DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, Wisconsin 53706

TECHNICAL REPORT No. 435

September 1974

REPRODUCING KERNEL HILBERT SPACES APPLIED TO BIVARIATE NATURAL SPLINE DENSITY ESTIMATION

by

Robert M. Kuhn
University of Wisconsin

Typist: Candy Smith

This work was supported by the Air Force Office of Scientific Research under Grant AFOSR 72-2363B.

Abstract

The purpose of this technical report is to introduce and describe the properties of a bivariate density estimation procedure which is a generalization of the univariate procedure described in [1]. The procedure can be described as follows.

(a) It is assumed that the bivariate density function has value zero outside the unit square. Divide one axis into $\ell_1 + 1$ and the other axis into $\ell_2 + 1$ equally spaced points. This causes the unit square to be broken up into ℓ equal size rectangles where ℓ equals ℓ_1 times ℓ_2.

(b) Define

$$\hat{F}(ih_1, jh_2) = a_{ij}$$

where

a_{ij} = the per cent of observations in A_{ij}

$A_{ij} = \{(t_1,t_2): t_1 \le ih_1 \text{ and } t_2 \le jh_2\}$

$h_i = 1/\ell_i$

(c) If $\hat{F}$ has bivariate density $\hat{f}(t_1,t_2)$ and marginal densities $\hat{f}_1(t_1)$ and $\hat{f}_2(t_2)$ then the density estimate can be chosen to uniquely minimize

$$\int_0^1 \left(\frac{d}{dt_1}\hat{f}_1(t_1)\right)^2 + \int_0^1 \left(\frac{d}{dt_2}\hat{f}_2(t_2)\right)^2 + \int_0^1\int_0^1 \left(\frac{d}{dt_1}\frac{d}{dt_2}\hat{f}(t_1,t_2)\right)^2 dt_1 dt_2$$

subject to:

$$\hat{F}(k_1h_1, k_2h_2) = a_{k_1,k_2} \text{ and } \hat{F}(0,t_2) = \hat{F}(t_1,0) = 0 \quad \forall k_1,k_2,t_1,t_2.$$

The first part of this paper is devoted to a solution to this problem, along with computational formulas.

The solution given above will turn out to be a unique natural bicubic spline of interpolation. It therefore will be seen that the above estimate also minimizes (not uniquely)

$$\int_0^1\int_0^1 \left(\frac{d}{dt_1}\frac{d}{dt_2}\hat{f}(t_1,t_2)\right)^2 dt_1 dt_2$$

subject to the above constraints.

It is shown that there is a constant K, independent of n, s, and s_2, such that

$$E|f(s_1,s_2) - \hat{f}(s_1,s_2)| \le K/n^{1/3}.$$

Finally, a computational formula is suggested for the situation in which the interpolation points are determined by data, rather than being equally spaced.

Introduction

This paper is divided into five sections. In the first section, the reproducing kernel solution to the bivariate density estimation problem is described. In the second section the reproducing kernel is computed and a computational formula obtained with the help of this kernel. This computational procedure is easy to extend to higher dimensions. In the third section, it is shown that the computed distribution estimate is the unique natural bicubic spline interpolating the data. Also, in the third section, it is shown that the problem described in this report differs from the density estimation problem of Schoenberg and de Boor [2] by some boundary conditions. Thus, the computational procedure described there could be applied to this problem with relatively minor modifications. Uniform bounds on the expected error are computed in section four. In the fifth and final section, density estimation will be discussed for the situation in which interpolation points are determined by the order statistics rather than being equally spaced.

The possibility of using splines to estimate two dimensional density functions was first discussed by Boneva, Kendall and Stefanov [3], section 4. In this reference, they mention a FORTRAN computer program called /SPLINE/CHART which is designed to enable a user to do rapid contour plotting of the density estimate. However, reference number [3] does not describe the method in enough detail to make evaluation possible.

Another computationally related procedure is due to Akima ([4] and [5]). Other multivariate density estimates are described in [6], [7], [8], [9], [10] and [11]. Glick [12] has in his bibliography some other references on the subject of multivariate density estimation.

The Theoretical Reproducing Kernel Solution

Let

$$\left.\begin{aligned} &0 = x_0 < x_1 < \ldots < x_{\ell_1} = 1 \\ &0 = y_0 < y_1 < \ldots < y_{\ell_2} = 1 \\ &A_{ij} = \{(t_1,t_2): t_1 \le x_i \text{ and } t_2 \le y_j\} \\ &a_{ij} = \text{the per cent of observations which are in } A_{ij} \end{aligned}\right\} \quad (1.1)$$

Let $\hat{F}(t_1,t_2)$ have bivariate density $\hat{f}(t_1,t_2)$ and marginal densities $\hat{f}_1(t_1)$ and $\hat{f}_2(t_2)$. In this section, we compute the solution to the following minimization problem.

$$\left.\begin{aligned} &\text{minimize } J(\hat{f}) \text{ subject to} \\ &\hat{F}(x_i,y_j) = a_{ij} \text{ and } \hat{F}(t_1,0) = \hat{F}(0,t_2) = 0 \\ &\forall t_1,t_2,\ i \text{ and } j \qquad \text{where} \\ &J(\hat{f}) = \int_0^1 \left(\frac{d}{ds}\hat{f}_1(s_1)\right)^2 + \int_0^1 \left(\frac{d}{dt}\hat{f}_2(t_2)\right)^2 + \int_0^1\int_0^1 \left(\frac{d}{ds}\frac{d}{dt}\hat{f}(s_1,t_2)\right)^2. \end{aligned}\right\}$$

(1.2)

In this first section, we derive a solution to this problem allowing for the possibility of unequal spacing. Later on, we specialize to the equal spacing case.

Let $L^2(D)$ be the set of measurable functions square integrable on the unit square. Use the definition of absolute continuity for functions of two variables given in Sard [13] pp. 534. Using equation (3.7) of Mansfield [14] we define C(D) to be the set of continuous functions on the unit square, D, and the Hilbert space H of functions satisfying

$$\left.\begin{aligned}
&\frac{\partial}{\partial t_1}^{i} \frac{\partial}{\partial t_2}^{j} F \;\varepsilon C(D) \text{ for } i\leq 1,\ j\leq 1 \\
&\frac{\partial}{\partial t_1} F(t_1,1) \text{ is absolutely continuous, } \frac{\partial^2}{\partial t_1^2} F(t_1,1)\varepsilon L^2(0,1) \\
&\frac{\partial}{\partial t_2} F(1,t_2) \text{ is absolutely continuous, } \frac{\partial^2}{\partial t_2^2} F(1,t_2)\varepsilon L^2(0,1) \\
&\frac{\partial}{\partial t_1} \frac{\partial}{\partial t_2} F \text{ is absolutely continuous } \left(\frac{\partial}{\partial t_1}\right)^2 \left(\frac{\partial}{\partial t_2}\right)^2 F \;\varepsilon L^2(D) \\
&F(0,t_2) = F(t_1,0) = 0
\end{aligned}\right\} \quad (1.3)$$

with norm given by (1.2) and

$$\langle F,F\rangle = J(f) + F(1,1)^2 \qquad (1.4)$$

In prop. 2.4 below, it will be shown that H is a Reproducing Kernel Hilbert Space (RKHS). By definition H is a RKHS of

functions of two variables if it has a reproducing kernel (RK). By definition, a function $Q(s_1,s_2,t_1,t_2)$ is the reproducing kernel for H if it has the following properties:

(a) $\qquad Q(s_1,s_2,\cdot,\cdot)\varepsilon H$

By this, we mean that, if we fix the values s_1 and s_2, then $Q(s_1,s_2,t_1,t_2)$ is a function of t_1 and t_2. This function of t_1 and t_2 is in H. (1.5)

(b) $\qquad \langle G,Q(s_1,s_2,\cdot,\cdot)\rangle = G(s_1,s_2)$.

This relationship holds for each G in H and each s_1 and s_2 in D.

If we assume that $f(s_1,s_2) = 0$ if s_1 and s_2 are not in D, this implies in particular (from (1.1) and (1.2) that

$$F(1,1) = 1.$$

Thus any solution to the problem described by (1.2) will have $F(1,1) = 1$ and hence (by (1.4))

$$\langle F,F\rangle = J(F) + 1.$$

This implies that (1.2) is equivalent to the following problem:

$$\text{minimize } \langle \hat{F},\hat{F}\rangle$$

subject to $\hat{F}(x_i,x_j) = a_{ij}$, $\hat{F}(t_1,0) = \hat{F}(0,t_2) = 0\ \forall t_1,t_2$

Using the same reasoning as in [1] proposition 1.2 we obtain

<u>Prop. 1.1</u> If Q is the reproducing kernel for H defined by (1.3), (1.4) and (1.5) then the solution to the problem defined by (1.2) is given by

$$\hat{F}(s_1,s_2) = \sum_{i=1}^{\ell_1}\sum_{j=1}^{\ell_2} d_{ij}' Q(x_i,y_j,s_1,s_2)$$

where

$$[d_{ij}'] = [Q(x_i,y_j,x_k,y_m)]^{-1}[a_{km}]$$

$$\hat{f}(s_1,s_2) = \frac{\partial}{\partial s_1}\frac{\partial}{\partial s_2}\hat{F}(s_1,s_2) = \sum_{i=1}^{\ell_1}\sum_{j=1}^{\ell_2} d_{ij}' \frac{\partial}{\partial s_1}\frac{\partial}{\partial s_2}Q(x_i,y_j,s_1,s_2).$$

