OPTIMAL CHOICE OF THE KERNEL FUNCTION FOR THE PARZEN KERNEL-TYPE DENSITY ESTIMATORS

by

Demetrios Kazakos

April 1975

Technical Report #7501

Optimal Choice of the Kernel Function for the Parzen Kernel-Type Density Estimators

by

Demetrios Kazakos

ABSTRACT

Let $W_p^{(2)}$ be the Sobolev space of probability density functions $f(X)$ whose first derivative is absolutely continuous and whose second derivative is in $L_p(-\infty, +\infty)$, for $p \in [1, +\infty]$.

Using an upper bound to the mean square error for a fixed X $E[f(X) - \hat{f}_n(X)]^2$, found by G. Wahba, where $\hat{f}_n(X)$ is the Parzen Kernel-type estimate of $f(X)$, we find the finite support Kernel function $K(X)$ that minimizes the said upper bound.

The optimal Kernel function is :

$$K(y) = (1+a^{-1})\ (2T)^{-1}\ [1 - T^{-a} |y|^{a}], \quad \text{for } |y| < T$$

where $[-T, T]$ is the support interval, and $a = 2 - p^{-1}$

Research supported by National Research Council

INTRODUCTION

Nonparametric estimation of probability density functions is based on the concept that the value of a density function at a continuity point can be estimated using the sample observations that fall within a small region around the point. Originally, Fix and Hodges [1] employed this concept. Rosenblatt [2] , Parzen [3] and Cacoullos [4] generalized these results and developed the Parzen Kernel class of density estimates.

These estimates were shown to be asymptotically unbiased, consistent in the mean square sense and uniformly consistent.

In a series of papers, Grace Wahba [5-7] recently compared the convergence properties of several density estimation methods.

Those are :

a) The ordinary histogram.

b) Certain orthogonal series estimates.

c) Kernel-type estimates.

d) Polynomial interpolation algorithm.

e) Interpolating spline methods with fixed and variable Knots.

Her conclusion on the convergence properties of the above methods, can be summarized as follows.

Let $W_p^{(m)}$ be the Sobolev space of functions whose first $m-1$ derivatives are absolutely continuous, and whose m th derivative is in L_p .

Let

$$\|f^{(m)}\|_p = \left| \int_{-\infty}^{+\infty} \left(f^{(m)}(\xi)\right)^p d\xi \right|^{1/p} \leq M$$

(where $f^{(m)}(X)$ is the mth derivative of $f(X)$)

for

$$1 \leq p < +\infty$$

and

$$\| f^{(m)} \|_{\infty} = \sup_{\xi} \left| f^{(m)} (\xi) \right| \leq M ,$$

for

$$p = +\infty$$

And let :

$$W_p^{(m)} (M) = \left\{ f ; \ f \in W_p^{(m)} , \ \| f^{(m)} \|_p \leq M \right\}$$

The functions in $W_p^{(m)} (M)$ may be thought of as possessing a certain minimal degree of smoothness, characterized by the parameters m, p, M.

Let $\hat{f}_n(X)$ be an estimate of $f(X)$, based on n independent observations from the density f.

Let

$$\varphi(m, p) = (2m - 2/p) / (2m + 1 - 2/p)$$

As measure of closeness of an estimate $\hat{f}_n(X)$ to the true density function is always taken here the mean square error :

$$E \left[f(X) - \hat{f}_n(X) \right]^2$$

for a fixed value of X.

Farrel, in [8], showed essentially that for any sequence of estimators, $\hat{f}_n(X)$, and for $f \in W_p^{(m)} (M)$, the mean square error cannot have a rate of convergence better than

$$n^{-\varphi (m, p)}$$

Specifically, if

$$\sup_{f \in W_p^{(m)}(M)} E\left[f(X) - \hat{f}_n(X) \right]^2 = b_n \, n^{-\varphi(m,p+\epsilon)}$$

where $\epsilon > 0$ is fixed but may be arbitrarily small, then

$$\liminf_{n \to \infty} b_n = D_o(\epsilon) > 0$$

where $D_o(\epsilon)$ is a constant depending on ϵ .

Nothing that $\varphi(m,p)$ is an increasing function of p , the conclusion follows that no rate of convergence better than

$$n^{-\varphi(m,p)}$$

can be achieved.

