ASYMPTOTIC INTEGRATED MEAN SQUARE ERROR USING POLYNOMIAL SPLINES[1]

BY

Girdhar G. Agarwal and W. J. Studden
Purdue University

Department of Statistics
Division of Mathematical Science
Mimeograph Series #78-20

August 1978

[1]This research was supported in part by NSF Grant No. MCS76-08235 A01.
AMS 1970 subject classification. Primary 62K05, G2J05, 62F10; Secondary 41A50, 41A60.

ASYMPTOTIC INTEGRATED MEAN SQUARE ERROR USING POLYNOMIAL SPLINES[1]

By Girdhar G. Agarwal and W. J. Studden

Department of Statistics

Purdue University

W. Lafayette, Indiana 47907

Abbreviated Title: Asymptotic IMSE using Splines.

Summary. Let S_k^d be the set of d-th order splines on [0,1] having k knots $\xi_1 < \xi_2 \cdots\cdots < \xi_k$. We consider the estimation of a sufficiently smooth response function g, using n uncorrelated observations, by an element s of S_k^d. For large n and k we have discussed the asymptotic behavior of the integrated mean square error (IMSE) for two types of estimators: (i) the least square estimator and (ii) a bias minimizing estimator. The asymptotic expression for IMSE is minimized with respect to three variables: (i) the allocation of observation (ii) the displacement of knots $\xi_1 < \cdots < \xi_k$ and (iii) number of knots.

---

[1]This research was supported in part by NSF Grant No. MCS76-08235 A01.

AMS 1970 subject classification. Primary 62K05, G2J05, 62F10; Secondary 41A50, 41A60.

Key words and phrases. Optimal design, least square estimator, bias minimizing estimator, approximation, B-splines, Bernoulli polynomial, L_2-projection operator.

1. INTRODUCTION

Let $g(x)$ be a function defined on the interval [0,1] such that $g \in C^d[0,1]$ i.e. g has d continuous derivatives. For each $x \in [0,1]$ a random variable y_x may be observed with mean $g(x)$ and constant variance σ^2. The problem to be considered is to estimate g using n uncorrelated observations. At each x_i, $i=1,\ldots,r$, $n_i = n\mu_i$ observations are taken. The probability measure assigning mass μ_i to the point $x_i (\Sigma\mu_i=1)$ is referred to as the design. The function $g(x)$ will be estimated by a function $s(x)$ in the class S_k^d. The set S_k^d is the collection of all polynomial splines of order d (degree d-1) having k knots $\xi_1 < \xi_2 \cdots < \xi_k$ in the interior of the interval [0,1]. That is $s(x)$ is a polynomial of degree d-1 on each interval (ξ_i, ξ_{i+1}) and belongs to $C^{d-2}[0,1]$. Fod d=1, S_k^d consists of functions which are constant on each interval (and suitably defined at each ξ_i). For d=2, S_k^d consists of functions s which are linear on each interval (ξ_i, ξ_{i+1}) and continuous on (0,1). The function $s(x) \in S_k^d$ has the representation

(1.1) $$s(x) = \sum_{i=1}^{k+d} \theta_i N_i(x)$$

where N_i are normalized B-splines. The polynomial splines and their B-spline basis will be discussed further in Section 2.

Let $\bar{y}_i$ denote the average of the n_i observations taken at x_i. Estimates which are linear in $\bar{y}' = (\bar{y}_1,\ldots,\bar{y}_r)$ will be used in nearly all cases. Thus the vector of parameters $\theta' = (\theta_1,\theta_2,\ldots,\theta_{k+d})$ will be estimated by

(1.2) $$\hat{\theta} = C\,\bar{y}$$

where C is a $(k+d) \times r$ matrix. As our crtierion for the goodness of the estimate we shall use an integrated mean square error (IMSE); the integration being taken with respect to a measure λ which has a continuous strictly positive density with respect to Lebesgue measure. Our estimate is then

$$N'(x)\hat{\theta} = N'(x)C\,\overline{y}, \tag{1.3}$$

where $N'(x) = (N_1(x), N_2(x), \ldots\ldots, N_{k+d}(x))$. The mean value of $N'(x)\hat{\theta}$ is $N'(x)C\,g_r$ where $g_r' = (g(x_1), \ldots, g(x_r))$. The variance is

$$E(N'(x)\hat{\theta} - N'(x)\,C\,g_r)^2 = (\sigma^2/n)\;N'(x)\;C\;D^{-1}(\mu)\;C'\;N(x)$$

where $D(\mu)$ is an $r \times r$ diagonal matrix with diagonal elements $\mu_1, \ldots \mu_r$. The mean square error is then variance plus squared bias and the integrated mean square error is

$$\begin{aligned} \text{IMSE} &= V + B \\ &= (\sigma^2/n)\text{Tr}\;C\;D^{-1}(\mu)\;C'\;M(\lambda) + \int_0^1 (g(x) - N'(x)\;C\;g_r)^2\,d\lambda(x) \end{aligned} \tag{1.4}$$

where $M(\lambda)$ is the $(k+d) \times (k+d)$ matrix

$$M(\lambda) = \int_0^1 N(x)\;N'(x)\;d\lambda(x). \tag{1.5}$$

Note that V and B denote the integrated variance and integrated squared bias respectively.

The purpose in studying (1.4) is to use it to adaptively estimate the function g by $\hat{\theta}'\,N(x)$.

The IMSE involves three variables (i) the design μ (ii) the knots $\xi_1 < \xi_2 \ldots\ldots < \xi_k$ and (iii) the estimate or choice of C. It is difficult to minimize the IMSE given in (1.4) directly with respect to these variables. The approach used is to first consider the asymptotic behavior of the IMSE for large n and k under some regularity conditions and then perform the minimization. The purpose of this paper is to prove some results concerning the asymptotic behavior of the IMSE.

The results presented here can be used in adaptively estimating a more or less arbitrary response function g. For detail see Agarwal and Studden (1978 b) where the details of an algorithm are presented for estimating g by linear spline functions. The algorithm is presented in such a way that explicit knowledge of the response function is not required. Examples are presented to illustrate the behavior of the algorithm.

In Section 3 and 4 we have discussed the asymptotic behavior of the IMSE for two kinds of estimator, namely, the least square estimator (LSE), and a bias minimizing estimator (BME). The minimization of the asymptotic expression for the IMSE is indicated in Section 5.