In this discussion $[d_{ij}]$ is a vector of length ℓ_1 times ℓ_2 and the Q matrix is a square matrix with columns and rows, both of length ℓ_1 times ℓ_2. If we write some of these quantities out, they appear as follows:

$$[Q(x_i,y_j,x_k,y_m)] = \begin{bmatrix} Q(x_1,y_1,x_1,y_1) \ldots Q(x_1,y_1,x_1,y_{\ell_2}) \ldots\ldots Q(x_1,y_1,x_{\ell_1},y_{\ell_2}) \\ \vdots \\ Q(x_1,y_{\ell_2},x_1,y_1) \\ Q(x_2,y_1,x_1,y_1) \\ \vdots \\ Q(x_2,y_{\ell_2},x_1,y_1) \\ \vdots \\ Q(x_{\ell_1},y_1,x_1,y_1) \\ \vdots \\ Q(x_{\ell_1},y_{\ell_2},x_1,y_1) \ldots \qquad \ldots Q(x_{\ell_1},y_{\ell_2},x_{\ell_1},y_{\ell_2}) \end{bmatrix}$$

$$[a_{ij}]^T = [a_{11} \ldots a_{1,\ell_2} a_{21} \ldots a_{2,\ell_2} \ldots\ldots a_{\ell_1,1} \ldots a_{\ell_1,\ell_2}].$$

In the sequel we use the following notation

$$D_{ij}g(s,t) = \frac{\partial^{i+j}}{\partial s^i \partial t^j}g(s,t), \quad \Delta_{00}a_{km} = a_{km}$$

$$\Delta_{i+1,j}a_{km} = \Delta_{ij}a_{k+1,m} - \Delta_{ij}a_{km}$$

$$\Delta_{i,j+1}a_{km} = \Delta_{ij}a_{k,m+1} - \Delta_{ij}a_{km} .$$

For equal spacing

$$\Delta_{i+1,j}g(s,t) = \Delta_{ij}g(s+h,t) - \Delta_{ij}g(s,t)$$

$$\Delta_{i,j+1}g(s,t) = \Delta_{ij}g(s,t+h) - \Delta_{ij}g(s,t) .$$

For example

$$\Delta_{12}g(s,t) = g(s+h,t+2h) - 2g(s+h,t+h) + g(s+h,t) - (g(s,t+2h) - 2g(s,t+h) + g(s,t)). \quad (1.6)$$

This notation will be extended to higher dimensions in an obvious manner.

Following in a manner similar to [1], we note that in the case of equal spacing, formulas will become simpler if we reformulate the problem described by (1.2) in terms of second differences. In terms of the notation defined by (1.6), the equivalent reformulated problem becomes:

$$\left.\begin{array}{l} \text{minimize } J(\hat{f}) \text{ subject to} \\ \hat{F}(t_1,0) = \hat{F}(0,t_2) = 0 \;\forall t_1,t_2,\ \hat{F}(1,1) = 1 \\ \Delta_{20}\hat{F}(x_i,1) = \Delta_{20}a_{i,\ell_2-1}, \Delta_{02}\hat{F}(1,y_j) = \Delta_{02}a_{\ell_1-1,j} \\ \text{and } \Delta_{22}\hat{F}(x_i,y_j) = \Delta_{22}a_{ij} \\ \text{where } 0\le i\le \ell_1-2,\ 0\le j\le \ell_2-2 \ . \end{array}\right\} \quad (1.7)$$

Now let

$$\left.\begin{array}{l} C_{ij} = \Delta_{22}a_{ij}/h_i{}^2h_j{}^2 \quad 0\le i\le \ell_1-2,\ 0\le j\le \ell_2-2 \\ C_{i,\ell_2-1} = \Delta_{20}a_{i,\ell_2-1}/h_i{}^2 \quad C_{\ell_1-1,j} = \Delta_{02}a_{\ell_1-1,j}/h_j \\ C_{\ell_1-1,\ell_2-1} = 1 \ . \end{array}\right\} \quad (1.8)$$

For the sequel we will assume equal spacing. Thus we will let

$$h_i = 1/\ell_i \text{ and } x_i = ih_1,\ y_i = ih_2 \ . \quad (1.9)$$

Let

$$\left.\begin{array}{l} \alpha_{ij}(s_1,s_2) = \Delta_{2200}Q(ih_1,jh_2,s_1,s_2)/h_1{}^2h_2{}^2 \\ \text{for } 0\le i\le \ell_1-2,\ 0\le j\le \ell_2-2 \\ \alpha_{i,\ell_2-1}(s_1,s_2) = \Delta_{2000}Q(ih_1,1,s_1,s_2)/h_1{}^2 \\ \alpha_{\ell_1-1,j}(s_1,s_2) = \Delta_{0200}Q(1,jh_2,s_1,s_2)/h_2{}^2 \\ \alpha_{\ell_1-1,\ell_2-1}(s_1,s_2) = Q(1,1,s_1,s_2) \end{array}\right\} \quad (1.10)$$

Now exactly as in [1] proposition 1.2 we obtain

Prop. 1.2 If Q is the reproducing kernel for H defined by (1.3) then the solution to the problem defined by (1.2) is given by

$$\hat{F}(s_1,s_2) = \sum_{i=0}^{\ell_1-1} \sum_{j=0}^{\ell_2-1} d_{ij}\alpha_{ij}(s_1,s_2) \text{ where}$$

$$[d_{ij}] = [\langle\alpha_{ij},\alpha_{km}\rangle]^{-1}[C_{ij}]$$

$$\hat{f}(s_1,s_2) = \sum_{i=0}^{\ell_1-1} \sum_{j=0}^{\ell_2-1} d_{ij}D_{11}\alpha_{ij}(s_1,s_2)$$

where α_{ij} and C_{ij} are defined by (1.8) and (1.10).

Notation In the above statement the quantity

$$\left.\begin{array}{l} [\langle\alpha_{ij},\alpha_{km}\rangle] \\ \text{is a matrix whose ijkn'th element is} \\ \langle\alpha_{ij},\alpha_{kn}\rangle \ . \end{array}\right\} \quad (1.1$$

This matrix is a square matrix with rows and columns both of length ℓ_1 times ℓ_2 where

$$0\le i\le \ell_1-1,\ 0\le j\le \ell_2-1 \ .$$

The matrix can be written out in exactly the same manner as the matrix of proposition 1.1.

This proposition will give the solution to the problem once the reproducing kernel is computed. This computation will be performed in the next section.

Computational Form using Reproducing Kernels

The goal of this section is to obtain a computational form of the density estimate. It will be shown first, in proposition 2.4, that the reproducing kernel for H, defined by equations (1.3), (1.4) and (1.5) has the form

$$Q(s_1,s_2,t_1,t_2) = K(s_1,t_1)K(s_2,t_2) \qquad (2.1)$$

where K(s,t) is defined by (2.12). This fact is used to obtain a computational form for the density estimate $\hat{f}(s_1,s_2)$. This computational form is given in Theorem 2.1. Following the proof of proposition 2.5 there is a short remark indicating how these results can be extended to higher dimensional space.

In the next section we will discuss how splines might be used for forming an alternative computational procedure.

In this section, tensor product notation will be useful. Let H_1 and H_2 be two separable Hilbert spaces with complete orthonormal systems (COS), $\{\ell_{1i}\}$ and $\{\ell_{2i}\}$ respectively. If f_1 and f_2 are functions in H_1 and H_2 respectively then define $h = f_1 \otimes f_2$ by the equation

$$h(s,t) = f_1(s)f_2(t) \ . \qquad (2.2)$$

Now, following Parzen [15] section 6 define the direct product by

$$H_1 \otimes H_2 = \{h : h = \textstyle\sum_{ij} a_{ij}\ell_{1i} \otimes \ell_{2j} \text{ with } \sum a_{ij}{}^2 < \infty\}$$

with inner product determined by the norm (2.3)

$$||\textstyle\sum_{ij} a_{ij}\ell_{1i} \otimes \ell_{2j}||^2 = \sum_{ij} a_{ij}{}^2$$

Prop. 21 If f_1 and g_1 are in H_1 and f_2 and g_2 are in H_2, then

$$<f_1 \otimes f_2, g_1 \otimes g_2> = <f_1,g_1>_1 <f_2,g_2>_2 \qquad (2.4$$

The inner product <,> is the only inner product which can be defined on the functions in $H_1 \times H_2$ to satisfy (2.4).

Proof If $f_i = \ell_{1i}$ and $g_j = \ell_{2j}$ then applying (2.3) directly yields

$$<\ell_{1i} \otimes \ell_{2j}, \ell_{1k} \otimes \ell_{2m}> = \delta_{ik}\delta_{jm} = <\ell_{1i},\ell_{1k}><\ell_{2j},\ell_{2m}>.$$

Thus (2.4) holds for elements of the COS.
Equation (2.4) can be shown in general using linearity and taking limits. Uniqueness is not difficult to show.

QED

Prop. 2.2 If H_1 and H_2 are reproducing kernel spaces then $H_1 \otimes H_2$ is a RKHS with reproducing kernel given by

$$Q(s_1,s_2,t_1,t_2) = K_1(s_1,t_1)K_2(s_2,t_2) \ . \qquad (2.$$

Proof From proposition 2.1 and equation (2.2) we obtain

$$\langle \ell_{1i} \otimes \ell_{2j}, K_1(s_1,\cdot) \otimes K_1(s_2,\cdot)\rangle = \langle \ell_{1i}, K_1(s_1,\cdot)\rangle\langle \ell_{2j}, K_2(s_2,\cdot)\rangle$$

$$= \ell_{1i}(s_1)\ell_{2j}(s_2) \; .$$

Since this result holds for each element of a complete orthonormal system for $H_1 \otimes H_2$, it can be shown using linearity and taking limits to hold for all $H_1 \otimes H_2$

$$\langle h, K_1(s_1,\cdot) \otimes K_2(s_2,\cdot)\rangle = h(s_1,s_2)$$

This plus the fact that $K_1(s_1,\cdot) \otimes K_2(s_2,\cdot)$ is in $H_1 \otimes H_2$ proves the result.

