13

Smoothing Splines

1. Introduction

In this section, we begin the study of nonparametric regression by way of smoothing splines. We wish to estimate the regression function f_o on a bounded interval, which we take to be $[\,0\,,\,1\,]$, from the data $y_{1,n}, \cdots, y_{n,n}$, following the model

$$y_{in} = f_o(x_{in}) + d_{in} \ , \quad i = 1, 2, \cdots, n \ . \tag{1.1}$$

Here, the x_{in} are design points (in this chapter, the design is deterministic) and $d_n = (d_{1,n}, d_{2,n}, \cdots, d_{n,n})^T$ is the random noise. Typical assumptions are that $d_{1,n}, d_{2,n}, \cdots, d_{n,n}$ are uncorrelated random variables, with mean 0 and common variance, i.e.,

$$\mathbb{E}[\,d_n\,] = 0 \quad , \quad \mathbb{E}[\,d_n d_n^T\,] = \sigma^2\, I \ , \tag{1.2}$$

where σ is typically unknown. We refer to this as the Gauss-Markov model, in view of the Gauss-Markov theorem for linear regression models. At times, we need the added condition that

$$\begin{aligned} &d_{1,n}, d_{2,n}, \cdots, d_{n,n} \quad \text{are iid and} \\ &\mathbb{E}[\,d_{1,n}\,] = 0 \quad , \quad \mathbb{E}[\,|\,d_{1,n}\,|^\kappa\,] < \infty \ , \end{aligned} \tag{1.3}$$

for some $\kappa > 2$. A typical choice is $\kappa = 4$.

A more restrictive but generally made assumption is that the d_{in} are iid normal random variables with mean 0 and again with the variance σ^2 usually unknown, described succinctly as

$$d_n \sim \text{Normal}(\,0\,, \sigma^2\, I\,) \ . \tag{1.4}$$

This is referred to as the Gaussian model.

Regarding the regression function, the typical *nonparametric* assumption is that f_o is smooth. In this volume, this usually takes the form

$$f_o \in W^{m,2}(0,1) \tag{1.5}$$

for some integer m, $m \geqslant 1$. Recall the definition of the Sobolev spaces $W^{m,p}(a,b)$ in (12.2.18). Assumptions of this kind are helpful when the

P.P.B. Eggermont, V.N. LaRiccia, *Maximum Penalized Likelihood Estimation*,
Springer Series in Statistics, DOI 10.1007/b12285_2,

data points are spaced close together, so that the changes in the function values $f_o(x_{in})$ for neighboring design points are small compared with the noise. The following exercise shows that in this situation one can do better than merely "connecting the dots". Of course, it does not say how much better.

(1.6) EXERCISE. Let $x_{in} = i/n$, $i = 1, 2, \cdots, n$, and let y_{in} satisfy (1.1), with the errors satisfying (1.2). Assume that f_o is twice continuously differentiable. For $i = 2, 3, \cdots, n-1$,
(a) show that

$$\tfrac{1}{4}\left\{ f_o(x_{i-1,n}) + 2\, f_o(x_{in}) + f_o(x_{i+1,n}) \right\} = f_o(x_{in}) + (1/n)^2\, f_o{}''(\theta_{in})$$

for some $\theta_{in} \in (\, x_{i-1,n}\,,\, x_{i+1,n}\,)$;
(b) compute the mean and the variance of

$$z_{in} = \tfrac{1}{4}\left\{\, y_{i-1,n} + 2\, y_{in} + y_{i+1,n} \,\right\} ;$$

(c) compare the mean squared errors

$$\mathbb{E}[\, |\, z_{in} - f_o(x_{in})\, |^2\,] \quad \text{and} \quad \mathbb{E}[\, |\, y_{in} - f_o(x_{in})\, |^2\,]$$

with each other (the case $n \to \infty$ is interesting).

In this chapter, we study the smoothing spline estimator of differential order m, the solution to

$$(1.7) \qquad \begin{array}{ll} \text{minimize} & \frac{1}{n} \sum\limits_{i=1}^{n} |\, f(x_{in}) - y_{in}\, |^2 + h^{2m}\, \|\, f^{(m)}\, \|^2 \\ \text{subject to} & f \in W^{m,2}(0,1)\ . \end{array}$$

(The factor $\frac{1}{n}$ appears for convenience; this way, the objective function is well-behaved as $n \to \infty$. The funny choice h^{2m} vs. h^2 or h is more convenient later on, although this is a matter of taste.) The solution is denoted by f^{nh}. The parameter h in (1.7) is the smoothing parameter. In this chapter, we only consider deterministic choices of h. Random (data-driven) choices are discussed in Chapter 18, and their effects on the smoothing spline estimator are discussed in Chapter 22.

The solution of (1.7) is a spline of polynomial order $2m$. In the literature, the case $m = 2$ is predominant, and the corresponding splines are called cubic splines. The traditional definition of splines is discussed in Chapter 19 together with the traditional computational details. The modern way to compute splines of arbitrary order is discussed in Chapter 20.

The following questions now pose themselves: Does the solution of (1.7) exist and is it unique (see § 3), and how accurate is the estimator (see § 4 and § 14.7)? To settle these questions, the reproducing kernel Hilbert space setting of the smoothing spline problem (1.7) is relevant, in which $W^{m,2}(0,1)$ is equipped with the inner products

$$(1.8) \qquad \langle\, f\,,\, g\, \rangle_{m,h} = \langle\, f\,,\, g\, \rangle + h^{2m}\, \langle\, f^{(m)}\,,\, g^{(m)}\, \rangle\ ,$$

where $\langle \cdot , \cdot \rangle$ is the usual $L^2(0,1)$ inner product. Then, $W^{m,2}(0,1)$ with the $\langle \cdot , \cdot \rangle_{m,h}$ inner product is a reproducing kernel Hilbert space with the reproducing kernel indexed by the smoothing parameter h. Denoting the reproducing kernel by $\mathcal{R}_{mh}(s, t)$, this then gives the reproducing kernel property

$$f(x) = \langle f , \mathcal{R}_{mh}(x , \cdot) \rangle_{m,h} , \quad x \in [0, 1] , \tag{1.9}$$

for all $f \in W^{m,2}(0,1)$ and all h, $0 < h \leqslant 1$.

The reproducing kernel shows up in various guises. For uniform designs and pure-noise data, the smoothing spline estimator is approximately the same as the solution ψ^{nh} of the semi-continuous version of the smoothing spline problem (1.7), viz.

$$\begin{array}{ll} \text{minimize} & \| f \|^2 - \frac{2}{n} \sum_{i=1}^{n} y_{in} f(x_{in}) + h^{2m} \| f^{(m)} \|^2 \\ \text{subject to} & f \in W^{m,2}(0,1) \ . \end{array} \tag{1.10}$$

The authors are tempted to call ψ^{nh} the C-spline estimator, C being short for “continuous”, even though ψ^{nh} is not a polynomial spline. The reproducing kernel now pops up in the form

$$\psi^{nh}(t) = \frac{1}{n} \sum_{i=1}^{n} y_{in} \mathcal{R}_{mh}(t , x_{in}) , \quad x \in (0,1) , \tag{1.11}$$

because $\mathcal{R}_{mh}(t,x)$ is the Green’s function for the Sturm-Liouville boundary value problem

$$\begin{array}{rcl} (-h^2)^m u^{(2m)} + u = w & , & t \in (0,1) , \\ u^{(\ell)}(0) = u^{(\ell)}(1) = 0 & , & m \leqslant \ell \leqslant 2m-1 \ . \end{array} \tag{1.12}$$

That is, the solution of (1.12) is given by

$$u(t) = \int_0^1 \mathcal{R}_{mh}(t , x) w(x) \, dx , \quad t \in [0, 1] \ . \tag{1.13}$$

With suitable modifications, this covers the case of point masses (1.11). In § 14.7, we show that the smoothing spline estimator is extremely well-approximated by

$$\varphi^{nh}(t) = \int_0^1 \mathcal{R}_{mh}(t,x) f_o(x) \, dx + \frac{1}{n} \sum_{i=1}^{n} d_{in} \mathcal{R}_{mh}(t , x_{in}) \tag{1.14}$$

for all $t \in [0, 1]$. In effect, this is the equivalent *reproducing* kernel approximation of smoothing splines, to be contrasted with the equivalent kernels of Silverman (1984). See also § 21.8. The reproducing kernel setup is discussed in § 2.