2. SPLINES AND B-SPLINES

Let

(2.1) $$(\xi_0 =) 0 < \xi_1 < \ldots\ldots\ldots < \xi_k < 1 (= \xi_{k+1})$$

be a subdivision of the interval [0,1] by k distinct points. These points are the "knots" of the spline function which is defined as follows: a spline function, $s \in S_k^d$, is a function which (i) in each open interval (ξ_{i-1}, ξ_i) for $i=1,\ldots,k+1$ is a polynomial of degree $\leq (d-1)$, (ii) has (d-2) continuous derivative in the open interval (0,1).

For each (fixed) set of knots of the form (2.1), the class S_k^d of such splines is a linear space of functions of dimension (k+d). A basis for this linear space is provided by B-splines, or Basic splines (Curry and Schoenberg (1966)). As well as being a powerful theoretical tool in spline theory, these elementary spline functions provide stable methods for computing with spline functions (see deBoor (1972)). One of the desirable property about B-splines is that their support consists of a small, fixed, finite number of intervals between knots.

For d=1, the $N_i(x)$ are simply the indicator functions on the intervals $(\xi_{i-1},\xi_i]$. For d=2 the support consists of two consecutive intervals (except for the first and last function) and on these intervals is given by

$$N_{i+1}(x) = \begin{cases} (x-\xi_{i-1})/(\xi_i-\xi_{i-1}) & \xi_{i-1} < x < \xi_i \\ (\xi_{i+1}-x)/(\xi_{i+1}-\xi_i) & \xi_i < x < \xi_{i+1} \end{cases}$$

For equally spaced knots the N_i are proportional to the density of the sum of d uniform random variables on (0,1) appropriately scaled and translated.

Explicit expressions for the B-splines will not be needed. For completeness we give a precise definition and list some of their properties below.

We write Π for the nondecreasing sequence $\{t_i\}_1^{k+2d}$ obtained from $\{\xi_i\}_0^{k+1}$ by repeating ξ_0 and ξ_{k+1} each exactly d times. The B-spline basis for the family S_k^d is formed by the following k+d normalized B-splines

(2.2) $$N_i(x) = (t_{i+d}-t_i)\ [t_i,\ldots\ldots,t_{i+d}](t-x)_+^{d-1}$$

i=1,...,k+d, where $[t_i,\ldots.,t_{i+d}]\phi$ denotes dth-order divided differences on the (d+1) points $t_i,\ldots,t_{i+d}$ of the function ϕ. For two or more than two coincident t_i's, the differences in (2.2) are taken to be confluent divided differences (cf. Milne-Thomson (1951)). The N_i are, apart from a constant factor, the B-splines of Curry and Schoenberg (1966).

The N_i defined in (2.2) satisfy

(2.3) $0 < N_i(x) \le 1$ for $x \in (t_i,t_{i+d})$ and $N_i(x) = 0$ otherwise;

(2.4) $\{N_i\}_{i=j}^{j+\ell}$ is linearly independent over $[t_{j+d-1};\ t_{j+\ell+1}]$ for any $\ell \ge d-1$ and any $1 \le j \le k+d-\ell$;

(2.5) $$\{N_i\}_{i=1}^{k+d} \text{ spans } S_k^d;$$

(2.6) $$\sum_{i=1}^{k+d} N_i(x) = 1 \text{ for all } x;$$

(2.7) $$\int_0^1 N_i(x) = (t_{i+d}-t_i)/d, \quad i=1,\ldots,k+d.$$

For (2.3), (2.5), (2.6) and (2.7) see Schoenberg (1966). DeBoor and Fix (1973) proved (2.4).

For $d > 1$, $N_i(x)$, as given by (2.2), are well-defined continuous functions. For d=1, (2.2) makes sense only for $x \neq t_j$, $1 \leq j \leq k+2d$, because of the jump discontinuity of $(t-x)_+^0$ at $t=x$. So in this case we assume the definition (2.2) to be augmented by the (admittedly arbitrary) demand that $N_i(x)$ be right continuous everywhere. Thus for d=1, we let

$$N_i(x) = \begin{cases} 1 & , t_i \leq x < t_{i+1} \\ 0 & , \text{otherwise.} \end{cases}$$

3. ASYMPTOTIC VALUE OF IMSE FOR LSE

In considering the asymptotic behavior of the IMSE, we shall be concerned with the sequences $T_k = \{\xi_0, \xi_1, \ldots, \xi_k, \xi_{k+1}\}$ of knots defined by

(3.1) $$\int_0^{\xi_i} p(x)dx = i/(k+1), \quad i=0,1,\ldots,k+1$$

where p(x) is a positive continuous density on [0,1]. Sacks and Ylvisaker (1970) call the sequence $\{T_k, k \geq 1\}$ so defined as a Regular Sequence generated by p {RS(p)}. We also assume that the design measure μ will either have a smooth density h or will be converging to such a design as k and n become large.

It will be convenient to introduce the following notation: for each fixed k, and i=1,...,k+1 let

$$\delta_i = \xi_i - \xi_{i-1}, \quad \delta = \max_i \delta_i, \quad \text{and } \Delta = \delta / \min_i \delta_i$$

Letting $0 < p_{min} = \min_x p(x) \le \max_x p(x) = p_{max}$, we see that

$$\Delta \le p_{max}/p_{min} \tag{3.2}$$

Also in view of the definition of t_j's in terms of ξ_i's we see that

$$\{\max_i(t_{i+d}-t_i)/\min_i(t_{i+d}-t_i)\} < d\,\Delta \le d\,p_{max}/p_{min}. \tag{3.3}$$

In this section we discuss the asymptotic behavior of the IMSE when the estimator used is the least square estimator (LSE).