QED

Given a COS, the matrix of a linear operator on a separable Hilbert space can be defined using equations which are completely analogous to the equations for the finite dimensional case (see Akhiezer and Glazman [16] volumn one, section 26, page 48). These equations will be applied to H_1, H_2 and $H_1 \otimes H_2$ using the complete orthonormal systems, $\{\ell_{1i}\}$, $\{\ell_{2j}\}$ and $\{\ell_{1i} \otimes \ell_{2j}\}$, in order to obtain the matrices which will be discussed below.

If B and B* are bounded linear operators on H_1 and H_2 then a new linear operator $B \otimes B^*$ can be defined on $H_1 \otimes H_2$ to satisfy

$$(B \otimes B^*)(f \otimes g) = Bf \otimes B^*g \quad \forall f \in H_1,\ g \in H_2 \quad . \tag{2.6}$$

If the matrix of B has elements, $\{b_{ij}\}$ and the matrix for B* has elements, $\{b_{ij}^*\}$ then it can be shown that the matrix for $B \otimes B^*$ has elements $\{b_{ij}b_{km}^*\}$. In finite dimensions this matrix is called the kroneker product and has the form:

$$[b_{ik}] \otimes [b_{jm}] = \begin{bmatrix} b_{11}b_{11}^* \cdots b_{11}b_{1M}^* \cdots b_{1K}b_{11}^* \cdots b_{1K}b_{1M}^* \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ b_{11}b_{J1}^* \\ \vdots \\ b_{21}b_{11} \\ \vdots \\ b_{21}b_{J1}^* \\ \vdots \\ b_{I1}b_{11}^* \qquad\qquad\qquad \vdots \\ b_{I1}b_{J1}^* \cdots \qquad\qquad b_{IK}b_{JM}^* \end{bmatrix} \tag{2.7}$$

It is not difficult to show that, if B and B* are invertible, then

$$(B \otimes B^*)^{-1} = (B^{-1}) \otimes (B^{*-1}) \tag{2.8}$$

The following interesting proposition is not difficult to prove.

<u>Prop. 2.3</u> Let $\{\alpha_1,\ldots,\alpha_I\}$ and $\{\beta_1,\ldots,\beta_J\}$ be sets of vectors in H_1 and H_2 respectively. Let:

$$V_1 = L\{\alpha_1,\ldots,\alpha_I\},\ V_2 = L\{\beta_1,\ldots,\beta_J\}$$
$$V = L\{\alpha_i \otimes \beta_j : 1 \le i \le I,\ 1 \le j \le J\}$$

By this we mean that V_1 is the subspace spanned by the vectors $\{\alpha_1,\ldots,\alpha_I\}$ V_2 and V are analogously defined by the above equations. Let P_1 be the projection operator from H_1 onto V_1. Let P_2 be similarly defined for V_2. Let P be the projection operator from H onto V. Then

$$P = P_1 \otimes P_2. \tag{2.9}$$

Proposition 2.2 of this paper and proposition 2.1 of [1] will be used to obtain the reproducing kernel of H, defined by equations (1.3) and (1.4). Proposition 2.1 of [1] will be restated here as

Lemma 2.1 Let

$$C^m = \{f : f^{(m)}(x) \text{ is absolutely continuous}\}$$

$$H_1 = \{h \varepsilon C' : h(0) = 0 \text{ and } h'' \varepsilon L_2(0,1)\} \tag{2.10}$$

with inner product defined by the norm

$$\langle h,h\rangle = h(1)^2 + \int_0^1 (h''(t))^2 dt \quad . \tag{2.11}$$

Then H_1 has reproducing kernel given by

$$\left.\begin{aligned} K(s,t) &= K'(s,t) + K''(s,t) \text{ where} \\ K'(s,t) &= K'(t,s) = \frac{-st^2}{2} - \frac{s^3}{6} \text{ for } s \leq t \\ K''(s,t) &= \frac{1}{6}(ts^3 + st^3) + \frac{4}{3}st \quad . \end{aligned}\right\} \tag{2.12}$$

The goal of the next section is to show that $H = H_1 \otimes H_2$ where H is defined by (1.3). This will enable us to use proposition 2.2 and lemma 2.1 to obtain the reproducing kernel (RK) for H.

Lemma 2.2 $\langle g_1 \otimes g_2, h_1 \otimes h_2\rangle = \langle g_1,h_1\rangle_1 \langle g_2,h_2\rangle_1$

where $\langle,\rangle$ is the inner product for H defined by (1.4).

Proof Equations (1.4), (1.6), (1.9) and (2.11) imply

$$\langle g_1 \otimes g_2, h_1 \otimes h_2\rangle = \int_0^1 g_1''(s_1)h_1''(s_1)g_2(1)h_2(1)ds_1$$

$$+ \int_0^1 g_1(1)h_1(1)g_2''(s_2)h_2''(s_2)ds_2 + \int_0^1\int_0^1 g_1''(s_1)h_1''(s_1)g_2''(s_2)h_2''(s_2)ds_1 ds_2$$

$$+ f_1(1)g_1(1)f_2(1)g_2(1)$$

$$= \Pi_{i=1}^2 \left(\int_0^1 g_i''(s_i)h_i''(s_i)ds_i + g_i(1)h_i(1)\right)$$

$$= \langle g_1,h_1\rangle_1 \langle g_2,h_2\rangle_1 \quad .$$

QED.

Prop. 2.4 The reproducing kernel (RK) for H, defined by (1.3), is

$$Q(s_1,s_2,t_1,t_2) = K(s_1,t_1)K(s_2,t_2)$$

where K(s,t) is defined by (2.12).

Proof This result is almost identical to Theorem 1 on page 118 of Mansfield [9]. The same method can be used for this result. Alternatively, with the help of proposition 2.1 and lemma 2.2, one can show that H, defined by (1.3), is equal to $H_1 \otimes H_1$ defined by (2.3) and (2.10). By lemma 2.2, we know K is the RK for H_1. Once this is known the result follows from proposition 2.2.

QED

Remarks Dr. Lois Mansfield has two other references giving, among other things reproducing kernels on bivariate spaces. See references [17] and [18]. Dr. Carl de Boor has an interesting discussion on the subject of tensor products and splines. See reference [2].

In the next section, the reproducing kernel, computed in proposition 2.4 will be used to obtain a computational form for the density estimate, which was given in proposition 1.2.

Proposition 2.5 The solution to the density estimation problem defined by (1.2) is given by

$$\left.\begin{aligned} \hat{F}(s_1,s_2) &= \sum_{i=0}^{\ell_1-1} \sum_{j=0}^{\ell_2-1} d_{ij}\rho_{1i}(s_1)\rho_{2j}(s_2) \\ \text{where} \quad [d_{ij}] &= ([<\rho_{1i},\rho_{1k}>_1]^{-1} \otimes [<\rho_{2j},\rho_{2m}>_1]^{-1})[C_{km}] \end{aligned}\right\} \quad (2.13)$$

and

$$\rho_{\mu\nu}(t) = \begin{cases} t((\nu+1)h_\mu-1) \text{ if } t\leq\nu h_\mu \\ ((t-\nu h_\mu)^3/6h_\mu^2) + t(h_\mu(\nu+1)-1) \text{ if } \nu h_\mu\leq t\leq(\nu+1)h_\mu \\ ((h_\mu(\nu+2)-t)^3/6h_\mu^2) + (t-1)h_\mu(\nu+1) \text{ if } (\nu+1)h_\mu\leq t\leq(\nu+2)h_\mu \\ (t-1)h_\mu(\nu+1) \text{ if } t\geq(\nu+2) \end{cases}$$

for $\mu = 1,2$, and $\nu = 0,1,\ldots,\ell_\mu-2$

$$\rho_{\mu,\ell_\mu-1}(t) = K(1,t) \quad (2.14)$$

where K is defined by (2.12); C_{ij} and h_μ are defined by equations (1.1), (1.8) and (1.9). The notation of equations (1.6), (1.11) and (2.7) is used. The density estimate is obtained by differentiation.

Proof Equation (2.14) of thispaper plus (1.6), (1.11) and (3.8) of [1] imply

$$\rho_{\mu\nu}(t) = \left\{\begin{aligned} &\Delta_{20}K(\nu h_\mu,t)/h_\mu^2 \quad \text{for } \nu = 0,1,\ldots,\ell_\mu-2 \\ &K(1,t) \text{ if } \nu = \ell_\mu-1 \end{aligned}\right\}. \quad (2.1$$

Equations (1.6), (1.10), and (2.15) plus proposition 2.4 can be combined to obtain

$$\alpha_{ij}(s_1,s_2) = \rho_{1i}(s_1)\rho_{2j}(s_2). \quad (2.1$$

Equation (2.16), proposition 2.1, plus equation (2.7) imply

$$\begin{aligned} [<\alpha_{ij},\alpha_{km}>]^{-1} &= [<\rho_{1i},\rho_{1k}>_1<\rho_{2j},\rho_{2n}>_1]^{-1} \\ &= ([<\rho_{1i},\rho_{1k}>_1]\otimes[<\rho_{2j},\rho_{2n}>_1])^{-1} \quad . \end{aligned}$$

Combine with (2.8) to obtain

$$[\langle\alpha_{ij},\alpha_{kn}\rangle]^{-1} = [\langle\rho_{1i},\rho_{1k}\rangle_1]^{-1} \otimes [\langle\rho_{2j},\rho_{2m}\rangle_1]^{-1} \quad (2.17)$$

Combine proposition 1.2 plus equations (2.16) and (2.17) to complete the proof.