In § 5, we discuss the need for boundary corrections and their construction by way of the Bias Reduction Principle of Eubank and Speckman (1990b). In §§ 6–7, we discuss the boundary splines of Oehlert (1992), which avoids rather than corrects the problem. Finally, in § 9, we briefly discuss the estimation of derivatives of the regression function.

EXERCISE: (1.6).

2. Reproducing kernel Hilbert spaces

Here we begin the study of the smoothing spline estimator for the problem (1.1)–(1.2). Recall that the estimator is defined as the solution to

$$(2.1)\qquad \begin{array}{ll} \text{minimize} & L_{nh}(f) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} |\, f(x_{in}) - y_{in} \,|^2 + h^{2m} \,\|\, f^{(m)} \,\|^2 \\ \text{subject to} & f \in W^{m,2}(0,1) \ . \end{array}$$

Here, h is the smoothing parameter and m is the differential order of the smoothing spline. The solution of (2.1) is a spline of polynomial order $2m$ (or polynomial degree $2m-1$). At times, we just speak of the *order* of the spline, but the context should make clear which one is meant.

The design points are supposed to be (asymptotically) uniformly distributed in a sense to be defined precisely in Definition (2.22). For now, think of the equally spaced design $x_{in} = t_{in}$ with

$$(2.2)\qquad t_{in} = \frac{i-1}{n-1} \ , \quad i = 1, 2, \cdots, n \ .$$

The first question is of course whether the point evaluation functionals $f \longmapsto f(x_{in})$, $i = 1, 2, \cdots, n$, are well-defined. This has obvious implications for the existence and uniqueness of the solution of (2.1). Of course, if these point evaluation functionals are well-defined, then we are dealing with reproducing kernel Hilbert spaces. In Volume I, we avoided them more or less (more!) successfully, but see the KLONIAS (1982) treatment of the maximum penalized likelihood density estimator of GOOD and GASKINS (1971) in Exercise (5.2.64) in Volume I. For spline smoothing, the use of reproducing kernel Hilbert spaces will have far-reaching consequences.

The setting for the problem (2.1) is the space $W^{m,2}(0,1)$, which is a Hilbert space under the inner product

$$(2.3)\qquad \langle\, f \,,\, \varphi \,\rangle_{W^{m,2}(0,1)} = \langle\, f \,,\, \varphi \,\rangle + \langle\, f^{(m)} \,,\, \varphi^{(m)} \,\rangle$$

and associated norm

$$(2.4)\qquad \|\, f \,\|_{W^{m,2}(0,1)} = \{\, \|\, f \,\|^2 + \|\, f^{(m)} \,\|^2 \,\}^{1/2} \ .$$

Here, $\langle \cdot \,,\, \cdot \rangle$ denotes the usual $L^2(0,1)$ inner product. However, the norms

$$(2.5)\qquad \|\, f \,\|_{m,h} = \{\, \|\, f \,\|^2 + h^{2m} \,\|\, f^{(m)} \,\|^2 \,\}^{1/2}$$

and corresponding inner products

$$\langle f, \varphi \rangle_{m,h} = \langle f, \varphi \rangle + h^{2m} \langle f^{(m)}, \varphi^{(m)} \rangle \tag{2.6}$$

are useful as well. Note that, for each $h > 0$ the norms (2.4) and (2.5) are equivalent, but not uniformly in h. (The "equivalence" constants depend on h.) We remind the reader of the following definition.

(2.7) Definition. Two norms $\| \cdot \|_U$ and $\| \cdot \|_W$ on a vector space V are equivalent if there exists a constant $c > 0$ such that

$$c \| v \|_U \leqslant \| v \|_W \leqslant c^{-1} \| v \|_U \quad \text{for all } v \in V .$$

(2.8) Exercise. Show that the norms (2.4) and (2.5) are equivalent.

We are now in a position to answer the question of whether the $f(x_{in})$ are well-defined for $f \in W^{m,2}(0,1)$ in the sense that $| f(x_{in}) | \leqslant c \| f \|_{m,h}$ for a suitable constant. This amounts to showing that $W^{m,2}(0,1)$ is a reproducing kernel Hilbert space; see Appendix 3, § 7, in Volume I.

In what follows, it is useful to introduce an abbreviation of the L^2 norm of a function $f \in L^2(0,1)$ restricted to an interval $(a,b) \subset (0,1)$,

$$\| f \|_{(a,b)} \stackrel{\text{def}}{=} \left\{ \int_a^b | f(x) |^2 \, dx \right\}^{1/2} , \tag{2.9}$$

but please do not confuse $\| \cdot \|_{(a,b)}$ (with parentheses) with $\| \cdot \|_{m,h}$ (without them).

(2.10) Lemma. *There exists a constant* c_1 *such that, for all* $f \in W^{1,2}(0,1)$, *all* $0 < h \leqslant 1$, *and all* $x \in [0, 1]$,

$$| f(x) | \leqslant c_1 h^{-1/2} \| f \|_{1,h} .$$

Proof. The inequality

$$| f(x) - f(y) | = \left| \int_y^x f'(t) \, dt \right| \leqslant | x - y |^{1/2} \| f' \|$$

implies that every $f \in W^{m,2}(0,1)$ is (uniformly) continuous on $(0,1)$.

Consider an interval $[a, a+h] \subset [0, 1]$. An appeal to the Intermediate Value Theorem shows the existence of a $y \in (a, a+h)$ with

$$| f(y) | = h^{-1/2} \| f \|_{(a,a+h)} .$$

From the inequalities above, we get, for all $x \in (a, a+h)$,

$$\begin{aligned} | f(x) | &\leqslant | f(y) | + | f(x) - f(y) | \\ &\leqslant h^{-1/2} \| f \|_{(a,a+h)} + h^{1/2} \| f' \|_{(a,a+h)} \\ &\leqslant h^{-1/2} \| f \| + h^{1/2} \| f' \| , \end{aligned}$$

and thus, after some elementary manipulations

$$| f(x) | \leqslant c\, h^{-1/2} \left\{ \| f \|^2 + h^2 \| f' \|^2 \right\}^{1/2}$$

with $c = \sqrt{2}$. Q.e.d.

(2.11) Lemma [Continuity of point evaluations]. *Let $m \geqslant 1$ be an integer. There exists a constant c_m such that, for all $f \in W^{m,2}(0,1)$, all $0 < h \leqslant 1$, and all $x \in (0,1)$,*

$$| f(x) | \leqslant c_m\, h^{-1/2} \| f \|_{m,h} \ .$$

The proof goes by induction on m, as per the next two lemmas.

