ASYMPTOTIC DESIGN AND ESTIMATION IN POLYNOMIAL SPLINE REGRESSION*

by

Girdhar G. Agarwal

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #78-11

June 1978

*This research was supported in part by Purdue University David Ross Fellowhsip and in part by NSF Grant No. MCS76-08235A01.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 General Background

Suppose that η is a response which depends on x, a set of q predictor or independent variables, and θ, a set of unknown parameters, through an unknown response function

$$\eta = g(x,\theta) = g(x) \tag{1.1.1}$$

where $x = (x_1, x_2, \ldots, x_q) \in \mathscr{X}$ a compact set. Usually the set θ of parameters is unknown and must be estimated from the experimental data. Typically such data will consist of n measured values of η corresponding to n specified combinations of the levels of x.

The response relationship is usually interpreted geometrically as representing a surface whose coordinates are the q+1 variables $x_1, x_2, \ldots, x_q$, and η. The n experimental combinations of the levels of x are represented by points in the space of the independent variables.

In observing the response η we assume that an additive experimental error, denoted by ε, exists so that, for each observation $y(x_i)$, $i = 1,2,\ldots,n$, we can write

$$y(x_i) = \eta_i + \varepsilon_i = g(x_i,\theta) + \varepsilon_i. \tag{1.1.2}$$

We assume that ε_i are uncorrelated and identically distributed with mean zero and a common variance σ^2 independent of x which is unknown. In many cases where the form of the true response function (1.1.1) is unknown, it is approximated by a polynomial function of as low an order as possible.

Now for further discussion we will assume that number of independent variables is one, i.e. $q = 1$. We can possibly extend our one dimensional results to higher dimensions, though these extensions will not be straight forward. We shall not discuss these here. For further discussion in multivariate case $(q > 1)$ we refer to Sewell (1972).

Suppose one is willing to entertain the linear model

$$y(x_i) = \theta' f(x_i) + \varepsilon_i \tag{1.1.3}$$

as an approximation of the unknown model (1.1.2), where $f' = (f_1, \ldots, f_m)$ denote an m-vector of continuous functions defined on $\mathscr{X} = [0,1]$ and $\theta' = (\theta_1, \ldots, \theta_m)$ is a m-vector of unknown parameters. The functions $f_1, \ldots, f_m$ are called the regression functions and are assumed known to the experimenter. If n experimental observations are to be obtained, we can express (1.1.3) in matrix form as

$$y = X\beta + \varepsilon. \tag{1.1.4}$$

The vector y is a nx1 vector of observations; X is a nxm matrix, with row i containing $f'(x_i)$; ε is an nx1 vector of uncorrelated random errors with mean zero and variance σ^2.

One of the main problems in the above setup is the estimation of the functions of the vector θ by means of uncorrelated observations $\{y(x_i)\}_{i=1}^n$. Given a specific function of θ and a criterion of what a good estimate is, the design problem is one of selecting the x_i's at which to experiment.

The design concept was generalized by Kiefer and Wolfowitz (1959) to allow for specifying a design as a probability measure μ on the Borel sets $\mathscr{B}$ of $\mathscr{X}$ where $\mathscr{B}$ includes all one point sets. Suppose we have an n-point design with n_i observations at x_i (note that $\Sigma n_i = n$). The design μ is such that $\mu_i = \mu(x_i) = 0$ if there are to be no observations at the point x_i, and such that $\mu_i = \mu(x_i) = n_i/n$ if there are to be $n_i > 0$ observations at the point x_i. For a discrete n-point design, μ takes on values which are multiples of 1/n, and defines an <u>exact</u> design on $\mathscr{X}$. Removing the restriction that μ be a multiple of 1/n, we can extend this idea to a design measure which satisfies, in general:

$$\left.\begin{aligned} &\mu(x) \geq 0, \quad x \in \mathscr{X} \\ &\int_{\mathscr{X}} \mu(dx) = 1 \end{aligned}\right\} \tag{1.1.5}$$

With respect to the model (1.1.4), we can define a matrix analogous to X'X for design μ. Let

$$m_{ij}(\mu) = \int_{\mathscr{X}} f_i(x) f_j(x) d\mu, \quad i,j = 1,2,\ldots,m \tag{1.1.6}$$

and

$$M(\mu) = [m_{ij}(\mu)].$$

The matrix $M(\mu)$ is called the information matrix of the design μ. Note that, for an exact design, $n\ M(\mu) = X'X$.

Extensive analysis and design methodology have been developed for the case where $g(x,\theta)$ is being approximated by a polynomial. In this case $f'(x) = (1,x,\ldots,x^{m-1})$ and θ would be a mx1 vector of polynomial coefficients. These coefficients are estimated from data by standard least squares methods. Various properties of the resulting fitted polynomials have been investigated, especially as the properties are influenced by the choice of experimental design. Initially, criteria for judging the goodness of designs were largely concerned with variance - either of the individual coefficients or of the fitted polynomial as a whole. The question of bias due to the inadequacy of the approximating polynomial was given somewhat secondary consideration. Box and Draper (1959, 1963) adopted mean square error integrated over some region of interest, R, as a basic criterion. This criterion involves both variance and bias.

A disadvantage of the polynomials in the context of curve fitting is their analyticity. That is, the behavior of a polynomial in an arbitrarily small region defines, through the concept of analytic continuity, its behavior everywhere. On the other hand, the spline functions possess the property of having local behavior that is less dependent on their behavior elsewhere. Spline functions are discussed in the next two sections (1.2 and 1.3).

1.2 Spline Functions

The term spline usually refers to a "piecewise polynomial". Here the space $\mathscr{X}$ is an interval [0,1]. The interval [0,1] is divided into k+1 segments by k "knots" $\xi_1, \xi_2, \ldots, \xi_k$ where $\xi_0 = 0 < \xi_1 < \ldots < \xi_k < 1 = \xi_{k+1}$. A function s(x) on [0,1] is called a spline if s(x) is equal to a polynomial on (ξ_i, ξ_{i+1}) (possibly different on each interval) and satisfies certain differentiability conditions at the points $\xi_1, \ldots, \xi_k$.

The simplest case stipulates that s(x) is linear on each interval (ξ_i, ξ_{i+1}), $i = 0,1,\ldots,k$ and is continuous at each ξ_i, $i = 1,2,\ldots,k$. We thus have a continuous polygonal line segment.

Generally a spline function of degree d with k prescribed knots $\xi_1 < \xi_2 < \ldots < \xi_k$ is a polynomial of degree at most d, in each of the intervals $(0,\xi_1)$, $(\xi_1,\xi_2), \ldots, (\xi_k,1)$ and has d-1 continuous derivatives at ξ_i. We can represent such a function in the form

$$s(x) = \sum_{i=0}^{d} \alpha_i x^i + \sum_{i=1}^{k} \beta_i (x-\xi_i)_+^d \tag{1.2.1}$$

where

$$(x-\xi_i)_+^d = \begin{cases} (x-\xi_i)^d, & x \geq \xi_i \\ 0 & x < \xi_i \end{cases} \qquad i = 1,2,\ldots,k.$$

The coefficient β_i represents the jump in the dth derivative at the point ξ_i. For d = 1, this is the change in the slope. Various degrees of differentiability can be allowed at ξ_i by using the terms $(x-\xi_i)_+^r$ for $r < d$, however, we shall not consider these here.

Because of their least oscillatory behavior spline functions can be used as approximating or interpolating functions. They can be used in approximating linear functionals, especially definite integrals, and as approximations to solution of ordinary differential equations. There is now considerable evidence that in many circumstances a spline function is a more adaptable approximating function than a polynomial involving a comparable number of parameters. This conclusion is based in part on actual numerical experience, and in part on the mathematical demonstrations that the solutions of a variety of problems of "best" approximation actually turn out to be spline functions. For example suppose we are given k data points

$$(\xi_i, y_1),\ (\xi_2, y_2), \ldots, (\xi_k, y_k) \qquad (1.2.2)$$

with distinct abscissas. If $1 \leq d \leq k$, the function s which interpolates the k data points (i.e. $s(\xi_i) = y_i$, $i = 1,2,\ldots,k$) and minimizes the integral

$$\int_0^1 (s^{(d)}(x))^2 dx$$

is a spline function of degree 2d-1 with knots at $\xi_1, \ldots, \xi_k$ and is such that in the two intervals $(0, \xi_1)$ and $(\xi_k, 1)$ it is given by some polynomial of degree d-1 (rather than 2d-1) or less. For more applications of spline functions and their "best" approximating properties see Greville (1969) and Schoenberg (1969).

1.3 Basis for the Spline Functions

In using the functions

$$1, x, \ldots, x^d, (x-\xi_1)_+^d, \ldots, (x-\xi_k)_+^d$$

as our basis in the spline problems, the linear systems arising tend to be very ill conditioned, and this may cause difficulty if the attempt is made to solve these systems directly in order to obtain the required parameters. The numerical instabilities encountered increases with the dimensions of the linear system involved and are related to the mathematical properties of the truncated power functions. The difficulties can be overcome by adopting a different basis for the classes of splines dealt with. One of the popular basis is the Lagrange functions $\ell_i(x)$ defined by the condition

$$\ell_i(\xi_j^*) = \delta_{ij} \qquad i,j = 1,2,\ldots,m$$

where $m = d+k+1$ and $\xi_1^* < \xi_2^* \ldots < \xi_m^*$ are m (= number of functions in basis) points in [0,1]. For example when d = 1, the k+2 points are $0 = \xi_0 < \xi_1 < \ldots < \xi_k < \xi_{k+1} = 1$, and $\ell_i(x)$ is the "roof function"

$$\ell_i(x) = \begin{cases} 0 & , \quad x \le \xi_{i-1} \\ (x-\xi_{i-1})/(\xi_i-\xi_{i-1}), & \xi_{i-1} \le x \le \xi_i \\ (\xi_{i+1}-x)/(\xi_{i+1}-\xi_i), & \xi_i \le x \le \xi_{i+1} \\ 0 & , \quad x \ge \xi_{i+1} \end{cases}$$

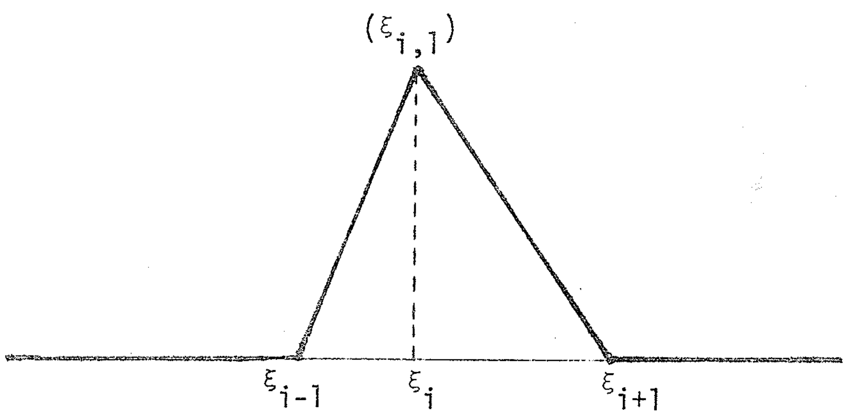
At the two end points $i = 0, k+1$,

$$\ell_0(x) = \begin{cases} 0 & , \quad x \geq \xi_1 \\ (\xi_1 - x)/(\xi_1 - \xi_0), & \xi_0 \leq x \leq \xi_1 \end{cases}$$

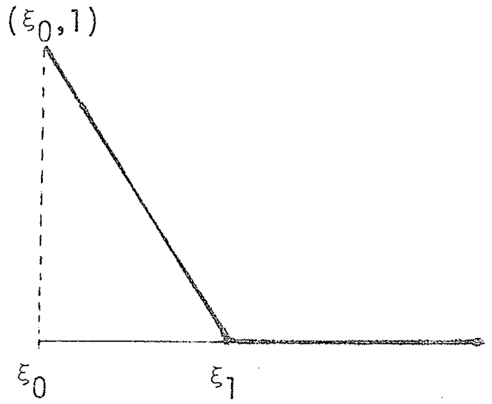
and

$$\ell_{k+1}(x) = \begin{cases} (x - \xi_k)/(\xi_{k+1} - \xi_k), & x \geq \xi_k \\ 0 & , \quad x \leq \xi_k \end{cases} .$$

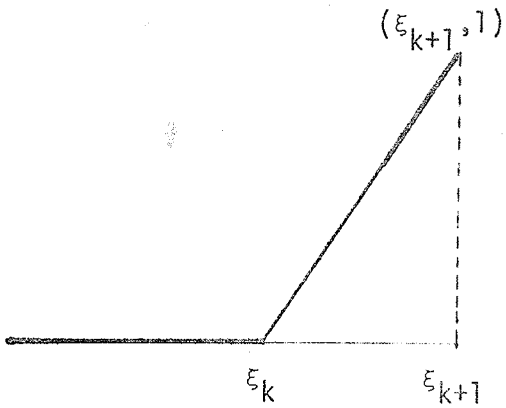
The graph of these functions are shown below:



Graph of $\ell_i(x)$, $1 \leq i \leq k$



Graph of $\ell_0(x)$



Graph of $\ell_{k+1}(x)$

1.4 Statement of the Problem and Outline

Let $g(x)$ be a function defined on the interval $[0,1]$ such that $g \in C^{d+1}[0,1]$ i.e. g has d+1 continuous derivatives. The function $g(x)$ will be approximated by a spline function $s(x)$ of degree d with k prescribed knots $\xi_1, \xi_2, \ldots, \xi_k$. From (1.2.1) $s(x)$ has the representation

$$s(x) = \sum_{i=0}^{d} \alpha_i x^i + \sum_{i=1}^{k} \beta_i (x-\xi_i)_+^d.$$

Note that in terms of the notation of section 1.1 here we have $m = k+d+1$, the vector $f(x)$ consists of the functions

$$1, x, \ldots, x^d, (x-\xi_1)_+^d, \ldots, (x-\xi_k)_+^d$$

and the vector θ of unknown parameters consists of the elements $\alpha_0, \alpha_1, \ldots, \alpha_d, \beta_1, \ldots, \beta_k$. Thus in terms of our previous notations (section 1.1) we can write $s(x)$ as

$$s(x) = \sum_{i=1}^{m} \theta_i f_i(x). \qquad (1.4.1)$$

Let $\bar{y}_{i.}$ denotes the average of $n_i = n\mu_i$ observations taken at x_i, $i = 1,2,\ldots,r$. We will use estimates which are linear in $\bar{y}' = (\bar{y}_1, \ldots, \bar{y}_r)$. Thus the vector of parameters $\theta' = (\theta_1, \ldots, \theta_m)$ will be estimated by

$$\hat{\theta} = C\,\bar{y} \qquad (1.4.2)$$

where C is a mxr matrix. Then

$$E(\hat{\theta}) = C\, g_r$$

where $g_r' = (g(x_1),\ldots,g(x_r))$. As indicated before, we shall use integrated mean square error (IMSE) as our criterion for the goodness of our estimate, the integration being taken with respect to Lebesgue or uniform measure on [0,1]. The mean square error is $E(f'(x)\hat{\theta} - g(x))^2$ = Variance + $(\text{Bias})^2$, where

$$\begin{aligned}\text{Variance} &= E(f'(x)\hat{\theta} - E(f'(x)\hat{\theta}))^2 \\ &= E(f'(x)(\hat{\theta}-E(\hat{\theta})))^2 \\ &= E(f'(x)(C\bar{y}-CE(\bar{y})))^2 \\ &= E\{(f'(x)C(\bar{y}-E(\bar{y})))(f'(x)C(\bar{y}-E(\bar{y})))'\} \\ &= f'(x)C\{E(\bar{y}-E(\bar{y}))(\bar{y}-E(\bar{y}))'\}C'f(x) \\ &= f'(x)C\,\text{Var}(\bar{y})C'f(x) \\ &= \frac{\sigma^2}{n} f'(x)CD_\mu^{-1}C'f(x) \qquad (1.4.3)\end{aligned}$$

and

$$\begin{aligned}(\text{Bias})^2 &= (g(x) - E(f'(x)\hat{\theta}))^2 \\ &= (g(x) - f'(x)Cg_r)^2. \qquad (1.4.4)\end{aligned}$$

In (1.4.3), D_μ is a rxr diagonal matrix with diagonal elements $\mu_1,\ldots,\mu_r$. The integrated mean square error (IMSE) is then

$$\begin{aligned}J &= \int_0^1 E(f'(x)\hat{\theta}-g(x))^2 dx \\ &= \frac{\sigma^2}{n}\,\text{Tr}\; CD_\mu^{-1}C'M_0 + \int(g(x)-f'(x)Cg_r)^2 dx \qquad (1.4.5) \\ &= V + B, \text{ say.}\end{aligned}$$

In (1.4.5) M_0 is the mxm matrix $\int f(x)f'(x)dx$. Note that V and B denote the integrated variance and the integrated squared bias respectively.

We propose to adaptively or sequentially estimate the function g(x) by $\hat{\theta}'f(x)$, attempting to minimize the IMSE using three "variables" (i) the design μ_i on x_i or where observations are chosen (ii) the estimator or choice of C and (iii) the choice of knots $\xi_1 < \xi_2 \ldots < \xi_k$. The approach used here will be to study the asymptotic behavior of (1.4.5) for large n and k.

In investigating the asymptotic behavior of IMSE we have assumed that the knots ξ_i, i = 1,2,...,k are chosen so that

$$i/(k+1) = \int_0^{\xi_i} p(x)dx, \qquad i = 1,2,\ldots,k$$

where p is some suitable chosen density. The design measure is assumed to have a smooth density h or will be converging to such a design as k and n become large. We suspect, in most cases, as $k \to \infty$, that

$$\text{IMSE} \approx \frac{ak}{n}\sigma^2 \int_0^1 \frac{p(x)}{h(x)}dx + \frac{b}{k^{2d+2}} \int_0^1 \frac{(g^{(d+1)}(x))^2}{(p(x))^{2d+2}}dx \qquad (1.4.6)$$

where a and b are some positive constants. In our thesis we have considered the cases d = 0 and d = 1 in detail and have found the exact value of constants a and b for these two cases.

In section 2.1 we minimize the IMSE (given in (1.4.5)) with respect to the choice of the estimator C, and in section 2.3 we minimize the asymptotic expression for the IMSE (see (1.4.6)) with respect to (i) k, the number of knots (ii) p(x), the displacement of knots and (iii) h(x), the distribution of observations. In section 2.2 we have discussed the asymptotic behavior of the IMSE when g(x) is being approximated by step functions.

The case, when g(x) is estimated by linear splines, is considered in Chapter III. In sections 3.1 and 3.2 we have considered the asymptotic behavior of integrated variance and integrated bias respectively. We have used the least square estimators and different choices of designs are considered. It is noticed that if the design is discrete, we get slightly smaller values for the variance term and slightly larger values for the bias term. On the other hand, the opposite happens if the design has a smooth density. Some other choices of estimators are considered in section 3.3. The results of these three sections (3.1, 3.2 and 3.3) indicate that in the asymptotic expression for the IMSE (see (1.4.6)), the constant a ranges between 2/3 and 1 and the constant b ranges between 1 and 6. In section 3.4, in order to facilitate the presentation, we have recalled the minimization results for linear spline case which were already obtained for splines of any degree d in section 2.3. There is a natural restriction on minimizing k (number of knots) that it should be less than or equal to n (number of observations). When this condition is violated, it is hard to characterize the exact solution for minimization problem. We have suggested some approximate solutions in section 3.5. To compare these solutions with exact solution some numerical examples are done.

In section 4.1, an algorithm is presented for fitting linear splines to any arbitrary response function g(x) using three different estimators. To illustrate the algorithm, two numerical examples are done in section 4.3.

In the Appendices, we have reviewed some results on matrix norm and Reproducing Kernel Hilbert Spaces. These are helpful in proving some of the results in our thesis.

Finally, let us mention some of the related work. Note that the second term in (1.4.5) is the square of L_2-error of approximation of the function g(x) by a spline function of degree d with k prescribed knots. The convergence rates of the L_p-error of best approximation to g(x) by polynomial splines have been investigated by many authors. Let S_k^d denote the class of spline functions of degree d with k prescribed (or fixed) knots $0 = \xi_0 < \xi_1 \ldots < \xi_k < \xi_{k+1} = 1$:

$$S_k^d = \{s(x) \in C^{d-1}[0,1] \mid s(x) \text{ is a polynomial of degree at most } d \text{ in each of the intervals } (\xi_i, \xi_{i+1}),\ i = 0,1,\ldots,k\}.$$

These spline functions were discussed in sections 1.2 and 1.3. Burchard (1974) has shown that if $g \in C^{d+1}[0,1]$, then for large k the distance in the L_p-norm $(0 < p \leq \infty)$ of g from S_k^d is bounded by M/k^{d+1}, with a much smaller M than in similar estimates for other processes of the approximation. Dodson (1972) has found the similar results for class of functions g other than in $C^{d+1}[0,1]$. On the basis of these asymptotic results he has presented an algorithm to produce sequentially a "good" knot-set. He means "good" in the sense that the sequence of errors of best approximations behave asymptotically like $O(1/k^{d+1})$.

A disadvantage with the class S_k^d is that it is not closed. A sequence of spline functions may have some knots coalescing in such a way that the limit function exists but does not have the required continuity of derivatives at the knots. In general, the coalescense of m knots reduces the number of continuous derivatives by m-1. We are thereby led to a larger set which turns out to be closed, the set $\bar{S}_k^d$ of extended spline functions:

$\bar{S}_k^d = \{s(x) |$ there exists $0 = \xi_0 < \xi_1 \ldots < \xi_{r+1} = 1$ and integers $m_1, \ldots, m_r$ with $1 \leq m_i \leq d+1$ and $\sum_{i=1}^{r} m_i = k$ such that $s(x)$ is a polynomial of degree at most d in each of the intervals (ξ_i, ξ_{i+1}) while $s \in C^{d-m_i}$ in an open neighborhood of ξ_i, $i = 1, \ldots, r\}$.

The point ξ_i is said to be a m_i-tuple knot. The first unusual result on the error of best approximation by extended splines is due to Rice (1969). He considered the function $g(x) = x^\alpha$ on $[0,1]$ for any value of α such that x^α is in $L_p(0,1)$, $1 \leq p \leq \infty$, and showed that

$$\mathrm{dist}_p(g, \bar{S}_k^d) \overset{\mathrm{def}}{=} \mathrm{Inf}\{||g-s||_p : s \in \bar{S}_k^d\} = O(1/k^{d+1}), \text{ as } k \to \infty.$$

In proving this result Rice has used a specific set of d-tuple knots which are selected according to a rule depending on α.

CHAPTER II

THEORETICAL RESULTS FOR STEP FUNCTION AND GENERAL CASE

2.1 General Case: Minimization of IMSE

In this section we shall consider the minimization of the IMSE for different choices of the estimator. 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 $g(x) = \sum_{i=1}^{m} \theta_i f_i(x)$ where $m = k+d+1$. The estimator $\hat{\theta} = C\bar{y}$ (see 1.4.2) is restricted to be unbiased. The unbiasedness of $\hat{\theta} = C\bar{y}$ restricts C so that

$$CF' = I \tag{2.1.1}$$

where F is the mxr matrix $F = (f(x_1),\ldots,f(x_r))$ and I is mxm identity matrix. Note that $f(x_j)$ for each $j = 1,2,\ldots,r$ is a column vector consisting of m elements $f_1(x_j),\ldots,f_m(x_j)$. The quantity V in (1.4.5) is then minimized by the usual least square estimator (LSE)

$$C = M^{-1}(\mu)FD_\mu. \tag{2.1.2}$$

Here μ represents the design measure placing mass μ_i on x_i, $i = 1,2,\ldots,r$, $M(\mu)$ is the mxm matrix $\int f(x)f'(x)d\mu(x)$, and $D(\mu)$ is the diagonal matrix with diagonal elements $\mu_1,\ldots,\mu_r$. The estimator $\hat{\theta}$ is then

$$\hat{\theta} = C\bar{y} = M^{-1}(\mu)\int f(x)\bar{y}_x d\mu(x) \qquad (2.1.3)$$

where $\bar{y}_x$ is the average of observations at point x.

Some authors including Box and Draper (1959) and Karson, Manson and Hader (1969) have proposed attaching more importance to the bias part B. For other discussions see Kiefer (1973).

If we minimize B in (1.4.5) separately with respect to Cg_r the minimizing value is easily seen to be

$$Cg_r = M_0^{-1}s \qquad (2.1.4)$$

where

$$s = \int f(x)g(x)dx \text{ and } M_0 = \int f(x)f'(x)dx. \qquad (2.1.5)$$

Now we will try to minimize the expression (1.4.5) as a whole.

Theorem 2.1.1: The matrix C, which minimizes the IMSE, differs from the bias minimizing estimator given in (2.1.4) by a factor of t, where t is a constant defined by

$$t = \frac{q}{\frac{\sigma^2}{n} + q}, \qquad q = g_r' D\mu g_r.$$

Remark: We shall note that the expression for minimizing C (see (2.1.8)) involves the unknown g. We will obtain an estimator close to the "optimal one" given in (2.1.8) by estimating g (e.g. using the LSE from (2.1.2)) and replugging into the expression for C in (2.1.8).

Proof:

In (1.4.5), we have

$$B = \int g^2(x)dx - 2\int g(x)f'(x)Cg_r dx + \int f'(x)Cg_r g_r' C'f(x)dx$$

$$= \int g^2(x)dx - 2s'Cg_r + \text{Tr } Cg_r g_r' \int f(x)f'(x)dx.$$

So, we can rewrite the expression (1.4.5) as

$$V + B = \text{Tr } CAC'M_0 - 2s'Cg_r + \int g^2(x)dx \qquad (2.1.6)$$

where $A = (g_r g_r' + (\sigma^2/n)D_\mu^{-1})$. The first two terms in (2.1.6) can be written as

$$\text{Tr}(C-M_0^{-1}sg_r'A^{-1})'M_0(C-M_0^{-1}sg_r'A^{-1})A - \text{Tr } A^{-1}g_r s'M_0^{-1}sg_r'. \qquad (2.1.7)$$

If we choose

$$C = M_0^{-1}sg_r'A^{-1} \qquad (2.1.8)$$

then any such choice clearly minimizes the expression in (2.1.7) and hence V+B. Now Cg_r is given by

$$Cg_r = M_0^{-1}sg_r'A^{-1}g_r.$$

To compare this Cg_r with (2.1.4) we can calculate A^{-1} using a standard lemma, e.g. see Federov (1972). The inverse of A is

$$A^{-1} = \left(\frac{n}{\sigma^2}\right)\left(I - \frac{D_\mu g_r g_r'}{\frac{\sigma^2}{n} + g_r'D_\mu g_r}\right)D_\mu$$

and

$$Cg_r = tM_0^{-1}s \qquad (2.1.9)$$

since

$$t = \frac{q}{\frac{\sigma^2}{n} + q} \quad \text{and} \quad q = g_r' D_\mu g_r .$$

Also the minimum value of V+B is given by

$$(V+B)_{min} = \int g^2(x)dx - \left(\frac{q}{\frac{\sigma^2}{n} + q}\right)(s'M_0^{-1}s).$$

Now we can see that the quantity Cg_r given in (2.1.9) is the same as that given for Cg_r in (2.1.4) except for the factor t. This factor compares the relative sizes of σ^2/n and the function g measured by $q = g_r' D_\mu g_r = \int g^2(x)d\mu(x)$.

Q.E.D.

2.2 Step Function: Asymptotic Value of IMSE

Here we shall consider the asymptotic behavior of the IMSE when g(x) is approximated by a step function s(x) having jumps (knots) at ξ_i, i = 1,...,k, i.e. s(x) is a spline function of degree d = 0 (see Section 1.4). So in accordance with section 1.4, here we shall assume that g(x) is continuously differentiable in the interval [0,1] and the function s(x) has the representation

$$s(x) = \alpha_0 + \sum_{i=1}^{k} \beta_i (x-\xi_i)_+^0 \qquad (2.2.1)$$

where $(x-\xi_i)_+^0 = 1$ if $x > \xi_i$ and zero otherwise. As explained in section 1.3, it is more convenient to use the Lagrange basis in (2.2.1). The Lagrange basis $\ell(x)$ consists of the functions $\ell_i(x)$ s.t. $\ell_i(\xi_j) = \delta_{ij}$, i,j = 1,2,...,k+1. The function $\ell_i(x)$ looks like as

$$\ell_i(x) = \begin{cases} 1, & \xi_{i-1} < x \le \xi_i \\ 0, & \text{otherwise.} \end{cases}$$

These functions are left continuous everywhere. In this form the coefficients θ_i in $s(x) = \sum_{i=1}^{k+1} \theta_i \ell_i(x)$ are simply the values of s on $\xi_{i-1} < x \le \xi_i$. Note that the vector θ is not the same as that of section 1.4, it is changed according to the change of basis.

In discussing the asymptotic behavior of the IMSE, we shall assume that the knot structure and the design μ show some regularity as k or n becomes large. In most cases the knots ξ_i, $i = 1,\ldots,k$ will be chosen so that

$$\int_0^{\xi_i} p(x)dx = i/(k+1), \quad i = 1,2,\ldots,k \tag{2.2.2}$$

where p is some suitable chosen density. 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.

Let $\{T_k\}$ be a sequence of partitions on [0,1]:

$$T_k: \quad 0 = \xi_0 < \xi_1 <\ldots< \xi_{k+1} = 1. \tag{2.2.3}$$

Set

$$\delta_\nu = \xi_\nu - \xi_{\nu-1}, \qquad \nu = 1,\ldots,k+1$$

and define the mesh of the partition to be

$$\text{mesh } T_k = \delta = \max_{1 \le i \le k+1} \delta_i .$$

We shall be concerned with sequences $\{T_k\}$ for which mesh $T_k \to 0$ as $k \to \infty$.

Our first theorem deals with the asymptotic expression for the integrated variance of LSE. From (1.4.5), the integrated variance is

$$V = \frac{\sigma^2}{n} \operatorname{tr} C\, D_\mu^{-1} C' M_0 \tag{2.2.4}$$

where now C is the (k+1)xr matrix and M_0 is the (k+1)x(k+1) matrix $\int_0^1 \ell(x)\ell'(x)dx$. If the least square estimator is used, then from (2.1.2)

$$C = M^{-1}(\mu) L D_\mu$$

where L is the (k+1)xr matrix $L = (\ell(x_1),\ldots,\ell(x_r))$ and $M(\mu)$ is the (k+1)x(k+1) matrix $\int \ell(x)\ell'(x)d\mu(x)$. Substituting this value of C in (2.2.4), we get

$$\begin{aligned} V &= \frac{\sigma^2}{n} \operatorname{tr} M^{-1}(\mu) L D_\mu L' M^{-1}(\mu) M_0 \\ &= \frac{\sigma^2}{n} \operatorname{tr} M^{-1}(\mu) M_0 \end{aligned} \tag{2.2.5}$$

since $\quad L D_\mu L' = M_\mu$.