Wahba essentially proved that all of the above density estimates (a), (b), (c), (d), (e) achieve the above convergence rate, in the following sense :

$$E\left[f(x) - \hat{f}_n(X) \right]^2 \leq D_j \, n^{-\varphi(m,p)}$$

where D_j is the constant corresponding to algorithm j , $j =$ (a), (b), (c), (d), (e) . For each type of algorithm, D_j , in general depends on (m, p, M) . For algorithm (c), D_c depends also on the Kernel function.

The above surprising result of Wahba leaves us in the dark with regard to what is the best density estimate.

The criterion of selection of a good algorithm is, therefore, achieving a small value for the corresponding D_j .

For algorithms (a), (b), (d), (e), the corresponding D_j's are pretty much fixed, depending only on (m, p, M, f).

But for algorithm (c), D_c is expressed in terms of the Kernel function. This fact gives us an added design flexibility, regarding the choice of the Kernel.

Until now, the choice of the Kernel function has been an ad hoc experiment.

Recently, I heuristically motivated solution was considered in [9], where the authors used a B-spline as a Kernel function. They argue that such a choice will roughly make D_c small.

The solution found here is superior to a B spline, being simpler and producing a faster rate of convergence.

The problem of finding the best Kernel function for Kernel-type density estimators has also been considered by Watson and Leadbetter [10].

They expressed the Fourier transform of the optimal Kernel function in terms of the characteristic function of the density.

Hence, their solution assumes a lot of knowledge about the probability density function to be estimated.

Epanechnikov [11] also studied the Kernel function optimization problem. He uses an approximation to the integrated mean square error between $f(X)$ and $\hat{f}_n(X)$, which he seeks to minimize.

However, he uses an unnecessary additional constraint on the Kernel function $K(y)$, namely :

$$\int_{-\infty}^{+\infty} y^2 \, K(y) \, dy = 1$$

The above condition is not required by the general Parzen or Kernel estimates, and is completely arbitrary.

Hence, the optimality of his solution is not established.

In section I of the present work, the results of Wahba are presented, and the motivation for the optimization problem is given.

In section II, an efficient solution to the optimization problem of section I is found.

I. Parzen Estimate and Its Performance Bound

We now assume that the true density function f is smooth enough to belong to $W_p^{(m)}(M)$:

$$f \in W_p^{(m)}(M) , \quad m \geq 2$$

Let $T(y)$ be a real valued function on $(-\infty, +\infty)$, satisfying the following conditions.

i) $$\sup_{-\infty < y < \infty} |T(y)| < \infty$$

ii) $$\int_{-\infty}^{+\infty} |T(y)| dy < \infty$$

iii) $$\lim_{y \to \infty} |y T(y)| = 0$$

iv) $$\int_{-\infty}^{+\infty} T(y) \, dy = 1$$

v) $$\int_{-\infty}^{+\infty} y^s \, T(y) \, dy = 0 , \quad s = 1, 2, \dots, m-1$$

vi) $$\int_{-\infty}^{+\infty} |y|^{m-1/p} \, |T(y)| \, dy < \infty$$

Let $(t_1, \dots, t_n)$ be a set of n independent observations distributed according to $f(X)$.

The Kernel-type (Parzen) estimate $\hat{f}_n(X)$, using as Kernel the function $T(y)$, is then given by

$$\hat{f}_n(X) = (n h_n)^{-1} \sum_{j=1}^{n} T\left[h_n^{-1} (X - t_j)\right]$$

where $h_n > 0$ is to be chosen so that $h_n \to 0$, $nh_n \to +\infty$

Let Λ be an upper bound to $f(X)$, i.e.

$$\sup_X f(X) \le \Lambda$$

Let

$$B_2 = \Lambda \int_{-\infty}^{+\infty} T^2(y)\, dy$$

Let

$$\frac{1}{q} + \frac{1}{p} = 1$$

Let

$$A_2 = \left[\int_{-\infty}^{+\infty} |T(y)|\, |y|^{m-1/p}\, dy \right]^2 \cdot$$

$$\cdot \left[(m-1)!\right]^{-2} \cdot \left[(m-1)\, q+1\right]^{-2/q}$$

An upper bound to the mean square error is :

$$E\left[f(X) - \hat{f}_n(X)\right]^2 \le M^2 A_2 h^{2m-2/p} +$$

$$+ B_2 (nh)^{-1} \qquad (1)$$

(where $h = h_n$)

