응용통계연구 제9권 2호
1994년 9월 pp. 323-333

Discrete Approximation to the Optimal Density in Moment Problems[1]

Changkon Hong[2]

Abstract

In this paper we present some approximation theorems related to the problem of finding optimal densities with prescribed moments. The implementation of the approximation theorems is to be done in some examples.

1. Introduction

Finding the smoothest function with various properties has recently been a popular topic in both mathematics and statistics (cf. Eubank (1988), Silverman (1986) and Wahba (1990)). In the function fitting problems, some difficulties arise from the given constraints (cf. Good and Gaskins (1971) and Tapia and Thompson (1990)). The problem of finding smooth densities with prescribed moments $c_1, \dots, c_n$ was studied by Hong (1992). The solution of the problem is to be obtained by minimizing the seminorm $\|f^{(m)}\|_{L_2}$ over Sobolev space W_m^2 under moment constraints (cf. Adams (1975)). If we let $J(f) = \|f^{(m)}\|_{L_2}$ and let $L_j f = \int_0^1 t^j f(t)\, dt$ then the problem can be formulated as follows;

$$\begin{aligned} &\text{Minimize} \quad J(f) \quad \text{on} \quad W_m^2 \\ &\text{subject to:} \quad L_j f = c_j, \ j = 0, \dots, n, \\ &\qquad \text{and } f(t) \ge 0 \quad \forall t \in [0,1], \end{aligned}$$

where $c_0 = 1$. The existence and uniqueness of the smoothest density with the moment constraints was proved and the characterization of the solution was also obtained in Hong (1992). He showed that on any interval where the solution f^* is positive, f^* agrees with a polynomial of degree $\le 2m+n$ and that $f^{*(2m)}$ agrees with a single polynomial of degree $\le n$ on its support. Furthermore, f is the unique solution if and only if f is nonnegative, satisfies the moment constraints, satisfies the boundary condition $f^{(i)}(0) = f^{(i)}(1) = 0$, $i = m, \dots, 2m-1$ and $f^{(m)}$ is of the form

1) This research was supported by the Dong_eui University Research Grant, 1993.
2) Department of Computer Science and Statistics, Dong_eui University, Pusan, 614-714, KOREA.

$$f^{(m)}(t)=\phi(t)+(-1)^m I_{m-1}(\xi)(t),$$

where $I_m(g)(t)=\int_0^t\int_0^{t_1}\cdots\int_0^{t_{m-1}} g(t_m)\,dt_m\,dt_{m-1}\cdots dt_1$, $\phi(t)$ is a polynomial of degree $\le m+n$ and ξ is a nondecreasing function which is constant on each interval where $f(t)>0$.

Using the characterization of the solution obtained in Hong (1992), we can find the exact minimizer. But it needs lots of work and in some sense it is by trial and error. It would be useful for finding the minimizer if we develop an automatic algorithm. In this paper we will achieve this by developing discrete approximation to the minimizer.

2. Discrete approximation

A penalty method can be hired to calculate the minimizer f^*. This method is often used to find solutions to optimization problems with equality constraints. If we let for $f\in W_m^2$

$$J_\alpha(f)=\sum_{j=1}^{n}(c_j-L_jf)^2+\alpha J(f),$$

then the penalty method leads a new optimization problem.

Problem (P_α):

$$\text{Minimize} \quad J_\alpha(f) \quad on \quad W_m^2$$
$$\text{subject to:} \quad L_0 f=1,$$
$$\text{and } f(t)\ge 0 \quad \forall t\in[0,1].$$

As α goes to 0, the penalty $\sum_{j=1}^{n}(c_j-L_jf)^2$ for the lack of fitness to the moment constraints becomes more important than the penalty $J(f)$ for the roughness. Let $M_1=\{f\in W_m^2 \mid L_0 f=1,\ f(t)\ge 0,\ \forall t\in[0,1]\}$. It can easily be shown that problem (P_α) is equivalent to the following problem.

$$\text{Minimize} \quad J(f) \quad on \quad M_1$$
$$\text{subject to:} \quad \sum_{j=1}^{n}(c_i-L_if)^2\le\rho(\alpha)$$

for some $\rho(\alpha)$ such that $\rho(\alpha)\to 0$ *as* $\alpha\to 0$.(cf. Schoenberg (1964)). Because of this equivalence between these two problems, we can find an approximation to f^* by solving problem (P_α) for sufficiently small α. The existence and uniqueness of the solution to problem (P_α) can readily be proved by the same method used in Hong (1992). Let f_α^* be the unique solution to problem (P_α).