3.1 Asymptotic Value of Variance

In the classical problem of regression theory, the analytic form of the function g(x) is supposed to be known. In our case g would be assumed to be of the form $g(x) = \sum_{i=1}^{k+d} \theta_i N_i(x)$. The estimator $\hat{\theta} = C\,\bar{y}$ is restricted to be unbiased. The unbiasedness of $\hat{\theta} = C\,\bar{y}$ restricts C so that

$$C\,F' = I$$

where F is the (k+d) × r matrix $F = (N(x_1),\dots,N(x_r))$ and I is the (k+d) × (k+d) identity matrix. The quantity V in (1.4) is then minimized by the usual least square estimate

$$C = M^{-1}(\mu)\;F\;D(\mu). \tag{3.4}$$

Here μ represents the design measure placing mass μ_i on x_i, i=1,...,r, $M(\mu)$ is the (k+d) × (k+d) matrix $\int N(x)\,N'(x)\,d\mu(x)$. The estimator $\hat{\theta} = C\,\bar{y}$ can then be represented by

$$\hat{\theta}_{LSE} = C\,\bar{y} = M^{-1}(\mu)\int N(x)\,\bar{y}_x\,d\mu(x) \tag{3.5}$$

where $\bar{y}_x$ is the average of the observations taken at the point x. The LSE estimator (3.5) gives a value of $V = (\sigma^2/n) \operatorname{Tr} M^{-1}(\mu)M(\lambda)$.

The results of this section indicates that as $k \to \infty$

(3.6) $$\operatorname{Tr} M^{-1}(\mu)\ M(\lambda) \approx ak \int (f(x)/h(x))p(x)dx$$

and

(3.7) $$B \approx (b/k^{2d}) \int \{(g^{(d)}(x))^2/(p(x))^{2d}\}f(x)dx$$

where the symbol $\approx$ indicates that the ratio tends to one. Here a and b are some constants, and f(x) is the density of the integrating measure λ, and h is the density of design measure μ. The integrated mean square error then becomes

(3.8) $$\text{IMSE} \approx (a\ k\ \sigma^2/n)\int(f(x)/h(x))p(x)dx + (b/k^{2d})\int\{(g^{(d)}(x))^2/(p(x))^{2d}\}f(x)dx.$$

The asymptotics found here are with respect to the number of knots k going to infinity. In practice the number of observations n must, of course, be at least (k+d) and will usually be increasing much faster than k. Some indication of why this is so is given in Section 5. In practice it may be worthwhile replacing k by k+d since in (3.6), for example, we would then get equality if a=1 and $f \equiv h$.

Theorem 3.1. Let $M(\phi)$ be the $(k+d) \times (k+d)$ matrix

(3.9) $$M(\phi) = \int_0^1 N(x)\ N'(x)\ \phi(x)dx$$

If ϕ and ψ and p are continuous strictly positive function defined on [0,1] and $\{T_k\}$ is RS(p) (see (3.1)) then

(3.10) $$\operatorname{Tr} M^{-1}(\phi)\ M(\psi) \approx k \int_0^1 (\psi(x)/\phi(x))p(x)dx.$$

(Remark. If the design measure μ and the integrating measure λ have continuous strictly positive densities h and f respectively, then $M(\mu) = \int N(x)\, N'(x)\, h(x)\, dx$ and $M(\lambda) = \int N(x)\, N'(x)\, f(x)\, dx$ and the asymptotics given in (3.6) will follow from the above theorem. We see that a=1 in (3.6) for a smooth design).

Proof. Let us write

$$M(\phi) = M_0\, D(\phi) - E(\phi) \tag{3.11}$$

where M_0 is given by (3.9) with $\phi\equiv 1$, $D(\phi)$ is the diagonal matrix with elements $\phi(\zeta_i)$, $i=1,2,\ldots,k+d$ and the error term $E(\phi)$ is defined through (3.11). The points $\zeta_1 < \zeta_2 < \ldots\ldots < \zeta_{k+d}$ are (k+d) arbitary points in [0,1] such that

$$\zeta_i \in \overline{\text{support}}\ N_i, \quad i=1,\ldots,k+d. \tag{3.12}$$

If we define

$$\zeta_i = \Big(\sum_{\ell=i+1}^{i+d-1} t_\ell\Big)/(d-1), \quad i=1,\ldots,k+d$$

We can see that these ζ_i's satisfy (3.12). Schoenberg (1966) calls these points as "nodes" and has used them in some other context. Using (3.9) and (3.11), we can write

$$M^{-1}(\phi)\, M(\psi) = [I-U]^{-1}[D^{-1}(\phi)D(\psi)-V] \tag{3.13}$$

where $U = D^{-1}(\phi)\, M_0^{-1}\, E(\phi)$, $V = D^{-1}(\phi)\, M_0^{-1}\, E(\psi)$ and I is $(k+d)\times(k+d)$ identity matrix. We want to expand $(I-U)^{-1}$ as a power series. This can be done if the matrix norm of U is less than one. In the following lemma we find $||U|| \stackrel{\text{def}}{=} \max_x(||Ux||/||x||)$, where vector norm $||x|| = (x'x)^{\frac{1}{2}}$.

<u>Lemma 3.2.</u> $||U|| < \alpha\, \omega(\phi,\delta)$, where α is a constant independent of k and

$\omega(\phi,\delta)$ is the modulus of continuity of ϕ and $\delta = \max_{1\le i\le k+1} (\xi_i - \xi_{i-1})$.

Proof of lemma. The proof consists of bounding the norms of M_0^{-1} and $E(\phi)$ Since M_0 is a positive definite matrix, it is easy to see that

$$(3.14) \qquad ||M_0^{-1}|| = (1/\ell_{min}),$$

where ℓ_{min} is the smallest latent (or characteristic) root of M_0 given by

$$(3.15) \qquad \ell_{min} = \min_x\{(x'M_0x)/(x'x)\}$$

Now to find a upper bound on $||M_0^{-1}||$ we use an inequality of deBoor (1973, p. 273). The inequality states that

$$(3.16) \qquad \rho^2(\gamma'\gamma) \le (\gamma'A\gamma) \le (\gamma'\gamma) \quad \text{for all } \gamma \in \mathbb{R}^{k+d}$$

where ρ is a constant independent of k and depends only on d, and matrix A, called as Gram matrix by deBoor, is related to matrix M_0 by

$$(3.17) \qquad M_0 = DAD$$

where D is the diagonal matrix with diagonal elements $\{(t_{i+d}-t_i)/d\}^{\frac{1}{2}}$, i=1,...,k+d. Using (3.14), (3.15), (3.16) and (3.17), we can show that

$$(3.18) \qquad ||M_0^{-1}|| \le (d/\rho^2)\{\min_{1\le i\le k+d} (t_{i+d}-t_i)\}^{-1}$$