(2.12) Interpolation Lemma. *Let $m \geqslant 1$ be an integer. There exists a constant $c_m \geqslant 1$ such that, for all $f \in W^{m+1,2}(0,1)$ and all $0 < h \leqslant 1$,*

(a) $$\| f^{(m)} \| \leqslant c_m\, h^{-m} \| f \|_{m+1,h} \ ,$$

and, with $\theta = 1/(m+1)$,

(b) $$\| f^{(m)} \| \leqslant c_m \| f \|^{\theta} \left\{ \| f \|_{W^{m+1,2}(0,1)} \right\}^{1-\theta} \ .$$

Note that the inequality (b) of the lemma implies that

$$\| f \|_{W^{m,2}(0,1)} \leqslant \widetilde{c}_m \| f \|^{\theta} \| f \|^{1-\theta}_{W^{m+1,2}(0,1)}$$

for another constant $\widetilde{c}_m$. So ignoring this constant, after taking logarithms, the upper bound on $\log \| f \|_{W^{m,2}(0,1)}$ is obtained by linear interpolation on $\log \| f \|_{W^{x,2}(0,1)}$ between $x = 0$ and $x = m+1$, hence the name.

Proof. From (a) one obtains that

$$\| f^{(m)} \| \leqslant c_m\, h^{-m} \| f \| + c_m\, h \| f \|_{W^{m+1}(0,1)} \ .$$

Now, take h such that $h^{m+1} = \| f \| \,/\, \| f \|_{W^{m+1,2}(0,1)}$ and (b) follows, for a possible larger constant c_m . (Note that indeed $h \leqslant 1$.)

The case $m = 1$ of the lemma is covered by the main inequality in the proof of Lemma (5.4.16) in Volume I. The proof now proceeds by induction.

Let $m \geqslant 1$. Suppose that the lemma holds for all integers up to and including m. Let $f \in W^{m+2,2}(0,1)$. Applying the inequality (a) with $m = 1$ to the function $f^{(m)}$ gives

(2.13) $$\| f^{(m+1)} \| \leqslant c_1\, h^{-1} \left(\| f^{(m)} \| + h^2 \| f^{(m+2)} \| \right) \ .$$

Now, apply the inequality (b) of the lemma for m , so

(2.14) $$\begin{aligned} \| f^{(m)} \| &\leqslant c_m \| f \|^{\,\theta} \left\{ \| f \|_{W^{m+1,2}(0,1)} \right\}^{1-\theta} \\ &\leqslant c_m \| f \| + c_m \| f \|^{\,\theta} \| f^{(m+1)} \|^{1-\theta} \ , \end{aligned}$$

since $(x+y)^\alpha \leqslant x^\alpha + y^\alpha$ for all positive x and y and $0 < \alpha \leqslant 1$. Substituting this into (2.13) gives, for suitable constants $\widetilde{c}_m$ and $\widetilde{c}_1$,

$$\text{(2.15)} \quad \| f^{(m+1)} \| \leqslant \widetilde{c}_m \, h^{-1} \| f \|^{\theta} \| f^{(m+1)} \|^{1-\theta} + \widetilde{c}_1 \, h^{-1} \left(\| f \| + h^2 \| f^{(m+2)} \| \right) .$$

Since $h \leqslant 1$, then $h^{-1} \leqslant h^{-m-1}$, so that

$$\begin{aligned} h^{-1} \left(\| f \| + h^2 \| f^{(m+2)} \| \right) &\leqslant h^{-m-1} \| f \| + h \| f^{(m+2)} \| \\ &\leqslant h^{-m-1} \| f \|_{m+2,h} . \end{aligned}$$

Substituting this into (2.15) gives

$$\text{(2.16)} \quad \| f^{(m+1)} \| \leqslant \widetilde{c}_m \, h^{-1} \| f \|^{\theta} \| f^{(m+1)} \|^{1-\theta} + \widetilde{c}_1 \, h^{-m-1} \| f \|_{m+2,h} .$$

This is an inequality of the form $x^p \leqslant a\,x + b$ with $p > 1$, which implies that $x^p \leqslant a^q + q\,b$, where $1/q = 1 - (1/p)$. See Exercise (4.10). This gives

$$\| f^{(m+1)} \| \leqslant (\widetilde{c}_m \, h^{-1})^{m+1} \| f \| + (m+1)\, \widetilde{c}_1 \, h^{-m-1} \| f \|_{m+1,h} .$$

This implies the inequality (a) for $m+1$. Q.e.d.

(2.17) Lemma. *Let $m \geqslant 1$ be an integer. There exists a constant k_m such that, for all $f \in W^{m+1,2}(0,1)$ and all $0 < h < 1$,*

$$\| f \|_{m,h} \leqslant k_m \| f \|_{m+1,h} .$$

Proof. Lemma (2.12) says that $h^m \| f^{(m)} \| \leqslant c_m \| f \|_{m+1,h}$. Now, squaring both sides and then adding $\| f \|^2$ gives the lemma, with $k_m^2 = 1 + c_m^2$. Q.e.d.

We now put all of the above together to show that the smoothing spline problem is "well-behaved" from various points of view.

Reproducing kernel Hilbert spaces. Lemma (2.11) shows that, for fixed $x \in [\,0\,,\,1\,]$, the linear functional $\ell(f) = f(x)$ is bounded on $W^{m,2}(0,1)$. Thus, the vector space $W^{m,2}(0,1)$ with the inner product (2.6) is a reproducing kernel Hilbert space and, for each $x \in (0,1)$, there exists an $\mathcal{R}_{m,h,x} \in W^{m,2}(0,1)$ such that, for all $f \in W^{m,2}(0,1)$,

$$f(x) = \left\langle \mathcal{R}_{m,h,x} \,,\, f \right\rangle_{m,h} .$$

It is customary to denote $\mathcal{R}_{m,h,x}(y)$ by $\mathcal{R}_{mh}(x,y)$. Applying the above to the function $f = \mathcal{R}_{mh}(y, \cdot\,)$, where $y \in [\,0\,,\,1\,]$ is fixed, then gives

$$\mathcal{R}_{mh}(y,x) = \left\langle \mathcal{R}_{mh}(x, \cdot\,) \,,\, \mathcal{R}_{mh}(y, \cdot\,) \right\rangle_{m,h} \quad \text{for all} \quad x \in [\,0\,,\,1\,] ,$$

whence $\mathcal{R}_{mh}(x,y) = \mathcal{R}_{mh}(y,x)$. Moreover, Lemma (2.11) implies that

$$\| \mathcal{R}_{mh}(x, \cdot\,) \|_{m,h}^2 = \mathcal{R}_{mh}(x,x) \leqslant c_m \, h^{-1/2} \| \mathcal{R}_{mh}(x, \cdot\,) \|_{m,h} ,$$

and the obvious conclusion may be drawn. We summarize this in a lemma.

(2.18) Reproducing Kernel Lemma. *Let $m \geqslant 1$ be an integer, and let $0 < h \leqslant 1$. Then $W^{m,2}(0,1)$ with the inner product $\langle \cdot , \cdot \rangle_{m,h}$ is a reproducing kernel Hilbert space, with kernel $\mathcal{R}_{mh}(x,y)$, such that, for all $f \in W^{m,2}(0,1)$ and all x,*

$$f(x) = \bigl\langle \mathcal{R}_{mh}(x, \cdot\,) \,,\, f \,\bigr\rangle_{m,h} \quad \textit{for all} \quad x \in [\,0\,,\,1\,]\ .$$

Moreover, there exists a c_m such that, for all $0 < h \leqslant 1$, and all x,

$$\| \, \mathcal{R}_{mh}(x, \cdot\,) \, \|_{m,h} \leqslant c_m \, h^{-1/2}\ .$$

Random sums. The reproducing kernel Hilbert space framework bears fruit in the consideration of the random sums

$$\tfrac{1}{n} \sum_{i=1}^{n} d_{in} \, f(x_{in})\ ,$$

where $f \in W^{m,2}(0,1)$ is random, i.e., is allowed to depend on the noise vector $d_n = (\, d_{1,n}, d_{2,n}, \cdots , d_{n,n}\,)$. In contrast, define the "simple" random sums

$$\mathfrak{S}^{nh}(x) = \tfrac{1}{n} \sum_{i=1}^{n} d_{in} \, \mathcal{R}_{mh}(x_{in}, \cdot\,)\ , \tag{2.19}$$

where the randomness of the functions f is traded for the dependence on a smoothing parameter.