2.1 Some approximation theorems

Here we will construct a discrete approximation to the minimizer f_α^* for problem (P_α),(whence to the minimizer f^*), by solving a finite dimensional version of problem (P_α). For given positive integer k, consider the uniform mesh $0=t_0<t_1< \cdots <t_k=1$, where $t_i=ih$ with $h=1/k$. Let $S_0(t_0,\cdots,t_k)$ and $S_1(t_0,\cdots,t_k)$ denote space of constant and linear spline functions on $[t_0,t_k]$ with knots $t_0,\cdots,t_k$, respectively. The m-th divided difference of f at the points $t_i, t_{i+1},\cdots, t_{i+m}$ is recursively given by

$$[t_i,\cdots,t_{i+m}]f = \frac{[t_{i+1},\cdots,t_{i+m}]f-[t_i,\cdots,t_{i+m-1}]f}{t_{i+m}-t_i},$$

with $[t_j]f = f(t_j)$. We develop some approximation theorems only for the case when $m=1$. For general m, parallel theorems can also be developed.

A discrete version of $\int_0^1 (f'(t))^2\,dt$ would be $\sum_{j=1}^{k}(f(t_j)-f(t_{j-1}))^2/h$. Let

$$G_{\alpha,h}=\sum_{i=1}^{n}(c_i-L_i f)^2+\alpha\frac{1}{h}\sum_{j=1}^{k}(f(t_j)-f(t_{j-1}))^2.$$

The following problem can be considered to be a finite dimensional version of problem (P_α), say problem (FP_α), when $m=1$.

Problem (FP_α):

Minimize $G_{\alpha,h}(s)$ on $S_0(t_0,\cdots,t_k)$

subject to: $L_0\, s=1$,

and $s(t)\le 0 \quad \forall t\in[0,1]$.

The existence and uniqueness of the solution to problem (FP_α) can easily be proved. We have the following theorem.

Theorem 2.1 Let $s_{\alpha,h}^*$ be the unique solution to problem (FP_α). Then $s_{\alpha,h}^*$ converges to f_α^* in the sup-norm, i.e.,

$$\| s_{\alpha,h}^*-f_\alpha^* \|_{sup}\to 0 \qquad as \qquad h\to 0,$$

where $\| g \|_{sup}=sup_{t\in[0,1]} | g(t) |$.

The idea used in Scott et al. (1980) was applied to prove this theorem. The proof of the theorem will follow from the following two lemmas.

Lemma 2.1 For each $\alpha>0$, it is possible to construct a family of nonnegative functions $s_{f_\alpha,h}$ in $S_0(t_0,\cdots,t_k)$ which integrate to 1 and satisfy

$$G_{\alpha,h}(s_{f_\alpha,h}) \rightarrow J_\alpha(f_\alpha^*) \qquad as \qquad h\rightarrow 0. \tag{2.1}$$

Proof Let

$$s_{f_\alpha,h}(t)=\frac{1}{h}\sum_{j=0}^{k-1}\left(\int_{t_j}^{t_{j+1}} f_\alpha^*(y)\,dy\right)\,\delta_{[t_j,t_{j+1}]}(t),$$

where $\delta_A(t)$ is the indicator function of A. Then $s_{f_\alpha,h}$ is obviously nonnegative and integrates to 1. By the mean value theorem, there exists an $x_j\in[t_j,t_{j+1}]$ such that

$$s_{f_\alpha,h}(t)=f_\alpha^*(x_j) \qquad \forall t\in[t_j,t_{j+1}].$$

By the fundamental theorem of calculus and the Hölder inequality, we get

$$|f_\alpha^*(t)-f_\alpha^*(x_j)| = \left|\int_{x_j}^{t} f_\alpha^{*\prime}(t)\,dt\right| \le h^{\frac{1}{2}}\,\|f_\alpha^{*\prime}\|_{L_2}\ .$$