Now we shall find a upper bound on $||E(\phi)||$. First of all since $E(\phi) = [e_{ij}(\phi)]$ is a (k+d) × (k+d) band matrix of bandwidth d-1 (i.e. $e_{ij}(\phi) = 0$ if $|i-j| > d-1$; i,j = 1,..,k+d) it is easy to check that

$$(3.19) \qquad ||E(\phi)|| \le (2d-1)^{\frac{1}{2}} \{\max_{1\le i\le k+d} \sum_{j=1}^{k+d} e_{ij}^2(\phi)\}^{\frac{1}{2}}$$

In view of (3.12), we have for i=1,..,k+d,

$$(3.20) \qquad \sum_{j=1}^{k+d} |e_{ij}(\phi)| < d\ \omega(\phi,\delta) \sum_{j=1}^{k+d} \int N_i(x)N_j(x)dx = \omega(\phi,\delta)(t_{i+d}-t_i).$$

The equality in (3.20) follows from (2.6) and (2.7). The equations (3.19) and (3.20) give

(3.21) $$||E(\phi)|| \le (2\ d-1)^{\frac{1}{2}}\omega(\phi,\delta) \max_{1\le i\le k+d} (t_{i+d}-t_i).$$

Finally since $||D^{-1}(\phi)|| = \{\min_i \phi(\zeta_i)\}^{-1}$, we may combine (3.18) and (3.21), to obtain

$$\begin{aligned} ||U|| &= ||D^{-1}(\phi)\ M_0^{-1}\ E(\phi)|| \\ &\le ||D^{-1}(\phi)||\ ||M_0^{-1}||\ ||E(\phi)|| \\ &< \frac{(2\ d-1)^{\frac{1}{2}} d}{\rho^2} \frac{\max_i (t_{i+d}-t_i)}{\min_i (t_{i+d}-t_i)} \frac{\omega(\phi,\delta)}{\min_i (\phi(\zeta_i))} \end{aligned}$$

In view of the quasi-uniformity condition (3.3) and the fact that ϕ is bounded below, it follows that

$$||U|| < \alpha\ \omega(\phi,\delta),$$

where the constant α does not depend on k. This proves the lemma.

Now since $\omega(\phi,\delta) \to 0$ as $k \to \infty$ (or $\delta \to 0$) we can make $\omega(\phi,\delta) < 1/\alpha$ and hence $||U|| < 1$. We can then invert (I-U) using a power series expansion,

$$\begin{aligned} (I-U)^{-1} &= I + U + U^2 + \ldots\ldots \\ &= I + W, \text{ say} \end{aligned}$$

where $W = \sum_{j=1}^{\infty} U^j$. Therefore from (3.13) and the above expansion,

(3.22) $$\begin{aligned} \text{Tr } M^{-1}(\phi)\ M(\psi) &= \text{Tr } D^{-1}(\phi)D(\psi) - \text{Tr } V \\ &\quad + \text{Tr } W\ D^{-1}(\phi)\ D(\psi) - \text{Tr } V\ W \end{aligned}$$

Now using the definition of the nodes ζ_i's and the mean value theorem in the expression (3.1), we see that the first term on the right of (3.22) divided by k (or k+d) will converge to the integral term in (3.10). Therefore our theorem will be proved if we show that, as $k \to \infty$

(i) $\operatorname{Tr} V = o(k)$

(ii) $\operatorname{Tr} W D^{-1}(\phi)D(\psi) = o(k)$

(iii) $\operatorname{Tr} VW = o(k)$.

Since $V = D^{-1}(\phi)\ M_o^{-1}\ E(\psi)$, from Lemma 3.2, we get

$$||V|| < \beta\ \omega(\psi,\delta)$$

where β is a constant independent of k and $\omega(\psi,\delta)$ is the modulus of continuity of ψ. Also $|\operatorname{Tr} V| < (k+d)\ ||V||$, where (k+d) is the order of matrix V, hence (i) holds.

Using the matrix norm properties, namely

$$||S + T|| \leq ||S|| + ||T||$$

and

$$||ST|| \leq ||S||\ ||T||$$

we can show that

(3.23) $$||W|| \leq ||U||/(1-||U||)$$

$$< \alpha\ \omega(\phi,\delta)/(1-\alpha\ \omega(\phi,\delta)),$$ by Lemma 3.2.

Now since

$$||\operatorname{Tr} W D^{-1}(\phi)\ D(\psi)|| < (k+d)\ ||W||\ ||D^{-1}(\phi)||\ ||D(\psi)||,$$

the relation (ii) holds in view of (3.23) and the fact that ϕ is bounded below and ψ is bounded above.

The proof for relation (iii) follows from the proof of (i) and (ii). Q.E.D.

3.2 Asymptotic Value of Bias

Let us recall that the bias term is

$$B = \int_0^1 (g(x)-N'(x)C\ g_r)^2 d\lambda(x).$$

If we use the LSE, then from (3.4) $C\ g_r = M^{-1}(\mu) \int N(x)\ g(x)\ d\mu(x)$. If the design measure μ and integrating measure λ have densities h and p respectively, then

$$B = \int_0^1 (g(x)-N'(x)\ M^{-1}(h) \int N(y)g(y)h(y)dy)^2 f(x)dx$$

where $M(\mu) = M(h) = \int N(x)\ N'(x)\ h(x)dx$. The main result of this section is the following theorem which yields the asymptotic expression for the bias as indicated in (3.7).

Theorem 3.3. Let $g \in C^d[0,1]$, μ and λ have continuous strictly positive densities h and f respectively. If LSE is used and $\{T_k\}$ is RS(p), then

$$\lim_{k\to\infty} k^{2d}\ B = (|B_{2d}|/(2d)!) \int_0^1 \{(g^{(d)}(x))^2/(p(x))^{2d}\}f(x)dx \tag{3.24}$$

where B_{2d} is 2d-th Bernoulli number (see Nörlund 1924 or Ghizzetti and Ossicini 1970).

The above theorem gives the value of the constant b in (3.7) as $b = |B_{2d}|/(2d)!$.

