delta v_i(t)dt$$

since τ has compact support $\mathcal{D}$ by Stocke's Theorem. Here Δ means $\Delta f = -(\partial^2 f/\partial x_1^2 + \partial^2 f/\partial x_2^2)$ that is the Laplacian operator.

So above expression is written as follows.

$$2\alpha \int_{\mathcal{D}} \tau(t)\Delta v_i(t)dt + \lambda \int_{\mathcal{D}} v_i(t) \cdot \tau(t)dt = \tau(x_i).$$

Hence,

$$2\alpha\Delta v_i(t) + \lambda v_i(t) = \delta(t - x_i), \text{ for } i = 1, ..., n \tag{3.2}$$

where $\delta(\cdot)$ is two dimensional Dirac delta function at the origin.

Equation (3.2) is equivalent to the following form.

$$\Delta v_0(t) + \frac{\lambda}{2\alpha} v_0(t) = \delta(t), \tag{3.3}$$

and $v_i(t) = v_0(t - x_i)$.

Equation (3.3) is well known as the modified Helmholtz's equation. Solving (3.3) , we obtain

$$v_0(t) = (1/2\pi)B_0\left(\sqrt{\frac{\lambda}{2\alpha}}||t||\right),$$

where

$$B_0(p) = \int_0^\infty \cos(p \sinh q)dq, \quad p > 0$$

is called the modified Bessel function. (see Arfken, 1985)

Now we consider the problem of maximizing $\log L_\lambda$ over the constraint set

$$S = \{v \in H(\mathcal{D}) : \int v(t)^2 dt = 1, v(t) \geq 0 \text{ for all } t \in \mathcal{D}\}.$$

For this, we first consider a constraint set S^* which includes S :

$$S^* = \{v \in H(\mathcal{D}) : \int v(t)^2 dt = 1, v(x_i) \geq 0, i = 1, ..., n\}.$$

Let v_λ^* be the maximizer over S^* (the existence and its uniqueness and again assured by Proposition 1). Then obviously $v_\lambda^*(x_i) > 0, i = 1, ..., n$. Hence it is an interior point of S^*. This means that the derivative of $\log L_\lambda$ at v_λ^* must be zero functional. So we obtain the form of v_λ^* by the following steps.

Consider

$$\log L_\lambda(v) = \sum_{i=1}^{n} \log v(x_i) - \phi_\lambda(v),$$

with $\phi_\lambda(v) =< v, v >_\lambda$. Replacing v by $v + k\tau$ and differentiating about k, we have

$$\begin{aligned}
\frac{d}{dk} \log L_\lambda(v) &= \sum_{i=1}^{n} \frac{\tau(x_i)}{(v + k\tau)(x_i)} - 2 < v, \tau >_\lambda + 2k < \tau, \tau >_\lambda \\
&= \sum_{i=1}^{n} \frac{\tau(x_i)}{v(x_i)} - 2 < v, \tau >_\lambda \text{ by plugging } k = 0 \\
&= \sum_{i=1}^{n} \frac{< v_i, \tau >_\lambda}{v(x_i)} - 2 < v, \tau >_\lambda \text{ by (3.1)} \\
&= < \sum_{i=1}^{n} \frac{v_i}{v(x_i)} - 2 < v, \tau >_\lambda .
\end{aligned}$$

Since the last equation is to be zero at $v = v_\lambda^*$, we obtain

$$v_\lambda^*(t) = \frac{1}{2} \sum_{i=1}^{n} \frac{v_i(t)}{v_\lambda^*(x_i)} > 0, \quad \text{for all } t \in \mathcal{D}.$$

Note that this v_λ^* belongs to S. Thus it is the maximizer over S too. This implies $v_\lambda^* \equiv v_\lambda$.

4. CONCLUSION

By Theorem 1 and 2, we obtained the MPLE for problem (2.1) . As the penalty function, we used

$$\phi(v) = 4\alpha \int_{R^2} |\nabla v^{1/2}|^2 dt, \ \alpha > 0.$$

We also showed that the MPLE is unique and positive in compact subset of R^2. Finally we obtained that its form is represented as the square of $v_\lambda(t)$ appeared in Theorem 2.

APPENDIX

Proof of Theorem 1. Note first that the Sobolev space $H(R^2)$ is a RKHS, and the set

$$\{v \in H(R^2) : \int_{R^2} v(t)^2 dt = 1, v(t) \geq 0 \text{ for all } t \in R^2\}$$

is a closed subset of $\{v \in H(R^2) : v(x_i) \geq 0, i = 1, ..., n\}$. Also, $\phi_\lambda(v)$ is equivalent to $||v||^2$. Theorem 1 now follows from Proposition 1.

REFERENCES

(1) Arfken, G. (1985). *Mathematical Methods for Physicists*, Academic Press.

(2) Devroye, L. and Gyorfi, L. (1985), Nonparametric Density Estimation : The L_1 view. New York : John Wiley.

(3) Good, I. J. and Gaskin, R. A. (1971). Nonparametric roughness penalties for probability density, *Biometrica*, **58**, 255-277.

(4) Izenman, A. J. (1991), recent Developments in Nonparametric Density Estimation, *Journal of the American Statistical Association*, **86**, 205-224.

(5) Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall.

(6) Tapia, R. A. and Thompson, J. R. (1978). *Nonparametric Probability Density Estimation*, Johns Hopkins University Press.

(7) Wegman, E. J. and Wright, I. W. (1983), Splines in Statistics, *Journal of the American Statistical Association*, **78**, 351-365.