If we define $k_n = nh$, the choice of k_n that minimizes the right hand side of (1), is

$$k_n = \left[(2m-2/p)^{-1} A_2^{-1} M^{-2} B_2\right]^{1/(2m+1-2/p)} \cdot$$

$$\cdot\ n^{\varphi(m,p)} \qquad (2)$$

with

$$\varphi(m,p) = (2m - 2/p) \;/\; (2m+1-2/p)$$

For $h = n^{-1}k_n$, and k_n given by (2), the final form of the upper bound to the mean square error is :

$$E\left[f(X) - \hat{f}_n(X)\right]^2 \leq D_2 \, n^{-\varphi(m,p)} \, [1+0(1)] \quad (3)$$

with

$$D_2 = (1+2a)\,(2a)^{-2a}\, M^{2/(1+2a)}\, \Lambda^{2a/(1+2a)} \,.$$

$$\cdot \left[(m-1)!\,[q(m-1)+1]^{1/q}\right]^{-2/(1+2a)} \,.$$

$$\cdot \left\{ \left[\int_{-\infty}^{+\infty} T^2(y)\,dy\right]^a \,.\right.$$

$$\left. \cdot \left[\int_{-\infty}^{+\infty} T(y) \quad y^a \; dy\right] \right\}^{2/(1+2a)}$$

and

$a = m - 1/p$ (hence $a > 0$)

It is seen that the only dependence of the bound on the Kernel function $T(y)$ is through the quantity :

$$G(T) = \int_{-\infty}^{+\infty} |T(y)|\,|y|^a\,dy \left[\int_{-\infty}^{+\infty} T^2(y)\,dy\right]^a$$

The choice of T that will minimize G under the constraints (i) - (vi) , is the optimal , in terms of minimizing the upper bound given.

We observe that if we require $T(y)$ to have a finite support and be bounded, then constraints (i), (ii), (iii), (iv), (vi) are satisfied. Only (v) is difficult to satisfy, the difficulty increasing with m.

m represents our knowledge of the degree of smoothness of $f(x)$.

The bound of inequality (3) is monotone decreasing with m. Furthermore, for $m \geq s$, $W_p^{(m)}(M) \subseteq W_p^{(s)}(M)$.

If we know the value of m, and it happens to be greater than 2, the bound will be tightest if we use the known value of m. But then, constraint (v) will be too restrictive.

If we use the value $m = 2$ at the bound, in spite of our knowledge that actually $m = m_o > 2$, the bound will still be valid, and (v) will be easier to satisfy.

In other words, we choose to relinguish some of our knowledge about the smoothness of $f(X)$ in exchange for relaxing constraint (v).

For $m = 2$, we have $a = 2 - 1/p$ and constraint (v) is satisfied, if $T(y)$ is an even function of y, i.e.

$$T(y) = T(-y)$$

In practice, knowledge of a large m, or large degree of smoothness of $f(X)$, is hard or impossible to come by. Hence, letting $m = 2$ is a practical advantage.

Let $[-T, T]$ be the support of $T(y)$.

A natural assumption for $T(y)$ is that it is nonnegative for $y \varepsilon [-T, T]$.

Let

$$T(y) = k^2(y)$$

The function $k^2(y)$ is even, and integrates to 1 :

$$\int_{-T}^{T} k^2(y)\, dy = 1$$

Hence,

$$\int_{0}^{T} k^2(y)\, dy = 1/2$$

Because of the evenness of $k^2(y)$, the quantity to be minimized, is :

$$G_T(k) = \int_{0}^{T} k^2(y)\; y^a \; dy \left[\int_{0}^{T} k^4(y)\, dy \right]^a$$

The choice of T is arbitrary for the time being. Increasing T implies larger amoothing internal, and more computations in the realization of the Kernel estimate. In the next section, T will be considered as one of the design parameters.

II. Optimal Choice of Kernel

We are interested in transforming the interval $[0, T]$ into $[0, 1]$, so that the optimization of G will be performed for the interval $[0, 1]$, independently of T.

Let

$$y = XT$$

We have :

$$G_T(k) = T^{2a+1} \int_0^1 k^2(XT)\ X^a\ dX \cdot \left[\int_0^1 k^4(XT)\ dX \right]^a$$

The constraint on k becomes :

$$\int_0^1 k^2(XT)\ dx = (2T)^{-1}$$

Let

$$k_o(X) = k(XT)\ , \quad b=(2T)^{-1}$$

Let

$$F(k_o) = \int_0^1 k_o^2(X)\ X^a\ dX \left[\int_0^1 k_o^4(X)\ dX \right]^a$$

$$1 < a < 2$$

We are seeking the minimization of $F(k_o)$, over $\{k_o(X), X \in [0, 1]\}$ under the constraint :

$$\int_0^1 k_o^2(X)\ dX = b$$

The dependence of the solution k_o on T is only through the constant b.