(2.20) Random Sum Lemma. *Let $m \geqslant 1$. For all $f \in W^{m,2}(0,1)$ and all noise vectors $d_n = (\, d_{1,n}, d_{2,n}, \cdots , d_{n,n}\,)$,*

$$\Bigl| \, \tfrac{1}{n} \sum_{i=1}^{n} d_{in} \, f(x_{in}) \, \Bigr| \leqslant \| \, f \, \|_{m,h} \; \| \, \mathfrak{S}^{nh} \, \|_{m,h}\ .$$

Moreover, if d_n satisfies (1.2), then there exists a constant c such that

$$\mathbb{E}[\, \| \, \mathfrak{S}^{nh} \, \|_{m,h}^2 \,] \leqslant c \, (nh)^{-1}$$

for all h, $0 < h \leqslant 1$, and all designs.

Proof. Since $f \in W^{m,2}(0,1)$, the reproducing kernel Hilbert space trick of Lemma (2.18) gives

$$f(x_{in}) = \bigl\langle \mathcal{R}_{mh}(x_{in}, \cdot\,) \,,\, f \,\bigr\rangle_{m.h}\ ,$$

and consequently

$$\tfrac{1}{n} \sum_{i=1}^{n} d_{in} \, f(x_{in}) = \bigl\langle \, \mathfrak{S}^{nh} \,,\, f \,\bigr\rangle_{m,h}\ ,$$

which gives the upper bound

$$\tfrac{1}{n} \sum_{i=1}^{n} d_{in} \, f(x_{in}) \leqslant \| \, f \, \|_{m,h} \, \| \, \mathfrak{S}^{nh} \, \|_{m,h}\ .$$

Note that all of this holds whether f is random or deterministic.

Now, one verifies that

$$\| \mathfrak{S}^{nh} \|^2 = n^{-2} \sum_{i,j=1}^{n} d_{in}\, d_{jn} \left\langle \mathcal{R}_{mh}(x_{in}, \,\cdot\,) \,,\, \mathcal{R}_{mh}(x_{jn}, \,\cdot\,) \right\rangle_{m,h} \,,$$

and so, under the assumption (1.2),

$$\mathbb{E}\Big[\| \mathfrak{S}^{nh} \|^2 \Big] = \sigma^2\, n^{-2} \sum_{i=1}^{n} \| \mathcal{R}_{mh}(x_{in}, \,\cdot\,) \|_{m,h}^2 \,.$$

The bound from the previous lemma on $\| \mathcal{R}_{mh}(x_{in}, \,\cdot\,) \|_{m,h}^2$ now completes the proof. Q.e.d.

(2.21) Exercise. Show that, under the assumptions of Lemma (2.20),

$$\sup \left\{ \frac{\left| \frac{1}{n} \sum_{i=1}^{n} d_{in}\, f(x_{in}) \right|}{\| f \|_{m,h}} \;\middle|\; \begin{array}{l} f \in W^{m,2}(0,1) \\ f \not\equiv 0 \end{array} \right\} = \| \mathfrak{S}^{nh} \|_{m,h} \,.$$

In other words, the supremum is attained by the solution of the pure-noise version of (1.10); i.e., with $y_{in} = d_{in}$ for all i.

Quadrature. The reproducing kernel Hilbert spaces setup of Lemma (2.18) shows that the linear functionals $\ell_{i,n}(f) = f(x_{in})$, $i = 1, 2, \cdots, n$, are continuous on $W^{m,2}(0,1)$ for $m \geqslant 1$. So the problem (2.1) starts to make sense. Along the same lines, we need to be able to compare

$$\frac{1}{n} \sum_{i=1}^{n} |\, f(x_{in}) \,|^2 \quad \text{and} \quad \| f \|^2$$

with each other, at least for $f \in W^{m,2}(0,1)$. In effect, this is a requirement on the design, and is a quadrature result for specific designs.

(2.22) Definition. We say that the design x_{in}, $i = 1, 2, \cdots, n$, is asymptotically uniform if there exists a constant c such that, for all $n \geqslant 2$ and all $f \in W^{1,1}(0,1)$,

$$\left| \frac{1}{n} \sum_{i=1}^{n} f(x_{in}) - \int_0^1 f(\,t\,)\, dt \right| \leqslant c\, n^{-1} \, \| f' \|_{L^1(0,1)} \,.$$

(2.23) Remark. The rate n^{-1} could be lowered to $\left(n^{-1} \log n \right)^{1/2}$ but seems to cover most cases of interest. Random designs require their own treatment; see Chapter 21.

(2.24) Lemma. *The design* (2.2) *is asymptotically uniform. In fact, for every* $f \in W^{1,1}(0,1)$,

$$\Big| \tfrac{1}{n} \sum_{i=1}^{n} f(t_{in}) - \int_0^1 f(t)\,dt \Big| \leqslant \tfrac{1}{n-1} \| f' \|_1 \ .$$

Proof. The first step is the identity,

$$\tfrac{1}{n} \sum_{i=1}^{n} c_{in} = \tfrac{1}{n-1} \sum_{i=1}^{n-1} \{ a_{in} c_{in} + b_{in} c_{i+1,n} \} \ , \tag{2.25}$$

for all c_{in}, $i = 1, 2, \cdots, n$, where $a_{in} = (n-i)/n$, $b_{in} = i/n$. Of course, we take $c_{in} = f(t_{in})$. Then, with the intervals $\omega_{in} = (t_{in}, t_{i+1,n})$,

$$\tfrac{1}{n-1} \{ a_{in} f(t_{in}) + b_{in} f(t_{i+1,n}) \} - \int_{\omega_{in}} f(t)\,dt =$$

$$a_{in} \int_{\omega_{in}} \{ f(t_{in}) - f(t) \}\,dt + b_{in} \int_{\omega_{in}} \{ f(t_{i+1,n}) - f(t) \}\,dt \ .$$

Now, for $t \in \omega_{in}$,

$$| f(t_{in}) - f(t) | = \Big| \int_t^{t_{in}} f'(s)\,ds \Big| \leqslant \int_{\omega_{in}} | f'(s) |\,ds \ ,$$

so, after integration over ω_{in}, an interval of length $1/(n-1)$,

$$\int_{\omega_{in}} | f(t_{in}) - f(t) |\,dt \leqslant \tfrac{1}{n-1} \int_{\omega_{in}} | f'(t) |\,dt \ .$$

The same bound applies to $\displaystyle\int_{\omega_{in}} | f(t_{i+1,n}) - f(t) |\,dt$. Then, adding these bounds gives

$$\Big| \tfrac{1}{n-1} \{ a_{in} f(t_{in}) + b_{in} f(t_{i+1,n}) \} - \int_{\omega_{in}} f(t)\,dt \Big| \leqslant \tfrac{1}{n-1} \int_{\omega_{in}} | f'(t) |\,dt \ ,$$

and then adding *these* over $i = 1, 2, \cdots, n-1$, together with the triangle inequality, gives the required result. Q.e.d.

(2.26) Exercise. Show that the design $t_{in} = i/(n+1)$, $i = 1, 2, \cdots, n$, is asymptotically uniform and likewise for $t_{in} = (i - \frac{1}{2})/n$.