Before proving Theorem 3.3, let us introduce some notation and describe two important results of Barrow and Smith (1978 a and 1978 b). These will be used in the proof of this theorem. Let $L_2^f = \{\psi | \int_0^1 \psi^2(x)f(x)dx < \infty\}$ denote the L_2 space corresponding to the measure f(x)dx with norm $||\cdot||_f$ and let P_k^f denote the orthogonal projection operator from L_2^f to S_k^d. The omission of the index f will correspond to Lebesgue measure.

Lemma 3.4 (Barrow and Smith 1978a). If $g \in C^d[0,1]$, p is continuous and strictly positive and $\{T_k\}$ is RS(p), then

$$\lim_{k\to\infty} k^{2d}\ ||g-P_k\ g||^2 = (|B_{2d}|/(2d)!)\int\{(g^{(d)}(x))^2/(p(x))^{2d}\}dx. \tag{3.25}$$

Note that the right side of the above expression differs from the right side of (3.24) by a factor of f in the integrand and that using the LSE with density h, the bias B is given by $||g-P_k^h g||_f^2$. It turns out that for h positive and continuous, the projection $P_k^h g$ is asymptotically independent of h as the number of knots $k \to \infty$. The error function $g-P_k^h g$ on each interval (ξ_i, ξ_{i+1}) begins to look proportion to a scaled version of the dth Bernoulli polynomial $B_d(x)$. A detailed discussion of these can be found in Schoenberg (1969), Ghizzetti and Ossicini (1970) or Nörlund (1924). We shall mention some of their properties momentarily.

To exploit the idea that locally the error $g-P_k g$ looks approximately like a Bernoulli polynomial, Barrow and Smith define a sequence of operator Q_k, such that $Q_k g \in S_k^d$ and is "close" to $P_k g$ in the sense that

(3.26) $$\lim_{k\to\infty} k^d\{||g-Q_k g|| - ||g-P_k g||\} = 0$$

Let

$$g(x) = \sum_{j=0}^{d} g^{(j)}(\xi_i)(x-\xi_i)^j/j! + o(\delta^d)$$

$$= \bar{g}(x) + o(\delta^d)$$

and denote $g^{(j)}(\xi_i)/j!$ by g_i^j. The spline $Q_k g = \sum_{\ell=1}^{k+d} a_\ell N_\ell$ is essentially characterized by the requirement that on every dth interval (ξ_i,ξ_{i+1})

(3.27) $$(\bar{g}-Q_k g)(x) = g_i^d \, B_d((x-\xi_i)/\delta_{i+1})\delta_{i+1}^d$$

Due to the fact that $Q_k g$ must have d-2 continuous derivatives at each ξ_i, the above equation cannot be made to hold on every interval (ξ_i,ξ_{i+1}) but only on every dth interval. For example if d=2 we approximate g by a continuous broken line segment. The error $g-P_k g$ is approximately $g''(\xi_i)$

times a scaled version of $B_2(x) = x^2-x+1/6$. One considers approximately the best line segment on every second interval and then joins the ends of these line segments on the intervals between. The polynomials $B_d(x)$ on (0,1) have leading coefficients one, satisfy

$$B_d^{(i)}(0) = B_d^{(i)}(1), \quad i=0,1,..,d-2$$

and minimize $\int_0^1 B_d^2(x)dx$.

The coefficients a_ℓ for $Q_k g = \sum_{\ell=1}^{k+d} a_\ell N_\ell$ can be determined explicitly by setting $\phi_{\ell,d}(s) = \prod_{r=1}^{d-1} (s-t_{\ell+r})$, $\gamma_{\ell,i}^j = \frac{(-1)^j j!}{(d-1)!} \phi_{\ell,d}^{(d-1-j)}(\xi_i)$ and

$$(3.28) \qquad a_\ell = \sum_{j=0}^{d-1} \gamma_{\ell,i}^j (g_i^j - g_i^d \, \delta_{i+1}^{d-j} \binom{d}{j} B_{d-j}), \quad \ell = i+1,..,i+d$$

By taking $i \equiv 0 \pmod d$ for sufficiently many i, all of the coefficients a_ℓ can be determined. For d=2 these coefficients turn out to be

$$a_{\ell+1} = g(\xi_\ell) - g''(\xi_\ell)(\delta_\ell^2/12) + o(\delta^2).$$

Barrow and Smith (1978b) have shown that the operator Q_k, defined by the above scheme, satisfy (3.26) and the following.

<u>Lemma</u> 3.5 (Barrow and Smith 1978b). Let $g \in C^d[0,1]$, and $\bar{\xi} \in [0,1)$. Let j be chosen so that $\xi_i \le \bar{\xi} < \xi_{j+1}$ and let $\delta_{j+1} = \xi_{j+1} - \xi_j$. Let

$$(3.29) \qquad R_k(y,\bar{\xi}) = k^d (g-Q_k\, g)(\xi_j + y\, \delta_{j+1}), \quad y \in [0,1)$$

and

$$K(y,\bar{\xi}) = (g^{(d)}(\bar{\xi})/(p(\bar{\xi}))^d)(B_d(y)/d!).$$

Then there exists a sequence of positive constants $\{\varepsilon_k\}_{k=1}^\infty$ tending to zero and which may be chosen independently of $\bar{\xi}$ such that

$$||R_k(\cdot,\bar{\xi}) - K(\cdot,\bar{\xi})||_\infty = \max_y |R_k(y,\bar{\xi}) - K(y,\bar{\xi})| < \varepsilon_k.$$

As indicated above this Lemma says, in essence, that for k sufficiently large, the error function $g-Q_k g$ is nearly equal (in a sup norm) to a properly scaled Bernoulli polynomial on each subinterval $[\xi_j,\xi_{j+1})$.

Proof of Theorem 3.3. First of all we note that the bias B can be written as

$$B = ||g-P_k^h g||_f^2 = ||g-P_k^f g||^2 + ||P_k^f g-P_k^h g||_f^2.$$

We will show that k^{2d} times the first term on the right converges to the right side of (3.24) and that the second term is $O(k^{-2d})$.