We also have the relationship :

$$G_T(k) = T^{2a+1} F(k_o)$$

The brute force minimization of $F(k_o)$ under the stated constraint is a hard, if not possible, task. Because the form of $F(k_o)$ does not make the problem amenable to calculus of variation techniques.

We are using, therefore, a more subtle method, better suited to the problem.

We add one more constraint on k_o .

We will seek the function k_o that minimizes $F(k_o)$ under the constraints

$$\left.\begin{aligned} \int_o^1 k_o^2(t)\, dt &= b \\ \int_o^1 k_o^4(t)\, dt &= q \end{aligned}\right\} \quad (c)$$

Furthermore, in order to comply with constraint (i), as well as for reasons of practicality in the implementation of the density estimator, we constrain $k_o^2(y)$ to be upper bounded by a constant M_o .

The functional F becomes :

$$F(k_o) = q^a \int_o^1 k_o^2(X)\, X^a\, dX$$

In Appendix I, it is shown that q can only take values in the interval $[b^2, bM_o]$.

Furthermore, for each $q \in [b^2, bM_o]$, there is at least one function $k_o^2(t)$ satisfying constraints (c) and bounded by M_o.

The minimization of F under constraints (c) is equivalent to the minimization of the functional

$$N(k_o) = \int_o^1 k_o^2(X)\ X^a\ dX$$

under the same constraints (c).

We note that X^a is a strictly monotone increasing function on $[0, 1]$.

We observe that a minimizing solution k_o has to be a strictly monotone decreasing function on $[0, 1]$

Because of the above observation, there are two possible forms that the optimal k_o may take :

(i) $k_o^2(t)$ = positive for all $t \in [0, 1]$

(ii) $k_o^2(t)$ = positive for $t \in [0, u]$ and zero for $t \in [u, 1]$

Clearly, a solution $k_o^2(X) = g(X)$ should satisfy the requirement that $g(X)$ be monotone decreasing.

If the solution is of type (i), then we must have $g(X) \geq 0$ for $X \in [0, 1]$.

If it is of type (ii), then $g(X) > 0$ for $X \in [0, u]$ and $g(u) = 0$.

In order to solve the constrained minimization problem, we use two Lagrange multipliers, A and B.

The idea is to solve the constrained minimization of $N(k_\circ)$ under constraints (c), and then by varying q in the interval $[b^2, bM_\circ]$, to generate all the solutions parametrically, in terms of q.

A final one-dimensional search will produce the minimizing q.

In order to consider all solutions, of type (i) and (ii), we construct the following functional :

$$L_u(k_\circ, A, B) = \int_\circ^u k_\circ^2(X)\ X^a\ dX - A\left[\int_\circ^u k^2(X)\ dX - b\right] + 2^{-1}B\left[\int_\circ^u k_\circ^4(X)\ dX - q\right],$$

where $0 < u \leq 1$.

A solution of type (i) will minimize the functional L_u for $u = 1$, and must be positive over the interval $[0, 1]$.

A solution of type (ii), will minimize L_u, be positive on $[0, u]$ and satisfy the condition $g(u) = 0$.

L_u can be written :

$$2L_u(k_\circ, A, B) = 2Ab - Bq - B^{-1}\int_\circ^u (A - X^a)^2\, dX + B\int_\circ^u \left[k^2(X) - B^{-1}(A - X^a)\right]^2 dX$$

Let

$$I_2(k_\circ) = \int_\circ^u \left[k^2(X) - B^{-1}(A - X^a)\right]^2 dX$$

According to the philosophy of the method of Lagrange multipliers, we must fix A,B, find in terms of A,B the function k_o^2 that achieves the unconstrained minimum of L_u, substitute it into the constraints (c), and solve for A,B in terms of b,q .

The minimizing k_o^2 should be monotone decreasing on $[0, u]$.

We assume for a moment that $k_o^2(X)$ can be unbounded, i.e., $M_o = +\infty$ (see Appendix) .

If B were negative, minimization of L_u is equivalent to unconstrained mazimization of $I_2(k_o)$.

This is achieved for $k_o^2 = +\infty$, which is an unacceptable solution.

Hence, the constant B must be positive.

For $B > 0$, the unconstrained minimum of L_u is equivalent to the unconstrained minimum of $I_2(k_o)$.