Using this inequality and the Hölder inequality again leads to

$$\begin{aligned}|L_i f_\alpha^* - L_i s_{f_\alpha,h}| &\le \left(\frac{1}{2i+1}\right)^{\frac{1}{2}}\left(\int_0^1 |f_\alpha^*(t)-s_{f_\alpha,h}(t)|^2\,dt\right)^{\frac{1}{2}}\\ &= \left(\frac{1}{2i+1}\right)^{\frac{1}{2}}\left(\sum_{j=0}^{1}\int_{t_j}^{t_{j+1}} |f_\alpha^*(t)-f_\alpha^*(x_j)|^2\,dt\right)^{\frac{1}{2}}\\ &\le \left(\frac{h}{2i+1}\right)^{\frac{1}{2}}\|f_\alpha^{*\prime}\|_{L_2}\ .\end{aligned}$$

It follows that

$$L_i s_{f_\alpha,h} \rightarrow L_i f_\alpha^* \qquad as \qquad h\rightarrow 0.$$

On the other hand, we already know that $f_\alpha^{*\prime}$ is (uniformly) continuous on [0,1]. Therefore, it can easily be shown that

$$\begin{aligned}\frac{1}{h}\sum_{j=0}^{k-1}[s_{f_\alpha,h}(t_{j+1})-s_{f_\alpha,h}(t_j)]^2 &= \frac{1}{h}\sum_{j=0}^{k-1}[f_\alpha^*(x_{j+1})-f_\alpha^*(x_j)]^2\\ &= \|f_\alpha^*\|_{L_2}^2+O(h)\ .\end{aligned}$$

This proves the lemma.

Lemma 2.2 For each $\alpha>0$, it is possible to construct a family of nonnegative functions $f^*_{\alpha,h}$ in W_1^2 which integrate 1 and satisfy

$$\| f^*_{\alpha,h} - s^*_{\alpha,h} \|_{sup} \to 0 \qquad as \qquad h \to 0 \tag{2.2}$$

and

$$G_{\alpha,h}(s^*_{\alpha,h}) \to J_\alpha(f^*_{\alpha,h}) \qquad as \qquad h \to 0. \tag{2.3}$$

Proof Let $f^*_{\alpha,h}$ be the linear interpolation of $s^*_{\alpha,h}$ at the points $t_0, t_0', \cdots, t'_{k-1}, t_k$, where $t'_j = \frac{1}{2}(t_j + t_{j+1})$. Then $f^*_{\alpha,h}$ is obviously nonnegative and integrates to 1. Recall that $s^*_{\alpha,h}$ is the unique minimizer for problem (FP_α), thus $G_{\alpha,h}(s^*_{\alpha,h}) \le G_{\alpha,h}(s_{f_\alpha,h})$. From Lemma 2.1 we know that $G_{\alpha,h}(s_{f_\alpha,h}) \to J_\alpha(f^*_\alpha)$, as $h \to 0$. All these facts together imply that

$$\sup_j \,| s^*_{\alpha,h}(t_{j+1}) - s^*_{\alpha,h}(t_j) | \to 0 \qquad as \qquad h \to 0. \tag{2.4}$$

Equation (2.2) follows from Equation (2.4) and the following obvious inequality

$$\| f^*_{\alpha,h} - s^*_{\alpha,h} \|_{sup} \le \sup_j \,| s^*_{\alpha,h}(t_{j+1}) - s^*_{\alpha,h}(t_j) | .$$

Now a straightforward calculation shows that

$$\int_0^1 (f^{*\prime}_{\alpha,h}(t))^2 dt = \frac{1}{h} \sum_{j=0}^{k-1} [s^*_{\alpha,h}(t_{j+1}) - s^*_{\alpha,h}(t_j)]^2 + O(h). \tag{2.5}$$

On the other hand, using Equation (2.4) leads to

$$L_i f^*_{\alpha,h} - L_i s^*_{\alpha,h} \to 0 \qquad as \qquad h \to 0. \tag{2.6}$$

Combining (2.5) with (2.6) gives Equation (2.3) and the lemma is proved.

Before we prove Theorem 2.1, we will show that the following claim is true.

Claim 2.1 The functional J_α is uniformly convex on the convex subset $M_1 = \{f \in W_1^2 \mid L_0 f = 1 \text{ and } f(t) \ge 0, \ \forall t \in [0,1]\}$.