QED

Remark Proposition 2.2 is obtained from proposition 1.2 as a result of substituting the reproducing kernel and computing as above. A similar result could have been obtained stating from proposition 1.1.

It is easy to see how equation (2.13) might be extended to higher dimensions. For example, in three dimensions:

$$F(s_1,s_2,s_3) = \sum_{i=1}^{\ell_1-1}\sum_{j=1}^{\ell_2-1}\sum_{k=1}^{\ell_3-1} d_{ijk}\rho_{1i}(s_1)\rho_{2j}(s_2)\rho_{3j}(s_3)$$

where

$$[d_{ijk}] = ([\langle\rho_{1i},\rho_{1j}\rangle_1]^{-1} \otimes [\langle\rho_{2k},\rho_{2m}\rangle_1]^{-1} \otimes [\langle\rho_{3\mu},\rho_{3\nu}\rangle_1]^{-1})[C_{ijk}]$$

and the quantity, C_{ijk}, could be defined in a manner analogous to the definition of C_{ij}.

In the next section the material from [1] is used to eliminate the matrix inversions required by equation (2.13).

Lemma 2.3 $\quad [\langle\rho_{\mu\nu},\rho_{\mu\omega}\rangle] = \begin{bmatrix} A_\mu & 0 \\ 0 & 1 \end{bmatrix}$

where

$$A_{\mu\nu\omega} = \begin{cases} 0 & \text{if } |\nu-\omega| > 1 \\ 2/3h_\mu & \text{if } \nu = \omega \\ 1/6h_\mu & \text{if } |\nu-\omega| = 1 \end{cases} \text{ and } A_\mu = [A_{\mu\nu\omega}] .$$

Proof From {4} lemma 3.8.

QED

Theorem 2.1 The solution to the density estimation problem defined by (1.2) is given by

$$\hat{F}(s_1,s_2) = s_1s_2 + 6s_1h_2\sum_{j,m\le\ell_2-2}\rho_{2j}(s_2)B_{2jm}C_{\ell_1-1,m}$$

$$+ 6h_1s_2\sum_{i,k\le\ell_1-2}\rho_{1i}(s_1)B_{1ik}C_{k,\ell_2-1}$$

$$+ 36h_1h_2\sum_{ijkm}\rho_{1i}(s_1)\rho_{2j}(s_2)B_{1ik}B_{2jm}C_{km}$$

where

$$\left.\begin{aligned} B_{\mu\nu\omega} &= -(n_\nu n_{\ell_\mu-\omega-2})/n_{\ell_\mu-1} \\ n_0 &= 1,\ n_1 = -4,\ n_\nu = -(4n_{\nu-1}+n_{\nu-2}) \end{aligned}\right\} \quad (2.18)$$

$$\hat{f}(s_1,s_2) = 1 + 6h_2\sum_{jm}\frac{\partial}{\partial s_2}\rho_{2j}(s_2)B_{2jm}C_{\ell_1-1,m} + 6h_1\sum_{ik}\frac{\partial}{\partial s_1}\rho_{1i}(s_1)B_{1ik}C_{k,\ell_2-1}$$

$$+ 36h_1h_2\sum_{ijkm}\frac{\partial}{\partial s_1}\rho_{1i}(s_1)\frac{\partial}{\partial s_2}\rho_{2j}(s_2)B_{1ik}B_{2jm}C_{kn} \quad (2.19)$$

where $\{C_{ij}\}$ is defined in equations (1.8), (1.6) and (1.1) and $\{\rho_{ij}\}$ is defined by equation (2.14).

Proof Let $A_\mu = [A_{\mu\nu\omega}]$, $B_\mu = [B_{\mu\nu\omega}]$, where $A_{\mu\nu\omega}$ and $B_{\mu\nu\omega}$ is defined by lemma 2.3 and equation (2.18). Then by [1] equations (3.3), (3.5) and (3.13) we have

$$A_\mu^{-1} = 6h_\mu B_\mu \quad . \quad (2.20)$$

Combine with proposition 2.5, lemma 2.3 plus equation (2.7) to obtain the result.

QED

Remark Theorem 2.1 gives a computational procedure for the density estimate, eliminating the need for matrix inversion. For computer programming efficiency, a calculation such as that of [1] equation (3.14) may be advisable.

Relationship with Splines

We start with a proposition designed to give the reader an intuitive feel for the estimate. We show that if it happens that we can write

$$\hat{F}(ih_1,jh_2) = a_i b_j$$

then our estimate is a product of two unidimensional natural cubic splines. Next we show that our estimate is the unique natural bicubic spline interpolating the data. We will then indicate the reason that this implies that our estimate also minimizes (not uniquely)

$$\int_0^1\int_0^1 \left(\frac{\partial}{\partial s_1}\frac{\partial}{\partial s_2} f(s_1,s_2)\right)^2 ds_1 ds_2$$

subject to the interpolation constraints. We next compare the density estimate in this paper to the procedure described by Schoenberg and de Boor [2]. It will be seen that the two estimates differ only in endpoint conditions. A computer program based on [2] is available from Schoenberg and de Boor, who are part of the faculty of the University of Wisconsin, Madison, Math. Research Center. It would probably not be

difficult to modify that program to compute the estimate described in this paper. This would provide an alternative computational procedure to the one described in this paper.

We start by proving that intuitive result described above.

Proposition 3.1 Suppose in the problem described by (1.1) and (1.2), we have the additional assumption that

$$\hat{F}(ih_1,jh_2) = a_{ij} = b_{1i}b_{2j} (\text{i.e. } a_{ij} = a_i b_j). \qquad (3.1)$$

Then the density estimate has the form

$$\hat{F}(s_1,s_2) = \hat{F}_1(s_1)\hat{F}_2(s_2)$$

where $\hat{F}_1(s_1)$ and $\hat{F}(s_2)$ are natural cubic splines with knots $\{ih_1\}$ and $\{jh_2\}$ respectively satisfying

$$\hat{F}(ih_1) = b_{1i} \text{ and } \hat{F}(ih_2) = b_{2i} .$$

Proof Equations (2.13) and (3.1) imply

$$\hat{F}(s_1,s_2) = \hat{F}_1(s_1)\hat{F}_2(s_2) \text{ where} \qquad (3.2)$$

$$\hat{F}_i(s_i) = \textstyle\sum_j d_{ij}^{!}\rho_{ij}(s_i) \text{ and}$$

$$d_{ij}^{!} = [\langle\rho_{ij},\rho_{km}\rangle]^{-1}[b_{km}] \qquad . \qquad (3.3)$$

However, equation (3.3) combined with [1], propositions 1.1 and 1.3 imply that $F_i(s_i)$ is the appropriate natural cubic spline. This fact combined with (3.2) completes the proof.

QED

Remark Although the above proof is given in terms of equal spacing, the extension to unequal spacing is simply a matter of change of notation.

Proposition 3.1 leads one to speculate about the possibility of applying the methods of the Schoenberg and de Boor paper [2] to the estimation problem described in this paper. To this end, the following proposition is of interest.

Prop. 3.2 $\hat{F}(s_1,s_2)$ is a natural bicubic spline satisfying

$$\hat{F}(x_i,x_j) = a_{ij} \quad (3.4)$$

Proof Equation (3.4) follows from (1.2). Also, propositions 1.1 and 2.4 plus equation (2.12) imply

$$\left.\begin{array}{l} \hat{F}_{ij} \text{ is continuous for } i \text{ and } j \text{ less than or equal to two:} \\ \hat{F}_{20}(0,y_j) = \hat{F}_{20}(1,y_j) = \hat{F}_{02}(x_i,0) = \hat{F}_{02}(x_i,1) = 0 \\ \hat{F}_{22}(0,0) = \hat{F}_{22}(0,1) = \hat{F}_{22}(1,0) = \hat{F}_{22}(1,1) = 0 \end{array}\right\} \quad (3.5)$$

$$\text{where } \hat{F}_{ij}(x,y) = \frac{\partial^{i+j}}{\partial x^i \partial y^j}\hat{F}(x,y). \quad (3.6)$$

Finally the same propositions plus equation (2.12) imply that $\hat{F}(x,y)$ is bicubic within each rectangular subsubdivision. In other words we can write

$$\left.\begin{array}{l} \hat{F}(x,y) = \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ijkm}x^i y^j \\ \text{for } x_k<x\leq x_{k+1} \text{ and } y_m<y\leq y_{m+1} \end{array}\right\}. \quad (3.7)$$

Equations (3.4), (3.5) and (3.7) imply that $\hat{F}(x,y)$ is a natural bicubic spline satisfying (3.9) (see Hall [19] equation (16)).

QED

Theorem 3.1 The density estimate described in this paper also minimizes the quantity

$$\int_0^1\int_0^1 \left(\frac{\partial}{\partial x}\frac{\partial}{\partial y}\hat{f}(x,y)\right)^2 dxdy \quad (3.8)$$

subject to the constraints.

Proof Using (3.6), note that

$$F_{22}(x,y) = f_{11}(x,y). \quad (3.9)$$

The result follows directly from proposition 3.2 and Ahlberg, Nilson and Walsh [20] theorem 7.6.1, page 243.