(2.27) Quadrature Lemma. *Let* $m \geqslant 1$. *Assuming the design is asymptotically uniform, there exists a constant* c_m *such that, for all* $f \in W^{m,2}(0,1)$, *all* $n \geqslant 2$, *and all* h, $0 < h \leqslant \frac{1}{2}$,

$$\Big| \tfrac{1}{n} \sum_{i=1}^{n} | f(x_{in}) |^2 - \| f \|^2 \Big| \leqslant c_m (nh)^{-1} \| f \|_{m,h}^2 \ .$$

Proof. Let $m \geqslant 1$. As a preliminary remark, note that, for $f \in W^{m,2}(0,1)$, we have of course that $\| f^2 \|_1 = \| f \|^2$ and that

$$\begin{aligned} \| (f^2)' \|_1 &= 2 \| f f' \|_1 \leqslant 2 \| f \| \| f' \| = 2 h^{-1} \| f \| \{ h \| f' \| \} \\ &\leqslant h^{-1} \{ \| f \|^2 + h^{-1} \| f' \|^2 \} = h^{-1} \| f \|_{1,h}^2 , \end{aligned}$$

where we used Cauchy-Schwarz and the inequality $2ab \leqslant a^2 + b^2$.

Then, for $n \geqslant 2$, by the asymptotic uniformity of the design,

$$\Bigl| \tfrac{1}{n} \sum_{i=1}^{n} | f(x_{in}) |^2 - \| f \|^2 \Bigr| \leqslant c n^{-1} \| (f^2)' \|_1 , \tag{2.28}$$

which by the above, may be further bounded by

$$c (nh)^{-1} \| f \|_{1,h}^2 \leqslant \widetilde{c} (nh)^{-1} \| f \|_{m,h}^2$$

for an appropriate constant $\widetilde{c}$, the last inequality by Lemma (2.17). This is the lemma. Q.e.d.

The following is an interesting and useful exercise on the multiplication of functions in $W^{m,2}(0,1)$.

(2.29) Exercise. (a) Show that there exists a constant c_m such that, for all h, $0 < h \leqslant \frac{1}{2}$,

$$\| f g \|_{m,h} \leqslant c_m h^{-1/2} \| f \|_{m,h} \| g \|_{m,h} \quad \text{for all} \quad f, g \in W^{m,2}(0,1) .$$

(b) Show that the factor $h^{-1/2}$ is sharp for $h \to 0$.

Exercises: (2.8), (2.21), (2.26), (2.29).

3. Existence and uniqueness of the smoothing spline

In this section, we discuss the existence and uniqueness of the solution of the smoothing spline problem. Of course, the quadratic nature of the problem makes life very easy, and it is useful to consider that first. Note that in Lemma (3.1) below there are no constraints on the design. We emphasize that, throughout this section, the sample size n and the smoothing parameter h remain fixed.

(3.1) Quadratic Behavior Lemma. *Let $m \geqslant 1$, and let φ be a solution of (2.1). Then, for all $f \in W^{m,2}(0,1)$,*

$$\tfrac{1}{n} \sum_{i=1}^{n} | f(x_{in}) - \varphi(x_{in}) |^2 + h^{2m} \| (f - \varphi)^{(m)} \|^2 =$$

$$\tfrac{1}{n} \sum_{i=1}^{n} \bigl(f(x_{in}) - y_{in} \bigr) \bigl(f(x_{in}) - \varphi(x_{in}) \bigr) + h^{2m} \langle f^{(m)} , (f - \varphi)^{(m)} \rangle .$$

PROOF. Since $L_{nh}(f)$ is quadratic, it is convex, and thus, see, e.g., Chapter 10 in Volume I or Chapter 3 in TROUTMAN (1983), it has a Gateaux variation (directional derivative) at each $\varphi \in W^{m,2}(0,1)$. One verifies that it is given by

$$(3.2) \quad \delta L_{nh}(\varphi\,,\, f-\varphi) = 2\,h^{2m} \left\langle\, \varphi^{(m)}\,,\, (f-\varphi)^{(m)} \,\right\rangle \;+ \\ \tfrac{2}{n} \sum_{i=1}^{n} \left(\varphi(x_{in}) - y_{in} \right) \left(f(x_{in}) - \varphi(x_{in}) \right) \,,$$

so that

$$(3.3) \quad L_{nh}(f) - L_{nh}(\varphi) - \delta L_{nh}(\varphi\,,\, f-\varphi) = \\ h^{2m} \,\|\, (f-\varphi)^{(m)} \,\|^2 + \tfrac{1}{n} \sum_{i=1}^{n} |\, f(x_{in}) - \varphi(x_{in}) \,|^2 \;.$$

In fact, this last result is just an identity for quadratic functionals. Now, by the necessary and sufficient conditions for a minimum, see, e.g., Theorem (10.2.2) in Volume I or Proposition (3.3) in TROUTMAN (1983), the function φ solves the problem (2.1) if and only if

$$(3.4) \qquad \delta L_{nh}(\varphi\,,\, f-\varphi) = 0 \quad \text{for all} \quad f \in W^{m,2}(0,1) \;.$$

Then, the identity (3.3) simplifies to

$$(3.5) \;\; \tfrac{1}{n} \sum_{i=1}^{n} |\, f(x_{in}) - \varphi(x_{in}) \,|^2 + h^{2m} \,\|\, (f-\varphi)^{(m)} \,\|^2 = L_{nh}(f) - L_{nh}(\varphi) \;.$$

Now, in (3.3), interchange f and φ to obtain

$$L_{nh}(f) - L_{nh}(\varphi) = -\tfrac{1}{n} \sum_{i=1}^{n} |\, f(x_{in}) - \varphi(x_{in}) \,|^2 - h^{2m} \,\|\, \bigl(\, f-\varphi \,\bigr)^{(m)} \,\|^2 + \\ \tfrac{2}{n} \sum_{i=1}^{n} \bigl(\, f(x_{in}) - y_{in} \,\bigr) \bigl(\, f(x_{in}) - \varphi(x_{in}) \,\bigr) + \\ 2\,h^{2m} \left\langle\, f^{(m)}\,,\, \bigl(\, f-\varphi \,\bigr)^{(m)} \,\right\rangle \;.$$

Finally, substitute this into (3.5), move the negative quadratics to the left of the equality, and divide by 2. This gives the lemma. Q.e.d.

(3.6) UNIQUENESS LEMMA. *Let $m \geqslant 1$, and suppose that the design contains at least m distinct points. Then the solution of* (2.1) *is unique.*

PROOF. Suppose φ and ψ are solutions of (2.1). Since $L_{nh}(\varphi) = L_{nh}(\psi)$, then, by (3.5),

$$\tfrac{1}{n} \sum_{i=1}^{n} |\, \varphi(x_{in}) - \psi(x_{in}) \,|^2 + h^{2m} \|\, \bigl(\, \varphi - \psi \,\bigr)^{(m)} \,\|^2 = 0 \;.$$

It follows that $\bigl(\, \varphi - \psi \,\bigr)^{(m)} = 0$ almost everywhere, and so $\varphi - \psi$ is a polynomial of degree $\leqslant m-1$. And of course

$$\varphi(x_{in}) - \psi(x_{in}) = 0 \;, \quad i = 1, 2, \cdots, n.$$

Now, if there are (at least) m distinct design points, then this says that the polynomial $\varphi - \psi$ has at least m distinct zeros. Since it has degree $\leqslant m-1$, the polynomial vanishes everywhere. In other words, $\varphi = \psi$ everywhere. Q.e.d.