Proof Uniform convexity of J_α is equivalent to uniformly positive definiteness of $\ddot{J_\alpha}$ relative to M_1, i.e., for each $f \in M_1$,

$$\ddot{J_\alpha}(f)(h,h) \ge C \| h \|^2_{W_1^2}, \qquad \forall h \in T(M_1, f),$$

for some $C>0$, where $T(S,x) = \{h \in W_1^2 \mid \exists \lambda > 0 \text{ such that } x + \lambda h \in S\}$, the cone tangent to S at x. Note that $T(M_1,f)$ is a subset of the set $M_0 = \{h \in W_1^2 \mid L_0 h = 0\}$. The second Gâteaux derivative of J_α is given by

$$\ddot{J_\alpha}(f)(h,h) = 2 \sum_{i=1}^n (L_i h)^2 + 2 \int_0^1 (h'(t))^2 dt .$$

And $h(1)=\int_0^1 t\, h'(t)dt$ on the set M_0. Using this and the Hölder inequality leads to

$$\| h \|^2_{W_1^2} \leq \frac{4}{3}\int_0^1 (h'(t))^2 dt$$

This proves the claim.

Now we prove Theorem 2.1.

Proof of Theorem 2.1 By the optimality of f_α^* and $s^*_{\alpha,h}$ with respect to problem (P_α) and problem (FP_α), respectively, we have

$$J_\alpha(f_\alpha^*) \leq J_\alpha(f^*_{\alpha,h}) \tag{2.7}$$

and

$$G_{\alpha,h}(s^*_{\alpha,h}) \leq G_{\alpha,h}(s_{f_\alpha^*,h}). \tag{2.8}$$

Combining (2.1), (2.3), (2.7) and (2.8), it follows that

$$J_\alpha(f^*_{\alpha,h}) \to J_\alpha(f_\alpha^*) \qquad as \qquad h \to 0. \tag{2.9}$$

By the uniform convexity of J_α for all $\beta \in (0,1)$ and for some C,

$$\beta J_\alpha(f^*_{\alpha,h}) + (1-\beta)J_\alpha(f_\alpha^*) - J_\alpha(\beta(f^*_{\alpha,h}) + (1-\beta)f_\alpha^*) \geq C(\beta(1-\beta)\, \| f^*_{\alpha,h} - f_\alpha^* \|^2_{W_1^2}\ .$$

By this inequality and the optimality of f_α^*, we get

$$J_\alpha(f^*_{\alpha,h}) - J_\alpha(f_\alpha^*) \geq C(1-\beta)\, \| f^*_{\alpha,h} - f_\alpha^* \|^2_{W_1^2}\ . \tag{2.10}$$

Combining (2.9) with (2.10), we see that $f^*_{\alpha,h}$ converges to f_α^* in W_1^2-norm. Since W_1^2 is a reproducing kernel Hilbert space (R.K.H.S.) of functions defined on a compact set [0,1], convergence in W_1^2-norm implies convergence in sup-norm. Therefore, from (2.2) and the triangle inequality we arrive at

$$\| s^*_{\alpha,h} - f_\alpha^* \|_{sup} \to 0 \qquad as \qquad h \to 0.$$

This proves the theorem.

2.2 Numerical implementation

For the case of $m=2$ the solution f_α^* will be approximated by using the linear spline and the discrete version of the roughness penalty $\| f' \|_{L_2}$. The problem will be formulated as follows:

$$\text{Minimize} \qquad G_{\alpha,h}(f) \qquad on \qquad S_1(t_0, \cdots, t_k)$$

$$\text{subject to:} \quad L_0 f = 1,$$

$$\text{and } f(t) \leq 0 \qquad \forall t \in [0,1].$$

In this case the objective functional $G_{\alpha,h}(f)$ is given by

$$G_{\alpha,h}(f) = \sum_{i=1}^{n}(c_i - L_i f)^2 + \alpha h \sum_{j=1}^{k-1}(2[t_{j-1}, t_j, t_{j+1}]f\ )^2 .$$

Let

$$p_i = f(t_i), \qquad i = 0, \dots, k.$$

Then

$$f(t) = \begin{cases} p_{j-1} + \frac{1}{h}(p_j - p_{j-1})(t - t_{j-1}) & \text{if } t \in [t_{j-1}, t_j) \\ 0 & \text{if } t \notin [t_0, t_k] \end{cases}$$
$$= \begin{cases} \frac{1}{h}(p_j - p_{j-1})t - (j-1)p_j + jp_{j-1} & \text{if } t \in [t_{j-1}, t_j) \\ 0 & \text{if } t \notin [t_0, t_k] . \end{cases}$$

and

$$L_i f = \sum_{j=1}^{k} \int_{t_{j-1}}^{t_j} t^i f(t)\ dt$$
$$= \sum_{j=1}^{k} \int_{(j-1)h}^{jh} [\frac{p_j - p_{j-1}}{h} t^{i+1} - ((j-1)p_j - jp_{j-1})t^i]\ dt .$$