For $B > 0$, $A < 0$, the minimizing solution is $k_o^2 = 0$, which is unacceptable.

Hence, we must have $B > 0$, $A > 0$.

Then the minimizing solution is :

$$k_o^2(X) = B^{-1}(A - X^a) \quad , \quad X \in [0, u]$$

It is an acceptable solution, because it is monotone decreasing.

First, we will consider solutions of type (i) .

The function $B^{-1}(A - X^a)$ should then satisfy constraints (c), and we should furthermore have

$$A \geq 1 \quad , \quad B > 0$$

Substituting the function $B^{-1}(A - X^a)$ into (c) and solving for A, B in terms of (a, q, b) , we find :

$$\left.\begin{aligned} B &= \left[(1+2a)^{-1} - (1+a)^{-2}\right]^{\frac{1}{2}} \left(q-b^2\right)^{-\frac{1}{2}} \\ A &= b\left(q-b^2\right)^{-\frac{1}{2}}\left[(1+2a)^{-1}-(1+a)^{-2}\right]^{\frac{1}{2}}+(1+a)^{-1} \end{aligned}\right\}(m)$$

The requirement that $B > 0$, is satisfied automatically.

Substituting the formula for A in the inequality $A \geq 1$, we find the condition :

$$q b^{-2} \leq 1 + (1+2a)^{-1}$$

Hence, in the parameter region

$$1 \leq q b^{-2} \leq 1 + (1+2a)^{-1}$$

We have a minimum solution of type (i), with A, B given by (m).

A minimizing solution of type (ii) must satisfy the constraints :

$$\left.\begin{aligned} \int_0^u k_o^2(X)\, dX &= b \\ \int_0^u k_o^4(X)\, dX &= q \\ A &= u^a \\ u &< 1 \end{aligned}\right\}(c')$$

Substituting the function $k_o^2(X) = (u^a - X^a) B^{-1}$ into (c'), we can solve for u, B :

$$\left.\begin{aligned} u &= b^2 q^{-1}\left[1 + (1+2a)^{-1}\right] \\ B &= a b^{-1} (1+a)^{-1} u^{a+1} \\ A &= u^a \end{aligned}\right\}(m')$$

The basic requirement for the existence of an optimizing solution of type (ii), is that

$$0 < u < 1$$

From (m′), this requirement is equivalent to

$$q\, b^{-2} \geq 1 + [1+2a]^{-1}$$

Hence, by considering solutions of type (i) and (ii), we have covered all of the q-parameter space.

We define a new parameter :

$$W = (q\, b^{-2} - 1)\ (1+2a)$$

In terms of the new parameter, w , we can summarize the solutions.

The optimizing solution k_o that minimizes the functional $N(k_o)$ under constraints (c), has the form $B^{-1}(A - X^a)$, and is of type (i) or (ii).

If $0 \leq W \leq 1$, then the solution is of type (i), and A,B are given by equations (m) .

If $W \geq 1$, then the solution is of type (ii), and A,B are given by equations (m′).

We will now express the minimum value of $F(k_o) = q^a N(k_o)$ as a function of W.

Let us define by $F(W)$ the minimum value of the functional $F(k_o)$ for a fixed q , equal to

$$q = b^2 \left[1 + W\ (1+2a)^{-1}\right]$$

If we substitute the optimal solutions of types (i), (ii) into $F(k_o)$, we find $F(W)$.

The result is :

$$b^{-1-2a}\,F(W) = \begin{cases} (1+a)^{-1}\left[1 - a(1+2a)^{-1}\,W^{\frac{1}{2}}\right]\cdot \left[1 + W(1+2a)^{-1}\right]^{a} & \text{for } 0 \leq W \leq 1 \\ (1+2a)^{-1}\left[1 + (1+2a)^{-1}\right]^{a} & \text{for } W \geq 1 \end{cases}$$

See Figure 3 .

The derivative of $F(W)$ is given by the following equation, for the region $W \epsilon [0.1]$:

$$2b^{-1-2a}\,a^{-1}\,(1+a)\,(1+2a)\,F'(W) =$$

$$= -\left[1 + W(1+2a)^{-1}\right]^{a-1}\,W^{-\frac{1}{2}}\,(1-W^{\frac{1}{2}})^{2}$$

Hence, $F'(W) < 0$ for $W \epsilon [0,1]$, and $F(W)$ is monotone decreasing in the region $[0,1]$.

Because of the form of $F(W)$, we note that for all $W \geq 1$, the optimal solution of type (ii) gives the same value to the functional F.