Theorem 2.2.1: Let the design measure μ have continuous density $h(x)$ such that $h(x) > 0$ for all $x \in [0,1]$. Let the knots be chosen using (2.2.2) where $p(x)$ is continuous on [0,1]. If the least square estimate (LSE) is used, then

$$\text{Tr } M^{-1}(\mu)M_0 \approx k \int_0^1 \frac{p(x)}{h(x)}\, dx \tag{2.2.6}$$

Proof: Since μ has density $h(x)$, we can express the information matrix $M(\mu)$ as

$$M(\mu) = \int_0^1 \ell(x)\ell'(x)h(x)dx.$$

Since $\ell_i(x)\ell_j(x) = 0$ if $i \neq j$ for all x, the matrices $M(\mu)$ and M_0 are diagonal matrices. The diagonal elements of $M(\mu)$ are $\int_0^1 \ell_i^2(x)h(x)dx$, $i = 1,2,\ldots,k+1$ and that of M_0 are $\int_0^1 \ell_i^2(x)dx$, $i = 1,2,\ldots,k+1$. Therefore

$$\text{Tr } M^{-1}(\mu)M_0 = \sum_{i=1}^{k+1} \frac{\int \ell_i^2(x)dx}{\int \ell_i^2(x)h(x)dx}$$

$$= \sum_{i=1}^{k+1} (\delta_i / \int_{\xi_{i-1}}^{\xi_i} h(x)dx)$$

where $\delta_i = \xi_i - \xi_{i-1}$. Now by the mean value theorem of integration $\int_{\xi_{i-1}}^{\xi_i} h(x)dx = h(\eta_i)\delta_i$ for some $\eta_i \in (\xi_{i-1}, \xi_i)$. Again using the mean value theorem we have from (2.2.2), $p(\gamma_i)\delta_i = (k+1)^{-1}$ for some $\gamma_i \in (\xi_{i-1}, \xi_i)$. Therefore,

$$\text{Tr } M^{-1}(\mu)M_0 = (k+1) \sum_{i=1}^{k+1} \frac{p(\gamma_i)}{h(\eta_i)} \delta_i .$$

In the limit as $k \to \infty$ in such a way that $\delta = \max_{1 \le i \le k+1} \delta_i \to 0$, the expression $\sum_{i=1}^{k+1} \frac{p(\gamma_i)}{h(\eta_i)} \delta_i$ tends to $\int_0^1 \frac{p(x)}{h(x)}\, dx$. Q.E.D.

The following version of "mean value theorem for integration" will be used very often in this and next chapter including the proof of next theorem.

Lemma 2.2.1: If ϕ and ψ are continuous and ϕ is positive on $[a,b]$, then there exists a point ξ such that $a \leq \xi \leq b$, and

$$\int_a^b \phi(x)\psi(x)dx = \psi(\xi)\int_a^b \phi(x)dx. \tag{2.2.7}$$

Proof: See Hardy (1943, p. 321).

We shall discuss, now, the asymptotic value of the bias term. The integrated bias term is

$$B = \int_0^1 (g(x) - \ell'(x)Cg_r)^2 \, dx.$$

Here in this section we are concerned with the LSE and if we assume that the design μ has the density $h(x)$, then

$$B = \int_0^1 (g(x) - \ell'(x)c_h)^2 dx$$

where $c_h = M^{-1}(\mu)\int_0^1 \ell(x)g(x)h(x)dx.$

We can break the bias term into two factors as follows:

$$B = B_1 + B_2$$

where $B_1 = \int_0^1 (g(x) - \ell'(x)c_g)^2 dx$ and $B_2 = \int_0^1 (\ell'(x)c_g - \ell'(x)c_h)^2 dx$, where $c_g = M_0^{-1}s$. Recall from section 2.1 that c_g is the vector which minimizes B (see (2.1.4)).

Theorem 2.2.2: Let the design μ have continuous density $h(x)$ such that $h(x) > 0$ for all $x \in [0,1]$. Let the knots satisfy (2.2.2), where $p(x)$ is continuous and positive on $[0,1]$. If the LSE is used then the bias term has the asymptotic expression

$$B \approx \frac{1}{12k^2} \int_0^1 \frac{(g'(x))^2}{(p(x))^2}\, dx. \tag{2.2.8}$$

Proof: The proof of this theorem can be divided into two parts. In the first part we shall show that

$$\lim_{k\to\infty} k^2 B_1 = \frac{1}{12} \int_0^1 \frac{(g'(x))^2}{p^2(x)}\, dx \tag{2.2.9}$$

and in the second part we shall show that $B_2 = o(k^{-2})$ as $k \to \infty$.

We have

$$B_1 = \int (g(x) - \ell'(x)c_g)^2 dx$$

where $c_g = M_0^{-1}s$, $M_0 = \int \ell(x)\ell'(x)dx$ and $s = \int \ell(x)g(x)dx$. Since M_0 is a diagonal matrix, the vector c_g looks like

$$c_g = \left(\frac{1}{\delta_1} \int_{\xi_0}^{\xi_1} g(y)dy, \ldots, \frac{1}{\delta_{k+1}} \int_{\xi_k}^{\xi_{k+1}} g(y)dy\right)' \tag{2.2.10}$$

and we can write

$$B_1 = \sum_{i=1}^{k+1} \int_{\xi_{i-1}}^{\xi_i} (g(x) - \ell'(x)c_g)^2 dx$$

$$= \sum_{i=1}^{k+1} \int_{\xi_{i-1}}^{\xi_i} (g(x) - \frac{1}{\delta_i} \int_{\xi_{i-1}}^{\xi_i} g(y)dy)^2 dx. \qquad (2.2.11)$$

From the mean value theorem,

$$g(y) = g(\xi_{i-1}) + g'(\eta_{y,i})(y-\xi_{i-1}) \qquad (2.2.12)$$

where $\eta_{y,i} \in (\xi_{i-1}, y)$. The subscript y in $\eta_{y,i}$ shows that point $\eta_{y,i}$ depends on y. Now

$$\int_{\xi_{i-1}}^{\xi_i} g(y)dy = g(\xi_{i-1})\delta_i + \int_{\xi_{i-1}}^{\xi_i} g'(\eta_{y,i})(y-\xi_{i-1})dy$$

$$= g(\xi_{i-1})\delta_i + g'(\rho_i)\int_{\xi_{i-1}}^{\xi_i} (y-\xi_{i-1})dy,$$

using Lemma 2.2.1, where $\rho_i \in (\xi_{i-1}, \xi_i)$. Therefore

$$\int_{\xi_{i-1}}^{\xi_i} g(y)dy = g(\xi_{i-1})\delta_i + g'(\rho_i)\frac{\delta_i^2}{2}. \qquad (2.2.13)$$

Now using this and substituting for g(x) from (2.2.12) in (2.2.11), we get

$$B_1 = \sum_{i=1}^{k+1} \int_{\xi_{i-1}}^{\xi_i} (g'(\eta_{x,i})(x-\xi_{i-1}) - g'(\rho_i)\frac{\delta_i}{2})^2 dx$$

$$= \sum_{i=1}^{k+1} \int_{\xi_{i-1}}^{\xi_i} (g'(\eta_{x,i}))^2 (x-\xi_{i-1})^2 dx + \sum_{i=1}^{k+1} \frac{(g'(\rho_i))^2 \delta_i^3}{4}$$

$$- 2 \sum_{i=1}^{k+1} \frac{g'(\rho_i)\delta_i}{2} \int_{\xi_{i-1}}^{\xi_i} g'(\eta_{x,i})(x-\xi_{i-1})dx.$$

Now using Lemma 2.2.1 twice, we get

$$B_1 = \sum_{i=1}^{k+1}(g'(\gamma_i))^2 \frac{\delta_i^3}{3} + \sum_{i=1}^{k+1} \frac{(g'(\rho_i))^2\delta_i^3}{4} - \sum_{i=1}^{k+1} g'(\rho_i)\delta_i \frac{g'(\rho_i)\delta_i^2}{2}$$

$$= \sum_{i=1}^{k+1} \frac{(g'(\gamma_i))^2\delta_i^3}{3} - \sum_{i=1}^{k+1} \frac{(g'(\rho_i))^2\delta_i^3}{4},$$

where $\gamma_i \in (\xi_{i-1},\xi_i)$. From (2.2.2), $p(\eta_i)\delta_i = (k+1)^{-1}$ for some $\eta_i \in (\xi_{i-1},\xi_i)$. So now

$$B_1 = \frac{1}{(k+1)^2} \cdot \frac{1}{3} \sum_{i=1}^{k+1} \frac{(g'(\gamma_i))^2}{p^2(\eta_i)} \delta_i - \frac{1}{(k+1)^2} \cdot \frac{1}{4} \sum_{i=1}^{k+1} \frac{(g'(\rho_i))^2}{p^2(\eta_i)} \delta_i .$$

Therefore

$$\lim_{k\to\infty} k^2 B_1 = \frac{1}{3} \int \frac{(g'(x))^2}{p^2(x)} dx - \frac{1}{4} \int \frac{(g'(x))^2}{p^2(x)} dx$$

$$= \frac{1}{12} \int \frac{(g'(x))^2}{p^2(x)} dx.$$

This proves the first part.

Now we shall prove that $B_2 = o(k^{-2})$ as $k \to \infty$. Here

$$B_2 = \int(\ell'(x)c_g - \ell'(x)c_h)^2 dx$$

$$= \sum_{i=1}^{k+1} \int_{\xi_{i-1}}^{\xi_i} (\ell'(x)c_g - \ell'(x)c_h)^2 dx$$

$$= \sum_{i=1}^{k+1} \int_{\xi_{i-1}}^{\xi_i} (c_{g,i} - c_{h,i})^2 dx$$

$$= \sum_{i=1}^{k+1} (c_{g,i} - c_{h,i})^2 \delta_i,$$

where $c_{g,i}$ and $c_{h,i}$ denote the ith element of the vector c_g and c_h

respectively. From (2.2.10),

$$c_{g,i} = \frac{1}{\delta_i} \int_{\xi_{i-1}}^{\xi_i} g(x)dx.$$

Also since $c_h = M^{-1}(\mu)\int \ell(x)g(x)h(x)dx$ and $M(\mu)$ is a $(k+1)x(k+1)$ diagonal matrix with diagonal elements as $\int \ell_i^2(x)h(x)dx$, $i = 1, 2, \ldots, k+1$, we have

$$c_{h,i} = \frac{\int_0^1 \ell_i(x)h(x)g(x)dx}{\int_0^1 \ell_i^2(x)h(x)dx}$$

$$= \frac{\int_{\xi_{i-1}}^{\xi_i} h(x)g(x)dx}{\int_{\xi_{i-1}}^{\xi_i} h(x)dx}$$

$$= \frac{\int_{\xi_{i-1}}^{\xi_i} h(x)(g(\xi_{i-1})+g'(\eta_{x,i})(x-\xi_{i-1}))dx}{\int_{\xi_{i-1}}^{\xi_i} h(x)dx}, \quad \text{using (2.2.12)}$$

$$= g(\xi_{i-1}) + \frac{\int_{\xi_{i-1}}^{\xi_i} h(x)g'(\eta_{x,i})(x-\xi_{i-1})dx}{\int_{\xi_{i-1}}^{\xi_i} h(x)dx}.$$

Now from (2.2.13) and the above equality, we get

$$c_{g,i}-c_{h,i} = g'(\rho_i)\,\frac{\delta_i}{2} - \frac{\int_{\xi_{i-1}}^{\xi_i} h(x)g'(\eta_{x,i})(x-\xi_{i-1})dx}{\int_{\xi_{i-1}}^{\xi_i} h(x)dx}$$

$$= \frac{\frac{g'(\rho_i)\delta_i}{2}\int_{\xi_{i-1}}^{\xi_i} h(x)dx - \int_{\xi_{i-1}}^{\xi_i} h(x)g'(\eta_{x,i})(x-\xi_{i-1})\ x}{\int_{\xi_{i-1}}^{\xi_i} h(x)dx},$$

where $\rho_i \in (\xi_{i-1},\xi_i)$.

Now the numerator of the above expression is equal to

$$\frac{g'(\rho_i)\delta_i}{2}\int_{\xi_{i-1}}^{\xi_i}(h(\xi_{i-1}) + h(x) - h(\xi_{i-1}))dx$$

$$- g'(\rho_i)\int_{\xi_{i-1}}^{\xi_i}(x-\xi_{i-1})(h(\xi_{i-1}) + h(x) - h(\xi_{i-1}))dx$$

$$- \int_{\xi_{i-1}}^{\xi_i}(g'(\eta_{x,i}) - g'(\rho_i))(x-\xi_{i-1})h(x)dx$$

$$= \frac{g'(\rho_i)\delta_i}{2}\int_{\xi_{i-1}}^{\xi_i}(h(x) - h(\xi_{i-1}))dx$$

$$- g'(\rho_i)\int_{\xi_{i-1}}^{\xi_i}(x-\xi_{i-1})(h(x)-h(\xi_{i-1}))dx$$

$$- \int_{\xi_{i-1}}^{\xi_i}(g'(\eta_{x,i}) - g'(\rho_i))(x-\xi_{i-1})h(x)dx.$$

Since h and g' are continuous on [0,1], the above expression in absolute value is less than

$$K_1 \frac{\delta_i}{2} \omega(h,\delta)\delta_i + K_1\omega(h,\delta) \frac{\delta_i^2}{2} + K_2\omega(g',\delta) \frac{\delta_i^2}{2}$$

$$= K_1\omega(h,\delta)\delta_i^2 + K_2\omega(g',\delta) \frac{\delta_i^2}{2}$$

where $\omega(h,\delta)$ and $\omega(g',\delta)$ are respectively the modulus of continuity of h and g', and K_1 and K_2 are positive constants such that $|g'(x)| < K_1$ and $h(x) < K_2$ for all $x \in [0,1]$. From the mean value theorem, $\int_{\xi_{i-1}}^{\xi_i} h(x)dx = h(\alpha_i)\delta_i$ for some $\alpha_i \in (\xi_{i-1},\xi_i)$ and therefore

$$|c_{g,i} - c_{h,i}| < \frac{K_1\omega(h,\delta)\delta_i + K_2\omega(g',\delta) \frac{\delta_i}{2}}{h(\alpha_i)} .$$

Given that $h(x) > 0$, we shall have $\frac{1}{h(x)} < K_3$ for some positive constant K_3. Therefore we have

$$B_2 = \sum_{i=1}^{k+1} (c_{g,i} - c_{h,i})^2\delta_i$$

$$< (K_3K_1\omega(h,\delta) + \frac{K_3K_2\omega(g',\delta)}{2})^2 \sum_{i=1}^{k+1} \delta_i^3$$

$$= (K_3K_1\omega(h,\delta) + \frac{K_3K_2\omega(g',\delta)}{2})^2 \sum_{i=1}^{k+1} \frac{\delta_i}{p^2(\eta_i)} \cdot \frac{1}{(k+1)^2} ,$$

applying the mean value theorem to (2.2.2), where $\eta_i \in (\xi_{i-1},\xi_i)$. Now letting $k \to \infty$ in such a way that $\delta \to 0$, we shall have $\lim_{k\to\infty} (k+1)^2B_2 = 0$ since $\omega(h,\delta)$ and $\omega(g',\delta)$ tends to zero as δ goes to zero. This proves the second part and hence the theorem. Q.E.D.

Thus here we have shown that if LSE is used then

$$\text{IMSE} \approx k \frac{\sigma^2}{n} \int_0^1 \frac{p(x)}{h(x)} + \frac{1}{12k^2} \int_0^1 \frac{(g'(x))^2}{(p(x))^2} \, dx \tag{2.2.14}$$

which shows that a = 1 and b = 1/12 in (1.4.6) for the case d = 0. Note the asymptotics found here are with the number of knots k going to infinity. The number of observations n should be at least k+1 and will usually be increasing much faster than k.

2.3. General Case: Minimization of Asymptotic Value of IMSE

As conjectured in section 1.4, the asymptotic value of the IMSE, when the function g(x) is estimated by splines of degree d, is

$$J = \text{IMSE} \approx a \frac{k}{n} \sigma^2 \int_0^1 \frac{p(x)}{h(x)} \, dx + \frac{b}{k^{2d+2}} \int_0^1 \frac{(g^{(d+1)}(x))^2}{(p(x))^{2d+2}} \, dx. \tag{2.3.1}$$

In the last section we actually proved our conjecture for the case d = 0 and found the value of the constants a and b for this case. Here we shall minimize the asymptotic value of the IMSE with respect to the three "variables" (i) k, the number of knots, (ii) p(x), the displacement of knots and (iii) h(x), the allocation of observations. Instead of doing this minimization for the particular case d = 0, we shall do it for the general case i.e. we shall minimize J given in (2.3.1). In the expression for J, only the first term contains the factor h(x). Using the Schwarz's inequality and the fact that h(x) is a density, we can show that the first term in (2.3.1) is minimized by

$$h(x) = p^{1/2}(x)/\int_0^1 p^{1/2}(y)dy. \tag{2.3.2}$$

Substituting this value of h(x) in (2.3.1) yields

$$J = \frac{ak\sigma^2}{n}\left(\int_0^1 p^{1/2}(x)dx\right)^2 + \frac{b}{k^{2d+2}}\int_0^1 \frac{(g^{(d+1)}(x))^2}{(p(x))^{2d+2}}\,dx. \tag{2.3.3}$$

Now the problem is reduced to minimize J with respect to p and k. It will be shown in the following theorem that the minimizing p and k are given by

$$p(x) = \frac{\left\{g^{(d+1)}(x)\right\}^{\frac{4}{(4d+5)}}}{\int_0^1 \left\{g^{(d+1)}(y)\right\}^{\frac{4}{(4d+5)}}dy} \tag{2.3.4}$$

and

$$k = \int_0^1 (g^{(d+1)})^{\frac{4}{4d+5}} \left\{\frac{nb(2d+2)}{a\sigma^2}\Big/\left\{\int_0^1 (g^{(d+1)})^{\frac{2}{4d+5}}\right\}\right\}^{\frac{1}{2d+3}}. \tag{2.3.5}$$

Theorem 2.3.1: The functional J given in (2.3.1) is absolutely minimized by h,p, and k given (2.3.2), (2.3.4), and (2.3.5) respectively provided k given in (2.3.5) is less than or equal to n.

Proof: Well, we have already shown the minimization of J with respect to h. Now differentiating the expression (2.3.3) with respect to k and equating it to zero, we get

$$\frac{a\sigma^2}{n}\left(\int \sqrt{p}\right)^2 - \frac{b(2d+2)}{k^{2d+3}}\int \frac{(g^{(d+1)})^2}{p^{2d+2}} = 0$$

which yields

$$k = \left\{ \frac{b(2d+2)\int \frac{(g^{(d+1)})^2}{p^{2d+2}}}{\frac{a\sigma^2}{n}\left(\int \sqrt{p}\right)^2} \right\}^{\frac{1}{(2d+3)}} . \tag{2.3.6}$$

We can verify that this k minimizes J for each p. Substituting this value of k in (2.3.3), we get

$$J = \rho \left\{ \int \frac{(g^{(d+1)})^2}{p^{2d+2}} \right\}^{\frac{1}{2d+3}} \left\{ \int \sqrt{p} \right\}^{\frac{4d+4}{2d+3}} \tag{2.3.7}$$

where $\rho = \left(\frac{2d+3}{2d+2}\right)\left(\frac{a\sigma^2}{n}\right)^{\frac{2d+2}{2d+3}} ((2d+2)b)^{\frac{1}{2d+3}}$. Finally, we have to minimize J given in (2.3.7) with respect to p(x). This can be done by using Holder's inequality

$$\int \psi\phi \le \left(\int \psi^\alpha\right)^{1/\alpha} \left(\int \phi^\beta\right)^{1/\beta}. \tag{2.3.8}$$

In (2.3.8), let us take $\psi^\alpha = \frac{(g^{(d+1)})^2}{p^{2d+2}}$, $\phi^\beta = p^{1/2}$. Now we want to choose α and β in such a way that $\psi\phi$ is independent of p and $1/\alpha + 1/\beta = 1$. This can be done by choosing $\alpha = 4d+5$, $\beta = (4d+5)/(4d+4)$, $\psi = \frac{(g^{(d+1)})^{2/4d+5}}{(p)^{(2d+2/4d+5)}}$, and $\phi = (p)^{(2d+2/4d+5)}$. Substituting these values in (2.3.8) gives

$$\int (g^{(d+1)})^{\frac{2}{4d+5}} \le \left\{ \int \frac{(g^{(d+1)})^2}{p^{2d+2}} \right\}^{\frac{1}{4d+5}} \left\{ \int \sqrt{p} \right\}^{\frac{4d+4}{4d+5}} .$$

In the above, equality holds if and only if $p(x) = c(g^{(d+1)}(x))^{\frac{4}{4d+5}}$, where c is a constant. This shows that J in (2.3.7) is minimized by $p = c(g^{(d+1)})^{\frac{4}{4d+5}}$, where $c = 1/\int(g^{(d+1)})^{\frac{4}{4d+5}}$ since p is a density. Now putting this value of p in (2.3.6) we get the desired result.

Q.E.D.

In the above, we have assumed that all the three variables, namely k, p and h, are unknown. We might confront the situations when one or two of these variables are known, e.g. we might be given the number of knots or the distribution of observations or both, and so on. To cope with these situations we shall now consider the following six other possible cases of the minimization of J:

(i) with respect to 'k' only

(ii) with respect to 'h(x)' only

(iii) with respect to 'k and h(x)'

(iv) with respect to 'k and p(x)'

(v) with respect to 'p(x)' only

(vi) with respect to 'p(x) and h(x)'.

This is not a natural ordering, but we are considering these cases in order of difficulty. In order to show the dependence of J on the variable (or variables) with respect to which it is being minimized, we shall express J as a function of that (or those) variable (or variables). For example in case (i) J will be denoted by J(k) while in case (iv) it will be denoted by J(k,p), and so on.

In the first four cases we have the exact solution for the minimization problem. In fifth case, J can be minimized under certain condition on h(x) and g(x) (see Theorem 2.3.2). For the last case, it is not possible to find the exact solution; only in a special case ($\sigma^2 = 0$), can we characterize the solution.

<u>Case (i)</u>:

Differentiating J(k), given in (2.3.1), with respect to k and equating it to zero, we get

$$\frac{a\sigma^2}{n} \int \frac{p}{h} - \frac{b(2d+2)}{k^{2d+3}} \int \frac{(g^{(d+1)})^2}{p^{2d+2}} = 0$$

which yields

$$k^{2d+3} = \frac{b(2d+2)\int \frac{(g^{(d+1)})^2}{p^{2d+2}}}{\frac{a\sigma^2}{n} \int \frac{p}{h}} . \qquad (2.3.9)$$

Since k has to be less than or equal to n (number of observations), and J(k) is a strictly convex function of k, the minimizing k is given by

$$k = \min\left[\left\{ \frac{b(2d+2)\int \frac{(g^{(d+1)})^2}{p^{2d+2}}}{\frac{a\sigma^2}{n} \int \frac{p}{h}} \right\}^{\frac{1}{(2d+3)}} , n\right] . \qquad (2.3.10)$$

<u>Case (ii)</u>:

We notice that in (2.3.1), the design factor h enters only in the variance term, i.e. the first term on right. Minimization of this term can be done using Schwarz's inequality. It is easily seen that the h(x) which minimizes J(h) is given by

$$h(x) = p^{1/2}(x)/\int_0^1 p^{1/2}(y)dy. \qquad (2.3.11)$$

This shows that h just depends upon p.

Case (iii):

We have

$$J(k,h) = \frac{ak}{n}\,\sigma^2 \int \frac{p}{h} + \frac{b}{k^{2d+2}} \int \frac{(g^{(d+1)})^2}{p^{2d+2}} \; .$$

If we minimize first with respect to k, then we get (see equation (2.3.9)),

$$k^{2d+3} = \frac{b(2d+2)\int \frac{(g^{(d+1)})^2}{p^{2d+2}}}{\frac{a\sigma^2}{n} \int \frac{p}{h}} \; . \qquad (2.3.12)$$

Also from case (ii), the minimizing h(x) is given by

$$h(x) = p^{1/2}(x)/\int p^{1/2}(y)dy. \qquad (2.3.13)$$

Substituting this value of h in (2.3.12), we get

$$k = \left\{ \frac{b(2d+2) \int \frac{(g^{(d+1)})^2}{p^{2d+2}}}{\frac{a\sigma^2}{n} (\int \sqrt{p})^2} \right\}^{1/(2d+3)} \; . \qquad (2.3.14)$$

As before, there is a restriction on k that it has to be less than or equal to n.

Case (iv):

We have

$$J(k,p) = \frac{ak\sigma^2}{n} \int \frac{p}{h} + \frac{b}{k^{2d+2}} \int \frac{(g^{(d+1)})^2}{p^{2d+2}} \quad . \qquad (2.3.15)$$

Minimizing first with respect to k, we get

$$k^{2d+3} = \frac{b(2d+2)\int \frac{(g^{(d+1)})^2}{p^{2d+2}}}{\frac{a\sigma^2}{n} \int \frac{p}{h}} \; .$$

Substituting this value of k in (2.3.15) gives

$$J(k,p) = \left(\frac{2d+3}{2d+2}\right)(b(2d+2))^{\frac{1}{2d+3}}\left(\frac{a\sigma^2}{n}\right)^{\frac{2d+2}{2d+3}}\left(\int \frac{p}{h}\right)^{\frac{2d+2}{2d+3}}\left(\int \frac{(g^{(d+1)})^2}{p^{2d+2}}\right)^{\frac{1}{2d+3}} .$$

Now the problem is to minimize the above functional with respect to 'p'. This can be done by using Holer's inequality, since

$$\int \frac{(g^{(d+1)})^{\frac{2}{2d+3}}}{h^{\frac{2d+2}{2d+3}}} \leq \left(\int \frac{p}{h}\right)^{\frac{2d+2}{2d+3}} \left(\int \frac{(g^{(d+1)})^2}{p^{2d+2}}\right)^{\frac{1}{2d+3}} .$$

In the above, the equality holds if

$$p(x) \propto \{(g^{(d+1)}(x))^2 h(x)\}^{1/(2d+3)} .$$

Thus the function J(k,p) is minimized by

$$p(x) = \frac{\{(g^{(d+1)}(x))^2 h(x)\}^{1/(2d+3)}}{\int \{(g^{(d+1)}(y))^2 h(y)\}^{1/(2d+3)} dy} \qquad (2.3.16)$$

and

$$k = \{\int\{(g^{(d+1)}(y))^2\ h(y)\}^{\frac{1}{2d+3}}dy\}\{nb(2d+2)/a\sigma^2\}^{\frac{1}{(2d+3)}}. \quad (2.3.17)$$

Case (v):

We have

$$J(p) = \frac{ak}{n}\sigma^2\int \frac{p}{h} + \frac{b}{k^{2d+2}}\int \frac{(g^{(d+1)})^2}{p^{2d+2}} . \quad (2.3.18)$$

Since $\int \frac{p}{h}$ is linear in p and $\int 1/p^{2d+2}$ is strictly convex in p, J(p) is strictly convex in p on the convex set U ={ p(x): p(x) > 0 for all $x \in [0,1]$ and $\int p(x)dx = 1$}. Here the minimization of J(p) will be done by using a variational argument on p. Since we want to minimize J subject to the constraint that $\int p(x)dx = 1$, we shall consider the minimization of the quantity

$$J(p) - \lambda\int p = c_1\int \frac{p}{h} + c_2\int \frac{(g^{(d+1)})^2}{p^{2d+2}} - \lambda\int p \quad (2.3.19)$$

where λ is the Lagrange multiplier, $c_1 = (a\sigma^2k)/n$, and $c_2 = (b/k^{2d+2})$.

Let $\bar{p}(x)$ denote the function that minimizes (2.3.19) and define

$$p(x) = \bar{p}(x) + \varepsilon\eta(x)$$

where $\eta(x)$ is an arbitrary continuous function for which $\int\eta(x)dx = 0$ and ε is an arbitrary parameter. Substituting this value of p in (2.3.19), we get

$$\bar{J}(\varepsilon) = c_1\int \frac{\bar{p}+\varepsilon\eta}{h} + c_2\int \frac{(g^{(d+1)})^2}{(\bar{p}+\varepsilon\eta)^{2d+2}} - \lambda\int \bar{p}+\varepsilon\eta. \quad (2.3.20)$$

A necessary condition for a minimum is, the vanishing of the first derivative of $\bar{J}$ with respect to ε at $\varepsilon = 0$; that is

$$\bar{J}'(0) = 0.$$

So we get

$$\int \left\{ \frac{c_1}{h} - \frac{(2d+2)c_2(g^{(d+1)})^2}{\bar{p}^{2d+3}} - \lambda \right\} \eta = 0. \tag{2.3.21}$$

Since η is an arbitrary function satisfying the condition that $\int \eta(x)dx = 0$, the equation (2.3.21) would imply that

$$\frac{c_1}{h} - \frac{(2d+2)c_2(g^{(d+1)})^2}{\bar{p}^{2d+3}} - \lambda = 0$$

which yields

$$\bar{p}(x) = \left\{ \frac{(2d+2)c_2(g^{(d+1)}(x))^2}{(c_1/h(x))-\lambda} \right\}^{\frac{1}{(2d+3)}} \tag{2.3.22}$$

where λ is to be found such that

$$\int_0^1 \bar{p}(x)dx = 1. \tag{2.3.23}$$

Since $\bar{J}'(0) = 0$ is only a necessary condition, $\bar{p}(x)$ given in (2.3.22) is an extremum (maximum, minimum, or stationary value) of $\bar{J}(\varepsilon)$ at $\varepsilon = 0$. In the following theorem we shall show that $\bar{p}(x)$ is actually an absolute (global) minima for $J(p)$.

Theorem 2.3.2: The functional $J = J(p)$ is absolutely and uniquely minimized by $\bar{p}$ given by (2.3.22) and (2.3.23) if

$$\max_{x\in[0,1]}\left\{\frac{c_1}{h(x)} - c_2(2d+2)(g^{(d+1)}(x))^2\right\} < \frac{c_1}{\max_{x\in[0,1]} h(x)}$$, where $c_1=(ak\sigma^2)/n$, and $c_2 = b/k^{2d+2}$.

Proof: First of all we note that $\bar{J}''(0) > 0$, where $\bar{J}(\varepsilon)$ is defined in (2.3.20). So $\bar{p}$ given in (2.3.22) is a local (relative) minima for $J(p)$. But the functional J is convex on the set $U = \{p: p(x) > 0$ for all $x \in [0,1]$, and $\int_0^1 p(x)dx = 1\}$, so a local minima will be also a global minima for J. To complete the proof of the theorem it is then sufficient to show that $\bar{p}$ given by (2.3.22) and (2.3.23) is an unique element of U. Let us define

$$\psi(\lambda) = \int_0^1 \left\{\frac{(2d+2)c_2(g^{(d+1)}(x))^2}{(c_1/h(x))-\lambda}\right\}^{\frac{1}{(2d+3)}} dx.$$

It is trivial to check that $\psi(\lambda)$ is a strictly increasing function of λ. Therefore the equation $\psi(\lambda) = 1$ will have an unique solution in λ. By mean value theorem of integration, for some $\rho \in [0,1]$,

$$\left\{\frac{(2d+2)c_2(g^{(d+1)}(\rho))^2}{(c_1/h(\rho))-\lambda}\right\}^{\frac{1}{(2d+3)}} = \psi(\lambda) = 1$$

which gives

$$\lambda = \frac{c_1}{h(\rho)} - (2d+2)c_2(g^{(d+1)}(\rho))^2. \tag{2.3.24}$$

We are given that

$$\max\left\{\frac{c_1}{h(x)} - c_2(2d+2)(g^{(d+1)}(x))^2\right\} < \frac{c_1}{\max h(x)} . \tag{2.3.25}$$

From (2.3.24) and (2.3.25), we have

$$\lambda < \frac{c_1}{\max h(x)}$$

which would imply that $\frac{c_1}{h(x)} - \lambda > 0$ for all x and hence $\bar{p}(x)$ given by (2.3.22) belongs to the class U. The uniqueness follows from the fact that the equation $\psi(\lambda) = 1$ has only one solution in λ.