After tedious calculation, we get

$$L_i f = a_i^T p,$$

where

$$a_i = \frac{h^{i+1}}{(i+1)(i+2)} \begin{pmatrix} 1 \\ \vdots \\ (j-1)^{i+2} - 2j^{i+2} + (j+1)^{i+2} \\ \vdots \\ (i+2-k)k^{i+1} + (k-1)^{i+2} \end{pmatrix}.$$

So we have

$$L f = A^T p ,$$

where

$$L = (L_0, \dots, L_n)^T$$
$$A = (a_0, \dots, a_n) .$$

Since

$$[t_{i-1}, t_i, t_{i+1}]f = \frac{1}{2h^2}(p_{i-1} - 2p_i + p_{i+1}) \ ,$$

we have

$$G_{\alpha,h}(f) = c^T_{\_0} c_{\_0} \ - \ 2c^T_{\_0} A^T_{\_0} p \ + \ p^T A_{\_0} A^T_{\_0} \ + \ \alpha \frac{1}{h^3} p^T H p \ ,$$

where

$$\begin{aligned} c_{\_0} &= (c_1, \cdots, c_n^T), \\ A_{\_0} &= (a_1, \cdots, a_n) \end{aligned}$$

and

$$H = \begin{bmatrix} 1 & -2 & 1 & 0 & & \cdots & & & & & & 0 \\ -2 & 5 & -4 & 1 & 0 & & \cdots & & & & & \\ 1 & -4 & 6 & -4 & 1 & 0 & & \cdots & & & & \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & & \cdots & & & \\ \vdots & & & & & & \vdots & & & & & \vdots \\ & & & \cdots & & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ & & & & \cdots & & 0 & 1 & -4 & 6 & -4 & 1 \\ & & & & & \cdots & & 0 & 1 & -4 & 5 & -2 \\ 0 & & & & & & \cdots & & 0 & 1 & -2 & 1 \end{bmatrix} .$$

Let

$$Q_{\alpha,h}(p) \ = \ -2c^T_{\_0} A^T_{\_0} p + p^T A_{\_0} A^T_{\_0} p + \alpha \frac{1}{h^3} p^T H p \ .$$

If p is determined, then f is completely determined. Since $f(t) \geq 0$ is equivalent to $p_i \geq 0, \ \forall i = 0, \ldots, k$, problem (FP_α) can be restated as the following $k+1$ dimensional optimization problem.

$$\begin{aligned} &\text{Minimize} \qquad G_{\alpha,h}(p) \qquad on \qquad R^{k+1} \\ &\text{subject to:} \qquad i)\ a_0^T p - 1 = 0 \\ &\qquad\qquad\qquad\quad ii)\ p_i \geq 0, \quad i = 0, \ldots, k \ . \end{aligned}$$

Example In this example we use the first two moments $c_1 = 1/5$, $c_2 = 1/16$. We could find values of p which exactly satisfies the moment constraints for reasonably small values of h. Therefore, we put $\alpha = 0$ and solve the following quadratic problem:

$$\begin{aligned} &\text{Minimize} \qquad p^T H p \qquad on \qquad R^{k+1} \\ &\text{subject to:} \qquad i)\ a_i^T p - c_1 = 0, \quad i = 0,1,2, \end{aligned}$$

$$ii)\ p_j \geq 0, \qquad j = 0, \ldots, k\ .$$

For the values k = 20, 50, 100, the fitted density $\hat{f}_0$ are obtained using the **IMSL** subroutine, **DQPROG** (cf. Powell (1983a, 1983b)). Here $\hat{f}_0$ is the linear interpolation of the points $(t_0, p_0^*), \ldots, (t_k, p_k^*)$, where p^* is the solution of the above quadratic problem. The following figure shows the exact solution $f^*(t)$ and approximation to it for the values of k = 20, 50, 100, respectively.