QED

Remark In the terminology of [20], since $\hat{F}(x,y)$ satisfies (3.4), (3.5) and (3.7) it follows that $\hat{F}(x,y)$ is a "type II' spline of interpolation of the values a_{ij}" (see [14]page 236). It seems possible to infer from the statement of theorem 7.6.1 of [20] that $\hat{F}(x,y)$ uniquely minimizes (3.8). This cannot be true. As a counter example let

$$\left.\begin{array}{l} F^*(x,y) = \hat{F}(x,y) + G(x,y) \\ \text{where } G(x,y) = x\sin(\frac{\pi y}{h_2}) \\ x_i = ih_1 \quad y_j = jh_2 . \end{array}\right\} \quad (3.10)$$

Then (3.4), (3.5) and (3.10) imply that (3.4) and (3.5) are satisfied by F^*. However (3.10) implies that

$$\int_0^1\int_0^1 \hat{F}_{22}(x,y)^2 dxdy = \int_0^1\int_0^1 F^*_{22}(x,y)^2 dxdy.$$

Thus $\hat{F}(x,y)$ is not the only function minimizing (3.8) subject to (3.4) and (3.5).

In defining a natural bicubic spline, Hall ([19] equation (16)) claims uniqueness based on Theorem 7.7.1 of [20]. In [20] the uniqueness proof used in theorem 7.7.1 is essentially the same as the apparently mistaken (in view of the above counter-example) uniqueness proof used in theorem 7.6.1. Thus we prefer to base our claim of uniqueness on the methods of de Boor [2] as described in the appendix of this paper.

Theorem 3.2 $\hat{F}(s_1,s_2)$ is the unique natural bicubic spline satisfying

$$\hat{F}(x_i,y_j) = a_{ij}.$$

Proof Proposition 3.2 implies that $\hat{F}$ is a bicubic spline satisfying the data. Uniqueness follows from appendix theorem A.1.

QED

Remark Schoenberg [2] describes two density estimates. The estimate most similar to the one in this paper is defined by [2] equation (5.10). In the terminology of this paper the corresponding distribution estimate described by that paper satisfies (3.4) and (3.7). However (3.5) of this paper is replaced by continuity plus the following condition

$$f(0,y) = f(1,y) = f(x,0) = f(x,1) = 0. \qquad (3.11)$$

Thus it seems probable that the computation methods used

there could be modified to apply to the problem described in this paper.

Remark It can be shown that (3.11) can be replaced by the conditions

$$\left.\begin{aligned} F_{10}(0,y_j) = F_{10}(1,y_j) = F_{01}(x_i,0) = F_{01}(x_i,1) = 0 \\ F_{11}(0,0) = F_{11}(0,1) = F_{11}(1,0) = F_{11}(1,1) = 0 \ . \end{aligned}\right\} \qquad (3.12)$$

In the terminology of [20] the Schoenberg and de Boor distribution estimate is the unique "type I' spline of interpolation of the values a_{ij}." The only difference between the Schoenberg and de Boor estimate and our estimate is that in the former case continuity and (3.12) hold instead of (3.5)

Error Bounds on the Density Estimate

In this section it will be shown first in Theorem 4.1 that

$$E|\hat{f}_n(s)-f(s)|^2 \le 2||F||^2(h_1+h_2) + 2K'/(nh_1h_2) \ . \qquad (4.1)$$

The first term on the right side of (4.1) is the "bias" error and the second term is the "variance" error. By choosing

$$h_i = K_i'''/n^{1/3}$$

we can obtain the result in Theorem 4.2 that

$$E|\hat{f}_n(s)-f(s)|^2 \leq K''/n^{1/3} \qquad (4.2)$$

where f_n is the two dimensional density estimate based on n observations.

The methods used will be similar to the methods used in the one dimensional case (see [1] section 4). Using methods similar to those used to derive equation (4.8) of [1] it is possible to obtain the following result

$$E|\hat{f}_n(s)-f(s)|^2 \leq E|D_{11}L(s,a^*)-f(s)|^2 + E|D_{11}L(s,a')|^2$$

$$\text{where } s = \{s_1,s_2\},\ a = [a_{ij}], a' = [a_{ij}],\ a^* = [a^*_{ij}] \qquad (4.3)$$

$L(s,a) = \hat{F}(s)$ defined by proposition 1.1 and D_{11} defined by equation (1.6), and

$$\left.\begin{aligned} a^*_{ij} &= F(ih_1,jh_2) \\ a'_{ij} &= a_{ij} - a^*_{ij} \quad \text{(see (1.1))} \end{aligned}\right\} \qquad (4.4)$$

It can be shown that

$$L(s,a') = L(s,a) - L(s,a^*) = \hat{F}(s) - E\hat{F}(s)$$

$$L(s,a^*) - F(s) = E\hat{F}(s) - F(s).$$

Thus, it can be seen by differentiation that the first term on the right hand side of (4.3) is the <u>bias</u> error and the

second term is the <u>variance</u>. The intuitive interpretation of these terms is similar to the intuitive interpretations in the one dimensional case (see [1] section 4).

The method used will be to find bounds on the variance and the bias terms separately. Those proofs which are similar to the corresponding proof in [1] will not be repeated here.

Since the bias term is simpler to examine, it will be examinated first.

<u>Lemma 4.1</u> Let g be any function in H. Then

$$D_{11}g(s_1,s_2) = \langle g, D_{1100}Q(s_1,s_2,\cdot,\cdot)\rangle .$$

<u>Proof</u> Apply (1.5) and use the notation of (1.6). Take limits in the usual manner using the fact that strong convergence implies weak convergence. See Kimeldorf and Wahba [10].

QED

<u>Lemma 4.2</u> $L(s,a^*)$ is the projection of F on to V where V is the subspace spanned by the vectors

$$Q(0,0,s_1,s_2), Q(0,h_2,s_1,s_2), Q(0,2h_2,s_1,s_2), \ldots, Q(0,1,s_1,s_2),$$

$$Q(h_1,0,s_1,s_2),$$

$$\vdots$$

$$Q(1,0,s_1,s_2), \ldots \qquad \ldots, Q(1,1,s_1,s_2)$$

Proof The proof is similar to proof of [1] lemma 4.4.

QED

Lemma 4.3 Let $R(s_1,s_2,.,.)$ be the projection of $D_{1100}Q(s_1,s_2,.,.)$ on to V. Then

$$|f(s_1,s_2)-L_{1100}(s,a^*)|\leq||F||^2||R(s_1,s_2,.,.)-D_{1100}Q(s_1,s_2,.,.)||^2 .$$

Proof The proof is similar to the proof of [1] lemma 4.5.

QED

Lemma 4.4 (Wahba [21])

$$||R(s_1,s_2,.,.) - D_{1100}Q(s_1,s_2,.,.)||^2 \leq h_1 + h_2 .$$

Proof By (1.6) and proposition 2.4 we have

$$D_{1100}Q(s_1,s_2,.,.) = D_{10}K(s_1,.) \otimes D_{10}K(s_2,.) \quad . \quad (4.5)$$

Let $R'(s,\cdot)$ be the projection of $D_{10}K(s,\cdot)$ on to the subspace spanned by

$$K(0,.),K(h_1,.),...,K(1,.) .$$

Let $R''(s,.)$ be similarly defined for the vectors

$$K(0,.),K(h_2,.),...,K(1,.) .$$

Then proposition 2.3 implies that

$$R(s_1,s_2,.,.) = R'(s_1,.) \otimes R''(s_2,.) . \quad (4.6)$$

Furthermore, reasoning as in [1] lemma A.8 it can be shown that:

$$\left.\begin{array}{l} ||R'(s,.) - D_{10}K(s,.)||^2 < h_1/3 \\ ||R''(s,.) - D_{10}K(s,.)||^2 < h_2/3 \end{array}\right\} \quad (4.7)$$

Using the notation of proposition 2.3 we obtain for f_i in H_i that

$$||f_1\otimes f_2-(P_1\otimes P_2)(f_1\otimes f_2)||^2=||f_1||^2||f_2||^2-||P_1f_2||^2||P_2f_2||^2$$
$$= ||f_1||^2(||f_2-P_2f_2||^2) + ||P_2f_2||^2(||f_1-P_1f||^2). \quad (4.8)$$

Equations (4.5), (4.6), (4.7) and (4.8) imply (since $||P_2f_2|| \leq ||f_2||$)

$$||D_{11}Q(s_1,s_2,.,.)-R(s_1,s_2,.,.)||^2\leq\frac{h_2}{3}||D_{10}K(s_1,.)||^2+\frac{h_1}{3}||D_{10}K(s_2,.)||$$

Combine this equation with equation (2.12) and [1] lemma 4.3 to obtain the result.

QED

The following proposition gives the uniform bound for the bias term.

Prop. 4.1 $|f(s_1,s_2) - D_{11}L(s,a^*)| \leq (h_1+h_2)||F||^2$.

Proof Apply lemmas 4.3 and 4.4.

QED

The variance calculation uses a substantial amount of material from the corresponding one dimensional calculation as given in [1]. This material is summarized in the next two lemmas.