Q.E.D.

Case (vi):

$$J(p,h) = \frac{ak}{n}\sigma^2 \int \frac{p}{h} + \frac{b}{k^{2d+2}} \int \frac{(g^{(d+1)})^2}{p^{2d+2}} . \tag{2.3.26}$$

For this case, in general we do not have an exact solution. However if $\sigma^2 = 0$, then the first factor on right in (2.3.26) is zero and the second factor is minimized if

$$p(x) \propto (g^{(d+1)}(x))^{\frac{2}{2d+3}} .$$

Thus if the first factor is small compared to the second factor, then we can find a local minimum for our problem. These will be discussed in the section 3.5.

Here we have discussed the minimization problem for spline functions of any degree d. For the case of step functions, we have to just put d = 0 in the above solutions of the minimization problem. In Theorem 2.3.1, we imposed a restriction that $k \leq n$. In the case when minimizing κ given by (2.3.5) becomes greater than n, we shall find local solutions (see section (3.5)).

CHAPTER III

THEORETICAL RESULTS FOR LINEAR SPLINE CASE

3.1. Asymptotic Value of Variance for LSE

Now we shall discuss the approximation of $g(x)$ by simple linear spline functions $s(x)$. Here we shall consider the asymptotic value of the integrated variance using the least square estimator for the different choices of design. Some other choices of the estimators will be considered in section 3.3. The function $s(x)$ is a continuous function composed of the straight line segments of the form

$$s(x) = \alpha_0 + \alpha_1 x + \sum_{i=1}^{k} \beta_i (x-\xi_i)_+ . \qquad (3.1.1)$$

As we did in the case of the step function (section 2.2) we shall use a Lagrange basis in (3.1.1) composed of linear spline functions $\ell_i(x)$, $i = 0,1,\ldots,k+1$ such that $\ell_i(\xi_j) = \delta_{ij}$, $i,j = 0,1,\ldots,k+1$. The graph of these ℓ_i's are shown in section 1.3. In this form the coefficient θ_i in $s(x) = \sum_{i=0}^{k+1} \theta_i \ell_i(x)$ are simply the values of s at $\xi_0,\xi_1,\ldots,\xi_{k+1}$. Here also we shall assume that the knots ξ_i, $i = 1,2,\ldots,k$ are chosen so that

$$\int_0^{\xi_i} p(x)dx = i/(k+1), \quad i = 1,2,\ldots,k \tag{3.1.2}$$

where p(x) is some smooth density.

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

$$\frac{n}{\sigma^2} V = \text{Tr } C \, D_\mu^{-1} C' M_0 \approx ak \int_0^1 \frac{p(x)}{h(x)} \, dx$$

for some suitable value of a.

Note that n appears in IMSE in (1.4.5) only in the denominator of the variance term. The asymptotics here and in the next sections are with the number of knots k going to infinity. In practice the number of observations n must, of course, be at least k+2 and will usually be increasing much faster than k.

Here also we shall be concerned with the sequences $\{T_k\}$ for which the mesh $\Gamma_k \to 0$ as $k \to \infty$. Under certain circumstances, here we shall also require the additional restriction on the partition T_k that

$$\frac{1}{\alpha} < \frac{\delta_i}{\delta_{i+1}} < \beta \ ; \quad i = 1,2,\ldots,k \tag{3.1.3}$$

where α and β are the positive constants. This restriction says that the ratio of the lengths of two neighboring intervals in the partition T_k should neither be too small nor be too large. In particular, we require this restriction in Theorem 3.1.2 and Lemma 3.1.1.

Our first theorem of this section deals with the asymptotic expression for the integrated variance of the LSE when we are using the design μ whose spectrum consists of the knot set $\xi_1, \xi_2, \ldots, \xi_k$ and the end points $\xi_0 = 0$ and $\xi_{k+1} = 1$. The weights at these points are given by,

$$\mu_\nu = \int_0^1 \ell_\nu(x)h(x)dx, \qquad \nu = 0,1,\ldots,k+1 \tag{3.1.4}$$

for some density $h(x)$. Unless otherwise mentioned we shall assume in most of the theorems of this and the next section that p and h are continuous on [0,1].

From (1.4.5), the integrated variance is

$$V = \frac{\sigma^2}{n} \text{ tr } CD_\mu^{-1}C'M_0 \tag{3.1.5}$$

where now C is a (k+2)xr matrix and M_0 is the (k+2)x(k+2) matrix $\int \ell(x)\ell'(x)dx$. If the least square estimator is used, then as shown in (2.2.5), we have

$$V = \frac{\sigma^2}{n} \text{ tr } M^{-1}(\mu)\, M_0 \tag{3.1.6}$$

where $M(\mu)$ is the (k+2)x(k+2) matrix $\int \ell(x)\ell'(x)d\mu(x)$.

Theorem 3.1.1: If the design μ and the knots are chosen using (3.1.4) and (3.1.2) respectively, where $h(x) > 0$ for all x, then for the LSE estimator, the asymptotic expression for Tr $M^{-1}(\mu)M_0$ is

$$\text{Tr } M^{-1}(\mu)M_0 \approx (2k/3)\int_0^1 \frac{p(x)}{h(x)}\, dx. \tag{3.1.7}$$

Proof: Since the design is concentrated at the knots, the matrix $M(\mu)$ is diagonal with diagonal elements μ_ν. Therefore

$$\mathrm{Tr}\; M^{-1}(\mu)M_0 = \sum_{\nu=0}^{k+1}(m_\nu/\mu_\nu)$$

where $m_\nu = \int \ell_\nu^2(x)dx = (\delta_\nu+\delta_{\nu+1})/3$, $\delta_\nu = \xi_\nu-\xi_{\nu-1}$. Also,

$$\mu_\nu = \int_0^1 \ell_\nu(x)h(x)dx$$

$$= \int_{\xi_{\nu-1}}^{\xi_\nu} \ell_\nu(x)h(x)dx + \int_{\xi_\nu}^{\xi_{\nu+1}} \ell_\nu(x)h(x)dx$$

$$= h(\xi_\nu)(\delta_\nu+\delta_{\nu+1})/2 + \int_{\xi_{\nu-1}}^{\xi_\nu} \ell_\nu(x)(h(x)-h(\xi_\nu))dx + \int_{\xi_\nu}^{\xi_{\nu+1}} \ell_\nu(x)(h(x)-h(\xi_\nu))dx.$$

So, we have

$$\frac{m_\nu}{\mu_\nu} = \frac{(\delta_\nu+\delta_{\nu+1})/3}{\left\{h(\xi_\nu)(\delta_\nu+\delta_{\nu+1})/2+\int_{\xi_{\nu-1}}^{\xi_\nu} \ell_\nu(x)(h(x)-h(\xi_\nu))dx+\int_{\xi_\nu}^{\xi_{\nu+1}} \ell_\nu(x)(h(x) - h(\xi_\nu))dx\right\}}$$

$$= \frac{(1/3)}{h(\xi_\nu)(1+\alpha_\nu)/2}$$

where

$$\alpha_\nu = \frac{2}{h(\xi_\nu)(\delta_\nu+\delta_{\nu+1})}\left[\int_{\xi_{\nu-1}}^{\xi_\nu} \ell_\nu(x)(h(x)-h(\xi_\nu))+\int_{\xi_\nu}^{\xi_{\nu+1}} \ell_\nu(x)(h(x)-h(\xi_\nu))dx\right].$$

Since $\delta_\nu = \xi_\nu-\xi_{\nu-1} < \delta$, $\delta_{\nu+1} = \xi_{\nu+1}-\xi_\nu < \delta$, we have $|h(x)-h(\xi_\nu)| < \omega(h,\delta)$ for $x \in [\xi_{\nu-1},\xi_\nu]$ and for $x \in [\xi_\nu,\xi_{\nu+1}]$. Letting $(1/h(x)) < K$ for some constant K, we get

$$|\alpha_\nu| < (2K/(\delta_\nu+\delta_{\nu+1}))\omega(h,\delta)[\int_{\xi_{\nu-1}}^{\xi_\nu} \ell_\nu(x)dx + \int_{\xi_\nu}^{\xi_{\nu+1}} \ell_\nu(x)dx]$$

$$= K\ \omega(h,\delta).$$

Since $\omega(h,\delta) \to 0$ as $\delta \to 0$, $|\alpha_\nu|$ can be made less than 1 for sufficiently small δ. Now we can write

$$m_\nu/\mu_\nu = (2/3h(\xi_\nu))(1+\alpha_\nu)^{-1}$$

$$= (2/3h(\xi_\nu))(1-\alpha_\nu+\alpha_\nu{}^2 - \ldots\ldots.)$$

$$\simeq (2/3h(\xi_\nu)) - (\alpha_\nu/(1+\alpha_\nu))(2/3h(\xi_\nu)).$$

Using the mean value theorem we have from (3.1.2), $p(\eta_\nu)\delta_\nu = 1/(k+1)$ for $\eta_\nu \in (\xi_{\nu-1},\xi_\nu)$. Now we have

$$(1/k+1)\text{Tr } M^{-1}(\mu)M_0 = (2/3)\sum_{\nu=0}^{k+1} \frac{p(\eta_\nu)}{h(\xi_\nu)}\ \delta_\nu - (2/3)\sum_{\nu=0}^{k+1} \frac{p(\eta_\nu)}{h(\xi_\nu)}\ \delta_\nu(\alpha_\nu/(1+\alpha_\nu)).$$

Since

$$(2/3)\sum_{\nu=0}^{k+1} \frac{p(\eta_\nu)}{h(\xi_\nu)}\ \delta_\nu(\alpha_\nu/(1+\alpha_\nu)) < (2/3)K\ \omega(h,\delta)\sum_{\nu=0}^{k+1} \frac{p(\eta_\nu)}{h(\xi_\nu)}\ \delta_\nu$$

and the quantity on right tends to zero as $\delta \to 0$, we have

$$\lim_{k\to\infty} (1/k+1)\ \text{Tr } M^{-1}(\mu)M_0 = (2/3)\int \frac{p(x)}{h(x)}\ dx.$$

Q.E.D.

It is a "well known" fact that given an arbitrary design μ, the value of Tr $M^{-1}(\mu)M_0$ can be decreased if we replace μ by a design concentrated at the knots $\xi_0,\xi_1,\ldots,\xi_{k+1}$ with weight $\mu_\nu = \int \ell_\nu(x)d\mu(x)$, $\nu = 0,1,\ldots,k+1$. For this result see Studden and Van Arman (1969). If μ has a given density $h(x)$, we expect the asymptotic value of V to be larger as we will show in Theorem 3.1.2. Discretizing μ appropriately decreases the variance term by a factor of 2/3 as shown in Theorem 3.1.1.

<u>Theorem 3.1.2</u>: Let μ has continuous density $h(x)$ such that $h(x) > 0$ for all $x \in [0,1]$. If the knots are chosen using (3.1.2) and the LSE is used, then

$$\operatorname{Tr} M^{-1}(\mu)M_0 \approx k\int \frac{p(x)}{h(x)}\,dx. \tag{3.1.8}$$

Proof: Since the design measure has the density $h(x)$, we can express the information matrix $M(\mu)$ as

$$M(\mu) = \int \ell(x)\ell'(x)h(x)dx.$$

To show its dependence on h, we shall denote the above matrix as $M(h)$ instead of $M(\mu)$ in the rest of the proof of this theorem.

The elements of the tridiagonal matrix $M(h)$ can be written as

$$m_{00}(h) = h(\xi_0)\delta_1/3 - b_{00}(h)$$

$$m_{k+1,k+1}(h) = h(\xi_{k+1})\delta_{k+1}/3 - b_{k+1,k+1}(h)$$

$$m_{\nu,\nu}(h) = h(\xi_\nu)(\delta_\nu+\delta_{\nu+1})/3 - b_{\nu,\nu}(h),\ \nu = 1,2,\ldots,k$$

$$m_{\nu,\nu+1}(h) = h(\xi_{\nu+1})\delta_{\nu+1}/6 - b_{\nu,\nu+1}(h),\ \nu = 0,1,\ldots,k$$

$$m_{\nu+1,\nu}(h) = h(\xi_\nu)\delta_{\nu+1}/6 - b_{\nu+1,\nu}(h),\ \nu = 0,1,\ldots,k$$

where $\delta_i = \xi_i - \xi_{i-1}$, $i = 1,2,\ldots,k+1$ and

$$b_{00}(h) = h(\xi_0)\delta_1/3 - \int \ell_0^2(x)h(x)dx$$

$$b_{k+1,k+1}(h) = h(\xi_{k+1})\delta_{k+1}/3 - \int \ell_{k+1}^2(x)h(x)dx$$

$$b_{\nu,\nu}(h) = h(\xi_\nu)(\delta_\nu+\delta_{\nu+1})/3 - \int \ell_\nu^2(x)h(x)dx,\ \nu = 1,2,\ldots,k$$

$$b_{\nu,\nu+1}(h) = h(\xi_{\nu+1})\delta_{\nu+1}/6 - \int \ell_\nu(x)\ell_{\nu+1}(x)h(x)dx,\ \nu=0,1,\ldots,k$$

$$b_{\nu+1,\nu}(h) = h(\xi_\nu)\delta_{\nu+1}/6 - \int \ell_{\nu+1}(x)\ell_\nu(x)h(x)dx,\ \nu = 0,1,\ldots,k.$$

With the above representation of the elements of $M(h)$, we can write down

$$M(h) = M_0D(h) - B(h)$$

where $M_0 = \int \ell(x)\ell'(x)dx$, $D(h) = \text{diag}(h(\xi_0), h(\xi_1),\ldots,h(\xi_{k+1}))$ and $B(h) = [b_{ij}(h)]$. Now

$$\begin{aligned} M^{-1}(h) &= [(I-B(h)D^{-1}(h)M_0^{-1})M_0D(h)]^{-1} \\ &= D^{-1}(h)M_0^{-1}(I-B(h)D^{-1}(h)M_0^{-1})^{-1}. \end{aligned} \tag{3.1.9}$$

Before going any further, we shall prove the following lemma which will enable us to complete the proof of the theorem.

Lemma 3.1.1: Let $F = B(h)D^{-1}(h)M_0^{-1}$, where $B(h)$, $D(h)$ and M_0 are as defined above and let the set of knots $\xi_1 < \xi_2 < \ldots\ldots < \xi_k$ satisfy (3.1.3). Then $||F|| = \max_i\{\sum_j|f_{ij}|\} < K\,\omega(h,\delta)$, where K is a positive constant and $\omega(h,\delta)$ is the modulus of continuity of h with respect to δ, where $\delta = \max_i(\xi_i-\xi_{i-1})$.

Proof: We want to find the bounds on the elements of matrix M_0 which will be done by using the Lemma A2 of the Appendix. For this purpose let us introduce the $(k+2)x(k+2)$ tridiagonal matrix A

$$A = \begin{bmatrix} 2 & 1 & 0 & & & 0 \\ \frac{\delta_1}{\delta_1+\delta_2} & 2 & \frac{\delta_2}{\delta_1+\delta_2} & & & \\ 0 & \frac{\delta_2}{\delta_2+\delta_3} & 2 & \ddots & & \\ \vdots & \vdots & & \ddots & & \\ & & & & 2 & \frac{\delta_{k+1}}{\delta_k+\delta_{k+1}} \\ 0 & 0 & & & 1 & 2 \end{bmatrix}$$

and rewrite M_0 in the form

$$M_0 = A'D_0$$

where $D_0 = \text{diag}(\delta_1/6,(\delta_1+\delta_2)/6,\ldots,(\delta_k+\delta_{k+1})/6,\delta_{k+1}/6)$. Let $A = [a_{ij}]$ and $A^{-1} = [a^{ij}]$. We notice that the elements of matrix A satisfy the conditions of Lemma A2 of the Appendix. Taking $n = k+2$ and $\beta_i = 2$, $i = 1,2,\ldots,k+2$ in that lemma, we get the bounds on the elements of matrix A^{-1} as below.

$$1/2 < a^{00} \leq 2/3 \tag{3.1.10}$$

$$1/2 < a^{ii} < 2/3, \quad i = 1,\ldots,k \tag{3.1.11}$$

$$1/2 < a^{k+1,k+1} \leq 2/3. \tag{3.1.12}$$

(Note that we do not have strict inequality on lower bounds of a^{00} and $a^{k+1,k+1}$ since $\alpha_1 = 1$ and $\gamma_n = 1$ here. See the note after Lemma A2). Also for the off-diagonal elements,

$$0 < (-1)^{i-j}a^{ij} < (4/3)(1/2^{|i-j|+1}),\ i\neq j,\ i,j = 0,1,\ldots,k+1. \quad (3.1.13)$$

Since we are interested in lower bounds only, we can replace for our purposes the inequalities (3.1.10) - (3.1.13) by one inequality:

$$0 < (-1)^{i-j}a^{ij} \leq (2/3)(1/2^{|i-j|}),\quad i,j = 0,1,\ldots,k+1. \quad (3.1.14)$$

Now we can write,

$$\begin{aligned} F &= B(h)D^{-1}(h)(A'D_0)^{-1} \\ &= B(h)D^{-1}(h)D_0^{-1}(A')^{-1} \\ &= B(h)D^{-1}(h)D_0^{-1}(A^{-1})' \\ &= C(h)(A^{-1})' \end{aligned} \quad (3.1.15)$$

where $C(h) = B(h)D^{-1}(h)D_0^{-1}$. The elements of the tridiagonal matrix $C(h)$ are as follows:

$$c_{00}(h) = 2-(6/\delta_1 h(\xi_0))\int \ell_0^2(x)h(x)dx$$

$$c_{k+1,k+1}(h) = 2-(6/\delta_{k+1} h(\xi_{k+1}))\int \ell_{k+1}^2(x)h(x)dx$$

$$c_{i,i}(h) = 2-(6/(\delta_i+\delta_{i+1})h(\xi_i))\int \ell_i^2(x)h(x)dx,\ i=1,2,\ldots,k$$

$$c_{10}(h) = 1-(6/\delta_1 h(\xi_0))\int \ell_0(x)\ell_1(x)h(x)dx$$

$$c_{k,k+1}(h) = 1-(6/\delta_{k+1} h(\xi_{k+1}))\int \ell_k(x)\ell_{k+1}(x)h(x)dx$$

$$c_{i,i+1}(h) = (\delta_{i+1}/(\delta_{i+1}+\delta_{i+2}))-(6/h(\xi_{i+1})(\delta_{i+1}+\delta_{i+2}))\int \ell_i(x)$$
$$\ell_{i+1}(x)h(x)dx,\ i=0,1,\ldots,k-1$$

$$c_{i+1,i}(h) = (\delta_{i+1}/(\delta_i+\delta_{i+1}))-(6/h(\xi_i)(\delta_i+\delta_{i+1}))\int \ell_i(x)\ell_{i+1}(x)$$

$$h(x)dx, \quad i = 1,2,\ldots,k.$$

Now we shall find the bounds on the elements of matrix C(h). We have for $i = 1,2,\ldots,k$,

$$c_{i,i}(h) = 2-(6/(\delta_i+\delta_{i+1})h(\xi_i))\int_{\xi_{i-1}}^{\xi_{i+1}} \ell_i^2(x)(h(x)-h(\xi_i))dx+(6/(\delta_i+\delta_{i+1}))$$

$$\int_{\xi_{i-1}}^{\xi_{i+1}} \ell_i^2(x)h(x)dx$$

$$= -(6/(\delta_i+\delta_{i+1})h(\xi_i))\int_{\xi_{i-1}}^{\xi_i} \ell_i^2(x)(h(x)-h(\xi_i))dx$$

$$-(6/(\delta_i+\delta_{i+1})h(\xi_i))\int_{\xi_i}^{\xi_{i+1}} \ell_i^2(x)(h(x)-h(\xi_i))dx.$$

But $|h(x)-h(\xi_i)| < \omega(h,\delta)$ in the intervals (ξ_{i-1},ξ_i) and (ξ_i,ξ_{i+1}), therefore

$$|c_{i,i}(h)| < (6/(\delta_i+\delta_{i+1})h(\xi_i))(\omega(h,\delta)\delta_i/3+\omega(h,\delta)\delta_{i+1}/3)$$

$$= 2\omega(h,\delta)/h(\xi_i) \qquad i = 1,2,\ldots,k.$$

Similarly, we can show that

$$|c_{00}(h)| < 2\omega(h,\delta)/h(\xi_0)$$

$$|c_{k+1,k+1}(h)| < 2\omega(h,\delta)/h(\xi_{k+1})$$

$$|c_{10}(h)| < \omega(h,\delta)/h(\xi_0)$$

$$|c_{k,k+1}(h)| < \omega(h,\delta)/h(\xi_{k+1}).$$

Also using (3.1.3), we can show that

$$|c_{i,i+1}(h)| < (\beta/\beta+1)(\omega(h,\delta)/h(\xi_{i+1})) < (\omega(h,\delta)/h(\xi_{i+1})),$$
$$i = 0,1,\ldots,k-1$$

and

$$|c_{i+1,i}(h)| < (\alpha/\alpha+1)(\omega(h,\delta)/h(\xi_i)) < (\omega(h,\delta)/h(\xi_i)), \ i=1,\ldots,k.$$

Now since $h(x)$ is continuous on $[0,1]$ and $h(x) > 0$ for all $x \in [0,1]$, we shall have $(1/h(x)) < K_1$ for all x, where K_1 is a positive constant. Now we can rewrite the bounds on elements of $C(h)$ as follows:

$$\begin{aligned} |c_{ii}(h)| &< 2K_1\omega(h,\delta), && i = 0,1,\ldots,k+1 \\ |c_{i,i+1}(h)| &< K_1\omega(h,\delta), && i = 0,1,\ldots,k \qquad (3.1.16) \\ |c_{i+1,i}(h)| &< K_1\omega(h,\delta), && i = 0,1,\ldots,k. \end{aligned}$$

Let matrix $F = [f_{ij}]$, then from (3.1.15)

$$f_{ij} = \sum_{\nu} c_{i\nu} a^{\nu j'}$$

where $[a^{\nu j'}] = (A^{-1})'$. Now since $a^{\nu j'} = a^{j\nu}$ and $c_{i\nu} = 0$ for $|i-\nu| > 1$, we have for $i = 1,2,\ldots,k$,

$$f_{ij} = c_{i,i-1}a^{j,i-1} + c_{ii}a^{ji} + c_{i,i+1}a^{j,i+1}$$

and

$$\begin{aligned} |f_{ij}| &< |c_{i,i-1}||a^{j,i-1}| + |c_{ii}||a^{ji}| + |c_{i,i+1}||a^{j,i+1}| \\ &< K_1\omega(h,\delta)[|a^{j,i-1}|+2|a^{ji}| + |a^{j,i+1}|], \text{ using (3.1.16).} \end{aligned}$$

Therefore

$$\sum_{j=0}^{k+1}|f_{ij}| < K_1\omega(h,\delta)[\sum_{j=0}^{k+1}|a^{j,i-1}| + 2\sum_{j=0}^{k+1}|a^{ji}| + \sum_{j=0}^{k+1}|a^{j,i+1}|]. \quad (3.1.17)$$

Now using the inequalities (3.1.14), we have

$$\begin{aligned}\sum_{j=0}^{k+1}|a^{j,i-1}| &= \sum_{j>i-1}(-1)^{j-(i-1)}a^{j,i-1} + \sum_{j\leq i-1}(-1)^{j-(i-1)}a^{j,i-1}\\ &< \sum_{j=i}^{k+1}(2/3)(1/2^{j-(i-1)}) + \sum_{j=0}^{i-1}(2/3)(1/2^{(i-1)-j})\\ &= (1/3)\sum_{j=0}^{k+1-i}(1/2^j) + (2/3)\sum_{j=0}^{i-1}(1/2^j)\\ &= 2-(2/3)\{(1/2)^{k+2-i} + (1/2)^{i-1}\}.\end{aligned}$$

Similarly,

$$\sum_{j=0}^{k+1}|a^{ji}| < 2-(2/3)\{(1/2)^{k+1-i} + (1/2)^{i}\}$$

and

$$\sum_{j=0}^{k+1}|a^{j,i+1}| < 2-(2/3)\{(1/2)^{k-i} + (1/2)^{i+1}\}.$$

Thus from (3.1.17) and above inequalities, we get

$$\sum_{j=0}^{k+1}|f_{ij}| < K_1\omega(h,\delta)[8-(3/2)\{(1/2)^{k-i} + (1/2)^{i-1}\}], \quad i = 1,2,\ldots,k$$

But $(1/2)^{k-i} + (1/2)^{i-1} > 1$ for $1 \leq i \leq k$. Hence

$$\max_{1\leq i\leq k}\{\sum_{j=0}^{k+1}|f_{ij}|\} < (13/2)K_1\omega(h,\delta).$$

Also for $i = 0$ and $k+1$, we can show that

$$\sum_{j=0}^{k+1} |f_{ij}| < (13/3)K_1\omega(h,\delta).$$

Thus

$$||F|| = \max_{0\le i\le k+1} \{\sum_{j=0}^{k+1} |f_{ij}|\}$$

$$< (13/2)K_1\omega(h,\delta)$$

$$= K\omega(h,\delta), \text{ say.}$$

This proves the lemma.

Now since $\omega(h,\delta) \to 0$ as $\delta \to 0$, we can make $\omega(h,\delta) < 1/K$ and hence $||F|| < 1$. We can then invert (I-F) using power series expansion,

$$(I-F)^{-1} = (I + F + F^2 + ...)$$

$$= (I+T), \text{ say}$$

where $T = \sum_{m=1}^{\infty} F^m$. Therefore from (3.1.9) and above expansion,

$$(1/k+2)\text{Tr } M^{-1}(\mu)M_0 = (1/k+2)\text{Tr}(D^{-1}(h)M_0^{-1}(I+T))M_0$$

$$= \{(1/k+2)\text{Tr } D^{-1}(h)M_0^{-1}M_0\} + \{(1/k+2)\text{Tr } D^{-1}(h)M_0^{-1} TM_0\}.$$

$$= \{(1/k+2)\text{Tr } D^{-1}(h)\} + \{(1/k+2)\text{Tr } D^{-1}(h)M_0^{-1}TM_0\}.$$

Now since $D^{-1}(h)$ is diagonal matrix with diagonal elements $1/h(\xi_i)$, we have for second term in above equality,

$$(1/k+2)\text{Tr } D^{-1}(h)M_0^{-1}TM_0 < (1/k+2)(\max_i 1/h(\xi_i))\text{Tr } M_0^{-1} TM_0$$

$$= (K_1/(k+2))\text{Tr } T$$

$$\therefore \quad |(1/k+2)\mathrm{Tr}\ D^{-1}(h)M_0^{-1}TM_0| < (K_1/(k+2))||T||.$$

From (A5) of the Appendix A, and above Lemma 3.3.1, we get $||T|| < K\omega(h,\delta)/(1-K\omega(h,\delta))$. Thus

$$1/(k+2)|\mathrm{Tr}\ D^{-1}(h)M_0^{-1}TM_0| < (K_1K\omega(h,\delta))/(1-K\omega(h,\delta)).$$

But $\omega(h,\delta) \to 0$ as $\delta \to 0$ since h is uniformly continuous on [0,1]. This shows that

$$(1/k+2)\mathrm{Tr}\ M^{-1}(\mu)M_0 = (1/k+2)\mathrm{Tr}\ D^{-1}(h) + o(1)$$

as $\delta \to 0$. From (3.1.2), $p(\eta_\nu)\delta_\nu = (k+1)^{-1}$ for some $\eta_\nu \in (\xi_{\nu-1},\xi_\nu)$, $\nu = 1,2,\ldots,k+1$. Therefore

$$(1/k+2)\mathrm{Tr}\ M^{-1}(\mu)M_0=((k+1)/(k+2))\sum_{\nu=0}^{k+1} \frac{p(\eta_\nu)}{h(\xi_\nu)}\,\delta_\nu + o(1)$$

or

$$(1/k)\mathrm{Tr}\ M^{-1}(\mu)M_0=((k+1)/k)\sum_{\nu=0}^{k+1} \frac{p(\eta_\nu)}{h(\xi_\nu)}\,\delta_\nu + o(1).$$

Letting $k \to \infty$ in such a way that $\delta \to 0$, the right hand side approaches $\int_0^1 \frac{p(x)}{h(x)}\,dx$ and the theorem is proved. Q.E.D.

3.2. Asymptotic Value of Bias for LSE

The results of this section indicates that the asymptotic value of bias, when the least square estimator is used, is

$$B \approx \frac{b}{720k^4}\int_0^1 \frac{(g''(x))^2}{p^4(x)}\,dx$$

as $k \to \infty$, where b is a positive constant. Let us recall that the bias term is (see (1.4.5)),

$$B = \int_0^1 (g(x) - \ell'(x)Cg_r)^2 dx.$$

If we use the least square estimator, then

$$Cg_r = M^{-1}(\mu)\int_0^1 \ell(x)g(x)d\mu(x).$$

If the design measure has the density h(x), then

$$c_h = Cg_r = M^{-1}(h)\int \ell(x)g(x)h(x)dx \tag{3.2.1}$$

where M(h) is the (k+2)x(k+2) matrix $\int \ell(x)\ell'(x)h(x)dx$. The bias term is then

$$B = \int_0^1 (g(x) - \ell'(x)c_h)^2 dx. \tag{3.2.2}$$

Recall that the minimum value of B with respect to $c = Cg_r$ for a fixed set of the knots occurs for $c_g = M_0^{-1}s$, where $s = \int \ell(x)g(x)dx$ (see (2.1.4)). Now the expression in (3.2.2) can be split as

$$B = B_1 + B_2 \tag{3.2.3}$$

where

$$B_1 = \int_0^1 (g(x) - \ell'(x)c_g)^2 dx \tag{3.2.4}$$

$$B_2 = \int_0^1 (\ell'(x)c_g - \ell'(x)c_h)^2 dx. \tag{3.2.5}$$

The first theorem of this section concerns the asymptotic behavior of the factor B_1 as $k \to \infty$.

Theorem 3.2.1: Let $g \in C^2[0,1]$ and B_1 be as defined in (3.2.4) and let the knots be chosen using (3.1.2), where p(x) is continuous and positive on [0,1]. Then

$$\lim_{k\to\infty} k^4 B_1 = \frac{1}{720}\int_0^1 \frac{(g''(x))^2}{p^4(x)}\,dx. \tag{3.2.6}$$

In the next theorem we have found the asymptotic expression for the bias term when the least square estimator is used and the design has a smooth density h(x).

Theorem 3.2.2: Let the design μ has a continuous density h(x) such that $h(x) > 0$ for all x. Let the knots satisfy (3.1.2), where p(x) is continuous and positive for all x. If the least square estimator is used, then the bias term has the asymptotic expression

$$B \approx \frac{1}{720k^4}\int_0^1 \frac{(g''(x))^2}{p^4(x)}\,dx. \tag{3.2.7}$$

It turns out as Theorems 3.2.2 and 3.2.3 indicate that the choice of c_g or c_h for suitable smooth h gives the same asymptotic expression for the bias term.

Note that the $(\nu+1)$th coordinate of c gives the value of the approximation $\ell'(x)c$ of g(x) at $x = \xi_\nu$. The $(\nu+1)$th coordinate of c_g is approximately

$$g(\xi_\nu) - (g''(\xi_\nu)\delta_\nu^2)/12. \tag{3.2.8}$$

If we use a design μ with the weights μ_ν on ξ_ν, the value for c is g_r and $\ell'(x)g_r$ simply interpolates g at the ξ_ν's and ignores the additional term involving g" in (3.2.8).