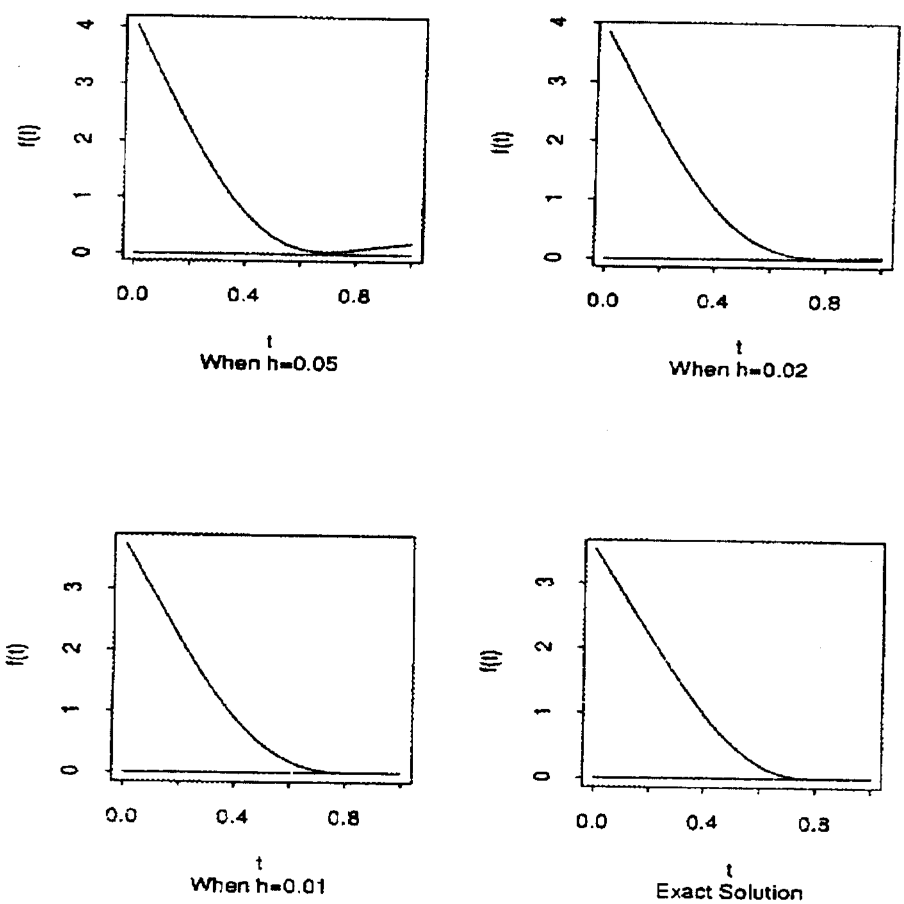

Figure. The exact solution and the approximated densities

References

[1] Adams, R.A. (1975). *Sobolev Spaces,* Academic Press, New York.

[2] Aronszajn, N. (1950). Theory of reproducing kernels, *American Mathematical Society:Transactions,* Vol. 68, 337-404.

[3] Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression,* M. Dekker, New York.

[4] Good, I.J. and Gaskins, R.A. (1971). Nonparametric roughness penalties for probability densities, *Biometrika,* Vol. 58, 255-277.

[5] Hong, C. (1992). Densities with optimal smoothness in moment problems, *Ph.D. Thesis, Purdue University.*

[6] Powell, M.J.D. (1983a). ZQPCVX a FORTRAN subroutine for convex quadratic programming. *DAMTP Report* NA17, Cambridge, England.

[7] Powell, M.J.D. (1983b). On the quadratic programming algorithm of Golbfarb and Idnani, *DAMTP Report* NA19, Cambridge, England.

[8] Schoenberg, I.J. (1964). Spline functions and the problem of graduation, *Proceedings, National Academy of Sciences.* USA 52; 947--950.

[9] Scott, D.W., Tapia, R.A., and Thompson, J.R. (1980). Nonparametric probability density estimation by discrete maximum penalized likelihood criteria, *Annals of Statistics,* Vol. 8, 820-832.

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

[11] Tapia, J.R. and Thompson, R.A. (1990). *Nonparametric Function Estimation, Modeling, and Simulation,* SIAM.

[12] Wahba, G. (1990). *Spline Models for Observational Data,* CBMS 59, SIAM.

적률 문제에 있어서의 최적 확률 밀도 함수의 이산적 근사[1]

홍창곤[2]

요약

본 논문에서는 주어진 n개의 적률을 갖는 최적의 확률 밀도 함수를 찾는 문제와 관련된 몇가지 근사 정리들을 제안하고 증명한다. 또한, 이 근사 정리들이 예를 통하여 수행될 것이다.

1) 이 연구는 1993 동의대학교 자체 학술 연구 조성비에 의하여 지원되었음.

2) (614-714) 부산직할시 진구 가야동, 동의대학교 전산통계학과.