Hence, we choose the solution $W = 1$ over all the solution of type (ii).

This is because the solution $u = W = 1$ has a better data-smoothing effect, hence its small sample performance should be superior to that of the solutions for $W > 1$.

As a conclusion, we see that the Kernel function

$$k_o^2(X) = b(1+a^{-1})(1-|X|^a), \quad |X| < 1$$

minimizes the given upper bound to the mean square error of the Parzen density estimate.

The resulting minimum value of F is :

$$F^\star = b^{1+2a} (1+2a)^{-1} \left[1 + (1+2a)^{-1}\right]^a$$

The optimal value of G_T is :

$$G_T^\star = T^{2a+1} F^\star = (Tb)^{2a+1} (1+2a)^{-1} \cdot$$

$$\cdot \left[1 + (1+2a)^{-1}\right]^a$$

or :

$$G_T^\star = (2+4a)^{-1} \left[4^{-1} + (4+8a)^{-1}\right]^a$$

It is interesting that the optimal G_T takes a value independent of the width $2T$ of the support of the Kernel function.

Therefore, all of the above results are independent of T.

The choice of T is hence left as an experimental task, because the actual performance of the density estimator does depend on T, in spite of the fact that the upper bound to the error variance is dependent of T.

CONCLUSIONS

Using as error criterion the mean square error of the estimate $\hat{f}_n(X)$ at a point X, and employing Wahba's upper bound, it was possible to find the optimal Kernel function with finite support that minimizes the mentioned upper bound. The simplicity of the optimal Kernel function is appealing both mathematically and computationally.

APPENDIX

Let M_o be an upper bound to the Kernel functions $K_o^2(y)$. In other words, for reasons of practicality and satisfaction of condition (i), the class of Kernel to be considered is constrained to be upper bounded by M_o.

$$k_o^2(y) \leq M_o \quad \text{for} \quad y \in [0, 1]$$

Hence, the upper bound to the parameter q is :

$$q \leq M_o b$$

The bound is achieved for

$$k_o^2(y) = \begin{cases} M_o & \text{for} \quad 0 < y < bM_o^{-1} \\ 0 & \text{for} \quad bM_o^{-1} < y < 1 \end{cases}$$

Next, we will seek a lower bound to q.

We will state the Hölder inequality, to be used for this purpose [12].

If $p, s \geq 0$, $p^{-1} + s^{-1} = 1$, and if the functions $f(t), g(t)$, $t \in [0, 1]$, are : $f(t) \in L^p[0, 1]$, $g(t) \in L^s[0, 1]$, then $f(t), g(t) \in L'[0, 1]$, and :

$$\int_o^1 \left| f(t) \; g(t) \right| \, dt \leq \left[\int_o^1 \left| f(t) \right|^p dt \right]^{1/p} \left[\int_o^1 \left| g(t) \right|^s dt \right]^{1/s}$$

Equality holds iff , for some nonzero constants X, y, we have

$$X \, |f(t)|^{p} = y \, |g(t)|^{s} , \quad \text{for almost all } t \in [0,1].$$

We apply now the above inequality for $g(t) = 1$, $f(t) = k_o^2(t) \; b^{-1}$, $s = p = 2$.

We have,then :

$$\int_o^1 b^{-1} \, k_o(t) \; dt \leq \left[\int_o^1 b^{-2} \; k_o^4(t) \; dt \right]^{\frac{1}{2}}$$

Hence,

$$1 \leq b^{-2} \int_o^1 k_o^4(t) \; dt = q \; b^{-2}$$

where q, b are as defined in constraints (c).

Equality is achieved only for $k_o^2(t) = 1$

Using the above results, the range of q is :

$$b^2 \leq q \leq b M_o$$

We will further show that for any q of the above interval, there is at least one function $k_o^2(y)$ satisfying constraints (c).

Let

$$z(t) = \begin{cases} b \;\; u^{-1} & \text{for} \quad 0 < t < u \\ 0 & \text{for} \quad u < t < 1 \end{cases}$$