Lemma 4.5 Let

$$[\bar{C}_{\mu\nu\omega}] = \bar{C}_\mu = [<\Delta_{10}K(\nu h_\mu,.),\Delta_{10}K(\omega h_\mu,.)>]^{-1}$$

then if $rh_\mu < s \leq (r+1)h_\mu$ we have for any positive sequence $\{\psi_i\}$ that

$$|\sum_{\nu\omega}\psi'_\nu\bar{C}_{\mu\nu\omega}\Delta_{10}K(\omega h_\mu,s)| \leq \frac{4}{h_\mu}\sum_\nu k_{r\nu}|\psi'_\nu| \quad . \qquad (4.9)$$

$$\left.\begin{aligned} \text{where } k_{\nu\omega} &= \begin{cases} \delta_{\nu\omega}\frac{1}{4} + k^*_{\nu\omega} & \text{if } |\tilde{K}_\omega| \leq |\tilde{K}_{\omega+1}| \\ \delta_{\nu\omega}\frac{1}{4} + k^*_{\nu,\omega+1} & \text{if } |\tilde{K}_\omega| > |\tilde{K}_{\omega+1}| \end{cases} \\ k^*_{\nu\omega} &= (1/2^{|\nu-\omega+1|+1}) + (1/2^{|\nu-\omega|+1}) \\ \tilde{K}_r &= \sum_{\nu\omega}\psi_\nu{}'C_{\mu\nu\omega}D_{02}\Delta_{10}K(\omega h_\mu,rh_\mu) \end{aligned}\right\} \quad . \qquad (4.10)$$

where $\delta_{\nu\omega} = \begin{cases} 1 & \text{if } \nu = \omega \\ 0 & \text{otherwise} \end{cases}$

and the notation of (1.6) and (1.11) is used.

Proof This is a direct consequence of [1] equations (4.20), the equation above (4.20), (4.21) and (4.15) plus [1] lemma 4.11.

QED

Lemma 4.6 Let $\phi_{ij} = \Delta_{11}a_{ij} = a_{i+1,j+1} - a_{i,j+1} - a_{i+1,j} + a_{ij}$

$$\phi^*_{ij} = \Delta_{11}a^*_{ij}, \qquad \phi'_{ij} = \Delta_{11}a'_{ij}$$

then $E\phi'^2_{ij} = \frac{1}{n}\phi^*_{ij}(1-\phi^*_{ij}) \leq \frac{3h_1h_2}{n}||F||$

where a_{ij} is defined by (1.1), $a^*_{ij} = F(ih_1,jh_2)$ and $a'_{ij} = a_{ij} - a^*_{ij}$ and h_1 and h_2 are less than or equal to 1.

Proof Using the analogy between the probability of being in an interval and the probability of being in a rectangle, we obtain in a manner similar to lemma 4.12 of [1] that

$$E|\phi'_{ij}|^2 = \frac{1}{n}\phi^*_{ij}(1-\phi^*_{ij}) \quad . \qquad (4.11)$$

also, as in lemma 4.12 of [1], and using proposition 2.4 of this paper we have

$$|\phi^*_{ij}|^2 \leq ||F||^2\Delta_{11}K(ih_1,ih_1)\Delta_{11}K(jh_2,jh_2) \leq 9h_1^2h_2^2||F||^2 \text{ if } h_\mu \leq 1.$$

Combine the above equation with (4.11) to complete the proof.

QED

The following proposition gives a bound for the variance term.

Prop. 4.2

$$E|D_{11}L(s,a')|^2 \leq 3(13)^4\frac{||F||}{nh_1h_2} \quad .$$

<u>Proof</u> Using the same method of proof as in proposition 2.5 we obtain

$$D_{11}L(s,a') = \sum_{i=0}^{\ell_1-1}\sum_{j=0}^{\ell_2-1}\sum_{k=0}^{\ell_2-1}\sum_{m=0}^{\ell_2-1}\phi'_{ij}\bar{C}_{1ik}\bar{C}_{2jm}\Delta_{10}K(kh_1,s_1)\Delta_{10}K(mh_2,s).$$

Let $$B_i = \sum_{jm}\phi'_{ij}\bar{C}_{2jm}\Delta_{10}K(kh_2,s_2).$$

Then applying lemma 4.5 twice we obtain from the above equations that

$$\left.\begin{aligned} |D_{11}L(s,a')| &= |\textstyle\sum_{ik}B_i\bar{C}_{1ik}\Delta_{10}K(kh_1,s_1)| \\ &\le \frac{4}{h_1}\textstyle\sum_i k_{qi}|B_i| \le \frac{16}{h_1h_2}\sum_{ij}k_{qi}k_{rj}|\phi'_{ij}| \\ &\text{where } qh_1 \le s_1 \le (q+1)h_1 \text{ and } rh_2 \le s_2 \le (r+1)h_2 \ . \end{aligned}\right\} \tag{4.12}$$

But direct caluclation from (4.10) implies

$$\sum_\omega k_{\nu\omega} \le \frac{13}{4} \quad . \tag{4.13}$$

Also Schwartz's inequality and lemma 4.6 imply

$$E(\phi'_{ij}\phi'_{km}) \le \frac{3h_1h_2}{n}||F|| \ . \tag{4.14}$$

Combine equations (4.12), (4.13) and (4.14) to obtain the result.

QED

We are now in a position to state the first of the two main theorems of this section.

<u>Theorem 4.1</u>

$$E|f(s_1,s_2) - \hat{f}(s_1,s_2)|^2 \le (h_1+h_2)||F||^2 + 3(13)^4\frac{||F||}{nh_1h_2}$$

where f and F are the density and distribution functions, $\hat{f}$ is the density estimate based on n observations and $||F||$ is defined by equations (1.4) and (1.2). It is assumed that f = 0 outside the unit square and that $||F||$ is well defined and finite.

<u>Proof</u> Combine equation (4.3) plus propositions 4.1 and 4.2 to obtain the result.

QED

<u>Lemma 4.7</u> The function

$$g(h) = \frac{K}{nh^2} + K'h \text{ has its minimum at } h^* = \left(\frac{2K}{nK'}\right)^{1/3} \text{ and}$$

$$g(h^*) = \left(K^{1/3}K'^{2/3}\ \frac{3}{4^{1/3}}\right)/n^{1/3} \ .$$

<u>Proof</u> By simple calculus.

QED

<u>Theorem 4.2</u> If we set

$$h_1 = h_2 = \left(\frac{3(13)^4}{||F||n}\right)^{1/3}$$

then we have

$$E|f(s_1,s_2) - \hat{f}(s_1,s_2)|^2 \le (3\,(13)^4||F||^5)^{1/3}/n^{1/3} \tag{4.15}$$

<u>Proof</u> Combine theorem 4.1 and lemma 4.7 to obtain the result.

QED

Remark In equation (4.15) the fact that $||F||$ is in the denominator confirms the intuitive fact that the more rough the distribution is, the finer the grid should be.

Estimation of the Density Function Using Order Statistics

A spline procedure for univariate density estimation using order statistics has been studied by Wahba [22]. In this section we will suggest, without going into complete detail, a formula for a two dimensional generalization of a univariate natural spline density estimate using order statistics.

To introduce the subject we define briefly a one dimensional estimate using observations

$$X_1', X_2', \ldots, X_n' \quad \text{where} \quad 0 \leq X_i' \leq 1.$$

We then let $\{X_i\}$ be the order statistics based on $\{X_i'\}$ and define

$$X_0 \equiv 0 \quad \text{and} \quad X_{n+1} \equiv 1.$$

Let

$N(X_i)$ = the number of observations that are less than or equal to X_i' (note that if all observations have different values than $N(X_i) = i$. This occurs with probability one.) (5.1)

$$F^*(X_i) = N(X_i)/(n+1) \quad F^*(X_0) = 0, \; F^*(X_{n+1}) = 1 \qquad (5.2)$$

For simplicity we assume that

$$n+1 = \ell k \text{ where } \ell \text{ and } k \text{ are fixed integers.} \qquad (5.3)$$

Then the natural spline density estimate is the unique solution to the problem of minimizing

$$\int_0^1 \left(\frac{d}{dx}\hat{f}(x)\right)^2$$

subject to

$$\hat{F}(X_{jk}) = F^*(X_{jk}) \quad 0 \le j \le \ell .$$

This completes our definition of the one dimension estimate. In the two dimensional case we have observations

$$\left.\begin{array}{l} Z'_1, Z'_2, \ldots, Z'_n \text{ where } Z'_i = (X'_i, Y'_i) \\ \text{where } 0 \le X'_i \le 1 \text{ and } 0 \le Y'_i \le 1 \end{array}\right\} . \quad (5.4)$$

Introduce the partial ordering

$$Z'_i \le Z'_j \text{ if } X'_i \le X'_j \text{ and } Y'_i \le Y'_j .$$

Define

$N(Z'_i)$ = the number of observations that are less than or equal to Z'_i

$$F^*(Z'_i) = \frac{N(Z_i)}{(n+1)} . \quad (5.5)$$

So far, the analogy to the one dimensional case seems to work well. However, in contrast to the one dimensional case, there seems to be no obvious way to order the data in the two dimensional case. One method, studied by Bock Ki Kim ([23] pp. 14 through 18 and 20 through 22),

using some ideas in Wald [24] and Gessaman [6], when adapted to the unit square, can be roughly described as follows: (a) Using (5.4) compute order statistics based on the first coordinate (i.e. based on $\{X_i\}$. Use these order statistics to compute ℓ_1 coordinates along the "X" axis, exactly as in the one dimensional case. In other words, the "X" coordinates are computed exactly as in the one dimensional case, completely ignoring the second coordinates $\{Y_i\}$. (b) The "X" coordinates just computed divide the unit square into $\ell_1 + 1$ vertical strips. Each vertical strip is then subdivided into $\ell_2 + 1$ horizontal subdivisions based on the "order" statistics of the "Y" coordinates of the observations that fall within each vertical strip. A histogram type estimate is obtained based on the empirical distribution function at the corners of each of the horizontal subdivisions within the vertical strips. This estimate has the advantage of being relatively easy to analyze by examination of the marginal density of X followed by examination of the conditional distribution of Y given X. It seems to have the disadvantage of not being symmetrical in X and Y. In other words, had we reversed the role of X and Y in the above procedure, we would have come out with a different estimate. However, there is no reason why the formulas given below could not be applied analogously to the above procedure.