Define an arbitrary subdivision of the interval [0,1] as

$$0 = \tau_1 < \tau_2 < \dots < \tau_{m+1} = 1.$$

Let $\rho = \max_j (\tau_{j+1}-\tau_j)$ and $w = \omega(f,\rho)$ be the modulus of continuity. Now

$$\text{(3.30)} \qquad k^{2d}||g-P_k^f g||_f^2 = k^{2d}\sum_{i=1}^{m}\int_{\tau_i}^{\tau_{i+1}}(g(x)-P_k^f g(x))^2 f(x)dx$$

$$\geq k^{2d}\sum_{i=1}^{m}\{f(\tau_i)-w\}\int_{\tau_i}^{\tau_{i+1}}(g(x)-P_k^f g(x))^2 dx$$

$$\geq k^{2d}\sum_{i=1}^{m}\{f(\tau_i)-w\}\int_{\tau_i}^{\tau_{i+1}}(g(x)-\overline{P}_{kp_i}g(x))^2 dx,$$

where $p_i = \int_{\tau_i}^{\tau_{i+1}}p(x)dx$, $\overline{P}_{kp_i}: \ L_2[\tau_i,\tau_{i+1}] \to S_k^d(\tau_i,\tau_{i+1})$ is the projection operator, and $S_{kp_i}^d(\tau_i,\tau_{i+1})$ is the class of spline functions of order d and kp_i knots. Define $\phi_i(x) = p(x)/p_i$, $x \in [\tau_i,\tau_{i+1}]$. If the knots of the class $S_{kp_i}^d(\tau_i,\tau_{i+1})$ are chosen according to the quantiles of ϕ_i, then in view of Lemma 3.4, we have

$$\text{(3.31)} \qquad \lim_{k\to\infty}(kp_i)^{2d}\int_{\tau_i}^{\tau_{i+1}}(g(x)-\overline{P}_{kp_i}g(x))^2 dx = \frac{|B_{2d}|}{(2d)!}\int_{\tau_i}^{\tau_{i+1}}\frac{(g^{(d)}(x))^2}{(\phi_i(x))^{2d}}dx$$

In (3.30) letting $k \to \infty$ and then using equality (3.31) and the fact that $f(\tau_i) > f(x)-w$ for $x \in (\tau_i,\tau_{i+1})$, we get

$$\lim_{k\to\infty} k^{2d}||g-P_k^f g||_f^2 \geq \frac{(1-2w)|B_{2d}|}{(2d)!} \int_0^1 \frac{(g^{(d)}(x))^2}{(p(x))^{2d}} f(x)dx.$$

In the above inequality let $\rho \to 0$. Since $w = \omega(f,\rho) \to 0$ as $\rho \to 0$, we have

$$\lim_{k\to\infty} k^{2d}||g-P_k^f g||_f^2 \geq C_f, \tag{3.32}$$

where $C_f = |B_{2d}|/(2d)! \int_0^1 \{(g^{(d)}(x))^2/(p(x))^{2d}\}f(x)dx$.

Let us consider, using (3.29)

$$k^{2d}||g-Q_k g||_f^2 = k^{2d} \int_0^1((g-Q_k g)(x))^2 f(x)dx$$

$$= \sum_{j=0}^{k} \delta_{j+1} \int_0^1 R_k^2(y,\xi_j)f(\xi_j+y\ \delta_{j+1})dy$$

By Lemma 3.5, this equals

$$\sum_{j=0}^{k} \delta_{j+1}[(g^{(d)}(\xi_j))^2/(p(\xi_j))^{2d}]\{\int_0^1(B_d(y)/d!)^2 f(\xi_j+y\ \delta_{j+1})dy + \beta_{j,k}\},$$

where $|\beta_{j,k}| < \alpha\ \varepsilon_k$, for some constant α which depends only on d, g and p. We also note that $f(\xi_j+y\ \delta_{j+1}) = f(\xi_j) + \gamma_j$, where $|\gamma_j| < \omega(f,\delta)$. Hence we have

$$k^{2d}||g-Q_k g||_f^2 = (\int_0^1(B_d(y)/d!)^2dy) \sum_{j=0}^{k} \delta_{j+1}((g^{(d)}(\xi_j))/(p(\xi_j))^d)^2 f(\xi_j) + o(1)$$

Let $k \to \infty$ in such a way that $\delta = \max \delta_j \to 0$, then

$$\lim_{k\to\infty} k^{2d}||g-Q_k g||_f^2 = C_f. \tag{3.33}$$

We use here the fact that

$$\int_0^1(B_d(y)/d!)^2dy = |B_{2d}|/(2d)!$$

In view of the inequality $||g-Q_k g||_f^2 \leq ||g-P_k^f g||_f^2$ it follows from (3.32) and

and (3.33), that

(3.34) $$\lim_{k\to\infty} k^{2d}||g-P_k^f g||_f^2 = C_f$$

and

(3.35) $$\lim_{k\to\infty} k^{2d}||Q_k g-P_k^f g||_f^2 = \lim_{k\to\infty} k^{2d}\{||g-Q_k g||_f^2-||g-P_k^f g||_f^2\} = 0$$

Since f and h are continuous and positive, the last equality implies that

$$\lim_{k\to\infty} k^{2d}||Q_k g-P_k^f g||_h^2 = 0$$

In a similar manner we can show that

(3.36) $$\lim_{k\to\infty} k^{2d}||Q_k g-P_k^h g||_f^2 = 0$$

Since

$$||P_k^h g-P_k^f g||_f \le ||Q_k g-P_k^f g||_f + ||Q_k g-P_k^h g||_f,$$

we have from (3.35) and (3.36),

(3.37) $$\lim_{k\to\infty} k^{2d}||P_k^h g-P_k^f g||_f^2 = 0$$

Finally, the theorem follows from (3.34) and (3.37). Q.E.D.

Here we have assumed that design measure μ has a smooth density. In a more practical situation the design measure μ will be discrete on a finite number of points. With some regularity conditions the IMSE is still of the form (3.8), however the constants a and b will be different. Generally a smoother design (resembling the uniform) will keep the bias term small and give slightly larger values for the variance term. An appropriate discrete design will give smaller values for the variance term. but will increase the bias. As an example in the linear spline case (i.e. d=2) see Agarwal (1978) and Agarwal and Studden (1978a).

4. ASYMPTOTIC VALUE OF IMSE FOR BME

Various authors, for example, Box and Draper (1959) and Karson, Manson and Hader (1969) have proposed attaching more importance to the bias part B.