(3.7) Existence Lemma. *Let $m \geqslant 1$. For any design, the smoothing spline problem* (2.1) *has a solution.*

Proof. Note that the functional L_{nh} is bounded from below (by 0), and so its infimum over $W^{m,2}(0,1)$ is finite. Let $\{ f_k \}_k$ be a minimizing sequence. Then, using Taylor's theorem with exact remainder, write

$$f_k(x) = p_k(x) + [\, T f_k^{(m)} \,](x) \ , \tag{3.8}$$

where p_k is a polynomial of order m, and for $g \in L^2(0,1)$,

$$Tg(x) = \int_0^x \frac{(x-t)^{m-1}}{(m-1)\,!}\, g(t)\, dt \ . \tag{3.9}$$

Note that the Arzelà-Ascoli theorem implies the compactness of the operator $T : L^2(0,1) \longrightarrow C[\,0\,,\,1\,]$.

Now, since without loss of generality $L_{nh}(\, f_k \,) \leqslant L_{nh}(\, f_1 \,)$, it follows that

$$\| \, f_k^{(m)} \, \|^2 \leqslant h^{-2m} \, L_{nh}(\, f_1 \,)$$

and so $\{ \, f_k^{(m)} \, \}_k$ is a bounded sequence in $L^2(0,1)$. Thus, it has a weakly convergent subsequence, which we denote again by $\{ \, f_k^{(m)} \, \}_k$, with weak limit denoted by φ_o. Then, by the weak lower semi-continuity of the norm,

$$\lim_{k\to\infty} \| \, f_k^{(m)} \, \|^2 \geqslant \| \, \varphi_o \, \|^2 \ . \tag{3.10}$$

Moreover, since T is compact, it maps weakly convergent sequences into strongly convergent ones. In other words,

$$\lim_{k\to\infty} \| \, T f_k^{(m)} - T\varphi_o \, \|_\infty = 0 \ . \tag{3.11}$$

Now, consider the restrictions of the f_k to the design points,

$$r_n f_k \stackrel{\text{def}}{=} \big(\, f_k(x_{1,n}), \, f_k(x_{2,n}), \, \cdots \, , \, f_k(x_{n,n}) \big) \ , \quad k = 1, 2, \cdots \ .$$

We may extract a subsequence from $\{ \, f_k \, \}_k$ for which the corresponding sequence $\{ \, r_n f_k \, \}_k$ converges in $\mathbb{R}^n$ to some vector v_o. Then it is easy to see that, for the corresponding polynomials,

$$\lim_{k\to\infty} p_k(x_{in}) = [\, v_o \,]_i - T\varphi_o(x_{in}) \ , \quad i = 1, 2, \cdots, n \ .$$

All that there is left to do is claim that there exists a polynomial p_o of order m such that

$$p_o(x_{in}) = [\, v_o \,]_i - T\varphi_o(x_{in}) \ , \quad i = 1, 2, \cdots, n \ ,$$

the reason being that the vector space $\{ r_n p : p \in \mathcal{P}_m \}$, where $\mathcal{P}_m$ is the vector space of all polynomials of order m, is finite-dimensional, and hence closed; see, e.g., HOLMES (1975).

So now we are in business: Define $\psi_o = p_o + T\varphi_o$, and it is easy to see that, for the (subsub) sequence in question,

$$\lim_{k\to\infty} L_{nh}(f_k) \geqslant L_{nh}(\psi_o) \ ,$$

so that ψ_o minimizes $L_{nh}(f)$ over $f \in W^{m,2}(0,1)$. Q.e.d.

(3.12) EXERCISE. The large-sample asymptotic problem corresponding to the finite-sample problem (2.1) is defined by

$$\begin{array}{ll} \text{minimize} & L_{\infty h}(f) \stackrel{\text{def}}{=} \| f - f_o \|^2 + h^{2m} \| f^{(m)} \|^2 \\ \text{subject to} & f \in W^{m,2}(0,1) \ . \end{array}$$

(a) Compute the Gateaux variation of $L_{\infty h}$, and show that

$$L_{\infty h}(f) - L_{\infty h}(\varphi) - \delta L_{\infty h}(\varphi\,,\, f-\varphi) = \| f - \varphi \|^2_{m,h} \ .$$

(b) Show that $L_{\infty h}$ is strongly convex and weakly lower semi-continuous on $W^{m,2}(0,1)$.

(c) Conclude that the solution of the minimization problem above exists and is unique.

(3.13) EXERCISE. Consider the C-spline estimation problem (1.10), repeated here for convenience:

$$\begin{array}{ll} \text{minimize} & \| f \|^2 - \frac{2}{n} \sum\limits_{i=1}^{n} y_{in}\, f(x_{in}) + h^{2m} \| f^{(m)} \|^2 \\ \text{subject to} & f \in W^{m,2}(0,1) \ . \end{array}$$

Show the existence and uniqueness of the solution of this problem. You should not need the asymptotic uniformity of the design.

As mentioned before, the convexity approach to showing existence and uniqueness is a heavy tool, but it makes for an easy treatment of convergence rates of the spline estimators, see § 4. It has the additional advantage that we can handle constrained problems without difficulty. Let $\mathcal{C}$ be a closed, convex subset of $W^{m,2}(0,1)$, and consider the problems

$$\text{(3.14)} \qquad \begin{array}{ll} \text{minimize} & \frac{1}{n} \sum\limits_{i=1}^{n} | f(x_{in}) - y_{in} |^2 + h^{2m} \| f^{(m)} \|^2 \\ \text{subject to} & f \in \mathcal{C} \ , \end{array}$$

and

$$(3.15)\qquad \begin{aligned} &\text{minimize} \quad \| f \|^2 - \tfrac{2}{n} \sum_{i=1}^{n} f(x_{in})\, y_{in} + h^{2m} \| f^{(m)} \|^2 \\ &\text{subject to} \quad f \in \mathcal{C} \ . \end{aligned}$$

(3.16) THEOREM. *The solution of the constrained smoothing spline problem* (3.13) *exists, and if there are at least* m *distinct design points, then it is unique. For the constrained problem* (3.14), *the solution always exists and is unique.*

(3.17) EXERCISE. Prove it!

Finally, we consider the Euler equations for the problem (2.1). One verifies that they are given by

$$(3.18)\qquad \begin{aligned} &(-h^2)^m\, u^{(2m)} + \tfrac{1}{n} \sum_{i=1}^{n} \bigl(u(x_{in}) - y_{in} \bigr)\, \delta(\,\cdot\, - x_{in}) = 0 \quad \text{in } (0,1) \ , \\ &u^{(k)}(0) = u^{(k)}(1) = 0 \ , \quad k = m, m+1, \cdots, 2m-1 \ . \end{aligned}$$

Here $\delta(\,\cdot\, - x_{in})$ is the unit point mass at $x = x_{in}$. (For the two endpoints, this requires the proper interpretation: Assume that they are moved into the interior of $[\,0\,,\,1\,]$ and take limits.) The boundary conditions in (3.18) go by the name of "natural" boundary conditions in that they are automagically associated with the problem (2.1). As an alternative, one could prescribe boundary values; e.g., if one knew $f_o^{(k)}(x)$, $k = 0, 1, \cdots, m-1$, at the endpoints $x = 0$, $x = 1$. In this case, the minimization in (2.1) could be further restricted to those functions f with the same boundary values, and the boundary conditions in (3.18) would be replaced by

$$(3.19)\qquad u^{(k)}(0) = f_o^{(k)}(0) \ , \ u^{(k)}(1) = f_o^{(k)}(1) \ , \quad 0 \leqslant k \leqslant m-1 \ .$$

(3.20) EXERCISE. (a) Verify that (3.18) are indeed the Euler equations for the smoothing spline problem (2.1) and that
(b) the unique solution of the Euler equations solves (2.1) and vice versa. [Hint: See § 10.5 in Volume I.]