Then,

$$\int_0^1 z(t)\ dt = b$$

$$\int_0^1 z^2(t)\ dt = b^2 u^{-1}$$

If we let

$$u = b^2 q^{-1} < 1$$

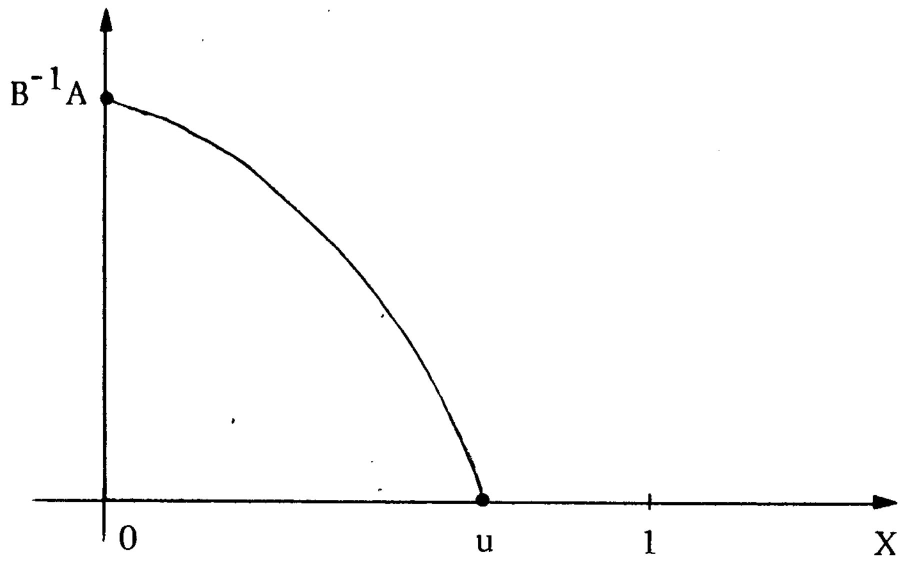
then

$$z(t) = \begin{cases} q\ b^{-1} & \text{for} \quad 0 < t < b^2 q^{-1} \\ 0 & \text{for} \quad b^2 q^{-1} < t < 1 \end{cases}$$

And the function $k_o^2(t) = z(t)$ satisfies constraints (c), and furthermore $z(t) < M_o$ for all $t \in [0, 1]$.

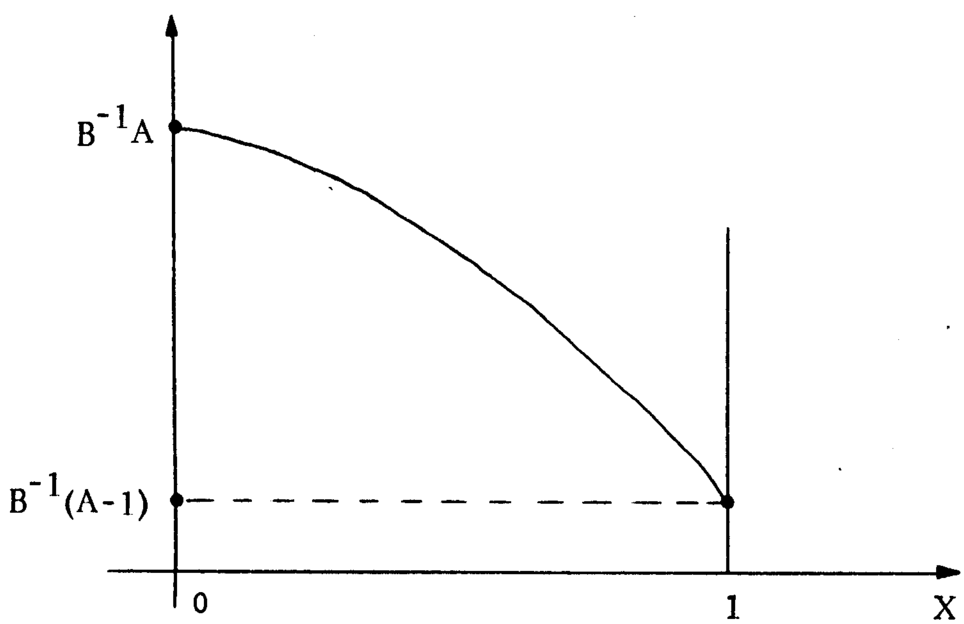
Hence, for each q $[b^2, bM_o]$, there is at least a function $k_o^2(t)$ satisfying constraints (c), that is also upper bounded by M_o.