In the following we describe another procedure due to Wahba [21]. Suppose we order the observations based on values of F^* as defined in (5.5). This would give a well defined ordering of the data, except for the fact that usually there are a number of "ties", that is, observations X_i and X_j satisfying

$$F^*(X_i) = F^*(X_j).$$

One way to decide how to order "tied" observations is to use the distance (e.g. the Euclidian distance) from the origin as the "tie breaking" criterion. Another procedure would be let "ties" be broken by a random number table. Suppose we decide on a tie breaking procedure. Then we have "order statistics" $\{Z_i\}$ satisfying

$$Z_i = (X_i, Y_i) \text{ with } F^*(Z_i) \le F^*(Z_j) \text{ if } i \le j$$

$$Z_0 \equiv (0,0)\ ,\ Z_{n+1} = (1,1),\ F^*(Z_0) = 0,\ F^*(Z_{n+1}) = 1\ .$$

As in the one dimensional case, we assume for simplicity that (5.3) holds. Then an estimate can be chosen to minimize

$$\int_0^1 \left(\frac{d}{dx}\hat{f}_x(x)\right)^2 + \int_0^1 \left(\frac{d}{dy}\hat{f}_y(y)\right)^2 + \int_0^1\int_0^1 \left(\frac{d}{dx}\frac{d}{dy}\hat{f}(x,y)\right)^2$$

subject to

$$\hat{F}(0,y) = \hat{F}(x,0) = 0$$

$$\hat{F}(Z_{jk}) = F^*(Z_{jk})\ 0 \le j \le\ \ .$$

Using proposition 2.4 we obtain the solution (as in proposition 1.1) to be

$$\hat{F}(s,t) = \sum_{i=0}^{\ell} d_i K(X_{ik},s)K(Y_{ik},t)$$

where the function K satisfies (2.12) and

$$[d_i] = [K(X_{ik},X_{jk})K(Y_{ik},Y_{jk})]^{-1}[F^*(X_{jk},Y_{jk})] \quad (5.7)$$

where $0 \le i \le \ell$ and $0 \le j \le \ell$.

It should be noted that the matrix that is inverted in (5.7) could be ill conditioned. Thus, care would be necessary in writing a computer program based on (5.7).

As noted earlier, an equation similar to (5.7) could easily be written using Bock Ki Kim's approach. It is conjectured that (5.7) could be simplified if Dr. Kim's approach were tried. The conjectured simplified form would consist of a standard univariate spline along the "X" axis multiplied by a "conditional" univariate spline along one of the $\ell_1 + 1$ vertical strips described earlier (i.e. the vertical strip which contains the "x" coordinate of the point at which the density function is being estimated). We conjecture further that uniform error bound calculations carried out by Wahba in [17] could be generalized to the two dimensional case using the simplified form just described.

Appendix

The purpose of this appendix is to use the ideas of deBoor [2] to prove that equations (3.4), (3.5) and (3.7) uniquely determine $\hat{F}(x,y)$. We start with a vector space V with S a subspace of V and Λ a space of linear functionals. The linear interpolation problem (LIP) determined by V,S and Λ is

Given $f \in V$ find $g \in S$ so that

$$\lambda f = \lambda g \quad \forall \lambda \in \Lambda \tag{A.1}$$

(see [2] pp. 42 and 43).

The LIP is correct if given any $f \in V$ there is a unique $g \in S$ satisfying (A.1). (See [2] pp. 43).

Lemma A.1 Let V_1 and V_2 be two vector spaces having subspaces S_1 and S_2 respectively. Let Λ_1 and Λ_2 be two spaces of linear functionals. Using (2.3) define

$$V = V_1 \otimes V_2,\ S = S_1 \otimes S_2 \text{ and } \Lambda = \Lambda_1 \otimes \Lambda_2.$$

Then if the LIP determined by V_i, S_i and Λ_i is correct for i equal to one and two then the LIP determined by V,S and Λ is correct.

Remark If $\lambda_i \varepsilon \Lambda_i$ and $g_i \varepsilon S_i$ then by definition (see (2.6))

$$(\lambda_1 \otimes \lambda_2)\cdot(g_1 \otimes g_2) = (\lambda_1 g_1)\cdot(\lambda_2 g_2) \ . \tag{A.2}$$

If Λ_i has basis $\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{i,\ell_i}$ then Λ has basis

$$\lambda_{11} \otimes \lambda_{21}, \ldots, \lambda_{1,\ell_1} \otimes \lambda_{21}, \ldots, \lambda_{11} \otimes \lambda_{2,\ell_2}, \ldots, \lambda_{1,\ell_1} \otimes \lambda_{2,\ell_2} \ .$$

where $\lambda_{1i} \otimes \lambda_{2j}$ is defined by (A.2).

Proof Lemma A.1 is part of the statement of Theorem A.2 of [2].

Lemma A.2 Let Π_i be a partition of [0,1] i.e.

$$0 = s_{i1} < s_{i2} < s_{i2} \cdots < s_{i,\ell_i} = 1 \ .$$

Let $S_k(\Pi_i)$ be the class of spline functions of degree k with knots $\{s_{ij}\}$. Let $S_{kk}(\Pi_1 \times \Pi_2)$ be the class of bivariate splines of degrees k in each variable with mesh points $\{(s_{1i}, s_{2j})\}$. Then

$$S_{k,k}(\Pi_1 \times \Pi_2) = S_k(\Pi_1) \otimes S_k(\Pi_2) \ .$$

Proof See de Boor [2] appendix equation (3.3).

Remark F is in $S_{k,k}(\Pi_1 \times \Pi_2)$ if it satisfies

(a) $\dfrac{\partial^{i+j}}{\partial x^i \partial x^j} F(x,y)$ is continuous for i and j both less than or equal to k-1 (A.3)

(b) Let $R_{ij} = [s_{1i}, s_{1i+1}] \otimes [s_{2j}, s_{2,j+1}]$. Then for $(x,y) \varepsilon R_{ij}$ we can write

$$F(x,y) = \sum_{p=0}^{k} \sum_{q=0}^{k} a_{ijpq} x^p y^q \tag{A.4}$$

Lemma A.3

Let

$$\left.\begin{aligned} \lambda_{ij} F &= F(s_{ij}) \quad 0 \le j \le \ell_1 \\ \lambda_{i,\ell_i+1} F &= F''(0), \ \lambda_{i,\ell_i+2} F = F''(1) \end{aligned}\right\} . \qquad (A.5)$$

Let

$$V_i = \{h \varepsilon C' : h'' \varepsilon L_2(0,1), h''(0) = h''(1) = 0\} \qquad (A.6)$$

Let Λ_i be the space spanned by $\lambda_{i0}, \lambda_{i1}, \ldots, \lambda_{i,\ell_1+2}$ then the LIP determined by $V_i, S(\Pi_i)$ and Λ_i is correct.

Proof By the definitions given above we need only show that

$$\forall h \varepsilon V_i [\exists \text{ unique } g \varepsilon S(\Pi_i) \ni [\lambda_{ij} h = \lambda_{ij} g]].$$

In other words

given $f \varepsilon V_i$ there is a unique $g \varepsilon S(\Pi_i)$ such that

$$h(s_{ij}) = g(s_{ij}) \quad 0 \le j \le \ell_i \qquad (A.7)$$

$$h''(0) = g''(0) = 0, \ F''(1) = g''(1) = 0 \ . \qquad (A.8)$$

But the requirement that $g \varepsilon S(\Pi_i)$ and satisfy (A.8) is simply the requirement that f be a natural spline. Thus our lemma is true if we can show that there is a unique natural spline satisfying (A.7). But the uniqueness of cubic natural spline interpolation is well known (see for example, Greville [25] theorem 6.2). QED

Before stating the main result we rewrite (3.4), (3.5) and (3.7) here for convenience

$$\hat{F}(x_i, y_j) = a_{ij} \qquad (A.9)$$

$$\hat{F}(x,y) \text{ satisfies (A.3)} \qquad (A.10)$$

$$\hat{F}_{20}(0,y_j) = \hat{F}_{20}(1,y_j) = \hat{F}_{20}(x_i,0) = \hat{F}_{02}(x_i,1) = 0 \qquad (A.11)$$

$$\hat{F}_{22}(0,0) = \hat{F}_{22}(1,0) = \hat{F}_{22}(1,0) = \hat{F}_{22}(1,1) = 0 \qquad (A.12)$$

$$\hat{F} \text{ satisfies (A.4) with } k = 3 \ . \qquad (A.13)$$

Theorem A.1 Equations (A.9) through (A.13) uniquely determine F.