Theorem 3.2.3: If the design μ has the weights μ_ν on the knots ξ_ν and the LSE is used, the asymptotic expression for the bias is six times as large as that given in (3.2.7), i.e.

$$B \approx \frac{1}{120k^4} \int_0^1 \frac{(g''(x))^2}{p^4(x)}\, dx. \tag{3.2.9}$$

Generally a smooth density for the design will keep the bias term small and give slightly larger values for the variance term. The opposite happens for an appropriate discrete design.

Now we shall prove the above three Theorems.

Proof of Theorem 3.2.1:

Let us denote by

S_k = Class of spline functions of degree one with k prescribed knots $0 = \xi_0 < \xi_1 < \ldots < \xi_k < \xi_{k+1} = 1$,

= Class of the continuous polygonal line segments with vertices at the points ξ_i, $i = 0,1,\ldots,k+1$,

and

U_k = Class of the functions $u(x)$ defined on $[0,1]$ such that $u(x)$ represents a line on each interval (ξ_i,ξ_{i+1}), $i = 0,1,\ldots,k$.

Note that $u(x) \in U_k$ need not be continuous at the knots.

Now we see that

$$\begin{aligned}\min_c \int_0^1 (g(x) - \ell'(x)c)^2 dx &= \int_0^1 (g(x)-\ell'(x)c_g)^2 dx \\ &= \min_{s\in S_k} \int_0^1 (g(x)-s(x))^2 dx \\ &\geq \min_{u\in U_k} \int_0^1 (g(x)-u(x))^2 dx \end{aligned} \tag{3.2.10}$$

since $S_k \subset U_k$. Now let us find an expression for $\min_{u \in U_k} \int_0^1 (g(x)-u(x))^2 dx$. Since $\int_0^1 (g(x)-u(x))^2 dx = \sum_{\nu=0}^{k} \int_{\xi_\nu}^{\xi_{\nu+1}} (g(x)-u(x))^2 dx$ and $u(x)$ is a line on each interval $(\xi_\nu, \xi_{\nu+1})$, the quantity $\min_{u \in U_k} \int_0^1 (g(x)-u(x))^2 dx$ is the sum of the (best) L_2-errors of approximation of $g(x)$ by the straight line on every interval $(\xi_\nu, \xi_{\nu+1})$, $\nu = 0,1,\ldots,k$. Let for $\nu = 0,1,\ldots,k$,

$$u(x) = a_\nu + b_\nu(x-\xi_\nu), \; x \in (\xi_\nu, \xi_{\nu+1})$$

then the (unique) solution of the problem of minimizing $\int_{\xi_\nu}^{\xi_{\nu+1}} (g(x)-a_\nu-b_\nu(x-\xi_\nu))^2 dx$ with respect to the constants a_ν and b_ν is

$$a_\nu^* = (4/\delta_{\nu+1})\int_{\xi_\nu}^{\xi_{\nu+1}} g(x)dx - (6/\delta_{\nu+1}^2)\int_{\xi_\nu}^{\xi_{\nu+1}} (x-\xi_\nu)g(x)dx \quad (3.2.11)$$

and

$$b_\nu^* = (12/\delta_{\nu+1}^3)\int_{\xi_\nu}^{\xi_{\nu+1}} (x-\xi_\nu)g(x)dx - (6/\delta_{\nu+1}^2)\int_{\xi_\nu}^{\xi_{\nu+1}} g(x)dx, \quad (3.2.12)$$

where $\delta_{\nu+1} = \xi_{\nu+1} - \xi_\nu$.

Since $g(x)$ is a function possessing the second derivative, we can express, using Taylor's theorem, $g(x)$ in terms of the values of g and its first and second derivatives at $x = \xi_\nu$. Here we shall use Taylor's theorem in a form given in the section 151 of Hardy (1943). We can write

$$g(x) = g(\xi_\nu) + (x-\xi_\nu)g'(\xi_\nu) + (x-\xi_\nu)^2(g''(\xi_\nu)+\eta_x)/2 \tag{3.2.13}$$

where $\eta_x \to 0$ as $x \to \xi_\nu$. We can also write (3.2.13) as

$$g(x) = g(\xi_\nu) + (x-\xi_\nu)g'(\xi_\nu) + ((x-\xi_\nu)^2 g''(\xi_\nu)/2) + o(x-\xi_\nu)^2. \tag{3.2.14}$$

Using (3.2.14) we want to find the quadrature formula for $\int_{\xi_\nu}^{\xi_{\nu+1}} g(x)dx$ and $\int_{\xi_\nu}^{\xi_{\nu+1}} (x-\xi_\nu)g(x)$. Since this involves the integration of "Order relations", let us prove the following result:

<u>Lemma 3.2.1</u>: Let $\phi = o(x-\xi_\nu)^q$ as $x \to \xi_\nu$ where $q > 0$, then $\int_{\xi_\nu}^{\xi_{\nu+1}} \phi(x)dx = o(\delta_{\nu+1}^{q+1})$ as $\delta_{\nu+1} \to 0$.

Proof: Since $\phi = o(x-\xi_\nu)^q$ as $x \to \xi_\nu$, given $\varepsilon > 0$ there exists an η so that $|\phi| < \varepsilon\,(x-\xi_\nu)^q$ for all x such that $|x-\xi_\nu| < \eta$, and hence

$$\begin{aligned}
\left|\int_{\xi_\nu}^{\xi_{\nu+1}} \phi(x)dx\right| &< \int_{\xi_\nu}^{\xi_{\nu+1}} |\phi(x)|dx \\
&< \varepsilon \int_{\xi_\nu}^{\xi_{\nu+1}} (x-\xi_\nu)^q dx \quad \text{if } \delta_{\nu+1} < \eta \\
&= (\varepsilon\, \delta_{\nu+1}^{q+1})/(q+1).
\end{aligned}$$

This proves the lemma.

Using the above lemma with $q = 2$ and $q = 3$ respectively, we can show that

$$\int_{\xi_\nu}^{\xi_{\nu+1}} g(x)dx = g(\xi_\nu)\delta_{\nu+1} + (g'(\xi_\nu)\delta_{\nu+1}^2/2) + (g''(\xi_\nu)\delta_{\nu+1}^3/6) + o(\delta_{\nu+1}^3) \tag{3.2.15}$$

and

$$\int_{\xi_\nu}^{\xi_{\nu+1}} (x-\xi_\nu)g(x)dx = (g(\xi_\nu)\delta_{\nu+1}^2/2)+(g'(\xi_\nu)\delta_{\nu+1}^3/3)+(g''(\xi_\nu)\delta_{\nu+1}^4/8) + o(\delta_{\nu+1}^4). \quad (3.2.16)$$

The value of the integrals obtained in (3.2.15) and (3.2.16) are substituted in (3.2.11) and (3.2.12) respectively, to get

$$a_\nu^* = g(\xi_\nu) - (g''(\xi_\nu)\delta_{\nu+1}^2/12) + \psi_1 \quad (3.2.17)$$

and

$$b_\nu^* = g'(\xi_\nu) + (g''(\xi_\nu)\delta_{\nu+1}/2) + \psi_2, \quad (3.2.18)$$

where $\psi_1 = o(\delta_{\nu+1}^2)$ and $\psi_2 = o(\delta_{\nu+1})$ as $\delta_{\nu+1} \to 0$. Now from (3.2.13), (3.2.17) and (3.2.18), we get

$$\int_{\xi_\nu}^{\xi_{\nu+1}} (g(x)-a_\nu^*-b_\nu^*(x-\xi_\nu))^2dx$$

$$= (1/144)\int_{\xi_\nu}^{\xi_{\nu+1}} \Big\{6(x-\xi_\nu)^2g''(\xi_\nu)-6(x-\xi_\nu)g''(\xi_\nu)\delta_{\nu+1}+g''(\xi_\nu)\delta_{\nu+1}^2+ 6(x-\xi_\nu)^2\eta_x-12(x-\xi_\nu)\psi_2-12\psi_1\Big\}^2dx$$

$$= (g''^2(\xi_\nu)\delta_{\nu+1}^5/720) + (1/4)\int_{\xi_\nu}^{\xi_{\nu+1}} \Big\{(x-\xi_\nu)^4\eta_x^2 + 4\psi_1^2 + 4(x-\xi_\nu)^2\psi_2^2\Big\}dx$$

$$+ (g''(\xi_\nu)/12)\int_{\xi_\nu}^{\xi_{\nu+1}} \Big\{6(x-\xi_\nu)^4\eta_x-6(x-\xi_\nu)^3\eta_x\delta_{\nu+1}+(x-\xi_\nu)^2\eta_x\delta_{\nu+1}^2 - 12(x-\xi_\nu)^2\psi_1$$

$$+ 12(x-\xi_\nu)\psi_1\delta_{\nu+1}-2\psi_1\delta_{\nu+1}^2-12(x-\xi_\nu)^3\psi_2$$

$$+ 12(x-\xi_\nu)^2\psi_2\delta_{\nu+1}-2(x-\xi_\nu)\psi_2\delta_{\nu+1}^2\Big\}dx.$$

Since g" is bounded on [0,1], we can show, using Lemma 3.2.1, that in the above expression on right all the terms except the first one are at least of order $o(\delta_{\nu+1}^5)$ as $\delta_{\nu+1} \to 0$. Therefore

$$\min_{a_\nu, b_\nu} \int_{\xi_\nu}^{\xi_{\nu+1}} (g(x)-a_\nu-b_\nu(x-\xi_\nu))^2 dx$$

$$=\int_{\xi_\nu}^{\xi_{\nu+1}} (g(x)-a_\nu^*-b_\nu^*(x-\xi_\nu))^2 dx$$

$$= (g''^2(\xi_\nu)\delta_{\nu+1}^5/720) + o(\delta_{\nu+1}^5) \text{ as } \delta_{\nu+1} \to 0.$$

From (3.1.2) and mean value theorem $p(\eta_\nu)\delta_{\nu+1} = (k+1)^{-1}$ for some $\eta_\nu \in (\xi_\nu, \xi_{\nu+1})$, hence we can write,

$$\int_{\xi_\nu}^{\xi_{\nu+1}} (g(x)-a_\nu^*-b_\nu^*(x-\xi_\nu))^2 dx$$

$$= (g''^2(\xi_\nu)/720(k+1)^4 p^4(\eta_\nu))\delta_{\nu+1} + o(k^{-5})$$

as $k \to \infty$, therefore

$$\min_{u \in U_k} \int_0^1 (g(x)-u(x))^2 = (1/720(k+1)^4) \sum_{\nu=0}^{k} (g''^2(\xi_\nu)/p^4(\eta_\nu))\delta_{\nu+1} + o(k^{-4}) \tag{3.2.19}$$

Let us now consider the L_2-approximation of g(x) by the specific member s*(x) of the class S_k. This s*(x) is a continuous line segment which has ordinates $g(\xi_\nu)-(g''(\xi_\nu)\delta_\nu^2/12)$ at the points ξ_ν, $\nu = 0,1,\ldots,k+1$. We set $\delta_0 = 0$. For $x \in (\xi_\nu, \xi_{\nu+1})$,

$$s^*(x) = \{g(\xi_{\nu+1})-(g''(\xi_{\nu+1})\delta_{\nu+1}^2/12)\}\{(x-\xi_\nu)/\delta_{\nu+1}\}$$

$$+\{g(\xi_\nu)-(g''(\xi_\nu)\delta_\nu^2/12)\}\{(\xi_{\nu+1}-x)/\delta_{\nu+1}\}. \tag{3.2.20}$$

Again since $g(x) \in C^2[0,1]$, the Lagrange's formula of interpolation (e.g. see p. 32, Whittaker and Robinson (1967)) gives

$$g(x) = (1/\delta_{\nu+1})\{g(\xi_{\nu+1})(x-\xi_\nu)+ g(\xi_\nu)(\xi_{\nu+1}-x)\}-g''(\eta_{\nu+1,x})(x-\xi_\nu)(\xi_{\nu+1}-x)/2 \qquad (3.2.21)$$

for $x \in (\xi_\nu,\xi_{\nu+1})$, where $\eta_{\nu+1,x} \in (\xi_\nu,\xi_{\nu+1})$. The subscript x in $\eta_{\nu+1,x}$ shows its dependence on x. Now from (3.2.20) and (3.2.21),

$$g(x)-s^*(x) = (1/12)(x-\xi_\nu)g''(\xi_{\nu+1})\delta_{\nu+1} + (1/12\delta_{\nu+1})(\xi_{\nu+1}-x)g''(\xi_\nu)\delta_\nu^2 - (1/2)(x-\xi_\nu)(\xi_{\nu+1}-x)g''(\eta_{\nu+1,x}).$$

Therefore

$$\int_{\xi_\nu}^{\xi_{\nu+1}} (g(x)-s^*(x))^2dx = (1/432)g''^2(\xi_{\nu+1})\delta_{\nu+1}^5+(1/432)g''(\xi_\nu)g''(\xi_{\nu+1})\delta_\nu^2\delta_{\nu+1}^3$$
$$+ (1/432)g''^2(\xi_\nu)\delta_\nu^4\delta_{\nu+1}+(1/4)\int_{\xi_\nu}^{\xi_{\nu+1}} g''^2(\eta_{\nu+1,x})(x-\xi_\nu)^2(\xi_{\nu+1}-x)^2dx$$
$$- (g''(\xi_{\nu+1})\delta_{\nu+1}/12)\int_{\xi_\nu}^{\xi_{\nu+1}} g''(\eta_{\nu+1,x})(x-\xi_\nu)^2(\xi_{\nu+1}-x)dx$$
$$- (g''(\xi_\nu)\delta_\nu^2/12\delta_{\nu+1})\int_{\xi_\nu}^{\xi_{\nu+1}} g''(\eta_{\nu+1,x})(x-\xi_\nu)(\xi_{\nu+1}-x)^2dx.$$

Now using Lemma 2.2.1 for the last three terms on right, we get

$$\int_{\xi_\nu}^{\xi_{\nu+1}} (g(x)-s^*(x))^2dx = (1/432)g''^2(\xi_{\nu+1})\delta_{\nu+1}^5+(1/432)g''(\xi_\nu)g''(\xi_{\nu+1})\delta_\nu^2\delta_{\nu+1}^3$$

$$+ (1/432)g''^2(\xi_\nu)\delta_\nu^4\delta_{\nu+1} + (1/120)g''^2(\alpha_{\nu+1})\delta_{\nu+1}^5$$

$$- (1/144)g''(\xi_{\nu+1})g''(\beta_{\nu+1})\delta_{\nu+1}^5$$

$$-(1/144)g''(\xi_\nu)g''(\gamma_{\nu+1})\delta_\nu^2\delta_{\nu+1}^3$$

where $\alpha_{\nu+1}$, $\beta_{\nu+1}$ and $\gamma_{\nu+1}$ are points in $(\xi_\nu,\xi_{\nu+1})$. Again using (3.1.2) and the mean value theorem, we get

$$\int_0^1(g(x)-s^*(x))^2dx$$

$$= (1/432(k+1)^4)\{\sum_{\nu=0}^{k}(g''^2(\xi_{\nu+1})/p^4(\rho_{\nu+1}))\delta_{\nu+1}+\sum_{\nu=0}^{k}(g''(\xi_\nu)g''(\xi_{\nu+1})/p^2(\rho_\nu)p^2(\rho_{\nu+1}))\delta_{\nu+1}\}$$

$$+ (1/432(k+1)^4)\{\sum_{\nu=0}^{k}(g''^2(\xi_\nu)/p^4(\rho_\nu))\delta_{\nu+1}+(37/10)\sum_{\nu=0}^{k}(g''^2(\alpha_{\nu+1})/p^4(\rho_{\nu+1}))\delta_{\nu+1}\}$$

$$- (1/144(k+1)^4)\{\sum_{\nu=0}^{k}(g''(\xi_{\nu+1})g''(\beta_{\nu+1})/p^4(\rho_{\nu+1}))\delta_{\nu+1}+\sum_{\nu=0}^{k}(g''(\xi_\nu)g''(\gamma_{\nu+1})/p^2(\rho_{\nu+1})p^2(\rho_\nu))\delta_{\nu+1}\}, \qquad (3.2.22)$$

where $\rho_\nu \in (\xi_{\nu-1},\xi_\nu)$ and $\rho_{\nu+1} \in (\xi_\nu,\xi_{\nu+1})$.

We have

$$\int_0^1(g(x)-s^*(x))^2dx \geq \min_{s\in S_k}\int_0^1 (g(x)-s(x))^2dx. \qquad (3.2.23)$$

Now from (3.2.10) and (3.2.23), we get

$$k^4 \int_0^1 (g(x)-s^*(x))^2 dx \geq k^4 \int_0^1 (g(x)-\ell'(x)c_g)^2 dx$$

$$\geq k^4 \min_{u \in U_k} \int_0^1 (g(x)-u(x))^2 dx. \qquad (3.2.24)$$

Letting $k \to \infty$ in such a way that $\delta = \max_\nu \delta_\nu \to 0$ and using (3.2.19) and (3.2.22), we get from (3.2.24),

$$\lim_{k\to\infty} k^4 \int_0^1 (g(x)-\ell'(x)c_g)^2 dx = \frac{1}{720} \int_0^1 \frac{(g''(x))^2}{p^4(x)} dx. \qquad \text{Q.E.D.}$$

Alternative Proof of Theorem 3.2.1:

We can prove Theorem 3.2.1 using the results of Appendix B and a theorem by Sacks and Ylvisaker (1970).

So far we were using the Lagrange basis $\ell(x)$ for the class S_k. Instead we can use the following basis:

$$1,\ (\xi_1-x)_+,\ (\xi_2-x)_+, \ldots\ldots (\xi_k-x)_+,\ (\xi_{k+1}-x)_+.$$

Then

$$\int (g(x) - \sum_i \alpha_i \ell_i(x))^2 dx$$

$$= \int (g(x) - \sum_i c_i (\xi_i - x)_+)^2 dx$$

$$= \int_0^1 (\int_0^1 (t-x)_+ g''(t)dt - \sum_i c_i(\xi_i-x)_+)^2 dx$$

$$= ||f - \sum_i c_i R(\cdot,\xi_i)||^2_{H(R)},$$

the last equality being obtained using the equation (B7) of the Appendix B, where now

$$f(s) = \int_0^1 R(s,t)g''(t)dt$$

and R(s,t) is as defined in (B3) of the Appendix B. Thus we have

$$\int_0^1 (g(x)-\ell'(x)c_g)^2 dx = \min_{c_i} ||f-\sum_i c_i R(\cdot,\xi_i)||^2_{H(R)}$$

$$= ||f-P_{T_k} f||^2_{H(R)}$$

where P_{T_k} is the projection operator defined on Hilbert space H(R) to the subspace spanned by $R(\cdot,t)$, $t \in T_k$. Now by a result given in section 3.3 of Sacks and Ylvisaker (1970) we have

$$\lim_{k\to\infty} k^4 ||f-P_{T_k} f||^2 = \frac{1}{720} \int_0^1 \frac{(g''(x))^2}{p^4(x)} dx. \qquad \text{Q.E.D.}$$

We shall use Theorem 3.2.1 in proving the next theorem. In proving Theorem 3.2.2, we also require a lower bound on the norm of the matrix $M_0 = \int \ell(x)\ell'(x)dx$ (see Appendix A for the definition and other results on matrix norm) which we will now find. We need the following restriction on the sequence $\{T_k\}$ of the partitions:

$$\max_{1\leq i\leq k+1} (\delta/\delta_i) \leq \beta < \infty. \qquad (3.2.25)$$

This is in addition to the assumption that the mesh $T_k \to 0$ as $k \to \infty$. Now since the main diagonal of M_0 is dominant (i.e. $|m_{ii}| > \sum_{j\neq i} |m_{ij}|$) we have from Lemma A1 of the Appendix,

$$||M_0^{-1}|| \leq [\min_{0\leq i\leq k+1} (m_{ii} - \sum_{j\neq i} m_{ij})]^{-1}$$

$$= [\min_{1\leq i\leq k} (\delta_i+\delta_{i+1})/6]^{-1}$$

$$< 3\beta/\delta, \qquad (3.2.26)$$

using (3.2.25).

Proof of Theorem 3.2.2:

We have

$$B = B_1 + B_2$$

where B_1 and B_2 are given by (3.2.4) and (3.2.5) respectively. Also from Theorem 3.2.1,

$$B_1 \approx \frac{1}{720k^4} \int_0^1 \frac{(g''(x))^2}{p^4(x)}\,dx.$$

To complete the proof of the theorem, it is sufficient, then, to show that $B_2 = o(\delta^4)$ as $\delta \to 0$. Let us define a (k+2)x1 vector c^* whose elements are given by

$$c_i^* = g(\xi_i) - (g''(\xi_i)\delta_{i+1}^2/12) \qquad i = 0,1,\ldots,k$$

and

$$c_{k+1}^* = g(\xi_{k+1}) - (g''(\xi_{k+1})\delta_{k+1}^2/12).$$

We have

$$s_h = \int \ell(x)g(x)h(x)dx = M(h)c_h; \; s = \int \ell(x)g(x)dx = M_0c_g.$$

Also define

$$s_h^* = M(h)c^*, \; s^* = M_0c^*.$$

We shall now evaluate the elements of the vector s using the interpolation formula,

$$g(x)=(1/\delta_{\nu+1})(g(\xi_{\nu+1})(x-\xi_\nu)+g(\xi_\nu)(\xi_{\nu+1}-x))-g''(\eta_{\nu+1,x})(x-\xi_\nu)(\xi_{\nu+1}-x)/2 \quad (3.2.27)$$

for $x \in (\xi_\nu,\xi_{\nu+1})$, where $\eta_{\nu+1,x} \in (\xi_\nu,\xi_{\nu+1})$. Using (3.2.27) and the Lemma 2.2.1 to find the integrals of the type $\int_{\xi_\nu}^{\xi_{\nu+1}} g''(\eta_{\nu+1,x})(x-\xi_\nu)(\xi_{\nu+1}-x)dx$, we get

$$s_0 = (g(\xi_0)\delta_1/3)+(g(\xi_1)\delta_1/6)-(g''(\rho_1)\delta_1^3/24)$$

$$s_\nu = (g(\xi_{\nu-1})\delta_\nu/6)+(g(\xi_\nu)(\delta_\nu+\delta_{\nu+1})/3)+(g(\xi_{\nu+1})\delta_{\nu+1}/6) - (g''(\rho_\nu)\delta_\nu^3/24)-(g''(\rho_{\nu+1})\delta_{\nu+1}^3/24), \quad \nu = 1,2,\dots,k$$

and

$$s_{k+1} = (g(\xi_{k+1})\delta_{k+1}/3) + (g(\xi_k)\delta_{k+1}/6)-(g''(\rho_{k+1})\delta_{k+1}^3/24),$$

for some $\rho_\nu \in (\xi_{\nu-1},\xi_\nu)$, $\nu = 1,2,\dots,k+1$. Now we can write down the elements of the difference vector $s-s^*$. For $\nu = 1,2,\dots,k$, $s_\nu-s_\nu^* = A_\nu + B_\nu$, where

$$A_\nu = (1/72)(g''(\xi_{\nu+1})-g''(\rho_\nu))\delta_\nu^3 + (1/36)(g''(\xi_\nu)-g''(\rho_\nu))\delta_\nu\delta_{\nu+1}^2 + (1/36)g''(\rho_\nu)\delta_\nu(\delta_{\nu+1}^2- \delta_\nu^2)$$

and

$$B_\nu = (1/36)(g''(\xi_\nu)-g''(\rho_{\nu+1}))\delta_{\nu+1}^3+(1/72)(g''(\xi_{\nu+1})-g''(\rho_{\nu+1}))\delta_{\nu+1}\delta_{\nu+2}^2 + (1/72)g''(\rho_{\nu+1})\delta_{\nu+1}(\delta_{\nu+2}^2 - \delta_{\nu+1}^2).$$

Again from the equation (3.1.2), $p(\alpha_\nu)\delta_\nu = (k+1)^{-1}$, for some $\alpha_\nu \in (\xi_{\nu-1},\xi_\nu)$, $\nu = 1,\dots,k+1$. Now

$$\delta_{\nu+1}-\delta_\nu = (k+1)^{-1}[(1/p(\alpha_{\nu+1}))-(1/p(\alpha_\nu))]$$
$$= (\delta_\nu/p(\alpha_{\nu+1}))(p(\alpha_\nu)-p(\alpha_{\nu+1}))$$
$$= (\delta_\nu/p(\alpha_{\nu+1}))\{(p(\alpha_\nu)-p(\xi_\nu))+(p(\xi_\nu)-p(\alpha_{\nu+1}))\}.$$

Therefore

$$|\delta_{\nu+1}-\delta_\nu| < 2\ \delta_\nu\ \omega(p,\delta)/p(\alpha_{\nu+1}) \qquad (3.2.28)$$

where $\omega(p,\delta)$ is the modulus of continuity of $p(x)$ with respect to $\delta = \max_i \delta_i$. Now let $|g''(x)| < K_1$ and $(1/p(x)) < K_2$ for all $x \in [0,1]$, where K_1 and K_2 are positive constants, then using (3.2.28),

$$|A_\nu| < \frac{\omega(g'',\delta)}{72}\ \delta_\nu^3 + \frac{\omega(g'',\delta)}{36}\ \delta_\nu\delta_{\nu+1}^2 + \frac{K_1\omega(p,\delta)\delta_\nu^2(\delta_\nu+\delta_{\nu+1})}{18K_2}$$

where $\omega(g'',\delta)$ is the modulus of continuity of $g''(x)$. Since $\omega(p,\delta)$ and $\omega(g'',\delta)$ goes to zero uniformly in δ, we have

$$A_\nu = o(\delta^3) \quad \text{as} \quad \delta \to 0.$$

Similarly, we can show that $B_\nu = o(\delta^3)$, therefore for $\nu = 1,2,\ldots,k$,

$$s_\nu - s_\nu^* = o(\delta^3).$$

It is easy to check that $(s_1-s_1^*) = o(\delta^3)$ and $(s_{k+1}-s_{k+1}^*) = o(\delta^3)$. Therefore

$$||M_0(c_g-c^*)|| = ||s-s^*||$$
$$= \max_i|s_i-s_i^*|$$
$$= o(\delta^3).$$

But in (3.2.26) we showed that $||M_0^{-1}|| < (3\beta/\delta)$, hence

$$||c_g - c^*|| \le ||M_0^{-1}||\ ||s-s^*|| = o(\delta^2) \tag{3.2.29}$$

as $\delta \to 0$.

Now we shall try to find a lower bound for the norm of the vector $c_h - c^*$. First, let us write down the elements of tridiagonal matrix $M(h)$.

$$m_{00}(h) = (1/3)h(\xi_0)\delta_1 - b_{00}(h)$$

$$m_{k+1,k+1}(h) = (1/3)h(\xi_{k+1})\delta_{k+1} - b_{k+1,k+1}(h)$$

$$m_{\nu,\nu}(h) = (1/3)h(\xi_\nu)(\delta_\nu + \delta_{\nu+1}) - b_{\nu,\nu}(h),\ \nu = 1,\dots,k$$

$$m_{\nu,\nu+1}(h) = (1/6)h(\xi_\nu)\delta_{\nu+1} - b_{\nu,\nu+1}(h),\ \nu = 0,1,\dots,k$$

$$m_{\nu+1,\nu}(h) = (1/6)h(\xi_{\nu+1})\delta_{\nu+1} - b_{\nu+1,\nu}(h),\ \nu = 0,1,\dots,k,$$

where

$$b_{00}(h) = -\int_{\xi_0}^{\xi_1} (h(x)-h(\xi_0))\ell_0^2(x)dx$$

$$b_{k+1,k+1}(h) = -\int_{\xi_k}^{\xi_{k+1}} (h(x)-h(\xi_{k+1}))\ell_{k+1}^2(x)dx$$

$$b_{\nu,\nu}(h) = -\int_{\xi_{\nu-1}}^{\xi_\nu} \ell_\nu^2(x)(h(x)-h(\xi_\nu))dx - \int_{\xi_\nu}^{\xi_{\nu+1}} \ell_\nu^2(x)(h(x)-h(\xi_\nu))dx$$

$$b_{\nu,\nu+1}(h) = -\int_{\xi_\nu}^{\xi_{\nu+1}} \ell_\nu(x)\ell_{\nu+1}(x)(h(x)-h(\xi_\nu))dx$$

and

$$b_{\nu+1,\nu}(h) = -\int_{\xi_\nu}^{\xi_{\nu+1}} \ell_\nu(x)\ell_{\nu+1}(x)(h(x)-h(\xi_{\nu+1}))dx.$$