The integrated squared bias B is minimized if

$$E(C\,\overline{y}) = M^{-1}(f)\int N(x)g(x)f(x)dx \tag{4.1}$$

where $M(f) = \int N(x)\ N'(x)f(x)dx$. In general a matrix C cannot be found for which (4.1) holds, so instead we try to find a C* such that

$$E(C^*\,\overline{y}) \overset{\sim}{-} M^{-1}(f)\int_0^1 N(x)g(x)f(x)dx. \tag{4.2}$$

The asymptotic is in the sense that $||E(C^*\,\overline{y}) - M^{-1}(f)\int N(x)g(x)f(x)dx||$ goes to zero as n (number of observations) tend to infinity, where the vector norm $||a|| \overset{\text{def}}{=} (a'a)^{\frac{1}{2}}$. We should emphasize here that k is fixed.

Let $L'(x) = (L_1(x),\ldots,L_r(x))$, where $L_i(x)$, $i=1,..,r$ are the normalized B-spline (Section 2) of order 2 with knots at observation points x_i, $i=2,..,r-1$. The $L_i(x)$ is a "roof-like" function which has a value one at x_i, goes linearly to zero at adjacent knots x_{i-1} and x_{i+1} and then remains zero. Let us define $\tilde{g}(x) = \sum_{i=1}^{r} g(x_i)\ L_i(x)$. Since $L_i(x_j) = \delta_{ij}$; $i,j=1,..r$, $\tilde{g}(x)$ interpolates g at x_i, $i=1,..,r$. As an approximation to g, the function $\tilde{g}$ satisfies the following two properties (e.g. see Prenter 1975).

(i) If g is continuous then $\tilde{g}$ converges to g as r(or n) $\to \infty$ in such a way that $\eta = \max_i(x_i-x_{i-1})$ tends to zero.

(ii) If g is twice continuously differentiable, then $||g-\tilde{g}||_\infty = \max_{x\in[0,1]} |g(x)-\tilde{g}(x)| \leq \alpha\ ||g''||_\infty\ \eta^2$, where α is a constant independent of n.

Now if we take

$$C^* = M^{-1}(f) \int_0^1 N(x)\ L'(x) f(x) dx \tag{4.3}$$

then in view of (i) and (ii) we see that C* of (4.3) satisfies (4.2). Hence our "bias minimizing" estimate (BME) is defined as

$$\hat{\theta}_{BME} = C^* \bar{y} = M^{-1}(f)(\int_0^1 N(x)\ L'(x)\ f(x)dx)\bar{y}. \tag{4.4}$$

In the following theorem we shall find the asymptotic expression for the variance term using the estimator $\hat{\theta}_{BME}$. This theorem concerns choosing the design μ to have weight μ_i on x_i such that

$$\mu_i = \int L_i(x)\ h(x)dx, \quad i=1,2,..,r \tag{4.5}$$

for some continuous strictly positive density h(x).

Theorem 4.1. If the estimator $\hat{\theta}_{BME}$, given in (4.4), is used, and the design is chosen using (4.5) and $\{T_k\}$ is RS(p), then

$$\lim_{k\to\infty} \lim_{n\to\infty} (nV/k\sigma^2) = \int (f(x)/h(x))p(x)dx. \tag{4.6}$$

Remark. We shall see later (Section 5) that the order of double limit in (4.6) is, in some sense, justified. Actually in considering the MLE we effectively let n approach infinity before k.

Proof. It is easy to check that V, the variance of $\hat{\theta}_{BME}$ is given by

$$\begin{aligned}(nV/\sigma^2) &= \text{Tr } C^*\ D^{-1}(\mu)\ C^{*\prime}\ M^{-1}(f) \\ &= \text{Tr } M^{-1}(f)(\int N(x)\ L'(x)f(x)dx)D^{-1}(\mu)(\int N(y)\ L'(y)f(y)dy)' \\ &= \text{Tr } M^{-1}(f)(\int \int N(x)\ N'(y) \sum_{i=1}^{r} \frac{L_i(x)L_i(y)}{\mu_i} f(x)f(y)dx\ dy)\end{aligned} \tag{4.7}$$

The proof of the Theorem will follow easily from the next Lemma.

Lemma 4.2. Let u(x) and v(x) be continuous functions defined on [0,1]. If $\eta = \max_i(x_i - x_{i-1}) \to 0$ as $n \to \infty$, we have

(4.8) $$\lim_{n\to\infty} \int_0^1 \int_0^1 u(x)v(x) \sum_{j=1}^{r} \frac{L_j(x)L_j(y)}{\mu_j} \, dx\, dy = \int_0^1 \frac{u(x)v(x)}{h(x)} dx.$$

The proof of the Lemma is deferred till the end of this section. Assuming for the present the truth of Lemma 4.2 we complete the proof of the theorem. Let $n \to \infty$ in (4.7) and then use (4.8) to get

$$\lim_{n\to\infty}(nV/\sigma^2) = \text{Tr } M^{-1}(f)\ M(f^2/h),$$

where $M(f^2/h) = \int N(x)\ N'(x)(f^2(x)/h(x))dx$. If we take $\phi\equiv f$ and $\psi\equiv f^2/h$ in Theorem 3.1, we then see that

$$\lim_{k\to\infty}(k^{-1} \text{ Tr } M^{-1}(f)\ M(f^2/h)) = \int(f(x)/h(x))p(x)dx.$$

which completes the proof of the theorem. Q.E.D.

Since the estimator $\hat{\theta}_{BME}$ satisfies (4.2), we can easily check that the bias term is asymptotically minimized, i.e.

(4.9) $$\lim_{k\to\infty} \lim_{n\to\infty} k^{2d} B = (|B_{2d}|/(2d)!)\int\{(g^{(d)}(x))^2/(p(x))^{2d}\}f(x)dx.$$

In the above we have suggested one choice for BME, but we can suggest some other choices, too, which would involve estimating g in $M^{-1}(f) \int_0^1 N(x)\ g(x)\ f(x)dx$.