The values $q = b^2$, $q = bM_o$ give us exactly one corresponding $k_o^2(t)$



Type (ii) Solution
$W \geq 1$

Figure 1



Type (i) Solution
$0 \leq W \leq 1$

Figure 2

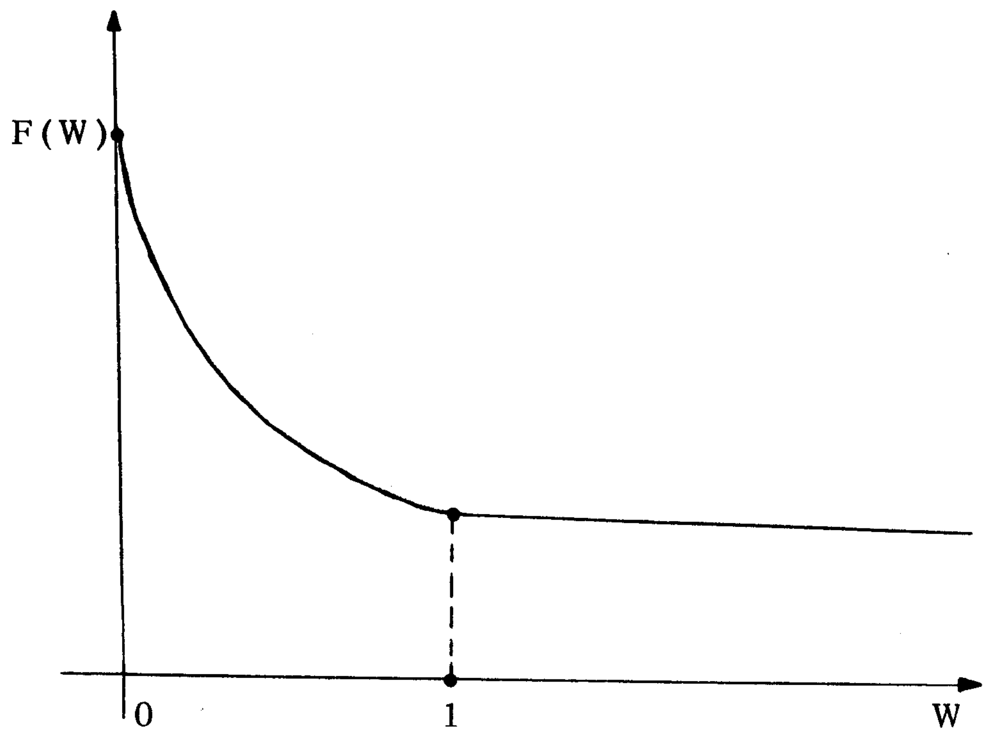


Figure 3

REFERENCES

[1] E. Fix and J. L. Hodges, "Discriminatory Analysis, Nonparametric Discrimination," U. S. Air Force School of Aviation Medicine, Randolph Field, Tex. Project 21-49-004, February 1951.

[2] M. Roseblatt, "Remarks on Some Nonparametric Estimates of a Density Function," Ann. Math. Stat., Vol. 27, p. 837, 1956.

[3] E. Parzen, "On Estimation of a Probability Density Function and Mode," Ann. Math. Stat., Vol. 33, p. 1065, 1962.

[4] T. Cacoullos, "Estimation of a Multivariate Density," Ann. Math. Stat., Vol. 18, p. 179, 1966.

[5] G. Wahba, "Optimal Convergence Properties of Variable Knot, Kernel, and Orthogonal Series Methods for Density Estimation," Ann. Math. Stat., Vol. 3, p. 15, January 1975.

[6] G. Wahba, "Interpolating Spline Methods for Density Estimation I. Equi-spaced Knots," Ann. Math. Stat., Vol. 3, p. 30, January 1975.

[7] G. Wahba, "Interpolating Spline Methods for Density Estimation II Variable Knots," Technical Report No. 337, Department of Statistics, University of Wisconsin, Madison, Wisconsin, June 1973.

[8] R. H. Farrel, "On Best Obtainable Asymptotic Rates of Convergence in Estimation of a Density Function at a Point," Ann. Math. Stat., Vol. 43, p. 170, 1972.

[9] J. O. Bennett, R. J. P. de Figueiredo, J. R. Thompson, "Classification by Means of B-Spline Potential Functions with

Application to Remote Sensing," Proceedings of Sixth Southeastern Symposium on System Theory, Baton Rouge, La.

[10] G. S. Watson and M. R. Leadbetter, " On the Estimation of the Probability Density," Ann. Math. Stat., Vol. 34, p. 480, 1962.

[11] V. A. Epanechnikov, "Nonparametric Estimation of a Multivariate Probability Density," Theory Probability Application (USSR), Vol. 14, p. 153, 1969.

[12] H.L.Royden,ch.6 "The Classical Banach Spaces" Real Analysis. The Macmillan co.,New York,1963 p.p.113