Using the above relations, we can evaluate the elements of the vector $s_h^* = M(h)c^*$. The interpolation formula (3.2.27) can be used to evaluate the elements of the vector s_h. These are as follows:

$$\begin{aligned} s_{h,0} = {} & (1/6)g(\xi_1)h(\xi_0)\delta_1+(1/3)g(\xi_0)h(\xi_0)\delta_1-(1/24)g''(\tau_1)h(\xi_0)\delta_1^3 \\ & + g(\xi_1)\int_{\xi_0}^{\xi_1} \ell_0(x)\ell_1(x)(h(x)-h(\xi_0))dx \\ & + g(\xi_0)\int_{\xi_0}^{\xi_1} \ell_0^2(x)(h(x)-h(\xi_0))dx \\ & - (\delta_1^2/2)\int_{\xi_0}^{\xi_1} g''(\eta_{1,x})\ell_0^2(x)\ell_1(x)(h(x)-h(\xi_0))dx, \end{aligned}$$

$$\begin{aligned} s_{h,\nu} = {} & (1/6)g(\xi_{\nu-1})h(\xi_\nu)\delta_\nu+(1/3)g(\xi_\nu)h(\xi_\nu)(\delta_\nu+\delta_{\nu+1})+(1/6)g(\xi_{\nu+1}) \\ & \qquad h(\xi_\nu)\delta_{\nu+1} \\ & - (1/24)g''(\tau_\nu)h(\xi_\nu)\delta_\nu^3-(1/24)g''(\tau_{\nu+1})h(\xi_\nu)\delta_{\nu+1}^3 \\ & + g(\xi_\nu)\int_{\xi_{\nu-1}}^{\xi_\nu} \ell_\nu^2(x)(h(x)-h(\xi_\nu))dx + g(\xi_{\nu-1})\int_{\xi_{\nu-1}}^{\xi_\nu} \ell_\nu(x) \\ & \qquad \ell_{\nu-1}(x)(h(x)-h(\xi_\nu))dx \\ & + g(\xi_{\nu+1})\int_{\xi_\nu}^{\xi_{\nu+1}} \ell_\nu(x)\ell_{\nu+1}(x)(h(x)-h(\xi_\nu))dx+g(\xi_\nu)\int_{\xi_\nu}^{\xi_{\nu+1}} \ell_\nu^2(x) \\ & \qquad (h(x)-h(\xi_\nu))dx \end{aligned}$$

$$- (\delta_\nu^2/2)\int_{\xi_{\nu-1}}^{\xi_\nu} g''(\eta_{\nu,x})\ell_\nu^2(x)\ell_{\nu-1}(x)(h(x)-h(\xi_\nu))dx$$

$$- (\delta_{\nu+1}^2/2)\int_{\xi_\nu}^{\xi_{\nu+1}} g''(\eta_{\nu+1,x})\ell_\nu^2(x)\ell_{\nu+1}(x)(h(x)-h(\xi_\nu))dx,$$

$$\nu = 1,2,\ldots,k$$

and

$$s_{h,k+1} = (1/3)g(\xi_{k+1})h(\xi_{k+1})\delta_{k+1}+(1/6)g(\xi_k)h(\xi_{k+1})\delta_{k+1}-(1/24)g''(\tau_{k+1})h(\xi_{k+1})\delta_{k+1}^3$$

$$+ g(\xi_{k+1})\int_{\xi_k}^{\xi_{k+1}} \ell_{k+1}^2(x)(h(x)-h(\xi_{k+1}))dx$$

$$+ g(\xi_k)\int_{\xi_k}^{\xi_{k+1}} \ell_{k+1}(x)\ell_k(x)(h(x)-h(\xi_{k+1}))dx$$

$$- (\delta_{k+1}^2/2)\int_{\xi_k}^{\xi_{k+1}} g''^2(\eta_{k+1,x})\ell_{k+1}^2(x)\ell_k(x)(h(x)-h(\xi_{k+1}))dx,$$

where $\tau_\nu \in (\xi_{\nu-1},\xi_\nu)$, $\nu = 1,2,\ldots,k+1$. Therefore the elements of the difference vector $s_h - s_h^*$ are as follows:

$$s_{h,\nu}-s_{h,\nu}^* = (1/72)h(\xi_\nu)\delta_\nu^3(g''(\xi_{\nu-1})-g''(\tau_\nu))+(1/36)h(\xi_\nu)\delta_\nu(\delta_{\nu+1}^2-\delta_\nu^2)g''(\xi_\nu)$$

$$+ (1/36)h(\xi_\nu)\delta_\nu^3(g''(\xi_\nu)-g''(\tau_\nu))+(1/36)h(\xi_\nu)\ \delta_{\nu+1}^3(g''(\xi_\nu) - g''(\tau_{\nu+1}))$$

$$+ (1/72)h(\xi_\nu)\delta_{\nu+1}(\delta_{\nu+2}^2-\delta_{\nu+1}^2)g''(\xi_{\nu+1})+(1/72)h(\xi_\nu)\delta_{\nu+1}^3(g''(\xi_{\nu+1})-g''(\tau_{\nu+1}))$$

$$-(\delta_\nu^2/2)\int_{\xi_{\nu-1}}^{\xi_\nu} g''(\eta_{\nu,x})\ell_\nu^2(x)\ell_{\nu-1}(x)(h(x)-h(\xi_\nu))dx$$

$$-(\delta_{\nu+1}^2/2)\int_{\xi_\nu}^{\xi_{\nu+1}} g''(\eta_{\nu+1,x})\ell_\nu^2(x)\ell_{\nu+1}(x)(h(x)-h(\xi_\nu))dx$$

$$+(g''(\xi_{\nu-1})\delta_\nu^2/12)\int_{\xi_{\nu-1}}^{\xi_\nu} \ell_{\nu-1}(x)\ell_\nu(x)(h(x)-h(\xi_\nu))dx$$

$$+(g''(\xi_\nu)\delta_{\nu+1}^2/12)\int_{\xi_{\nu-1}}^{\xi_\nu} \ell_\nu^2(x)(h(x)-h(\xi_\nu))dx$$

$$+(g''(\xi_\nu)\delta_{\nu+1}^2/12)\int_{\xi_\nu}^{\xi_{\nu+1}} \ell_\nu^2(x)(h(x)-h(\xi_\nu))dx$$

$$+(g''(\xi_{\nu+1})\delta_{\nu+2}^2/12)\int_{\xi_\nu}^{\xi_{\nu+1}} \ell_\nu(x)\ell_{\nu+1}(x)(h(x)-h(\xi_\nu))dx$$

Now let $h(x) < K_3$, then using (3.2.28) and a similar inequality for $|\delta_{\nu+2}-\delta_{\nu+1}|$, and the bounds on p, g" and h, we get

$$\begin{aligned}|s_{h,\nu} - s^*_{h,\nu}| < {} & \frac{K_3\omega(g'',\delta)}{72}\delta_\nu^3 + \frac{K_3K_1\omega(p,\delta)}{18K_2}\delta_\nu^2(\delta_\nu+\delta_{\nu+1}) \\ & + \frac{K_3\omega(g'',\delta)}{36}\delta_\nu^3 + \frac{K_3\omega(g'',\delta)}{36}\delta_{\nu+1}^3 \\ & + \frac{K_3K_1\omega(p,\delta)}{36K_2}\delta_{\nu+1}^2(\delta_{\nu+1}+\delta_{\nu+2}) + \frac{K_3\omega(g'',\delta)}{72}\delta_{\nu+1}^3 \\ & + \frac{K_1\omega(h,\delta)}{24}\delta_\nu^3 + \frac{K_1\omega(h,\delta)}{24}\delta_{\nu+1}^3 \\ & + \frac{K_1\omega(h,\delta)}{72}\delta_\nu^3 + \frac{K_1\omega(h,\delta)}{36}\delta_\nu\delta_{\nu+1}^2\end{aligned}$$

$$+ \frac{K_1\omega(h,\delta)\delta^3_{\nu+1}}{36} + \frac{K_1\omega(h,\delta)}{72}\delta_{\nu+1}\delta^2_{\nu+2},$$

where $\omega(h,\delta)$ is the modulus of continuity of $h(x)$ with respect to $\delta = \max_i \delta_i$ on $[0,1]$. Now since $\omega(h,\delta)$, $\omega(g'',\delta)$ and $\omega(p,\delta) \to 0$ as $\delta \to 0$, we have for $\nu = 1,2,\ldots,k$

$$s_{h,\nu} - s^*_{h,\nu} = o(\delta^3) \quad \text{as } \delta \to 0.$$

It is easy to show that for $\nu = 0$ and $k+1$ also the above order relations hold. Therefore

$$||s_h - s_h^*|| = o(\delta^3). \tag{3.2.30}$$

During the proof of the Theorem 3.1.2, we showed that

$$M^{-1}(h) = D^{-1}(h)M_0^{-1}(I+T)$$

where $D(h) = \text{diag}(h(\xi_0),h(\xi_1),\ldots,h(\xi_{k+1}))$ and T is a $(k+2)x(k+2)$ matrix such that $||T|| < K\omega(h,\delta)/(1-K\omega(h,\delta))$. Hence using the equations (A3) and (A4) of the Appendix,

$$\begin{aligned} ||M^{-1}(h)|| &< ||D^{-1}(h)||\ ||M_0^{-1}|| + ||D^{-1}(h)||\ ||M_0^{-1}||\ ||T|| \\ &< (3\beta K_4/\delta) + \{3\beta K_4 K\omega(h,\delta)/\delta(1-K\omega(h,\delta))\} \end{aligned} \tag{3.2.31}$$

where $(1/h(x)) < K_4$. Therefore from (3.2.30) and (3.2.31),

$$\begin{aligned} ||c_h - c^*|| &= ||M_h^{-1}(s_h - s_h^*)|| \\ &= o(\delta^2) \end{aligned} \tag{3.2.32}$$

as $\delta \to 0$. Now (3.2.29) and (3.2.32) gives

$$||c_g - c_h|| < ||c_g - c^*|| + ||c_h - c^*||$$

$$= \psi_1, \text{ say} \qquad (3.2.33)$$

where $\psi_1 = o(\delta^2)$ as $\delta \to 0$.

Now

$$B_2 = \int(\ell'(x)c_g - \ell'(x)c_h)^2 dx$$

$$= (c_g - c_h)'M_0(c_g - c_h).$$

Thus from (3.2.33),

$$B_2 < (k+2)||c_g - c_h|| \; ||M_0(c_g - c_h)||$$

$$< (k+2)||c_g - c_h||^2 ||M_0||$$

$$< (k+2)\delta \; \psi_1^2,$$

since $||M_0|| < \delta$. Note that $(k+2)$ is the number of elements in the vector $c_g - c_h$. From (3.1.2) write $(k+1)^{-1} = p(\eta_\nu)\delta_\nu$ and then use (3.2.25) to get that

$$B_2 = o(\delta^4),$$

since $\psi_1^2 = o(\delta^4)$. Q.E.D.

Proof of Theorem 3.2.3:

Since we are using the LSE, the bias term is

$$B = \int_0^1 (g(x) - \ell'(x)M^{-1}(\mu)\int\ell(x)g(x)d\mu(x))^2 dx.$$

But here the design μ is concentrated at the knots, therefore the matrix $M(\mu)$ is diagonal with the diagonal elements μ_ν and

$c = M^{-1}(\mu)\int \ell(x)g(x)d\mu(x) = g_r$. Then

$$B = \int_0^1 (g(x) - \sum_{\nu=0}^{k+1} \ell_\nu(x)g(\xi_\nu))^2 dx$$

$$= \sum_{i=1}^{k+1} \int_{\xi_{i-1}}^{\xi_i} (g(x)- \sum_{\nu=0}^{k+1} \ell_\nu(x)g(\xi_\nu))^2 dx$$

$$= \sum_{i=1}^{k+1} \int_{\xi_{i-1}}^{\xi_i} (g(x)-g(\xi_{i-1})\ell_{i-1}(x)-g(\xi_i)\ell_i(x))^2 dx \qquad (3.2.34)$$

since in interval (ξ_{i-1},ξ_i) only $\ell_{i-1}(x)$ and $\ell_i(x)$ are non-zero. Using the interpolation formula (3.2.27), we can evaluate each integral in (3.2.34) and get

$$B = \sum_{i=1}^{k+1} \frac{(g''(\eta_i))^2}{120} \delta_i^5.$$

From (3.1.2), $p(\gamma_i)\delta_i = (k+1)^{-1}$ for $\gamma_i \in (\xi_{i-1},\xi_i)$. Hence

$$B = \frac{1}{120} \frac{1}{(K+1)^4} \sum_{i=1}^{k+1} \frac{(g''(\eta_i))^2}{p^4(\gamma_i)} \delta_i.$$

If $k \to \infty$ in such a way that $\delta = \max_i \delta_i \to 0$, then $\sum_{i=1}^{k+1} \frac{g''(\eta_i)}{p^4(\gamma_i)} \delta_i \to \int \frac{(g''(x))^2}{p^4(x)} dx$ and therefore we have

$$B \approx \frac{1}{120k^4} \int_0^1 \frac{(g''(x))^2}{p^4(x)} dx.$$

Q.E.D.

3.3 Asymptotic Value of IMSE for Estimates other Than LSE

Reasonable choices for C other than the LSE can be obtained by trying to find C so that the expected value of $C\bar{y}$ is equal to (i) $M_0^{-1}s$ or (ii) $tM_0^{-1}s$ (see (2.1.9)) or (iii) the $(\nu+1)$st coordinate of the expected value of $C\bar{y}$ is approximately equal to the expression in (3.2.8).

As a suggestion for an estimator in case (iii) let us take the observations on the knots and choose the $(\nu+1)$st coordinate of $C\bar{y}$ to be

$$\bar{y}_\nu - ((\bar{y}_{\nu+1} - 2\bar{y}_\nu + \bar{y}_{\nu-1})/\delta_\nu^2)(\delta_\nu^2/12) \qquad (3.3.1)$$

to approximate $g(\xi_\nu) - g''(\xi_\nu)(\delta_\nu^2/12)$. If we choose C in this way, we see that the bias term is asymptotically minimized and therefore the asymptotic expression for the bias will be given by (3.2.7). The asymptotic expression for the variance of this estimator is found in the following Theorem 3.3.1. Note that here the weights at the knots are given by (3.1.4).

Theorem 3.3.1: Let the estimate C be chosen so that the $(\nu+1)$st coordinate of $C\bar{y}$ is given by (3.3.1) and let the observations be taken at the knots with weights given by (3.1.4). Then the asymptotic expression for the variance is

$$V \approx \frac{23}{27}\frac{\sigma^2 k}{n}\int_0^1 \frac{p(x)}{h(x)}\,dx. \qquad (3.3.2)$$

Proof: The νth element of the estimate $C\bar{y}$ is given by

$$\bar{y}_\nu - (1/12)(\bar{y}_{\nu+1} - 2\bar{y}_\nu + \bar{y}_{\nu-1}), \quad \nu = 0,1,\ldots,k+1.$$

To make sense for $\nu = 0$ and $\nu = k+1$, we set $\bar{y}_{-1} = 0$ and $\bar{y}_{k+2} = 0$. Now the matrix C is a tridiagonal matrix with the diagonal elements equal to (14/12), and off-diagonal elements equal to (-1/12). Thus we have

$$\begin{aligned}\mathrm{Tr}\ CD_\mu^{-1}C'M_0 \\ &= (1/144)\sum_{i=0}^{k}\{(189\delta_{i+1}/3\mu_i)-(2\delta_{i+1}/\mu_{i+1})+(\delta_i/3\mu_{i+1})\} \\ &\quad + (1/144)\sum_{i=1}^{k+1}\{(189\delta_i/3\mu_i)-(2\delta_i/\mu_{i-1})+(\delta_{i+1}/3\mu_{i-1})\}.\end{aligned}$$

Note that in the above equality we have assumed $\delta_0 = 0$, $\delta_{k+2} = 0$. Again we can write

$$\begin{aligned}\mathrm{Tr}\ CD_\mu^{-1}C'M_0 &\simeq (23/27)\sum_i(\delta_i/\mu_i) \\ &\simeq (23/27)\sum_i(1/h(\xi_i)),\end{aligned}$$

since $\mu_i \simeq h(\xi_i)\delta_i$. Using (3.1.2), we have $p(\eta_i)\delta_i = (k+1)^{-1}$. Therefore

$$(k+1)^{-1}\mathrm{Tr}\ CD_\mu^{-1}C'M_0 = (23/27)\sum_i(p(\eta_i)/h(\xi_i))\delta_i. \tag{3.3.3}$$

As $k \to \infty$ in such a way that $\delta = \max_i \delta_i$ tends to zero, the right hand side of (3.3.3) goes to $(23/27)\int(p(x)/h(x))dx$. Therefore we have

$$\begin{aligned}V &= (\sigma^2/n)\mathrm{Tr}\ CD_\mu^{-1}C'M_0 \\ &\approx \frac{23}{27}\frac{\sigma^2 n}{k}\int_0^1 \frac{p(x)}{h(x)}\,dx.\end{aligned}$$

Q.E.D.

In the following we consider an estimate for the case (i) i.e. an estimate for which the bias term is minimized. For further discussion we shall refer to such estimates as KMH estimates. The initials stand for Karson, Manson and Hader (1969) who used the approach of minimizing the bias. Recall that we have $n_i = n\mu_i$ observations at each x_i, $i = 1,2,\ldots,r$. Let $L'(x) = (L_1(x), L_2(x),\ldots, L_r(x))$, where $L_i(x)$, $i = 1,\ldots,r$ are the linear spline functions such that $L_i(x_j) = \delta_{ij}$, $i,j = 1,\ldots,r$. Note that the functions $L(x)$ defined here are different from the functions $\ell(x)$ used before (section 3.1 and 3.2). $L(x)$ consits of r elements and its ith element $L_i(x)$ has support on the interval (x_{i-1},x_{i+1}), while $\ell(x)$ contains $(k+2)$ elements and its ith element $\ell_i(x)$ has support on the interval (ξ_{i-1},ξ_{i+1}). Given any function $f(x)$ defined on $[0,1]$, let $\tilde{f}(x)$ denotes the first degree spline which interpolates $f(x)$ at x_i, $i = 1,2,\ldots,r$. We can then represent the interpolating function $\tilde{f}(x)$ in terms of the 'roof or triangle-shaped functions' (see section 1.3) L_i's as

$$\tilde{f}(x) = \sum_{i=1}^{r} f(x_i)L_i(x). \tag{3.3.4}$$

It is a well known results (e.g. see Prenter (1975)) that if f is continuous, then $\tilde{f}$ converges to f pointwise as $\eta = \max_i(x_i-x_{i-1})$ goes to zero and if f is in $C^2[0,1]$ then

$$\begin{aligned} ||f-\tilde{f}||_\infty &= \max_{x\in[0,1]} |f(x)-\tilde{f}(x)| \\ &\leq \text{const.}||f''||_\infty \eta^2 . \end{aligned} \tag{3.3.5}$$

These properties of $\tilde{f}$ will be kept in mind in finding KMH estimates. Recall that we started with $\hat{\theta} = C\bar{y}$. Here we want a C such that

$$E(C\bar{y}) = M_0^{-1} \int_0^1 \ell(x)g(x)dx. \qquad (3.3.6)$$

It is hard to find a C for which (3.3.6) holds, so instead we try to find a C such that

$$E(C\bar{y}) \simeq M_0^{-1} \int_0^1 \ell(x)g(x)dx. \qquad (3.3.7)$$

The asymptotic is in the sense that $||E(C\bar{y})-M_0^{-1} \int_0^1 \ell(x)g(x)dx||$ tends to zero as $r \to \infty$, where the vector norm $||\ ||$ is the sup norm (see Appendix A). Let $\tilde{g}(x) = \sum_{i=1}^{r} g(x_i)L_i(x)$. Since $g \in C^2[0,1]$, using the inequalities (3.2.26) and (3.3.5) we can easily see that

$$\begin{aligned} &||M_0^{-1} \int_0^1 \ell(x)\tilde{g}(x)dx - M_0^{-1} \int_0^1 \ell(x)g(x)dx|| \\ &= ||M_0^{-1} \int_0^1 \ell(x)(\tilde{g}(x)-g(x))dx|| \\ &\le ||M_0^{-1}||\ ||\int_0^1 \ell(x)(\tilde{g}(x)-g(x))dx|| \\ &< \text{Const.}||g''||_\infty \eta^2. \end{aligned}$$

The expression on right of the above inequality goes to zero as $\eta = \max(x_i - x_{i-1}) \to 0$. We notice that,

$$\begin{aligned} E\{(M_0^{-1} \int_0^1 \ell(x)L'(x)dx)\bar{y}\} &= (M_0^{-1} \int_0^1 \ell(x)L'(x)dx)g_r \\ &= M_0^{-1} \sum_{i=1}^{r} \int_0^1 \ell(x)L_i(x)g(x_i)dx \\ &= M_0^{-1} \int_0^1 \ell(x)\tilde{g}(x)dx. \end{aligned}$$

Therefore, in order that (3.3.7) holds, we should choose C as:

$$C = M_0^{-1} \int_0^1 \ell(x)L'(x)dx. \tag{3.3.8}$$

We shall use this as one possible estimator in Chapter 4 while discussing an algorithm and some numerical examples. In the following theorem we shall find the asymptotic expression for the variance term using the estimator (3.3.8). This theorem concerns choosing the design μ to have weight μ_i on x_i such that

$$\mu_i = \int L_i(x)h(x)dx, \quad i = 1,2,\ldots,r \tag{3.3.9}$$

for some continuous density h(x) which is positive for all x.

<u>Theorem 3.3.2</u>: If the estimator C, given in (3.3.8), is chosen, the design and the knots are chosen using (3.3.9) and (3.1.2) respectively, where $h(x) > 0$ for all x, then

$$V \approx \frac{\sigma^2 k}{n} \int_0^1 \frac{p(x)}{h(x)}\, dx. \tag{3.3.10}$$

Proof: We have from (1.4.5) and (3.3.8),

$$\begin{aligned} V &= (\sigma^2/n)\text{Tr } CD_\mu^{-1}C'M_0 \\ &= (\sigma^2/n)\text{Tr } M_0^{-1} \int_0^1 \ell(x)L'(x)dxD_\mu^{-1} \int_0^1 L(y)\ell'(y)dy \\ &= (\sigma^2/n)\text{Tr} \int_0^1 \int_0^1 \ell(x)\ell'(y) \sum_{i=1}^{r} \frac{L_i(x)L_i(y)}{\mu_i}\, dxdy\, M_0^{-1}. \end{aligned} \tag{3.3.11}$$

We will use the following lemma to complete the proof of this theorem.

<u>Lemma 3.3.1</u>: 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 $r \to \infty$, we have

$$\lim_{r\to\infty} \int_0^1 \int_0^1 u(x)v(y) \sum_{i=1}^{r} \frac{L_i(x)L_i(y)}{\mu_i} \, dydx = \int_0^1 \frac{u(x)v(x)}{h(x)} \, dx, \qquad (3.3.12)$$

The proof of the lemma will be deferred till the end of this section.

Using the above lemma, we can write from (3.3.11),

$$V \approx (\sigma^2/n) \mathrm{Tr}\, M(h^{-1}) M_0^{-1}$$

where $M(h^{-1})$ is the $(k+2)x(k+2)$ matrix $\int \frac{\ell(x)\ell'(x)}{h(x)} dx$. As we did in Theorem 3.1.2, we can show that

$$M(h^{-1}) = M_0 D(h^{-1}) - B(h^{-1})$$

where $D(h^{-1})$ is a diagonal matrix with diagonal elements $(h^{-1}(\xi_0), h^{-1}(\xi_1), \ldots, h^{-1}(\xi_{k+1}))$ and $B(h^{-1})$ is a tridiagonal matrix with its elements as follows:

$$b_{00}(h^{-1}) = (1/h(\xi_0)) \int_{\xi_0}^{\xi_1} (h(x) - h(\xi_0))(\ell_0^2(x)/h(x)) dx$$

$$b_{k+1,k+2}(h^{-1}) = (1/h(\xi_{k+1})) \int_{\xi_k}^{\xi_{k+1}} (h(x) - h(\xi_{k+1}))(\ell_{k+1}^2(x)/h(x)) dx \quad .$$

$$b_{\nu,\nu}(h^{-1}) = (1/h(\xi_\nu)) \int_{\xi_{\nu-1}}^{\xi_{\nu+1}} (h(x) - h(\xi_\nu))(\ell_\nu^2(x)/h(x)) dx, \quad \nu = 1, \ldots, k$$

$$b_{\nu,\nu+1}(h^{-1}) = (1/h(\xi_{\nu+1})) \int_{\xi_\nu}^{\xi_{\nu+1}} (h(x) - h(\xi_\nu))(\ell_\nu(x)\ell_{\nu+1}(x)/h(x)) dx, \quad \nu = 0, 1, \ldots, k$$

and

$$b_{\nu+1,\nu}(h^{-1}) = (1/h(\xi_\nu))\int_{\xi_\nu}^{\xi_{\nu+1}} (h(x)-h(\xi_\nu))(\ell_\nu(x)\ell_{\nu+1}(x)/h(x))dx,$$

$$\nu = 0,1,\ldots,k.$$

Let $\min_{x\in[0,1]} h(x) = K$. The constant K will be positive since $h(x) > 0$ on $[0,1]$. Now we get the inequalities

$$|b_{\nu,\nu}(h^{-1})| < \omega(h,\delta)(\delta_\nu+\delta_{\nu+1})/3K^2$$

$$|b_{\nu,\nu+1}(h^{-1})| < \omega(h,\delta)\delta_{\nu+1}/6K^2$$

$$|b_{\nu+1,\nu}(h^{-1})| < \omega(h,\delta)\delta_{\nu+1}/6K^2,$$

where $\omega(h,\delta)$ is the modulus of continuity of h. Now since $B(h^{-1})$ is a tridiagonal matrix, the above bounds on the elements of $B(h^{-1})$ will yields

$$||B(h^{-1})|| = \max_\nu(\sum_j |b_{\nu,j}(h^{-1})|)$$

$$< \omega(h,\delta)\delta/K^2, \tag{3.3.13}$$

where $\delta = \max_\nu \delta_\nu$. Therefore

$$(k+2)^{-1}\mathrm{Tr}\ M_0^{-1}\ M(h^{-1}) = (k+2)^{-1}\mathrm{Tr}\ M_0^{-1}(M_0D(h^{-1})-B(h^{-1}))$$

$$= (k+2)^{-1}\mathrm{Tr}\ D(h^{-1})-(k+2)^{-1}\mathrm{Tr}M_0^{-1}B(h^{-1}).$$

Now using the bounds on the norm of the matrices M_0^{-1} (see (3.2.26)) and $B(h^{-1})$ (see (3.3.13)), we have for the second term in the above equality,

$$|(k+2)^{-1}\mathrm{Tr}\; M_0^{-1}B(h^{-1})| < (k+2)^{-1}(k+2)||M_0^{-1}B(h^{-1})||$$
$$< ||M_0^{-1}||\;||B(h^{-1})||$$
$$< 3\beta\omega(h,\delta)/K^2.$$

But $\omega(h,\delta) \to 0$ as $\delta \to 0$ since h is uniformly continuous on [0,1]. Thus

$$(k+2)^{-1}\mathrm{Tr}\; M_0^{-1}M(h^{-1}) = (k+2)^{-1}\mathrm{Tr}\; D(h^{-1}) + o(1)$$

as $\delta \to 0$. From (3.1.2), $p(\eta_\nu)\delta_\nu = (k+1)^{-1}$ for some $\eta_\nu \in (\xi_{\nu-1},\xi_\nu)$. Therefore

$$(k+2)^{-1}\mathrm{Tr}\; M_0^{-1}M(h^{-1}) = (k+1/k+2)\sum(p(\eta_\nu)/h(\xi_\nu))\delta_\nu + o(1).$$

Letting $k \to \infty$ in such a way that $\delta \to 0$, the right hand side approaches $\int \frac{p(x)}{h(x)}\,dx$. Q.E.D.

As above, we can suggest some other choices for $C\bar{y}$ which would involve approximating g in $M_0^{-1}s$ where $s = \int \ell(x)g(x)dx$ (e.g. see the remark following Theorem 2.1.1).

Proof of Lemma 3.3.1:

Let us denote by I, the double integral on left of (3.3.12). Since $L_j(x)$ $(2 \leq j \leq r-1)$ has support on the interval (x_{j-1},x_{j+1}), $L_1(x)$ and $L_r(x)$ has support on (x_1,x_2) and (x_{r-1},x_r) respectively, we can express the integral I as

$$I = \sum_{j=1}^{r} I_j, \qquad (3.3.14)$$

where

$$I_1 = (1/\mu_1)\int_{x_1}^{x_2}\int_{x_1}^{x_2} u(x)v(y)L_1(x)L_1(y)dydx$$

$$I_r = (1/\mu_r)\int_{x_{r-1}}^{x_r}\int_{x_{r-1}}^{x_r} u(x)v(y)L_r(x)L_r(y)dydx$$

and

$$I_j = (1/\mu_j)\int_{x_{j-1}}^{x_{j+1}}\int_{x_{j-1}}^{x_{j+1}} u(x)v(y)L_j(x)L_j(y)dydx, \; j = 2,\ldots,r-1.$$

Writing $u(x) = u(x_j) + (u(x)-u(x_j))$ for $x \in (x_{j-1},x_{j+1})$ and $v(y) = v(x_j) + (v(y)-v(x_j))$ for $y \in (x_{j-1},x_{j+1})$, we get for $j = 2,\ldots,r-1$,

$$I_j = I_j^{(1)} + I_j^{(2)} + I_j^{(3)} + I_j^{(4)}, \text{ where}$$

$$I_j^{(1)} = (1/4\mu_j)u(x_j)v(x_j)(\eta_j+\eta_{j+1})^2$$

$$I_j^{(2)} = (1/2\mu_j)v(x_j)(\eta_j+\eta_{j+1})\int_{x_{j-1}}^{x_{j+1}} (u(x)-u(x_j))L_j(x)dx$$

$$I_j^{(3)} = (1/2\mu_j)u(x_j)(\eta_j+\eta_{j+1})\int_{x_{j-1}}^{x_{j+1}} (v(y)-v(x_j))L_j(y)dy$$

$$I_j^{(4)} = (1/\mu_j)\left\{\int_{x_{j-1}}^{x_{j+1}} (u(x)-u(x_j))L_j(x)dx\right\}\left\{\int_{x_{j-1}}^{x_{j+1}} (v(y)-v(x_j))L_j(y)dy\right\}.$$

From (3.3.9), for $2 \leq j \leq r-1$,

$$\mu_j = \int L_j(x)h(x)dx$$

$$= (1/2)h(x_j)(\eta_j+\eta_{j+1})(1+\alpha_j), \qquad (3.3.15)$$

where

$$\alpha_j = \{2/h(x_j)(\eta_j + {}_{j+1})\}\int_{x_{j-1}}^{x_{j+1}} (h(x)-h(x_j))L_j(x)dx.$$

Substituting this value of μ_j in the expression for $I_j^{(1)}$ yields

$$I_j^{(1)} = (1/2h(x_j))u(x_j)v(x_j)(\eta_j+\eta_{j+1})(1+\alpha_j)^{-1}.$$

Since $\min_{x\in[0,1]} h(x) = K > 0$, we get the following bound on α_j,

$$|\alpha_j| \leq \omega(h,\eta)/K. \tag{3.3.16}$$

Since $\omega(h,\eta) \to 0$ as $\eta \to 0$, it follows that

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

Let $\omega(u,\eta)$ and $\omega(v,\eta)$ be the modulus of continuity of $u(x)$ and $v(x)$ respectively on $[0,1]$. Using (3.3.15) for μ_j, it is easy to check that

$$\begin{aligned}
\Big|\sum_{j=2}^{r-1} I_j^{(2)}\Big| &< (1/2)\omega(u,\eta)\sum_{j=2}^{r-1} \frac{v(x_j)}{h(x_j)} (\eta_j+\eta_{j+1})(1+\alpha_j)^{-1} \\
\Big|\sum_{j=2}^{r-1} I_j^{(3)}\Big| &< (1/2)\omega(v,\eta)\sum_{j=2}^{r-1} \frac{u(x_j)}{h(x_j)} (\eta_j+\eta_{j+1})(1+\alpha_j)^{-1} \qquad (3.3.17) \\
\Big|\sum_{j=2}^{r-1} I_j^{(4)}\Big| &< (1/2)\omega(u,\eta)\omega(v,\eta)\sum_{j=2}^{r-1} \frac{1}{h(x_j)} (\eta_j+\eta_{j+1})(1+\alpha_j)^{-1}.
\end{aligned}$$

Since $\omega(v,\eta)$ and $\omega(u,\eta)$ tends to zero as $\eta \to 0$, it follows from (3.3.16) that each of the terms in (3.3.17) is of the order $o(1)$ as $\eta \to 0$, and therefore

$$\sum_{j=2}^{r-1} I_j = \sum_{j=2}^{r-1} I_j^{(1)} + \sum_{j=2}^{r-1} I_j^{(2)} + \sum_{j=2}^{r-1} I_j^{(3)} + \sum_{j=2}^{r-1} I_j^{(4)}$$

$$= (1/2)\sum_{j=2}^{r-1} \frac{u(x_j)v(x_j)}{h(x_j)} (\eta_j+\eta_{j+1}) + o(1). \qquad (3.3.18)$$

Since functions u and v are bounded above and h is bounded below, it is easy to verify that

$$|I_1| < \text{Const. } \eta_1, \text{ and } |I_r| < \text{Const. } \eta_r. \qquad (3.3.19)$$

Combining the results in (3.3.18) and (3.3.19), we finally get

$$I = \sum_{j=1}^{r} I_j$$

$$= (1/2)\sum_{j=2}^{r-1} \frac{u(x_j)v(x_j)}{h(x_j)} (\eta_j+\eta_{j+1}) + o(1) \text{ as } \eta \to 0.$$

This proves our lemma. Q.E.D.