Proof of Lemma 4.2. Let us denote by I, the double integral on left of (4.8). Since $L_j(x)$ has support on the interval (x_{j-1}, x_{j+1}), we can express the integral I as

$$I = \sum_{j=1}^{r} \mu_j^{-1} \int_{x_{j-1}}^{x_{j+1}} \int_{x_{j-1}}^{x_{j+1}} u(x)v(y)\ L_j(x)\ L_j(y)\ dxdy,$$

where $x_0=x_1=0$ and $x_{r+1}=x_r=1$. By use of the mean value theorem, we get

$$I = \sum_{j=1}^{r} \{\mu_j^{-1}u(x_j)v(x_j) \int_{x_{j-1}}^{x_{j+1}} \int_{x_{j-1}}^{x_{j+1}} L_j(x)\ L_j(y)\ dx\ dy + \gamma_j\}$$

where $|\gamma_j| < \alpha(\omega(u,\eta) + \omega(v,\eta))$ where α depends only on u and v, and $\omega(u,\eta)$ and $\omega(v,\eta)$ are the modulus of continuity of u and v. From (4.5), for $1 \le j \le r$,

$$\mu_j = (\tfrac{1}{2})\ h(x_j)\ (\eta_j+\eta_{j+1})\ (1+\tau_j)$$

where $\eta_j = x_j - x_{j-1}$, $j=2,\ldots,r$, $\eta_1=\eta_{r+1}=0$, and $|\tau_j| < \rho\ \omega(h,\eta)$ where constant ρ depends only on h. Therefore now I equals

$$\sum_{j=1}^{r} \{u(x_j)v(x_j)/h(x_j)\}\{(\eta_j+\eta_{j+1})/2\} + o(1)$$

Now the proof of Lemma follows since this sum is a Riemann sum for the integral on right of (4.8).

5. MINIMIZING IMSE

The results of last two section indicate that

$$\text{(5.1)}\qquad \text{IMSE} \approx (k\sigma^2/n) \int (f(x)/h(x))p(x)dx + (b/k^{2d}) \int \{(g^{(d)}(x))^2/(p(x))^{2d}\}f(x)dx$$

where $b = |B_{2d}|/(2d)!$. This asymptotic value depends on the three "variables" (i) k, the number of knots, (ii) p(x), the displacement of knots and (iii) h(x), the allocation of observations. Note that the choice of estimator has been eliminated. The results of minimizing the asymptotic value of the IMSE are given in the following theorem.

Theorem 5.1. The IMSE given in (5.1) is absolutely minimized by h, p and k given as follows:

$$\text{(5.2)}\qquad h(x) = \alpha_{g,f}\{(f(x))^{2d+1}(g^{(d)}(x))^2\}^{1/(4d+1)},$$

$$\text{(5.3)}\qquad p(x) = \beta_{g,f}\{f(x)(g^{(d)}(x))^4\}^{1/(4d+1)},$$

(5.4) $$k = \beta_{g,f}^{-1}[(2\ b\ dn/\sigma^2)\alpha_{g,f}]^{1/(2d+1)}$$

where $\alpha_{g,f}^{-1} = \int_0^1\{(f(x))^{2d+1}(g^{(d)}(x))^2\}^{1/(4d+1)}\ dx$, and $\beta_{g,f}^{-1} = \int_0^1\{f(x)(g^{(d)}(x))^4\}^{1/(4d+1)}dx$.

For proof of this theorem we refer to the proof a theorem in Section 3 of Agarwal and Studden (1978b) in which parallel results are proved for the case $f(x)\equiv 1$.

The knot displacement in (5.3) indicates that the knots should be placed where $f(x)(g^{(d)}(x))^4$ is large. Using (5.2) and (5.3) (or going back to 5.1)). we see that $h \propto \sqrt{fp}$ so that h is usually more dispersed than p.

Equation (5.4) indicates that k is decreasing in σ and of order $n^{1/(2d+1)}$. For example for d=2 this gives $n \propto k^5$. This indicates that there should generally be many more observations than knots. This order relation justifies to some extent the order of the double limit in (4.6) and (4.9) of the last section.

REFERENCES

[1] Agarwal, G. G. (1978). Asymptotic design and estimation in polynomial spline regression. Technical Report No. 78-11, Purdue University.

[2] Agarwal, G. G. and Studden, W. J. (1978a). Asymptotic design and estimation using linear splines. To appear in Comm. Statist.

[3] Agarwal, G. G. and Studden, W. J. (1978b). An algorithm for selection of design and knots in the response curve estimation by spline functions. Technical Report No. 78-15, Purdue University.

[4] Barrow, D. L. and Smith, P. W. (1978a). Asymptotic properties of best $L_2[0,1]$ approximation by splines with variable knots. To appear in Quart. of Appl. Math.

[5] Barrow, D. L. and Smith, P. W. (1978b). Efficient L_2 approximation by splines. Submitted to Numer. Math.

[6] deBoor, C. (1972). On calculating with B-splines. J. Approximation Theory., 6, 50-62.

[7] deBoor, C. (1973). The quasi-interpolant as a tool in elementary spline theory. In Approximation Theory (G. G. Lorentz, ed.) Academic Press, New York.

[8] deBoor, C. and Fix, G. J. (1973). Spline approximation by quasi-interpolants. J. Approximation Theory, 8, 19-45.

[9] Box, G. E. P. and Draper, N. R. (1959). A basis for selection of response surface design. J. Amer. Statist. Assoc. 54, 622-654.

[10] Curry, H. B. and Schoenberg (1966). On Polya frequency functions IV: The fundamental spline functions and their limits. J. Analyse. Math. 17, 71-107.

[11] Ghizzetti, A. and Ossicini, A. (1970). Quadrature formulae. Academic Press, New York.

[12] Karson, M. J., Manson, A. R. and Hader, R. J. (1969). Minimum bias estimation and experimental design for response surfaces. Technometrics. 11, 461-475.

[13] Milne-Thomson, L. M. (1951). The calculus of finite differences. Macmillan, London.

[14] Nörlund, N. E. (1924). Vorlesungen über differenzenrechnung. Verlag Von Julius Springer, Berlin.

[15] Prenter, P. M. (1975). Splines and variational methods. John Wiley, New York.

[16] Sacks, J. and Ylvisaker, D. (1970). Design for regression problems with correlated errors III. Ann. Math. Statist. 41, 2057-2074.

[17] Schoenberg, I. J. (1966). On spline functions. MRC Technical Summary Report No. 625, Univ. of Wisconsin-Madison.

[18] Schoenberg, I. J. (1969). Monosplines and quadrature formulae. In Theory and Applications of Spline Functions (T.N.E. Greville, ed.) Academic Press, New York.