(3.21) EXERCISE. (a) Show that the Euler equations for the C-spline problem discussed in (3.15) are given by

$$\begin{aligned} &(-h^2)^m\, u^{(2m)} + u = \tfrac{1}{n} \sum_{i=1}^{n} y_{in}\, \delta(\,\cdot\, - x_{in}) \quad \text{in } (0,1) \ , \\ &u^{(k)}(0) = u^{(k)}(1) = 0 \ , \quad k = m, m+1, \cdots, 2m-1 \ . \end{aligned}$$

(b) Verify that the solution is given by

$$\psi^{nh}(\,t\,) = \tfrac{1}{n} \sum_{i=1}^{n} y_{in}\, \mathcal{R}_{mh}(x_{in}, t\,) \ , \quad t \in [\,0\,,\,1\,] \ .$$

(c) Show that the unique solution of the Euler equations solves (2.1) and vice versa. [Hint: § 10.5.]

EXERCISES: (3.12), (3.13), (3.17), (3.20), (3.21).

4. Mean integrated squared error

We are now ready to investigate the asymptotic error bounds for the smoothing spline estimator. We recall the model

$$y_{in} = f_o(x_{in}) + d_{in} \ , \quad i = 1, 2, \cdots, n \ , \tag{4.1}$$

in which the noise vector $d_n = (d_{1,n}, d_{2,n}, \cdots, d_{n,n})^T$ satisfies the Gauss-Markov conditions

$$\mathbb{E}[\, d_n \,] = 0 \ , \quad \mathbb{E}[\, d_n \, d_n^T \,] = \sigma^2 I \ , \tag{4.2}$$

and f_o is the function to be estimated. The design is supposed to be asymptotically uniform; see Definition 2.22. Regarding the unknown function f_o, we had the assumption

$$f_o \in W^{m,2}(0,1) \ . \tag{4.3}$$

The smoothing spline estimator, denoted by f^{nh}, is the solution to

$$\begin{array}{ll} \text{minimize} & L_{nh}(f) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} |\, f(x_{in}) - y_{in} \,|^2 + h^{2m} \,\|\, f^{(m)} \,\|^2 \\ \text{subject to} & f \in W^{m,2}(0,1) \ . \end{array} \tag{4.4}$$

It is useful to introduce the abbreviation ε^{nh} for the error function,

$$\varepsilon^{nh} \equiv f^{nh} - f_o \ . \tag{4.5}$$

(4.6) THEOREM. *Let $m \geqslant 1$. Suppose the Markov conditions (4.1) and (4.2) hold and that $f_o \in W^{m,2}(0,1)$. If x_{in}, $i = 1, 2, \cdots, n$, is asymptotically uniform, then for all $n \geqslant 2$ and all h, $0 < h \leqslant \frac{1}{2}$, with $nh \to \infty$,*

$$\zeta^{nh} \,\|\, f^{nh} - f_o \,\|_{m,h}^2 \leqslant \Big\{ \,\|\, \mathfrak{S}^{nh} \,\|_{m,h} + h^m \,\|\, f_o^{(m)} \,\| \,\Big\}^2 \ ,$$

where $\zeta^{nh} \to 1$. Here, $\mathfrak{S}^{nh}$ is given by (2.19).

(4.7) COROLLARY. *Under the same conditions as in the previous theorem, for $h \asymp n^{-1/(2m+1)}$ (deterministically),*

$$\mathbb{E}[\,\|\, f^{nh} - f_o \,\|_{m,h}^2 \,] = \mathcal{O}\big(n^{-2m/(2m+1)} \big) \ .$$

Proof of Theorem (4.6). The approach to obtaining error bounds is via the Quadratic Behavior Lemma (3.1) for $L_{nh}(f)$. This gives the equality

$$\text{(4.8)}\quad \tfrac{1}{n}\sum_{i=1}^{n} |\,\varepsilon^{nh}(x_{in})\,|^2 + h^{2m}\,\|\,(\varepsilon^{nh})^{(m)}\,\|^2 = \tfrac{1}{n}\sum_{i=1}^{n} d_{in}\,\varepsilon^{nh}(x_{in}) - h^{2m}\,\langle\, f_o^{(m)}\,,\,(\varepsilon^{nh})^{(m)}\,\rangle\;.$$

Of course, first we immediately use Cauchy-Schwarz,

$$-\,\langle\, f_o^{(m)}\,,\,(\varepsilon^{nh})^{(m)}\,\rangle \leqslant \|\,f_o^{(m)}\,\|\,\|\,(\varepsilon^{nh})^{(m)}\,\|\;.$$

Second, by the Random Sum Lemma (2.20), the random sum in (4.8) may be bounded by $\|\,\varepsilon^{nh}\,\|\,\|\,\mathfrak{S}^{nh}\,\|_{m,h}$.

Third, by the Quadrature Lemma (2.27), the sum on the left of (4.8) may be bounded from below by

$$\zeta^{nh}\,\|\,\varepsilon^{nh}\,\|_{m,h}^2 \leqslant \tfrac{1}{n}\sum_{i=1}^{n} |\,\varepsilon^{nh}(x_{in})\,|^2 + h^{2m}\|\,(\varepsilon^{nh})^{(m)}\,\|^2\;,$$

with $\zeta^{nh} = 1 - c_m(nh)^{-1}$. So, under the stated conditions, then $\zeta^{nh} \to 1$.

It follows from (2.10) that then

$$\text{(4.9)}\qquad \zeta^{nh}\,\|\,\varepsilon^{nh}\,\|_{m,h}^2 \leqslant \|\,\varepsilon^{nh}\,\|_{m,h}\,\Big\{\,\|\,\mathfrak{S}^{nh}\,\|_{m,h} + h^m\,\|\,f_o^{(m)}\,\|\,\Big\}\;,$$

where we used that $h^{2m}\,\|\,(\varepsilon^{nh})^{(m)}\,\| \leqslant h^m\,\|\,\varepsilon^{nh}\,\|_{m,h}$. The theorem follows by an appeal to the following exercise. Q.e.d.

(4.10) Exercise. Let a and b be positive real numbers, and let $p > 1$. If the nonnegative real number x satisfies $x^p \leqslant a\,x + b$, then

$$x^p \leqslant a^q + q\,b\;,$$

in which q is the dual exponent of p; i.e., $(1/p) + (1/q) = 1$.

(4.11) Exercise. Show that the bounds of Theorem (4.6) and Corollary (4.7) apply also to $\frac{1}{n}\sum_{i=1}^{n} |\,f^{nh}(x_{in}) - f_o(x_{in})\,|^2$.

The above is a concise treatment of the smoothing spline problem. The reader should become very comfortable with it since variations of it will be used throughout the text.

Can the treatment above be improved? The only chance we have is to avoid Cauchy-Schwarz in

$$-\,\langle\, f_o^{(m)}\,,\,(\varepsilon^{nh})^{(m)}\,\rangle \leqslant \|\,f_o^{(m)}\,\|\,\|\,(\varepsilon^{nh})^{(m)}\,\|\;,$$

following (4.8). Under the special smoothness condition and natural boundary conditions,

$$\text{(4.12)}\quad f_o \in W^{2m,2}(0,1)\;,\quad f^{(\ell)}(0) = f^{(\ell)}(1) = 0\;,\quad m \leqslant \ell \leqslant 2m-1\;,$$

this works. Results like this go by the name of superconvergence, since the accuracy is much better than guaranteed by the estimation method.