3.4. Minimization of Asymptotic Value of IMSE

As shown in the last three sections, the asymptotic expression for the IMSE for the case d = 1 is

$$J \approx (a\sigma^2 k/n)\int_0^1 \frac{p(x)}{h(x)} + (b/720k^4)\int_0^1 \frac{g''^2(x)}{p^4(x)}\, dx \qquad (3.4.1).$$

where a ranges between 2/3 and 1 and b ranges from 1 to 6. The least square estimator with design measure concentrated on the knots gave values of a = 2/3 and b = 6, while the LSE for a smooth density h gave values a = 1 and b = 1. We have already considered the minimization problem in Section 2.3. So for our purposes here we have to just put d = 1 in all the solutions we got there for seven different mnimimzation problems. In particular, when J is minimized with respect to h, k and p simultaneously, the minimizing h, k and p are as follows:

$$h(x) = p^{1/2}(x)/\int p^{1/2}(y)dy \tag{3.4.2}$$

$$k = \int g''^{4/9}\{nb/180a\sigma^2 \int g''^{2/9}\}^{1/5} \tag{3.4.3}$$

$$p(x) = (g''(x))^{4/9}/\int(g''(y))^{4/9}dy. \tag{3.4.4}$$

Some restriction must be imposed on k given in (3.4.3) since we want at least $k \leq n$. Thus we see that J as a function of k, p and h is minimized by k, p and h given in (3.4.3), (3.4.4), and (3.4.2) respectively. The function p in (3.4.4) indicates that knot should be placed where g" is large. The relation $h \propto \sqrt{p}$ indicates that h should move away from p becoming more uniform. Finally k is generally increasing in g", decreasing in σ and of the order $n^{1/5}$. This last order relation tells to take many more observations than knots. This is due to the fact that V increases linearly in k while the bias B decreases with k^{-4}.

3.5 Local Solutions for Minimization Problems

Let us recall Theorem 2.3.1. There, we imposed the restriction that the minimizing k should be less than or equal to n. Here we shall discuss the case when k given by (2.3.5) becomes greater than n. For simplicity, we shall consider here the linear spline case but the results can be extended to the spline functions of any degree d. For the linear spline case (d = 1) this minimizing k is given in (3.4.3). The restriction $k \leq n$ means, then, that

$$\int g''^{4/9}\{nb/180a\sigma^2 \int g''^{2/9}\}^{1/5} \leq n$$

or

$$\sigma^2 \geq \sigma^2_{crit} \tag{3.5.1}$$

where $\sigma^2_{crit} = \{b(\int g''^{4/9})^5/180an^4 \int g''^{2/9}\}$.

In the following we shall find some approximate solutions for the minimization problem when the condition (3.5.1) is not satisfied. We note that J given in (3.4.1) is convex as a function of k, so when (3.5.1) is not satisfied, we take the minimizing k to be equal to n. Replacing k by n in (3.4.1), we get

$$J = a\sigma^2 \int \frac{p(x)}{h(x)}\, dx + (b/720n^4) \int \frac{(g''(x))^2}{(p(x))^4}\, dx. \tag{3.5.2}$$

As a function of h, J is minimized by $h(x) = p^{1/2}(x)/\int p^{1/2}(y)dy$. Substituting this value of h in (3.5.2), we get

$$J = a\sigma^2 \left(\int_0^1 \sqrt{p}\right)^2 + (b/720n^4)\int (g''^2/p^4). \tag{3.5.3}$$

We have to minimize this J with respect to p. Now when $\sigma^2 = 0$, the function J is minimized by $p(x) \alpha (g''(x))^{2/5}$ and when $\sigma^2 \geq \sigma^2_{crit}$, the function is minimized by $p(x) \alpha (g''(x))^{4/9}$. Therefore when $0 < \sigma^2 < \sigma^2_{crit}$, we suggest that the minimizing $p = p^*$ should be given by

$$p^*(x) = (1-\gamma)p^{(0)}(x) + \gamma p^{(1)}(x), \tag{3.5.4}$$

where $p^{(0)}(x) = (g'')^{2/5}/\int(g'')^{2/5}$, $p^{(1)}(x) = (g'')^{4/9}/\int(g'')^{4/9}$ and $\gamma = \sigma^2/\sigma^2_{crit}$.

p* is one of the approximate solution which will be investigated in examples 3.5.1 and 3.5.2.

Let f_1 and f_2 be two functions defined on [0,1] and both have a second derivative at $x_0 \in [0,1]$. Also let f_1 has a minimum at x_0, then for some arbitrary but small ε, the quadratic

$$Q(x) = f_1(x_0)+\varepsilon f_2(x_0)+(x-x_0)\varepsilon f_2'(x_0)+(1/2)(x-x_0)^2(f_1''(x_0)+\varepsilon f_2''(x_0))$$

is a good approximation to $f_1(x)+\varepsilon f_2(x)$ for x near x_0. The function Q(x) is minimized at

$$x_0^L = x_0 - \varepsilon f_2'(x_0)(f_1''(x_0)+\varepsilon f_2''(x_0))^{-1}.$$

We shall say that x_0^L is a local minimum for $f_1(x)+\varepsilon f_2(x)$.

We shall use the above idea to find an approximate minimum for J when $0 < \sigma^2 < \sigma^2_{crit}$. In the expression for J, the integrals will be replaced by Riemann sums. Let $0 = Z_0 < Z_1 < Z_2 <,..< Z_m = 1$ be an equally spaced partition of [0,1] and let $Z_j^* = (Z_{j-1}+Z_j)/2$, $j = 1,2,\ldots,m$. Also let $p_j = p(Z_j^*)$, $g_j = g(Z_j^*)$, $g_j'' = g''(Z_j^*)$, etc. for $j = 1,2,\ldots,m$. Now for large m, J in (3.5.3) can be replaced by

$$J = J(p_1,\ldots,p_m) = a\sigma^2\Big(\frac{1}{m}\sum_1^m \sqrt{p_j}\Big)^2 + (b/720n^4)\Big(\frac{1}{m}\sum_1^m \frac{g_j''^2}{p_j^4}\Big). \qquad (3.5.5)$$

Now we shall find the local minimum of J when (i) σ^2 is closer to 0 (i.e. $0 < \sigma^2 < \sigma^2_{crit}/2$) and (ii) σ^2 is closer to σ^2_{crit}

(i.e. $(\sigma^2_{crit}/2) < \sigma^2 < \sigma^2_{crit}$).

<u>Case 1</u>: Local minimum for $J(\underset{\sim}{p})$ when $0 < \sigma^2 < \sigma^2_{crit}/2$.

The problem of minimizing J given in (3.5.5) subject to the condition that $0 < p_j$ and $(1/m)\Sigma p_j = 1$ is equivalent to minimize

$$J(\underset{\sim}{p}) = J(p_1,\ldots,p_m,\lambda)$$

$$= a\sigma^2(\frac{1}{m}\sum\sqrt{p_j})^2 + \frac{b}{720n^4}(\frac{1}{m}\sum\frac{g_j''^2}{p_j^4}) + \lambda(\frac{1}{m}\sum p_j - 1)$$

$$= \sigma^2 J_1(\underset{\sim}{p}) + J_2(\underset{\sim}{p}), \qquad (3.5.6)$$

where

$$J_1(\underset{\sim}{p}) = (a/m^2)(\sum\sqrt{p_j})^2$$

and

$$J_2(\underset{\sim}{p}) = (b/720mn^4)\sum(g_j''^2/p_j^4) + (\lambda/m)(\sum p_j - m).$$

Note that λ is Lagrange's multiplier. The function $J_2(\underset{\sim}{p})$ is minimized at point $\underset{\sim}{p}^{(0)} = (p_1^{(0)},\ldots,p_m^{(0)}, \lambda^{(0)})'$, where

$$\left.\begin{aligned} p_j^{(0)} &= (mg_j''^{2/5}/\sum g_j''^{2/5}), \quad j = 1,\ldots,m \\ \lambda^{(0)} &= (b/180n^4m^5)(\sum g_j''^{2/5})^5 \end{aligned}\right\} \qquad (3.5.7)$$

Let us assume that function $J(\underset{\sim}{p})$ is twice continuously differentiable. Using Taylor expansion of $J(\underset{\sim}{p})$ about the point $\underset{\sim}{p}^{(0)}$, we get

$$J(\underset{\sim}{p}) = \sum_{i=0}^{2}\frac{1}{i!}\left[\sum_{j=1}^{m}(p_j - p_j^{(0)})\frac{\partial}{\partial p_j} + (\lambda-\lambda^{(0)})\frac{\partial}{\partial\lambda}\right]^i(\sigma^2 J_1(\underset{\sim}{p}^{(0)}) + J_2(\underset{\sim}{p}^{(0)})).$$

(3.5.8)

Now using the fact that

$$[\sum_{j=1}^{m}(p_j-p_j^{(0)})\frac{\partial}{\partial p_j}+(\lambda-\lambda^{(0)})\frac{\partial}{\partial\lambda}]J_2(\underset{\sim}{p}^{(0)}) = 0,$$

the expression (3.5.8) is simplified to

$$J(\underset{\sim}{p}) = \sigma^2 J_1(\underset{\sim}{p}^{(0)}) + J_2(\underset{\sim}{p}^{(0)}) + \sigma^2(\underset{\sim}{p}-\underset{\sim}{p}^{(0)})'\underset{\sim}{c}$$
$$+ (1/2)(\underset{\sim}{p}-\underset{\sim}{p}^{(0)})'(A^{(1)}+\sigma^2 A^{(2)})(\underset{\sim}{p}-\underset{\sim}{p}^{(0)}) \qquad (3.5.9)$$

where $\underset{\sim}{c} = (c_i)$ is a (m+1)x1 vector, $A^{(1)} = [a_{ij}^{(1)}]$ and $A^{(2)} = [a_{ij}^{(2)}]$ are (m+1)x(m+1) matrices with their elements defined as follows:

$$c_i = \frac{\partial}{\partial p_i} J_1(\underset{\sim}{p}^{(0)}) = (a/m^2)(\sum \sqrt{p_j^{(0)}})(1/\sqrt{p_i^{(0)}}), \quad i = 1,\ldots,m$$

$$c_{m+1} = \frac{\partial}{\partial\lambda} J_1(\underset{\sim}{p}^{(0)}) = 0,$$

$$a_{ij}^{(1)} = \frac{\partial^2}{\partial p_i \partial p_j} J_2(\underset{\sim}{p}^{(0)}) = 0, \quad i \neq j, \; i,j = 1,\ldots,m$$

$$a_{m+1,j}^{(1)} = a_{j,m+1}^{(1)} = \frac{\partial^2}{\partial p_j \partial\lambda} J_2(\underset{\sim}{p}^{(0)}) = (1/m), \; j = 1,\ldots,m$$

$$a_{ii}^{(1)} = \frac{\partial^2}{\partial p_i^2} J_2(\underset{\sim}{p}^{(0)}) = (b/36mn^4)\{g_i''^2/(p_i^{(0)})^6\}, \; i = 1,\ldots,m$$

$$a_{m+1,m+1}^{(1)} = \frac{\partial^2}{\partial\lambda^2} J_2(\underset{\sim}{p}^{(0)}) = 0$$

$$a_{ij}^{(2)} = \frac{\partial^2}{\partial p_i \partial p_j} J_1(\underset{\sim}{p}^{(0)}) = (a/2m^2)(1/\sqrt{p_i^{(0)}p_j^{(0)}}), \; i \neq j, \; i,j=1,\ldots,m$$

$$a_{m+1,j}^{(2)} = a_{j,m+1}^{(2)} = \frac{\partial^2}{\partial p_j \partial\lambda} J_1(\underset{\sim}{p}^{(0)}) = 0, \; j = 1,\ldots,m$$

$$a_{ii}^{(2)} = \frac{\partial^2}{\partial p_i^2} J_1(\underset{\sim}{p}^{(0)}) = (a/2m^2 p_i^{(0)})(1-\sum\sqrt{p_j^{(0)}}\ /\sqrt{p_i^{(0)}}),\ i = 1,\ldots,m$$

$$a_{m+1,m+1}^{(2)} = \frac{\partial^2}{\partial \lambda^2} J_1(\underset{\sim}{p}^{(0)}) = 0.$$

We can rewrite $J(\underset{\sim}{p})$ given in (3.5.9) as

$$J(\underset{\sim}{p}) = \sigma^2 J_1(\underset{\sim}{p}^{(0)}) + J_2(\underset{\sim}{p}^{(0)}) - (\sigma^4/2)(\underset{\sim}{c}'A^{-1}\underset{\sim}{c})$$

$$+ (1/2)[\underset{\sim}{u} + \sigma^2 A^{-1}\underset{\sim}{c}]'A[\underset{\sim}{u}+\sigma^2 A^{-1}\underset{\sim}{c}],$$

where $\underset{\sim}{u} = \underset{\sim}{p}-\underset{\sim}{p}^{(0)}$, $A = A^{(1)} + \sigma^2 A^{(2)}$. This expression shows that J is minimized when

$$\underset{\sim}{u} + \sigma^2 A^{-1}\underset{\sim}{c} = \underset{\sim}{0}.$$

So the local minimum p^{L_1}, when σ^2 is close to zero, is given by,

$$\underset{\sim}{p}^{L_1} = \underset{\sim}{p}^{(0)} - \sigma^2 A^{-1}\underset{\sim}{c}. \tag{3.5.10}$$

<u>Case 2</u>: Local minimum for $J(\underset{\sim}{p})$ when $(\sigma_{crit}^2/2) < \sigma^2 < \sigma_{crit}^2$.

If $\sigma^2 = \sigma_{crit}^2$, the function $J(\underset{\sim}{p})$ is minimized at the point $\underset{\sim}{p}^{(1)} = (p_1^{(1)},\ldots,p_m^{(1)},\lambda^{(1)})'$ where $\tilde{\lambda}^{(1)} = 0$ and

$$p_j^{(1)} = (mg_j''^{4/9}/\sum g_j''^{4/9}),\quad j = 1,\ldots,m. \tag{3.5.11}$$

Here we shall expand $J(\underset{\sim}{p})$ in Taylor series about the point $\underset{\sim}{p}^{(1)}$. Proceeding in a similar way as in Case 1, we get the local minimum $\underset{\sim}{p}^{L_2}$ as

$$\underset{\sim}{p}^{L_2} = \underset{\sim}{p}^{(1)} - C^{-1}\underset{\sim}{a} \tag{3.5.12}$$

where $C = [c_{ij}]$ is a (m+1)x(m+1) matrix and $\underset{\sim}{a} = (a_i)$ is a (m+1)x1 vector defined as follows:

$$a_i = \frac{\partial}{\partial p_i} J(\underset{\sim}{p}^{(1)}) = (a\sigma^2/m^2)(\sum \sqrt{p_j^{(1)}}/\sqrt{p_i^{(1)}}) + (\lambda^{(1)}/m)$$

$$- (b/180mn^4)\{g_i''^2/(p_i^{(1)})^5\}, \quad i = 1,\ldots,m$$

$$a_{m+1} = \frac{\partial}{\partial \lambda} J(\underset{\sim}{p}^{(1)}) = 0,$$

$$c_{ij} = \frac{\partial^2}{\partial p_i \partial p_j} J(\underset{\sim}{p}^{(1)}) = (a\sigma^2/2m^2\sqrt{p_i^{(1)} p_j^{(1)}}), \quad i \neq j, \ i,j = 1,\ldots,m$$

$$c_{j,m+1} = c_{m+1,j} = \frac{\partial^2}{\partial \lambda \partial p_j} J(\underset{\sim}{p}^{(1)}) = (1/m), \quad j = 1,\ldots,m$$

$$c_{jj} = \frac{\partial^2}{\partial p_j^2} J(\underset{\sim}{p}^{(1)}) = (a\sigma^2/2m^2 p_j^{(1)})(1-\sum \sqrt{p_i^{(1)}}/\sqrt{p_j^{(1)}})$$

$$+ (b/36mn^4)\{g_j''^2/(p_j^{(1)})^6\}, \quad j = 1,\ldots,m$$

and

$$c_{m+1,m+1} = \frac{\partial^2}{\partial \lambda^2} J(\underset{\sim}{p}^{(1)}) = 0.$$

It is to be noted that the approximate solutions suggested above works good if σ^2_{crit} is not far away from zero. Since σ^2_{crit} is of order n^{-4}, hence for large n it will be close to zero. In some cases even for small n, σ^2_{crit} is close to zero. For example if $g(x) = x^q$, $q \geq 2$, we have

$$\sigma^2_{crit} = \{729bq^2(q-1)^2(2q+5)/20n^4a(4q+1)^5\}.$$

If we take b = a = 1, for p = 3, 4, 5, 6, the respective values of $(n^4\ \sigma^2_{crit})$ are .039, .048, .054 and .057. So here even for n = 10, $\sigma^2_{crit} \approx 5\times10^{-6}$ which is fairly close to zero.

Now we shall consider two examples and see how do these solutions compare to the correct solution obtained by directly minimizing $J(\underset{\sim}{p})$.

Example 3.5.1.

Let $g(x) = x^4$, n = 15, a = b = 1, $\sigma^2 = 2.5\times10^{-7} \leq \sigma^2_{crit} = 9.5\times10^{-7}$. The problem is to minimize J given in (3.5.5) or equivalently to minimize

$$J(\underset{\sim}{p}) = \frac{1}{m}\,\sigma^2\left(\frac{1}{m}\sum\sqrt{p_j}\right)^2 + \frac{1}{720n^4}\left(\frac{1}{m}\sum\frac{g''^2_j}{p_j^4}\right) + \lambda\left(\frac{1}{m}\sum p_j - 1\right)$$

with respect to p_j's. The (m+1) equations (for getting (m+1) unknown namely m p_j's and one λ) are

$$(\sigma^2/m)(\textstyle\sum\sqrt{p_j}/\sqrt{p_i}) - (g''^2_i/180n^4p_i^5) + \lambda = 0,\ i = 1,\dots,m$$

and

$$(1/m)\textstyle\sum p_j - 1 = 0. \tag{3.5.13}$$

Since $g(x) = x^4$, $g''_i = g''(Z^*_i) = 12Z^{*2}_i$, i = 1,2,...,m. The (m+1) equations in (3.5.13) can be solved to get a correct solution of the problem of minimization of $J(\underset{\sim}{p})$ we shall denote this solution by $\underset{\sim}{p}^{**}$.

For m = 5 and m = 10, the (m+1) equations in (3.5.13) were solved using an efficient routine for solving non-linear equations. Here since σ^2 is closer to 0, the local solution $\tilde{p}^{L_1}$ was obtained using (3.5.10) and solution p* was obtained using (3.5.4). These results are shown in the following table.

Table 1: Comparison of Local and Exact Solution - I

m = 5			m = 10		
p*	p**	p^{L_1}	p*	p**	p^{L_1}
.27372	.27104	.26900	.15527	.15175	.14856
.67504	.67538	.67540	.38249	.38068	.37941
1.02770	1.02920	1.02988	.58202	.58175	.58147
1.35581	1.35697	1.35775	.76756	.76834	.76864
1.66772	1.66741	1.66796	.94389	.94526	.94590
			1.11343	1.11499	1.11580
			1.27764	1.27904	1.27991
			1.43747	1.43842	1.43926
			1.59362	1.59386	1.59459
			1.74660	1.74588	1.74645

Both p* and p^{L_1} are close to the correct solution p**, but p^{L_1} is a little bit better than p*. In the above we have considered the case when σ^2 is close to 0. Now we will consider the case when σ^2 is closer to σ^2_{crit}. We take $\sigma^2 = 6.4 \times 10^{-7}$, every other quantity is same. We calculated the correct solution p** and approximate solution p* in a way similar to the above. But now the local solution p^{L_2} was obtained using (3.5.12). The results are as follows:

Table 2: Comparison of Local and Exact Solution - I

m = 5			m = 10		
p^*	p^{**}	p^{L_2}	p^*	p^{**}	p^{L_2}
.25692	.25424	.25298	.14211	.13893	.13785
.65929	.65953	.65908	.36417	.36222	.36102
1.02258	1.02421	1.02472	.56449	.56406	.56331
1.36574	1.36705	1.36789	.75361	.75433	.75413
1.69545	1.69496	1.69532	.93526	.93670	.93698
			1.11137	1.11307	1.11371
			1.28308	1.28463	1.28545
			1.45117	1.45220	1.45300
			1.61619	1.61635	1.61692
			1.77857	1.77751	1.77763

Both p^* and p^{L_2} are close to p^{**} but p^{L_2} is closer to p^{**} than p^* is to p^{**}. We also tried $\sigma^2 = 4.75\times10^{-7}$ which is midway between 0 and σ^2_{crit}. In this case solution p^* and p^{L_2} were a little bit better than p^{L_1}.

Example 3.5.2.

In this example, we shall consider an exponential function. Let $g(x) = e^x + x^3 + x^2$, $a = b = 1$, $n = 15$, $\sigma^2 = 2.5\times10^{-7} < \sigma^2_{crit} = 4.6\times10^{-6}$. The correct solution $\tilde{p}^{**}$ was obtained by solving (m+1) nonlinear equations, solution $\tilde{p}^*$ was obtained using (3.5.4) and since σ^2 is close to zero, the local solution $\tilde{p}^{L_1}$ was obtained using (3.5.10). The results for m = 5 and 10 are shown in the following table. As the table indicates the solution p^{**} and p^{L_1} are almost same and p_j^*'s differs from p_j^{**}'s only in fifth decimal place.

Table 3: Comparison of Local and Exact Solution - II

m = 5			m = 10		
p^*	p^{**}	p^{L_1}	p^*	p^{**}	p^{L_1}
0.79782	0.79788	0.79788	0.76651	0.76656	0.76654
0.91083	0.91089	0.91089	0.82817	0.82823	0.82823
1.00943	1.00946	1.00946	0.88453	0.88459	0.88460
1.09905	1.09901	1.09902	0.93689	0.93695	0.93695
1.18287	1.18275	1.18275	0.98613	0.98617	0.98617
			1.03289	1.03291	1.03291
			1.07767	1.07766	1.07766
			1.12085	1.12080	1.12080
			1.16274	1.16265	1.16265
			1.20361	1.20347	1.20347

Now we took $\sigma^2 = 4.0 \times 10^{-6}$ which is close to σ^2_{crit}. Here the solutions p^* and p^{**} were obtained as before but here the local solution p^{L_2} was obtained using (3.5.12). The results are shown in the following table:

Table 4: Comparison of Local and Exact Solution - II

m = 5			m = 10		
p^*	p^{**}	p^{L_2}	p^*	p^{**}	p^{L_2}
0.78129	0.78143	0.78140	0.74792	0.74802	0.74798
0.90249	0.90266	0.90266	0.81365	0.81380	0.81378
1.00936	1.00944	1.00946	0.87412	0.87429	0.87429
1.10730	1.10723	1.10724	0.93060	0.93076	0.93077
1.19956	1.19925	1.19924	0.98397	0.98410	0.98411
			1.03489	1.03495	1.03497
			1.08384	1.08382	1.08383
			1.13121	1.13109	1.13110
			1.17733	1.17709	1.17709
			1.22246	1.22209	1.22208

Here also we see that although p^* is close to p^{**} but local solution p^{L_2} is much closer to correct solution p^{**}.

As these examples indicate the approximate solution p^* and the local solution p^{L_1} and p^{L_2} are giving fairly good solution of minimization of IMSE in the case $0 < \sigma^2 < \sigma^2_{crit}$. However the local solutions p^{L_1} and p^{L_2} do relatively a little bit better than p^*.

CHAPTER IV

ALGORITHM AND NUMERICAL RESULTS

4.1. Algorithm for Estimating a Function

Now we shall discuss how the theoretical results obtained in the previous chapters can be exploited in adaptively estimating a more or less arbitrary response function g. To make the matters simple, in this chapter, we shall estimate g by the simple linear spline function. Let us recall the results of the minimization of the asymptotic value of IMSE (See section 3.4). The minimizing h, p and k are given by

$$h(x) \propto (g''(x))^{2/9} \tag{4.1.1}$$

$$p(x) \propto g''(x)^{4/9} \tag{4.1.2}$$

and

$$k = \int g''^{4/9} \left\{ \frac{n}{180\sigma^2} \frac{1}{\int g''^{2/9}} \right\}^{1/5}. \tag{4.1.3}$$

Firstly note that here we have taken a = b = 1 in the equation (3.4.3), and secondly we have assumed that all the three variables, namely, k, p and h are unknown in the asymptotic value of J, the IMSE. We are discussing the algorithm for this general case, since the cases in which one of the variables (k, p or h) is known or two of them are known, the same algorithm can be used in a much more simplified form.

We consider the following iterative procedure. We are given $n^{(0)}$ observations on g which are distributed among $r^{(0)}$ points $x_1^{(0)},\ldots,x_{r^{(0)}}^{(0)}$ with $n_i^{(0)}$ observations at $x_i^{(0)}$, $i = 1,2,\ldots,r^{(0)}$, so that $\sum_{i=1}^{r^{(0)}} n_i^{(0)} = n^{(0)}$. Let y_{ij} denote the jth observation $(j = 1,2,\ldots,n_i^{(0)})$ at $x_i^{(0)}$. We begin with a knot-set

$$\pi^{(0)}:\ 0 = \xi_0^{(0)} < \xi_1^{(0)} <\ldots< \xi_k^{(0)} < \xi_{k+1}^{(0)} = 1.$$

This partition $\pi^{(0)}$ is an initial guess perhaps based on some information about the function g being estimated. In the absence of information about g, a uniform knot spacing may be used. Let us now spell out the steps involved in this iterative procedure and then we shall explain the implementation of these steps.

I. Estimate $g(x)$, $g''(x)$.

II. Estimate σ^2.

III. Find $\hat{J}$, an estimate of IMSE.

IV. Find $\hat{h}$, an estimate of the best design h, using (4.1.1).

V. Add, say, $r^{(1)}$ points of observation $x_1^{(1)},\ldots,x_{r^{(1)}}^{(1)}$. Let $n^{(1)}$ more observations are taken. Distribute $n^{(0)} + n^{(1)}$ observations among $r^{(0)} + r^{(1)}$ points $x_1^{(0)},\ldots,x_{r^{(0)}}^{(0)}$, $x_1^{(1)},\ldots,x_{r^{(1)}}^{(1)}$ so that the displacement of the combined set of observations resembles $\hat{h}$ (obtained in IV) as clearly as possible.

VI. Estimate $g(x)$ and $g''(x)$.

VII. Estimate σ^2.

VIII. Find $\hat{k}$, the estimate of k, using (4.1.3).

IX. Find $\hat{p}$, the estimate of knot displacement p, using (4.1.2).

X. The number $\hat{k}$ (obtained in VIII) of knots are chosen.

XI. Estimate $g(x)$, $g''(x)$.

XII. Find $\hat{J}$, the estimate of IMSE.

The steps IV through XII is called a cycle. The cycle is repeated unless a termination is encountered. We shall talk about the termination criterion later. Now we discuss the implementation of the above algorithm.

Step I:

To estimate g we have used three different kind of estimators. These will be discussed in section 4.2. The estimation of $g''(x)$ can be done in many different ways. We indicate three of them:
(a) Find least square estimate of $g(x)$ by a cubic splines and then take the second derivative of the fitted cubic spline to get an estimate of $g''(x)$. There are some standard routines for approximating g by a cubic spline , e.g. see de Boor and Rice (1968).
(b) g" can be estimated on intervals centered at the knots by a simple second difference with estimated values of g at adjacent knots.
(c) Let $s_0(x) = \hat{g}(x)$ be the linear spline approximation to $g(x)$ which is obtained previously using one of the three estimators. We hope that the approximation s_0 contains enough information about

g so that a reasonable approximation to g" may be obtained from it. We can not use s_0'' as an approximation to g" for the obvious reason that s_0'' is zero except at the knots, where it does not exist. As suggested by Dodson (1972) we first find a broken line approximation of g'; its derivative may then be used as the desired approximation to g".

For this purpose s_0', the derivate of s_0, is used. Note that s_0' is constant on the intervals $(\xi_{i-1}^{(0)},\xi_i^{(0)})$, $i = 1,\ldots,k+1$. We put a breakpoint τ_i at the center of each interval $[\xi_{i-1}^{(0)},\xi_i^{(0)}]$, $i = 1,\ldots,k+1$; and define a broken line function $\tilde{s}$ on $\{\tau_i\}$ so as to interpolate s_0' at each τ_i. In the first and last intervals, $[0,\tau_1]$ and $[\tau_{k+1},1]$, the continuation of the broken line in the adjacent interior interval is used. Finally, with $\tilde{s}$ so defined, we use the step function $\hat{g}'' \overset{\text{def}}{=} \tilde{s}'$ as an estimate to g".

Step II:

At any stage (or iteration) we are trying to fit the model

$$y = X\beta + \varepsilon$$

to the given data, where y is nx1 vector of observations, X is nxm design matrix, β is mx1 vector of unknown parameters and ε is nx1 vector of random errors. From the least square regression theory we know that:

$$\hat{\sigma}_e^2 = \text{Residual sum of square}$$

$$= \frac{y'y-y'X(X'X)^{-1}X'y}{n-m}$$

is an estimate of σ^2 if the model is correct but not otherwise. If a prior estimate of σ^2 is available we can see (or test by an F-test) whether or not the residual mean square is significantly greater than this prior estimate. If it is significantly greater we say that there is lack of fit and we would reconsider the model. If no prior estimate of σ^2 is available, but repeat measurements of y (i.e. two or more measurements) have been made at the same value of x, we can use the mean square for pure error defined by

$$\hat{\sigma}_{p.e} = \left(\sum_{i=1}^{r} \sum_{j=1}^{n_i} (y_{ij}-\bar{y}_i)^2\right) / \left(\sum_{i=1}^{r} n_i - r\right)$$

as an estimate of σ^2.

Step III:

This will be discussed in section 4.2.

Step IV:

We have obtained $\hat{g}''$ in Step I. So from (4.1.1), we get

$$\hat{h}(x) = (\hat{g}''(x))^{2/9} / \int (\hat{g}''(y))^{2/9} dy.$$

Step V:

(a) Find $r^{(0)} + r^{(1)}$ points t_i's according to the quantiles of $\hat{h}$, i.e. t_i is such that

$$\int_0^{t_i} \hat{h}(x)dx = \frac{i-1}{r^{(0)}+r^{(1)}-1}, \quad i = 1,\ldots,r^{(0)}+r^{(1)}.$$

(b) Among the $r^{(0)} + r^{(1)}$ t_i's find the ones which are close to $x_j^{(0)}$, $j = 1,2,\ldots,r^{(0)}$, the remaining $r^{(1)}$ t_i's will be the points $x_j^{(1)}$, $j = 1,\ldots,r^{(1)}$ which are to be added at this stage. We arrange these $r^{(0)} + r^{(1)}$ points in increasing order and denote the ordered set as $x_1,\ldots,x_r$, where $r = r^{(0)} + r^{(1)}$.

(c) $\hat{h}$ is a continuous design, we have to discretize it. To do this, we find $H(x) = \int_0^x \hat{h}(t)dt$ and approximate it by a distribution function $u(x)$ having jumps at x_i, $i = 1,\ldots,r$. We take u to be an uniform approximation of H i.e. u is the solution of the problem

$$\min \max_{x \in [0,1]} |H(x)-u(x)| \tag{4.1.4}$$

where minimum is taken over all step functions having jumps at x_i, $i = 1,2,\ldots,r$. Thus u is given by

$$u(x) = \begin{cases} 0 & , \quad x < 0 \\ \frac{1}{2}(H(x_{i-1})+H(x_i)), & x_{i-1} \leq x < x_i, \ i = 2,\ldots,r \\ 1 & , \quad x \geq 1. \end{cases}$$

Note that $x_1 = 0$ and $x_r = 1$. So now we have the discrete design μ as below:

$$\mu = \begin{cases} x_1,\ldots,x_r \\ \mu_1,\ldots,\mu_r, \end{cases}$$

where $\mu_1 = u(x_1)$, $\mu_i = u(x_i) - u(x_{i-1})$, $i = 2,\ldots,r$.