Proof Let $V = V_1 \otimes V_2$, $\Lambda = \Lambda_1 \otimes \Lambda_2$ where V_i and Λ_i are defined by the statement of lemma A.3 plus equation A.2. Then lemmas A.1, A.2 and A.3 imply that the LIP determined by $V, S_{3,3}(\Pi_1 \times \Pi_2)$ and Λ is correct. This implies that

$$\forall h \varepsilon V [\exists \text{ unique } g \varepsilon S_{3,3}(\Pi_1 \times \Pi_2) \ni [(\lambda_{1i} \otimes \lambda_{2j}) h = (\lambda_{1i} \otimes \lambda_{2j}) g]]$$
$$\text{for } 0 \le i \le \ell_1 + 2 \text{ and } 0 \le j \le \ell_2 + 2. \qquad (A.14)$$

Choose any $h \varepsilon V$ satisfying

$$h(x_i, y_j) = a_{ij} \ . \qquad (A.15)$$

Then (A.10), (A.13), (A.3) and (A.4) imply

$$\hat{F} \varepsilon S_{3,3}(\Pi_1 \times \Pi_2) \ . \qquad (A.16)$$

Equations (A.5), (A.9) and (A.15) imply

$$(\lambda_{1i} \otimes \lambda_{2j})\hat{F} = (\lambda_{1i} \otimes \lambda_{2j})h \text{ for } 0 \le i \le \ell_1,\ 0 \le j \le \ell_2. \quad (A.17)$$

Equations (A.5), (A.6), (A.11) and (A.15) imply

$$(\lambda_{1i} \otimes \lambda_{2j})\hat{F} = (\lambda_{1i} \otimes \lambda_{2j})h \text{ for } i = \ell_1 + 1, \ell_1 + 2 \text{ and } 0 \le j \le \ell_2$$
$$\text{and } 0 \le i \le \ell_1 \text{ and } j = \ell_2 + 1,\ \ell_2 + 2 \quad (A.18)$$

Similarly from (A.12) we have

$$(\lambda_{1i} \times \lambda_{2j})\hat{F} = (\lambda_{1i} \times \lambda_{2j})h \text{ for } i = \ell_1 + 1,\ \ell_1 + 2$$
$$\text{and } j = \ell_2 + 1,\ \ell_2 + 2 \ . \quad (A.19)$$

Equations (A.17), (A.18) and (A.19) imply (A.9), (A.11) and (A.12) are equivalent to the equation

$$(\lambda_{1i} \otimes \lambda_{2j})\hat{F} = (\lambda_{1i} \otimes \lambda_{2j})h \text{ for } 0 \le i \le \ell_1 + 2,\ 0 \le j \le \ell_2 + 2\ . \quad (A.20)$$

The proof follows directly from equations (A.14), (A.16) and (A.20).

QED

REFERENCES

(1) Kuhn, R. M. (1974), Reproducing Kernels and Natural Spline Density Estimation, University of Wisconsin, Madison, Statistics Department Technical Report No. 411.

(2) Schoenberg, I. J. (1972). "Splines and Histograms." Univ. of Wisc. Math. Research Ctr. Summary Report No. 1273. (with an appendix by Carl de Boor); also published in Spline Functions and Approximation Theory, Proceedings of the Symposium at the Univ. of Alberta in 1972. Meir, A. and Sharma, A., Editors. Publ. 1973 by Berkhaeuser Verlag, Basel and Stuttgart. 277-358.

(3) Boneva, L., Kendall, D. and Stefanov, I. (1971). "Spline Transformations: Three New Diagnostic Aids for the Statistical Data Analyst." J. Royal Statist. Society, Series B. 33, 1-70

(4) Akima, H. (1974). "A Method of Bivariate Interpolation and Smooth Surface Fitting Based on Local Procedures." Communications of the ACM. 17, 18-20.

(5) Akima, H. (1974). Algorithm 474, FORTRAN, "Bivariate Interpolation and Smooth Surface Fitting Based on Local Procedures," Communications of the ACM. 17, 26-31.

(6) Gessaman, M. P. (1970). "A Consistent Nonparametric Multivariate Density Estimator Based on Statistically Equivalent Blocks." Annals of Math. Stat. 41, 1344-1346.

(7) Anderson, T. W. (1966). "Some Nonparametric Multivariate Procedures Based on Statistically Equivalent Blocks," Multivariate Analysis. (proceedings of an International Symposium in Dayton, Ohio, in 1965). P.R. Krishnaiah, Editor. Academic Press, New York and London. 5-27.

(8) Murthy, V. K. (1966). "Nonparametric Estimation of Multivariate Densities with Applications." Multivariate Analysis, 43-56. (proceedings of an International Symposium in Dayton, Ohio, in 1965). Krishnaiah, R., Editor. Academic Press, New York and London.

(9) Loftsgaarden, D. O. and Quesenberry, C.P. (1965). "A Nonparametric Estimate of a Multivariate Density Function." Annals of Math. Stat. 36, 1049-1051.

(10) Nadaraya, E. A. (1964). "Estimation of a Bivariate Probability Density." Soobshch Akad Nauk Gruz SSR. 36, 267-268.

(11) Epanechnikov, V. A. (1967). "Non-Parametric Estimation of a Multivariate Probability Density." Theory of Probability and Its Applications. (translation of a Russian journal). 14, 153-158.

(12) Glick, N. (1972). "Sample-Based Classification Procedures Derived from Density Estimators." JASA, 67, 116-122.

(13) Sard, A. (1963). Linear Approximation. Amer. Math. Soc., Providence, Rhode Island.

(14) Mansfield, L. E. (1971). "On the Optimal Approximation of Linear Functions in Spaces of Bivariate Functions." Siam. J. Numer. Anal. 8, 115-126.

(15) Parzen, E., (1971). "Statistical Inference on Time Series by RKHS Methods." Proceedings 12th Biennial Seminar of the Canadian Mathematical Congress. Ronald R. Pyke, Editor. 1-37.

(16) Akhiezer, N. I. and Glazman, I. M. (1961). Theory of Linear Operators in Hilbert Space. (translated from Russian by W. Nestell). Frederick Ungar Publ. Co., New York.

(17) Mansfield, L. E. (1972). "Optimal Approximation and Error Bounds in Spaces of Bivariate Functions." Journal of Approximation Theory. 5, 77-96.

(18) Mansfield, L. E. (1972). "On the Variational Characterization and Convergence of Bivariate Splines." Numer. Math. 20, 99-114.

(19) Hall, C. A. (1973). "Natural Cubic and Bicubic Spline Interpolation," Siam J. Numer. Anal. 10, 1055-1060.

(20) Ahlberg, J. H., Nilson, E. N. and Walsh, J. L. (1967). The Theory of Splines and Their Applications. Academic Press, New York and London.

(21) Wahba, G., personal communication

(22) Wahba, G. (1974). "Interpolating Spline Methods for Density Estimation II. Variable Knots." Univ. of Wisc., Madison, Stat. Dept. Tech. Report No. 337.

(23) Kimeldorf, G. and Wahba, G. (1971). "Some Results on Tchebycheffian Spline Functions." Journal of Mathematical Analysis and Applications. 33, 82-95.

(24) Wald, A. (1943). "An Extension of Wilkes' Method for Setting Tolerance Limits." Annals of Math. Stat. 14, 45-55.

(25) Greville, T. N. E. (1969). "Introduction to Spline Functions." Theory and Application of Spline Functions. Greville, T. N. E., Editor. Univ. of Wisc. Math Research Center Seminar in 1968. Academic Press, New York and London. 1-35.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	Technical Report No. 435
2. GOVT ACCESSION NO.	
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle)	Reproducing Kernel Hilbert Spaces Applied To Bivariate Natural Spline Density Estimation
5. TYPE OF REPORT & PERIOD COVERED	
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s)	Robert M. Kuhn
8. CONTRACT OR GRANT NUMBER(s)	Grant AFOSR 72-2363B
9. PERFORMING ORGANIZATION NAME AND ADDRESS	Department of Statistics, University of Wisconsin, Madison, Wisconsin
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
11. CONTROLLING OFFICE NAME AND ADDRESS	Air Force Office of Scientific Research, 1400 Wilson Blvd., Arlington, VA.
12. REPORT DATE	September 1974
13. NUMBER OF PAGES	48
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	
15. SECURITY CLASS. (of this report)	Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report)	Distribution of this document is unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)	
18. SUPPLEMENTARY NOTES	None
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)	Bivariate density function; Bivariate natural spline; Interpolation points

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of this technical report is to introduce and describe the properties of a bivariate density estimation procedure which is a generalization of the univariate procedure described in [1]. The procedure can be described as follows.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

(a) It is assumed that the bivariate density function has value zero outside the unit square. Divide one axis into $\ell_1 + 1$ and the other axis into $\ell_2 + 1$ equally spaced points. This causes the unit square to be broken up into ℓ equal size rectangles where ℓ equals ℓ_1 times ℓ_2.

(b) Define

$$\hat{F}(ih_1, jh_2) = a_{ij}$$

where

$$a_{ij} = \text{the per cent of observations in } A_{ij}$$
$$A_{ij} = \{(t_1,t_2):\ t_1 \le ih_1 \text{ and } t_2 \le jh_2\}$$
$$h_i = 1/\ell_i$$

(c) If $\hat{F}$ has bivariate density $\hat{f}(t_1,t_2)$ and marginal densities $\hat{f}_1(t_1)$ and $\hat{f}_2(t_2)$ then the density estimate can be chosen to uniquely minimize

$$\int_0^1 (\tfrac{d}{dt_1}\hat{f}_1(t_1))^2 + \int_0^1 (\tfrac{d}{dt_2}\hat{f}_2(t_2))^2 + \int_0^1\int_0^1 (\tfrac{d}{dt_1}\tfrac{d}{dt_2}\hat{f}(t_1,t_2))^2 dt_1 dt_2$$

subject to:

$$\hat{F}(k_1h_1, k_2h_2) = a_{k_1,k_2} \text{ and } \hat{F}(0,t_2) = \hat{F}(t_1,0) = 0 \ \forall k_1,k_2,t_1,t_2.$$

The first part of this paper is devoted to a solution to this problem, along with computational formulas.

The solution given above will turn out to be a unique natural bicubic spline of interpolation. It therefore will be seen that the above estimate also minimizes (not uniquely)

$$\int_0^1\int_0^1 (\tfrac{d}{dt_1}\tfrac{d}{dt_2}\hat{f}(t_1,t_2))^2 dt_1 dt_2$$

subject to the above constraints.

It is shown thatthere is a constant K, independent of n, s_1, and s_2, such that

$$E|f(s_1,s_2) - \hat{f}(s_1,s_2)| \le K/n^{1/3}.$$

Finally, a computational formula is suggested for the situation in which the interpolation points are determined by data, rather than being equally spaced.