(d) Now we have to allocate $n^{(0)} + n^{(1)} = n$ observations according to the design μ. This can be done by a scheme given in section 3.1 of Federov (1972). We allocate $[(n-r)\mu_i]^+$ observations to point x_i, $i = 1,2,\ldots,r$, where $[c]^+$ indicate the smallest integer satisfying the inequality $[c]^+ \geq c$. It is trivial to check that $\sum_{i=1}^{r} [(n-r)\mu_i]^+ \leq n$. So the problem is "how to allocate the remaining unrealized observations $n' = n - \sum_{i=1}^{r} [(n-r)\mu_i]^+$". Federov has suggested that these n' observations can be added one-by-one up to the point where

$$(n-r)\mu_i \geq [(n-r)\mu_i]^+ - \frac{1}{2}. \tag{4.1.5}$$

But we will show in an example below that this is not always possible. We shall, therefore, distribute n' remaining observations randomly among r points. The above scheme for the distribution of observations works well if r is very small compared to n. Note that we should make sure that we took at least $n_i^{(0)}$ observations at points $x_i^{(0)}$, $i = 1,\ldots,r^{(0)}$.

Example: Let $r = 11$, $n = 75$ and the weight distribution is $\mu_1 = .05$, $\mu_2 = .1$, $\mu_3 = .1$, $\mu_4 = .1$, $\mu_5 = .05$, $\mu_6 = .115$, $\mu_7 = .085$, $\mu_8 = .085$, $\mu_9 = .13$, $\mu_{10} = .085$, and $\mu_{11} = .1$. If we allocate $[(n-r)\mu_i]^+$ observations to x_i, the distribution of observations looks like $n_1 = 4$, $n_2 = 7$, $n_3 = 7$, $n_4 = 7$, $n_5 = 4$, $n_6 = 8$, $n_7 = 6$, $n_8 = 6$, $n_9 = 9$, $n_{10} = 6$, and $n_{11} = 7$; and these add up to 71, so we are left with 4 observations. But we check that there is no point satisfying (4.1.5), where these 4 remaining observations can be allocated.

Steps VI and VII can be implemented in a way similar to steps I and II respectively.

Step VIII:

Once we get the estimates of σ^2 and g", we can immediately find $\hat{k}$ from (4.1.3),

$$\hat{k} = \int \hat{g}''^{4/9} \left\{ \frac{n}{180\hat{\sigma}^2} \frac{1}{\int \hat{g}''^{2/9}} \right\}^{1/5}$$

$\hat{k}$ obtained above need not be an integer, so we round it off to make it an integer.

Step IX:

We have from (4.1.2),

$$\hat{p}(x) = \frac{(\hat{g}''(x))^{4/9}}{\int (\hat{g}''(y))^{4/9} dy} .$$

Step X:

We obtain the partion $\pi^{(1)}$ consisting of the $\hat{k}$ knots $\xi_1^{(1)}, \ldots, \xi_k^{(1)}$. The $\xi_i^{(1)}$'s are obtained from the integral relationship

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

$\xi_0^{(1)}$ and $\xi_{k+1}^{(1)}$ represent the end points.

Steps XI and XII have been already discussed.

Let us make a few remarks concerning the algorithm. The estimates $\hat{h}$, $\hat{p}$ and $\hat{k}$ depend on the estimate of g" which is discussed in Step I. We have indicated three methods of finding an approximation to g" and there are many other ways in which this can be done. Because of its simplicity we shall usually adopt the method indicated in (c) of Step I. There the estimate of g" is obtained from the linear spline approximation s_0. This s_0 should therefore be a good approximation to g, whatever that may mean. Thus, generally, the starting partition $\pi^{(0)}$ must consist of a large number of knots and they must be distributed through the interval in a reasonable way. It is conceivable that a very bad choice of the partition $\pi^{(0)}$ could result in a "misleading" spline approximation s_0 and that the resulting partition $\pi^{(1)}$ be still worse. We did not confront such cases of instability during the testing of the algorithm.

The estimates $\hat{h}$, $\hat{p}$, and $\hat{k}$ obtained here are based on the asymptotic results, so if $\hat{k}$ obtained from (4.1.3) is small, it might not be a very good estimate of the number of knots. However it has been found useful to do the following. Instead of finding a partition consisting of $\hat{k}$ knots, we find five different partitions consisting of $\hat{k}$-2, $\hat{k}$-1, $\hat{k}$, $\hat{k}$+1, and $\hat{k}$+2 knots respectively. In each partition the knots are chosen according to the quantiles of same $\hat{p}$. We find the estimate of g(x) and the corresponding integrated mean square error (IMSE) using the knot sets of each of the five partitions. Now select the partition for which the IMSE is minimum. We call the points of this partition to be "good" knot set. It has also been found useful to iterate the algorithm a few times with a fixed number of knots. This allows the algorithm to base its resulting "good" set of knots on a "good" set of the same size. One or two iterations has usually been satisfactory.

Termination Criteria: We use here two termination criteria for the algorithm. The first is a simple bound on the number of complete cycles, i.e. we shall perform no more than "m" cycles. Usually m = 3 or 4 is found to be a reasonable number. The second criteria is based on the test for the "lack of fit". This test can be performed using $\hat{\sigma}_e^2$ and $\hat{\sigma}_{p.e}^2$ (see Step II). If there is no lack of fit in the model, we stop. Other than the above two we can use some other termination criteria depending on our problem, e.g. if the IMSE at any stage is not decreased much compared to the IMSE at the previous stage, then also we can stop.

4.2. Estimates and Their IMSE

We will use here three estimators, least square estimate (LSE) and two "bias minimizing" (or KMH) estimators. The least square estimate is (see (2.1.3)),

$$\hat{\theta}_{LSE} = M^{-1}(\mu)\int \ell(x)\bar{y}_x d\mu(x), \qquad (4.2.1)$$

where $M(\mu) = \int \ell(x)\ell'(x)d\mu(x)$, and $\ell(x)$ is the basis for linear splines with k knots $\xi_1,\ldots,\xi_k$. One of the bias minimizing estimator is (see (3.3.8))

$$\hat{\theta}^{(1)}_{KMH} = M_0^{-1}\int \ell(x)L'(x)\bar{y}dx \tag{4.2.2}$$

where $M_0 = \int \ell(x)\ell'(x)dx$, $\bar{y}' = (\bar{y}_{1.},\ldots,\bar{y}_r)$ is the vector of means of observation, and $L(x)$ is the basis for linear splines with knots at x_i, the point of observation. The other bias minimizing estimator, we shall use here, is

$$\hat{\theta}^{(2)}_{KMH} = M_0^{-1} \sum_{i=1}^{r} \ell(x_i)\bar{y}_{i.}\eta_i, \tag{4.2.3}$$

where $\eta_i = (x_i - x_{i-1})$. The reason for using $\hat{\theta}^{(2)}_{KMH}$ as a bias minimizing estimator is as follows. Note that $E(\hat{\theta}^{(2)}_{KMH}) = M_0^{-1} \sum_{i=1}^{r} \ell(x_i)g(x_i)\eta_i$, and asymptotically it would look like $M_0^{-1}\int \ell(x)g(x)dx$. This is the property which we want the bias minimizing estimator to satisfy. For further reference we shall call estimates given in (4.2.2) and (4.2.3) as KMH(1) and KMH(2) respectively.

For calculating the LSE, we should first check the nonsingularity of the information matrix $M(\mu)$. This can be done by using a result of Karlin and Ziegler (1966, Theorem 2). From this result we derive that $M(\mu)$ will be nonsingular if and only if

$$t_1 < \xi_1,$$
$$t_2 < \xi_2,$$
$$\xi_{i-2} < t_i < \xi_i, \quad i = 3,\ldots,k,$$

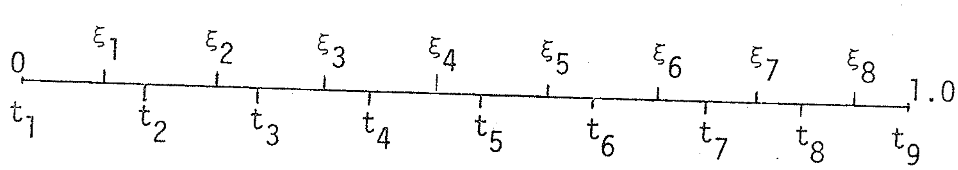
$$\xi_{k-1} < t_{k+1},$$

and

$$\xi_k < t_{k+2}.$$

Where $t_1 < t_2 < \ldots < t_{k+2}$ corresponds to the observation points and $\xi_1 < \xi_2 < \ldots < \xi_k$ corresponds to the knots. The condition says that there should be at least k+2 distinct points of observation (i.e. $r \geq k+2$) and they should be distributed in a manner such that there is at least one observation between any two alternate knots. Let us consider an example to make this condition clear.

Example: $\xi_1 = 0.11$, $\xi_2 = 0.22$, $\xi_3 = .34$, $\xi_4 = .48$, $\xi_5 = .63$, $\xi_6 = .72$, $\xi_7 = .81$, $\xi_8 = .91$.
$t_1 = 0.0$, $t_2 = .125$, $t_3 = .25$, $t_4 = .375$, $t_5 = .5$, $t_6 = .625$, $t_7 = .75$, $t_8 = .875$, $t_9 = 1.00$.



Here we see that $t_1 < \xi_1$, $t_2 < \xi_2$, $\xi_{i-2} < t_i < \xi_i$, for $i = 3,\ldots,8$. We need at least two observations in the interval $(\xi_7, 1.0]$ to make the matrix $M(\mu)$ nonsingular, but there is only one observation at t_9 which is in $(\xi_7, 1.0]$, therefore the matrix will be singular.

We do not have to worry about any such verification in calculating the estimates KMH(1) and KMH(2). Actually we do not even need the restriction that $r(\geq k+2)$. However, the calculations for finding these two bias minimizing estimators are harder than for finding the least square estimator.

Finding the IMSE: The integrated mean square (IMSE) for the LSE is

$$J = \int E(g(x)-\ell'(x)M^{-1}(\mu)\int \ell(x)\bar{y}_x d\mu(x))^2 dx$$
$$= V+B,$$

where

$$V = \frac{\sigma^2}{n} \operatorname{Tr} M^{-1}(\mu)M_0, \tag{4.2.4}$$

and

$$B = \int (g(x)-\ell'(x)M^{-1}(\mu)\int \ell(x)g(x)d\mu(x))^2 dx. \tag{4.2.5}$$

The only unknown parameter in V is σ^2, so to find an estimate of V we just replace σ^2 by $\hat{\sigma}^2$ in (4.2.4). If g(x) is known we can evaluate the integral in (4.2.5) to get B. If the form of g(x) is unknown, which is usually the case, we can find an estimate of B as follows. First using the trapezoidal rule we replace the integral in (4.2.5) by a summation:

$$B \approx \sum_{i=2}^{r} \frac{1}{2}(x_i - x_{i-1})\{(g(x_i)-u(x_i))^2 + (g(x_{i-1})-u(x_{i-1}))^2\}$$

where

$$u(x_i) = \ell'(x_i)M^{-1}(\mu)\int \ell(x)g(x)d\mu(x), \quad i = 1,\dots,r.$$

Now we replace $g(x_i)$ by $\bar{y}_{i.}$, mean of observations at x_i to get an estimate of B as

$$\hat{B} = \sum_{i=2}^{r} \frac{1}{2}(x_i - x_{i-1})\{(\bar{y}_i - \bar{u}(x_i))^2 + (\bar{y}_{i-1} - \bar{u}(x_{i-1}))^2\},$$

where

$$\bar{u}(x_i) = \ell'(x_i)M^{-1}(\mu)\int \ell(x)\bar{y}_x d\mu(x), \quad i=1,2,\dots,r.$$

The above method gives a good estimate of B if we have many more observations than the number of knots.

We can similarly find the estimate of IMSE for other two estimates.

4.3. Numerical Results for Fitting a Curve

Here we shall illustrate the algorithm described in section 4.1 by two numerical examples.

All computations were carried out on a CDC 6500-systems computer. Single precision arithmetic, good to approximately 10 decimal places, was used. We tried to use the double precision arithmetic, but the gain due to the double precision arithmetic over the single precision arithmetic was almost negligible for our calculation purposes. The system of equations, involving the positive definite matrix A, were solved with a routine which uses the Cholesky decomposition of A into LL' where L is the lower triangular matrix.

The digital data were plotted using a Model 936 Calcomp Digital Incremental Plotter.

Example 1:

The first example is a fit to the 'measurements' of a cubic function

$$g(x) = x(3x-1)(3x-2) \tag{4.3.1}$$

for $x \in [0,1]$.

We simulated data errors by adding to $g(x_i)$ a number sampled from the standard Normal Distribution with mean zero and variance 1. To begin we took three equally spaced knots and took five observations at each knot and the end points for a total of 25 observations. We performed four cycles with number of observations n = 75, 150, 250 and 375 respectively. At the end of the fourth cycle, following results were obtained for the three estimates:

LSE	knot-set:	0.0	0.146	0.617	0.832	1.000
	$\hat{\theta}_{LSE}$:	0.135	0.172	-.261	0.435	2.186
KMH(1)	knot-set:	0.0	0.146	0.617	0.832	1.000
	$\hat{\theta}^{(1)}_{KMH}$:	0.24	0.089	-0.225	0.445	2.182
KMH(2)	knot-set:	0.0	0.141	0.576	0.801	1.000
	$\hat{\theta}^{(2)}_{KMH}$:	0.263	0.120	-0.279	0.391	2.014

The function g(x), and the linear spline fits due to these three estimates are shown in Figure 1. The breaks (joints) in the graph of the three estimates are the knots. The algorithm has chosen the knots at the points where curve is taking turn, and it seems to be reasonable.

At each stage a comparison is made with a fourth design and estimate. This was done using a comparable number of knots and observations. Equal number of observations were taken at the end points and at a set of equally spaced knots. These results are

Curves representation

(1) Solid line (———) g(x)

(2) Dotted line (·····) LSE

(3) Small dashed line (-----) KMH(1) estimate

(4) Big dashed line (— — — —) KMH(2) estimate

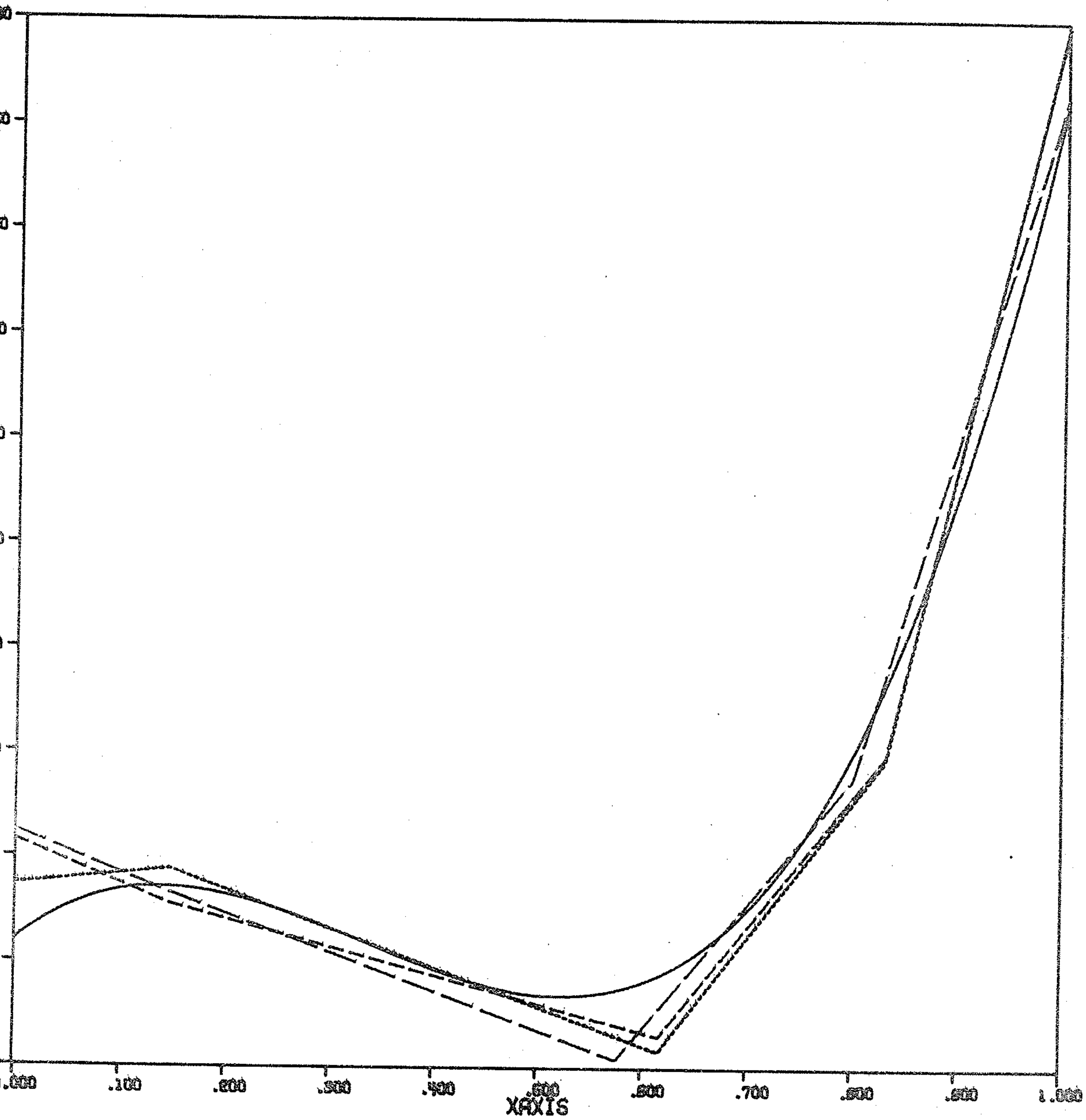


Figure 1

Function and three estimates at the end of fourth cycle shown in Table 5. In this table Estimate 1, 2 and 3 corresponds to least square estimate, KMH(1) estimate and KMH(2) estimate, while estimate 4 corresponds to uniform design with equally spaced knots. For the first three estimates cycle zero represents the start of the algorithm (e.g. here in each case we started with three equally spaced knots and 25 observations distributed equally among end points and knots). We have left blank the column corresponding to 'cycle number' for fourth estimates since we did not use any algorithm for this estimate, so there was no cycle involved in this case.

As expected, the three estimates, namely LSE, KMH(1) and KMH(2), did better in overall IMSE than the uniform design with equally spaced knots. The LSE did a good job in reducing the variance, while KMH(1) and KMH(2) estimates did a comparable job in reducing the bias. Actually, we notice that LSE has done fairly good job in reducing the bias also. Recall that (section 3.2) that for a continuous design h the LSE did a comparable job in asymptotically minimizing B. In the present example the number of knots was small and the design became dispersed over the interval at the second, third and fourth cycle.

In the above, at each cycle, for each of the three estimates (LSE, KMH(1), and KMH(2)), $g''(x)$ was obtained as the second derivative of the cubic spline approximation of $g(x)$ using the current knot set. This cubic spline approximation was found using a program (deBoor and Rice, 1968). As indicated in

Table 5: Comparison of the Four Estimates

Cycle Number	k	n	Integrated Variance V	Integrated Bias B	Integrated Mean Square Error IMSE
			Table 5(a): Estimate 1		
0	3	25	1.333×10^{-1}	1.0076×10^{-2}	1.4341×10^{-1}
1	2	75	4.9680×10^{-2}	7.1080×10^{-3}	5.6788×10^{-2}
2	3	150	2.4948×10^{-2}	6.2794×10^{-3}	3.1227×10^{-2}
3	3	250	1.9747×10^{-2}	1.2596×10^{-3}	2.1007×10^{-2}
4	3	375	1.1875×10^{-2}	9.7437×10^{-4}	1.2849×10^{-2}
			Table 5(b): Estimate 2		
0	3	25	1.3333×10^{-1}	1.0076×10^{-2}	1.4341×10^{-1}
1	2	75	4.6564×10^{-2}	7.3542×10^{-3}	5.3918×10^{-2}
2	2	150	2.6265×10^{-2}	5.6764×10^{-3}	3.1941×10^{-2}
3	3	250	2.0660×10^{-2}	1.0242×10^{-3}	2.1684×10^{-2}
4	3	375	1.3049×10^{-2}	9.5018×10^{-4}	1.3999×10^{-2}
			Table 5(c): Estimate 3		
0	3	25	8.9955×10^{-2}	3.7902×10^{-2}	1.2786×10^{-1}
1	3	75	4.8175×10^{-2}	4.8374×10^{-3}	5.3012×10^{-2}
2	3	150	3.0157×10^{-2}	2.4437×10^{-3}	3.2601×10^{-2}
3	3	250	1.9056×10^{-2}	1.9658×10^{-3}	2.1022×10^{-2}
4	3	375	1.4101×10^{-2}	1.4477×10^{-3}	1.5549×10^{-2}
			Table 5(d): Estimate 4		
-	2	75	3.5088×10^{-2}	3.0688×10^{-2}	6.5776×10^{-2}
-	2	150	1.7544×10^{-2}	3.0688×10^{-2}	4.8232×10^{-2}
-	3	250	1.3333×10^{-2}	1.0076×10^{-2}	2.3409×10^{-2}
-	3	375	8.8889×10^{-3}	1.0076×10^{-2}	1.8965×10^{-2}

Section 4.1 (Step I (c)), $g''(x)$ can also be estimated by the use of linear spline fitted at each stage. We used this later method of estimation of $g''(x)$ for the case of least square estimate, and the following results were obtained.

Table 6: IMSE for Two Different Procedure of Estimation of $g''(x)$.

Cycle Number	k	n	V	B	IMSE
0	3	25	1.3333×10^{-1}	1.0076×10^{-2}	1.4341×10^{-1}
1	2	75	4.6118×10^{-2}	2.6721×10^{-3}	4.8790×10^{-2}
2	2	150	2.4397×10^{-2}	2.4300×10^{-3}	2.6827×10^{-2}
3	2	250	1.5555×10^{-2}	2.3206×10^{-3}	1.7876×10^{-2}
4	2	375	9.9606×10^{-3}	2.0282×10^{-3}	1.1988×10^{-2}

Comparing the results in the above Table 6 with the corresponding results given for LSE in Table 5(a), we see that there is not much difference in two methods of estimating $g''(x)$.

Example 2:

In the previous example $g''(x)$ was linear. Now we consider a function for which $g''(x)$ varies by a large amount in the interval [0,1].

$$g(x) = [(0.1)^2+(2x-0.3)^2]^{-1} + [(0.12)^2+(2x-1.2)^2]^{-1}. \quad (4.3.2)$$

The data errors are simulated in the way similar to Example 1. Here also we started with three equally spaced knots and took five observations at each knot and the end points. We call this as cycle zero. We stopped after three cycles for each of the three estimates, namely LSE, KMH(1) and KMH(2). Here, as well as in the previous example, a number of internal iteration within each cycle were done to select the number and displacement of the knots. The linear spline fits obtained at the end of cycle zero, cycle one, and cycle two for each of the three estimates (LSE, KMH(1), and KMH(2)) are shown in Figure 2, Figure 3, and Figure 4 respectively. The results of final i.e. third cycle are shown in Figure 5, Figure 6, and Figure 7. We can see from these figures that the number of knots is getting larger and linear spline fit is getting better at each cycle for each of the three estimates. For estimate KMH(1), there is not much improvement in the result of the third cycle (Figure 6) over the result of the second cycle (Figure 3), since we already got a good fit at the end of second cycle. The KMH(2) estimate shows some irregularity at the first cycle, but the fit improved in later cycles.

We calculated the integrated mean square error for all the three estimates at the end of each cycle. These results are shown in Table 7. For each of the three estimates (or procedures), $\hat{g}''(x)$ was obtained by the use of linear spline fit corresponding to that estimate (or procedure). So, here $\hat{g}''(x)$ depends upon the particular estimate (or procedure) being used and hence at any cycle, the number of knots $(\hat{k})$, the displacement of knots $(\hat{p}(x))$, and the allocation of observation $(\hat{h}(x))$ will depend upon the particular procedure. Therefore, even we are using the same number of observations for each of the three procedures at any cycle, the

Curves representation

(1) Solid line (———) g(x)

(2) Big dashed line (— — — —) LSE at end of cycle zero

(3) Small dashed line (------) LSE at end of cycle one

(4) Dotted line (·······) LSE at end of cycle two.

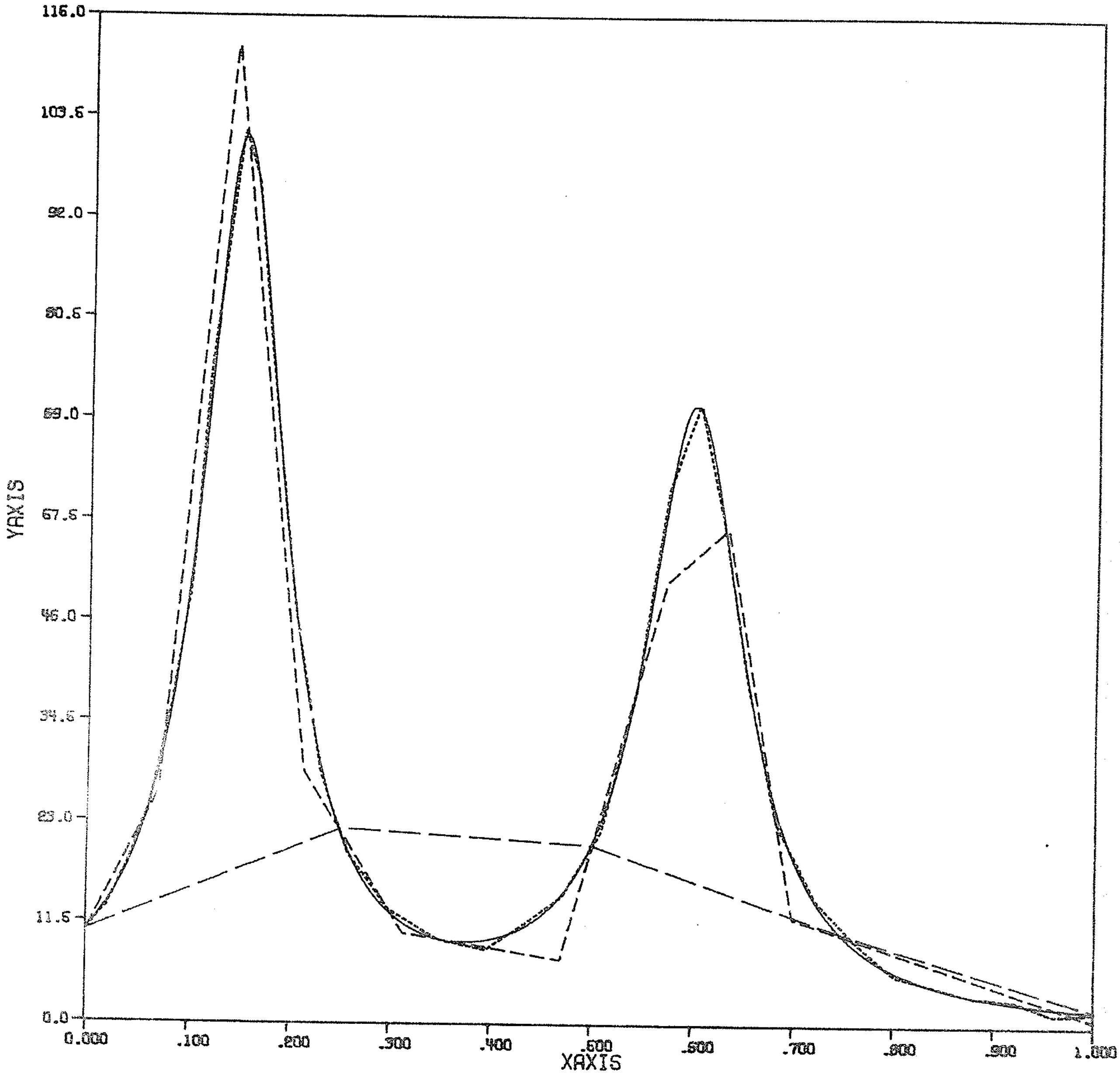


Figure 2

Function and LSE at the end of cycle zero, one and two

Curves representation

(1) Solid line (———) g(x)

(2) Big dashed line (— — — —) KMH(1) Est. at end of cycle zero

(3) Small dashed line (-----) KMH(1) Est. at end of cycle one

(4) Dotted line (······) KMH(1) Est. at end of cycle two.

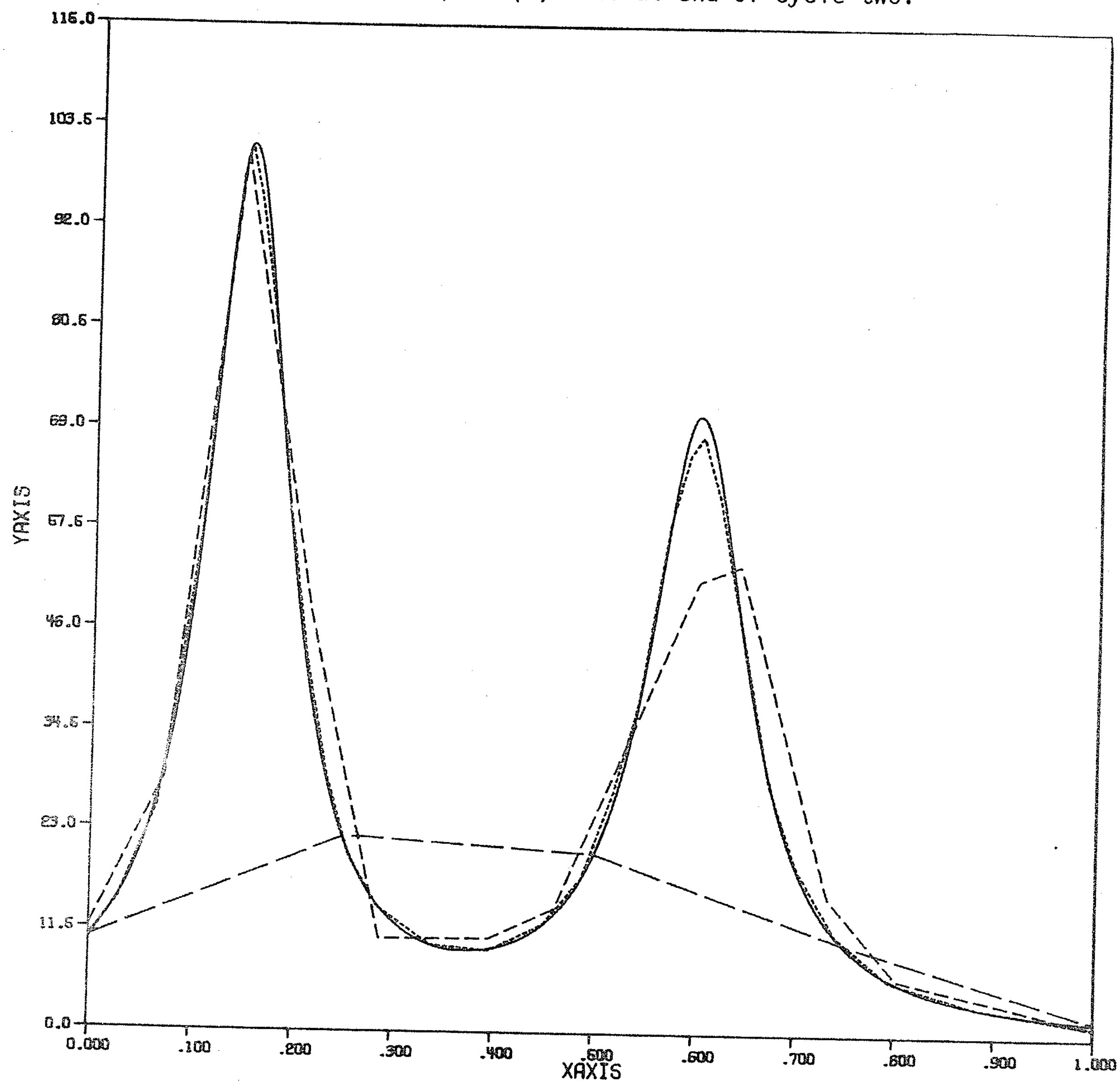


Figure 3

Function and KMH(1) estimate at the end of cycle zero, one and two

Curves representation

(1) Solid line (———) g(x)

(2) Big dashed line (— — — — —) KMH(2) Est. at the end of cycle zero

(3) Small dashed line (-------) KMH(2) Est. at the end of cycle one

(4) Dotted line (·······) KMH(2) Est. at the end of cycle two

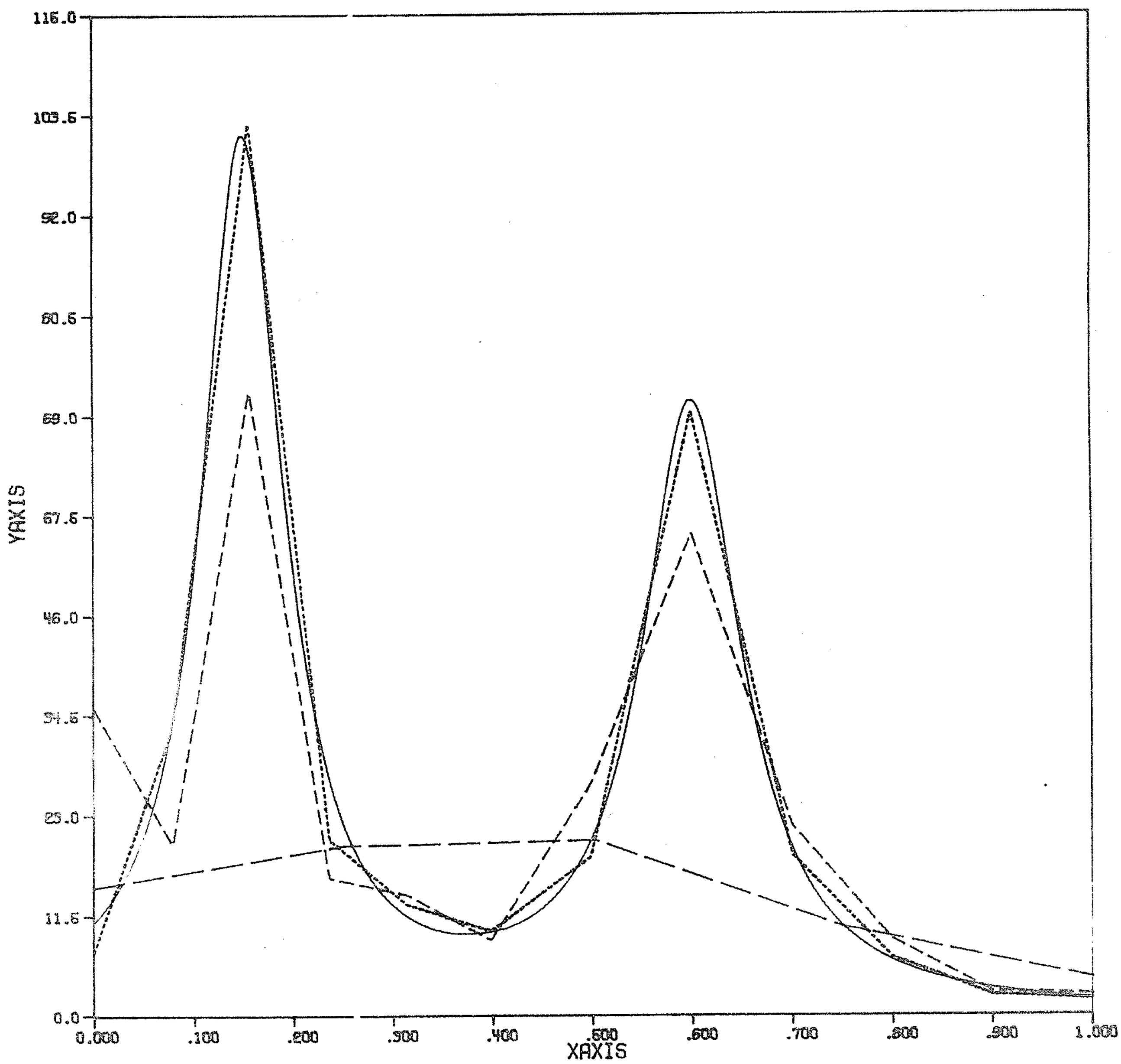


Figure 4

Function and KMH(2) estimate at the end of cycle zero, one and two

Curves representation

(1) Solid line (g(x))

(2) Dashed line (LSE)

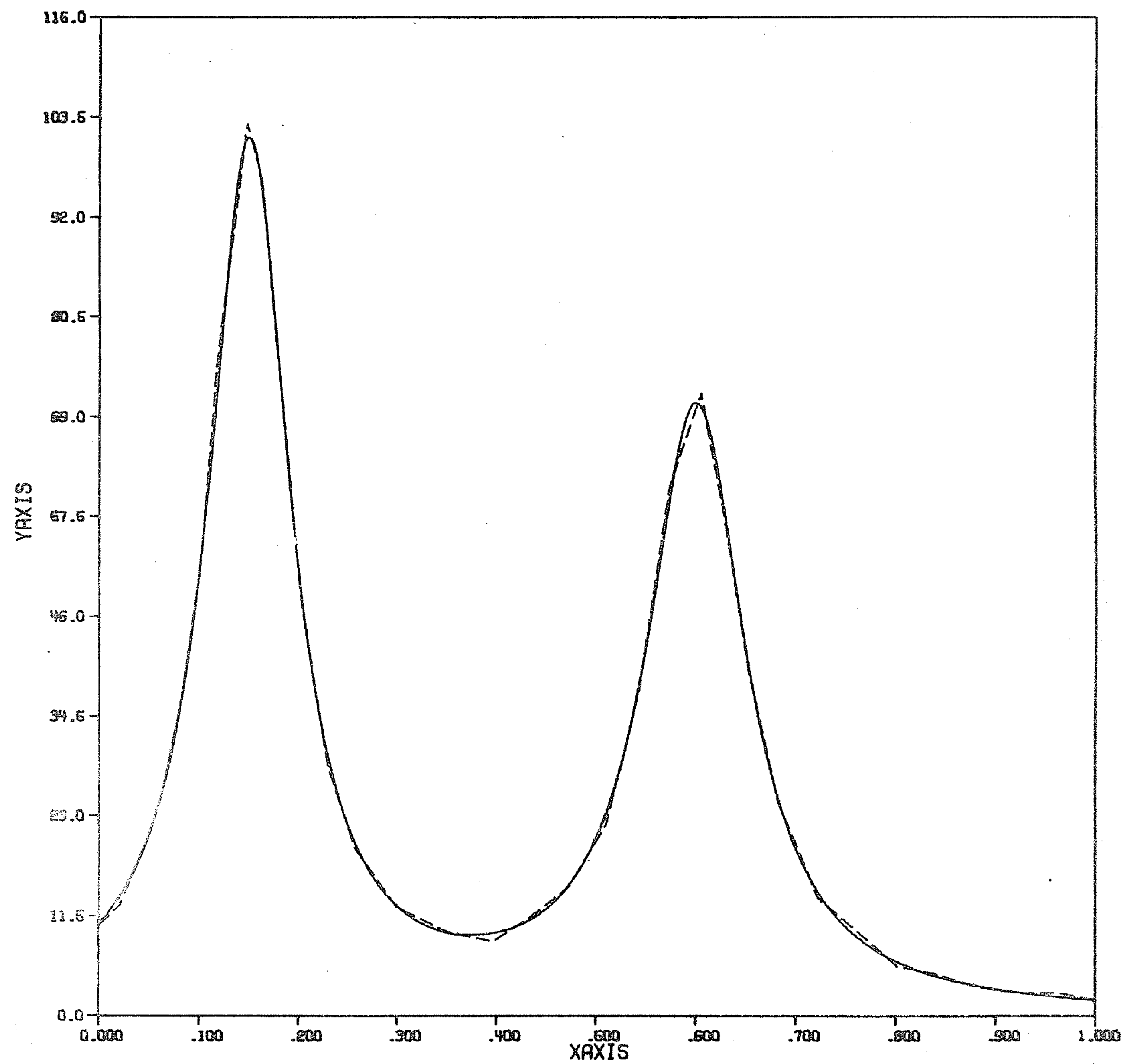


Figure 5

Function and LSE at the end of cycle three

Curve representation

(1) Solid line (g(x))

(2) Dashed line (KMH(1) Est.)

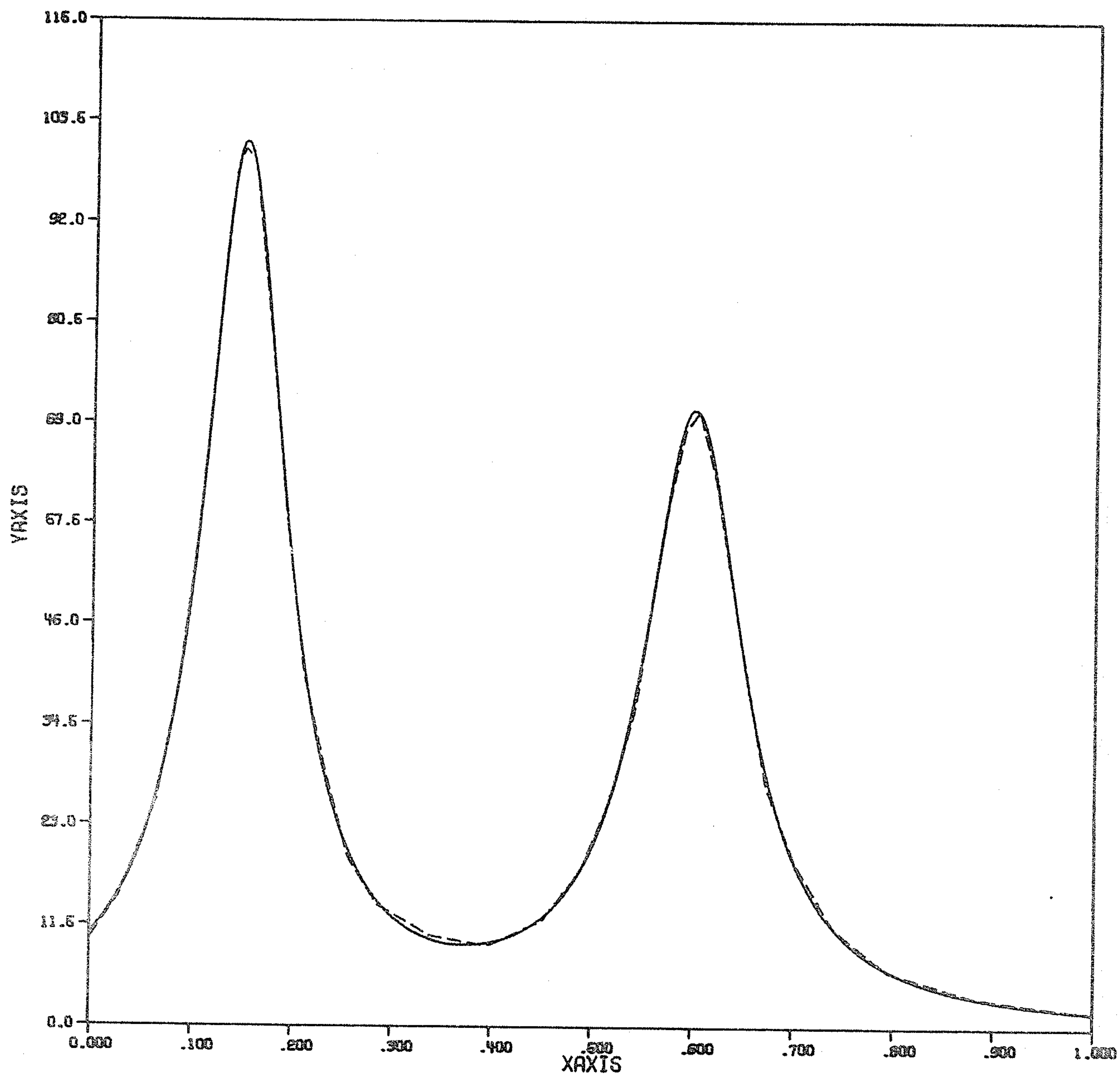


Figure 6

Function and KMH(1) estimate at the end of cycle three

Curve representation

(1) Solid line (g(x))

(2) Dashed line (KMH(2) Est.)

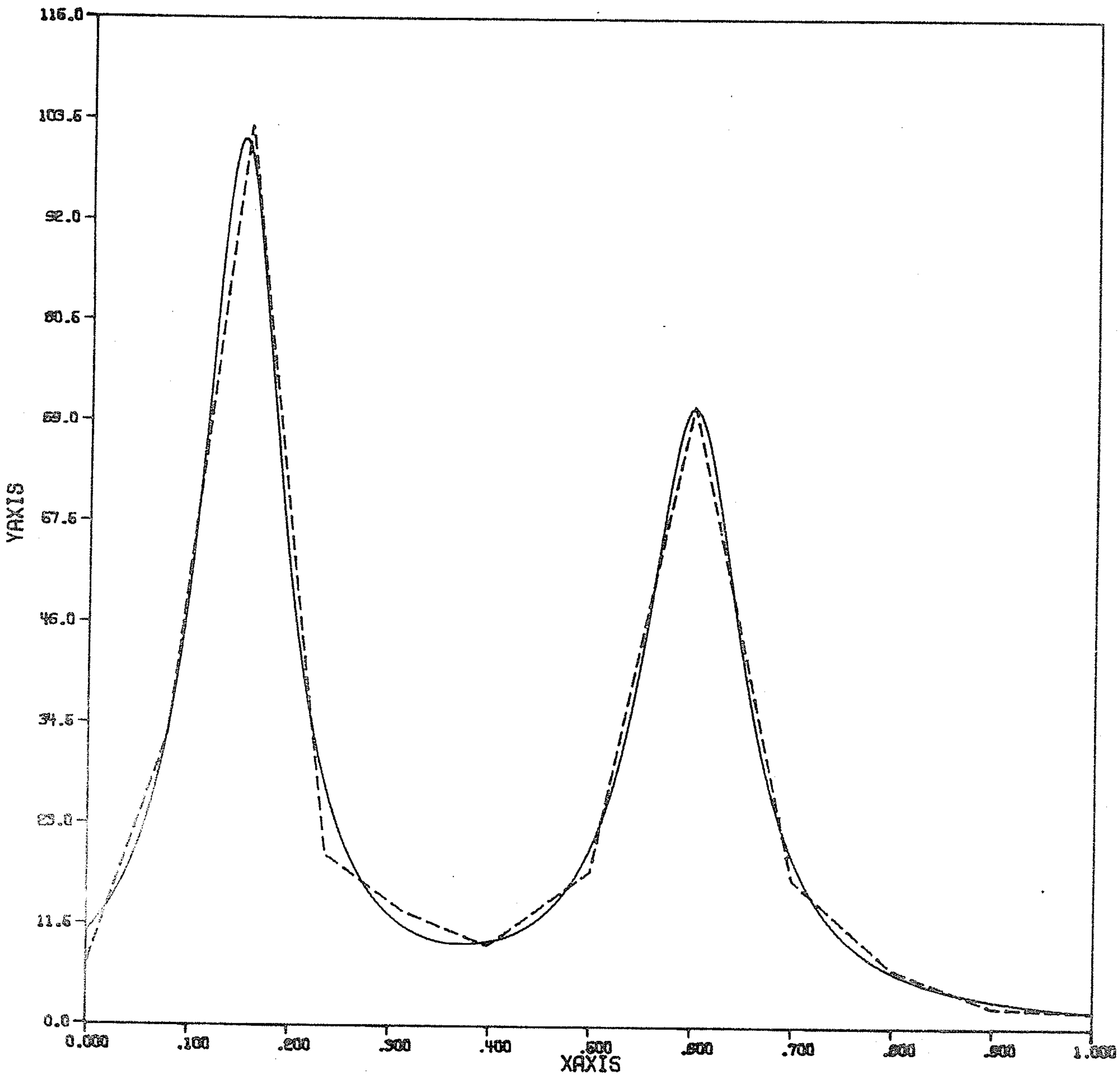


Figure 7

Function and KMH(2) estimate at the end of cycle three comparison of integrated variance, or integrated bias or integrated mean square error of the three estimates does not seem reasonable. For the same reason we did not compare these estimates with the estimate (so-called fourth estimate in the previous example) which was based on uniform design with equally spaced knots.

Table 7: IMSE for three estimates

Cycle Number	k	n	V	B	IMSE
			Least Square Estimate		
0	3	25	8.0173×10^{-2}	6.8436×10^{2}	6.8444×10^{2}
1	8	75	1.2075×10^{-1}	2.6718×10^{1}	2.6839×10^{1}
2	31	150	3.8867×10^{-1}	4.2081×10^{-1}	8.0948×10^{-1}
3	31	250	1.5688×10^{-1}	2.6611×10^{-1}	4.2299×10^{-1}
			KMH(1) Estimate		
0	3	25	8.0173×10^{-1}	6.8436×10^{2}	6.8444×10^{2}
1	11	75	8.1564×10^{-2}	3.5166×10^{1}	3.5248×10^{1}
2	30	150	1.6147×10^{-1}	6.1210×10^{-1}	7.7357×10^{-1}
3	30	250	1.0558×10^{-1}	6.6987×10^{-2}	1.7257×10^{-1}
			KMH(2) Estimate		
0	3	25	5.4090×10^{-2}	6.8037×10^{2}	6.8042×10^{2}
1	10	75	6.0871×10^{-2}	1.1663×10^{2}	1.1669×10^{2}
2	10	150	7.6050×10^{-2}	6.4454×10^{0}	6.5215×10^{0}
3	10	250	4.2817×10^{-2}	5.8441×10^{0}	5.8869×10^{0}

Remarks

(1) It is found that among the bias minimizing estimator, KMH(1) is better than KMH(2).

(2) The number of knots, displacement of knots, and allocation of observation depends upon the approximation to g". As stated earlier, there are many ways in which an approximation to g" may be obtained from a spline approximation to g. We have chosen the one which is simple to calculate and computationally inexpensive.

In the future, efforts could profitably be expended in the investigation of more complicated approaches, yielding perhaps the better estimate of p(x), h(x) and k. In addition, the stability of the method could be studied with the hope of showing some condition on the initial cycle which is sufficient to guarantee proper behavior of the algorithm.

BIBLIOGRAPHY

Aronszajn, N. (1950). Theory of reproducing kernels. Trans. Amer. Math. Soc., Vol. 68, 337-404.

de Boor, C. and Rice, J. R. (1968). Least squares cubic spline approximation I - Fixed knots. Purdue University Computer Science Technical Report, No. 20.

de Boor, C. and Rice, J. R. (1968). Least squares cubic spline approximation II - Variable knots. Purdue University Computer Science Technical Report, No. 21.

Box, G. E. P. and Draper, N. R. (1959). A basis for selection of a response surface design. JASA, Vol. 54, 622-654.

Box, G. E. P. and Draper, N. R. (1963). Choice of a second order rotable design. Biometrika, Vol. 50, 335-352.

Burchard, H. G. (1974). Splines (with optimal knots) are better. Applicable Analysis, Vol. 3, 309-319.

Dodson, D. S. (1972). Optimal order approximation by polynomial spline functions. Ph.D. Thesis, Purdue University.

Federov, V. V. (1972). Theory of Optimal Experiment. Translated by Studden, W. J. and Klimko, E. M., Academic Press, New York.

Greville, T. N. E. (1969). Theory and Application of Spline Functions. Academic Press, New York.

Hardy, G. H. (1943). A Course of Pure Mathematics. The Macmillan Company, New York.

Karlin, S. and Ziegler, Z. (1966). Chebyshevian spline functions. SIAM J. Numer. Anal., Vol. 3, No. 3, 514-543.

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

BIBLIOGRAPHY

Kershaw, D. (1970). Inequalities on the elements of the inverse of a certain tridiagonal matrix. Mathematics of Computation, Vol. 24, No. 109, 155-158.

Kiefer, J. (1973). Optimum Design for fitting biased multiresponse surfaces. Proceedings of the Third Symposium on Multivariate Analysis, Wright State University, Dayton, Ohio. Edited by P. R. Krishnaiah, 287-97.

Kiefer, J. and Wolfowitz, J. (1959). Optimum design in regression problems. Ann. Math. Statist., Vol. 30, 271-294.

Parzen, E. (1960). Regression analysis of continuous parameter time series, II. Stanford University Applied Mathematics and Statistical Laboratories Technical Report, No. 35.

Prenter, P. M. (1975). Splines and Variational Methods. John Wiley and Sons, New York.

Rice, J. R. (1969). On the degree of convergence of nonlinear spline approximation, in Approximation Theory with Special Emphasis on Spline Functions (I. J. Schoenberg, ed.) Academic Press.

Sacks, J. and Ylvisaker, D. (1966). Designs for regression problems with correlated errors, Ann. Math. Statist., Vol. 37, 66-89.

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

Schoenberg, I. J. (1969). Approximation with Special Emphasis on Spline Functions. Academic Press, New York.

Sewell, E. D. (1972). Automatic generation of triangulations for piecewise polynomial approximation. Ph.D. Thesis, Purdue University.

Studden, W. J. and VanArman, D. J. (1969). Admissible designs for polynomial spline regression. Ann. Math. Statist., Vol. 40, 1557-1569.

Taylor, A. E. (1958). Introduction to Functional Analysis. John Wiley, New York.

Wahba, G. (1971). On the regression design problem of Sacks and Ylvisaker. Ann. Math. Statist., Vol. 42, 1035-1053.

Waugh, F. V. (1950). Inversion of the Leontief matrix by power series. Econometrica, Vol. 18, 142-154.

Whittaker, Sir Edmund and Robinson, G. (1967). The Calculus of Observations. Dover Publication, New York.

APPENDICES

APPENDIX A: MATRIX NORM

To prove some asymptotic results in the section 3.1 for the linear spline case, we have used the matrix power series. For this purpose we need the concept of the matrix norm. If A is a mxm matrix and we take the norm on the space of n-tuples $x = (x_1,\ldots,x_n)$ to be the sup norm,

$$||x|| = \max_{1\leq i\leq n} |x_i|,$$

then the induced norm on the matrix A [Taylor (1958, Chapter III)] is the row-max norm,

$$||A|| = \sup_{x\neq 0} \frac{||Ax||}{||x||} = \max_i \sum_j |a_{ij}|. \tag{A1}$$

If $||A|| < 1$, we can write

$$(I-A)^{-1} = \sum_{k=0}^{\infty} A^k = I+A+A^2 +\ldots \tag{A2}$$

For more detail see Waugh (1950). It is easy to check that the matrix norm (A1) satisfies the two essential properties,

$$||A + D|| \leq ||A|| + ||D|| \tag{A3}$$

and

$$||AD|| \leq ||A|| \; ||D||. \tag{A4}$$

If $C = \sum_{n=k}^{\infty} A^n$ exists, then using (A3) and (A4) it is easy to show that

$$||C|| \leq ||A^k||/(1-||A||) \text{ for all } k \geq 1. \qquad \text{(A5)}$$

Lemma A1: Assume that the main diagonal of matrix A is dominant (i.e. $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$), then

$$||A^{-1}|| \leq \{\min_i [|a_{ii}| - \sum_{j \neq i} |a_{ij}|]\}^{-1}.$$

Proof: For a given x, choose k such that $||x|| = |x_k|$. Then

$$\begin{aligned} ||y|| = ||Ax|| &= \max_i |\sum_{j=1}^{n} a_{ij} x_j| \\ &\geq |\sum_{j=1}^{n} a_{kj} x_j| \\ &\geq \Big| |a_{kk} x_k| - |\sum_{j \neq k} a_{kj} x_j| \Big| \\ &= |a_{kk} x_k| - |\sum_{j \neq k} a_{kj} x_j| \\ &\geq \{|a_{kk}| - \sum_{j \neq k} |a_{kj}|\} ||x|| \\ &\geq \min_i \{|a_{ii}| - \sum_{j \neq i} |a_{ij}|\} ||x||. \end{aligned}$$

Since a matrix with the main diagonal elements dominant is non-singular, A^{-1} exists and $x = A^{-1}y$, we obtain the bound on $||A^{-1}||$:

$$\begin{aligned} ||A^{-1}|| &= \sup_{y \neq 0} \frac{||A^{-1}y||}{||y||} \\ &\leq \{\min_i [|a_{ii}| - \sum_{j \neq i} |a_{ij}|]\}^{-1}. \end{aligned}$$

Q.E.D.

In the following lemma, which is due to Kershaw (1970), the inequalities are obtained for the elements in the inverse of a tridiagonal matrix with the positive off-diagonal elements.

<u>Lemma A2</u>: Let the matrix is

$$A = \begin{bmatrix} \beta_1 & 1-\alpha_1 & 0 & \cdots & 0 & 0 \\ \alpha_2 & \beta_2 & 1-\alpha_2 & \cdots & 0 & 0 \\ 0 & \alpha_3 & \beta_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & & 1-\alpha_{n-1} \\ 0 & 0 & 0 & & \alpha_n & \beta_n \end{bmatrix}$$

where $0 < \alpha_r < 1$, $r = 1,2,\ldots,n$ and $\beta_r\beta_{r+1} > 1$, $r=1,\ldots,n-1$.

Let $A^{-1} = [a^{rs}]$. Then the following inequalities hold:

$$1 < \beta_s\, a^{ss} < \gamma_s/(\gamma_s-1),\ s = 1,2,\ldots,n$$

$$0 < (-1)^{r-s} a^{rs} \prod_{t=t_1}^{t_2} \beta_t < \gamma_s/(\gamma_s-1),\ r,s=1,2,\ldots,n,\ r\neq s$$

where $t_1 = \min(r,s)$, $t_2 = \max(r,s)$, and

$$\gamma_s = \min(\beta_{s-1}\beta_s, \beta_s\beta_{s+1}),\quad s = 2,\ldots,n-1,$$

with $\gamma_1 = \beta_1\beta_2$, $\gamma_n = \beta_{n-1}\beta_n$.

Proof: See Kershaw (1970).

<u>NOTE</u>: The conditions on α_1, α_n can be relaxed to $0 \leq \alpha_1 < 1$, $0 < \alpha_n \leq 1$, in which case $1 < a^{ss}\beta_s \leq \gamma_s/(\gamma_{s-1})$ for $s = 1,n$.

APPENDIX B: SOME BACKGROUND RESULTS

In this section we present some results which have been used in proving a theorem in section 3.2.

Consider a linear regression model in which one may observe a stochastic process Y having the form

$$Y(t) = \theta f(t) + Z(t), \tag{B1}$$

$t \in [0,1]$. θ is an unknown constant, $f(t)$ is a known function and Z is assumed to have a representation

$$Z(t) = \int_0^1 X(u)(t-u)_+^0 \, du \tag{B2}$$

where $(t-u)_+^0 = 1$ if $t \geq u$ and zero otherwise, and $\{X(t),\ t \in [0,1]\}$ is the Brownian motion process with the mean value function zero and the covariance kernel $K(s,t) = \min(s,t)$. It follows that $Z^{(1)} = X$ (in the quadratic mean) and

$$\begin{aligned} R(s,t) &= E(Z(s)Z(t)) \\ &= \int_0^1 \int_0^1 (s-u)_+^0 \, (t-v)_+^0 \, K(u,v)dudv. \qquad \text{(B3)} \\ &= \int_0^1 (s-u)_+(t-u)_+du. \end{aligned}$$

Let H(R) and H(K) denote the Reproducing Kernel Hilbert spaces (see Aronszajn (1950) for details) with the kernels R and K respectively. We denote the inner product in H(R) by the symbol $<\cdot,\cdot>_{H(R)}$ and similarly for H(K). Let $\mathscr{M}$ denote the L_2- space generated by the random variables $\{Z(t),\ t \in [0,1]\}$. Since $Z^{(1)}(t) = X(t)$ in the quadratic mean, $\mathscr{M}$ is also the L_2-space generated by $\{X(t),\ t \in [0,1]\}$. Since $R(s,t) = <R(\cdot,s),\ R(\cdot,t)>_{H(R)} = E(Z(s)Z(t))$ = inner product defined on $\mathscr{M}$, there exists a congruence (a one-one inner product preserving linear mapping) ψ from H(T) onto $\mathscr{M}$ satisfying

$$\psi(R(\cdot,t)) = Z(t). \tag{B4}$$

For the details see Parzen (1960). Every random variable U in $\mathscr{M}$ may be written as

$$U = \psi(g)$$

for some function g in H(R). We note that D: $H(R) \to H(K)$ is an isomorphism since $E(UDZ(\cdot)) = DE[UZ(\cdot)]$. Hence

$$||f||^2_{H(R)} = ||f'||^2_{H(K)}$$

for any function f in H(R), where $||\cdot||_{H(R)}$ and $||\cdot||_{H(K)}$ denote the norms in H(R) and H(K) respectively. Let us assume that

$$f(s) = \int_0^1 R(s,t)\rho(t)dt \tag{B5}$$

where ρ is a continuous function. It is easy to check that $f \in H(R)$. Let $T_n = \{t_0,t_1,\ldots,t_n\}$, $0 \le t_0 <\ldots< t_n \le 1$ and

$g_{T_n}(\cdot)$ is a function in $L(T_n)$, the linear space spanned by $\{R(\cdot,t), t \in T_n\}$, then $g_{T_n}(\cdot)$ has the representation $\sum_{j=0}^{n} c_j R(\cdot,t_j)$. Since $||f-g_{T_n}||^2_{H(R)} = ||(f-g_{T_n})'||^2_{H(K)}$, taking the derivative of f in (B5) and $R(\cdot,t)$ in (B3), we get

$$||f(s) - \sum_{j=0}^{n} c_j R(s,t_j)||^2_{H(R)}$$

$$= ||\int_0^1 (\frac{t^2}{2} - \frac{(t-s)_+^2}{2})\rho(t)dt - \sum_{j=0}^{n} c_j (\frac{t_j^2}{2} - \frac{(t_j-s)_+^2}{2})||^2_{H(K)}. \quad (B6)$$

Since the covariance kernel $K(s,t) = u(s)v(t)$, $s < t$ where $u(s)=s$ and $v(t) = 1$, the norm in $H(K)$ can be computed by a result given in Sacks and Ylvisaker (1966, page 86). Using this result, we get

$$||f - \sum_{j=0}^{n} c_j R(\cdot,t_j)||^2_{H(R)}$$

$$= \int_0^1 (\int_0^1 (t-s)_+ \rho(t)dt - \sum_{j=0}^{n} c_j (t_j-s)_+)^2 ds. \quad (B7)$$

The above relation holds for any real constants c_j's.

Note that the equality (B7) can also be obtained by using a result given in Wahba (1971).