(4.13) SUPER CONVERGENCE THEOREM. *Assume the conditions of Theorem* (4.6). *If the regression function* $f_o \in W^{2m,2}(0,1)$ *satisfies the natural boundary conditions* (4.12), *then*

$$\| f^{nh} - f_o \|_{m,h}^2 \leqslant \Big\{ \| \mathfrak{S}^{nh} \|_{m,h}^2 + h^{2m} \| f_o \|_{W^{2m,2}(0,1)} \Big\}^2 ,$$

and for $h \asymp n^{-1/(4m+1)}$ *(deterministically),*

$$\mathbb{E}[\, \| f^{nh} - f_o \|_{m,h}^2 \,] = \mathcal{O}\big(n^{-4m/(4m+1)} \big) .$$

PROOF. The natural boundary conditions (4.11) allow us to integrate by parts m times, without being burdened with boundary terms. This gives

$$- \big\langle f_o^{(m)} , (\varepsilon^{nh})^{(m)} \big\rangle = (-1)^{m+1} \big\langle f_o{}^{(2m)} , \varepsilon^{nh} \big\rangle \leqslant \| f_o{}^{(2m)} \| \, \| \varepsilon^{nh} \| ,$$

and, of course, $\| \varepsilon^{nh} \| \leqslant \| \varepsilon^{nh} \|_{m,h}$. Thus, in the inequality (4.9), we may replace $h^m \| f_o^{(m)} \|$ by $h^{2m} \| f_o^{(2m)} \|$, and the rest follows. Q.e.d.

A brief comment on the condition (4.12) in the theorem above is in order. The smoothness assumption $f_o \in W^{2m,2}(0,1)$ is quite reasonable, but the boundary conditions on f_o are inconvenient, to put it mildly. In the next two sections, we discuss ways around the boundary conditions. In the meantime, the following exercise is useful in showing that the boundary conditions of Theorem (4.13) may be circumvented *at a price* (viz. of periodic boundary conditions).

(4.14) EXERCISE. Let $m \geqslant 1$ and $f_o \in W^{2m,2}(0,1)$. Prove the bounds of Theorem (4.13) for the solution of

$$\begin{aligned} &\text{minimize} \quad && L_{nh}(f) \stackrel{\text{def}}{=} \tfrac{1}{n} \sum_{i=1}^{n} | f(x_{in}) - y_{in} |^2 + h^{2m} \| f^{(m)} \|^2 \\ &\text{subject to} \quad && f \in W^{m,2}(0,1) , \\ & && \text{and for} \quad k = 0, 1, \cdots, m-1 , \\ & && f^{(k)}(0) = f_o^{(k)}(0) , \ f^{(k)}(1) = f_o^{(k)}(1) . \end{aligned}$$

The following exercise discusses what happens when the boundary conditions in Theorem (4.13) are only partially fulfilled. This finds a surprising application to boundary corrections; i.e., for obtaining estimators for which the conclusions of Theorem (4.13) remain valid. See § 5.

(4.15) EXERCISE. Let $1 \leqslant k \leqslant m$. Suppose that $f_o \in W^{m+k}(0,1)$ satisfies

$$f_o^{(\ell)}(0) = f_o^{(\ell)}(1) = 0 , \quad \ell = m, \cdots, m+k-1 .$$

Show that $\| f^{nh} - f_o \|^2 \leqslant \{ \| \mathfrak{S}^{nh} \|_{m,h} + h^{m+k} \| f_o^{(m+k)} \| \}^2$.

We finish with some exercises regarding constrained estimation and the C-spline problem (1.10).

(4.16) Exercise. (a) Derive the error bounds of Theorem (4.7) and Corollary (4.8) for the constrained smoothing spline problem

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{n} \sum_{i=1}^{n} | f(x_{in}) - y_{in} |^2 + h^{2m} \| f^{(m)} \|^2 \\ &\text{subject to} \quad f \in \mathcal{C} \ , \end{aligned}$$

where $\mathcal{C}$ is a closed and convex subset of $W^{m,2}(0,1)$. Assume that $f_o \in \mathcal{C}$. (b) Do likewise for the constrained version of (1.10).

(4.17) Exercise. Show that the error bounds of Theorems (4.6) and (4.13) also apply to the solution of the C-spline estimation problem (1.10).

An alternative approach. We now consider an alternative development based on the observation that there are three sources of "error" in the smoothing spline problem (4.4). The obvious one is the noise in the data. Less obvious is that the roughness penalization is the source of bias, and finally there is the finiteness of the data. Even if the data were noiseless, we still could not estimate f_o perfectly due to the finiteness of the design. We need to introduce the finite noiseless data problem,

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{n} \sum_{i=1}^{n} | f(x_{in}) - f_o(x_{in}) |^2 + h^{2m} \| f^{(m)} \|^2 \\ &\text{subject to} \quad f \in W^{m,2}(0,1) \ , \end{aligned} \tag{4.18}$$

as well as the large-sample asymptotic noiseless problem,

$$\begin{aligned} &\text{minimize} \quad \| f - f_o \|^2 + h^{2m} \| f^{(m)} \|^2 \\ &\text{subject to} \quad f \in W^{m,2}(0,1) \ . \end{aligned} \tag{4.19}$$

In the exercise below, we (i.e., you) will analyze these problems.

The following simple exercise is quite useful.

(4.20) Exercise. Show that, for all real numbers A, B, a, b

$$| A - b |^2 - | A - a |^2 + | B - a |^2 - | B - b |^2 = 2\,(a-b)(A-B) \ .$$

(4.21) Exercise. Let $f_o \in W^{m,2}(0,1)$. Let f^{nh} be the solution of (4.4) and f_{hn} the solution of (4.18). Show that, for $nh \to \infty$ and h bounded,

$$\| f^{nh} - f_{hn} \|_{m,h} \leqslant \| \mathfrak{S}^{nh} \|_{m,h} \ ,$$

with $\mathfrak{S}^{nh}$ as in (2.19).

The bias due to the finiteness of the data is considered next.

(4.22) EXERCISE. Let $f_o \in W^{m,2}(0,1)$. Let f_h be the solution of (4.19) and f_{hn} the solution of (4.18). Show that, for a suitable constant c, as $h \to 0$ and nh large enough,

$$\| f_{hn} - f_h \|_{m,h} \leqslant c\,(nh)^{-1} \| f_h - f_o \|_{m,h} \; .$$

(4.23) EXERCISE. Show that the solution f_h of (4.19) satisfies

$$\| f_h - f_o \|_{m,h}^2 \leqslant h^{2m} \| f_o^{(m)} \|^2 \; .$$

We may now put these exercises together.

(4.24) EXERCISE. Prove Theorem (4.6) using Exercises (4.21)–(4.23).

(4.25) EXERCISE. Prove the analogue of Theorem (4.6) for the C-spline estimator of (1.10) straightaway (or via the detour).

EXERCISES: (4.10), (4.11), (4.14), (4.15), (4.16), (4.17), (4.20), (4.21), (4.22), (4.23), (4.24), (4.25).

5. Boundary corrections

In this section and the next, we take a closer look at the smoothness and boundary conditions (4.12), repeated here for convenience:

$$f_o \in W^{2m,2}(0,1) \; , \tag{5.1}$$

$$f_o^{(\ell)}(0) = f_o^{(\ell)}(1) \; , \quad m \leqslant \ell \leqslant 2m-1 \; . \tag{5.2}$$

In Theorem (4.13), we showed that, under these circumstances, the smoothing spline estimator f^{nh} of order $2m$ (degree $2m-1$) has expected error

$$\mathbb{E}[\, \| f^{nh} - f_o \|^2 \,] = \mathcal{O}\big(n^{-4m/(4m+1)} \big) \; , \tag{5.3}$$

at least for $h \asymp n^{1/(4m+1)}$ (deterministic), and that the improvement over the bounds of Corollary (4.7) is due to *bias reduction*. The variance part remains unchanged. It follows from STONE (1982) (see the discussion in § 12.3) that (5.3) is also the asymptotic lower bound. See also RICE and ROSENBLATT (1983). However, away from the boundary, (5.3) holds regardless of whether (5.2) holds. Thus, the question is whether one can compute *boundary corrections* to achieve the global error bound (5.3).

Returning to the conditions (5.1)–(5.2), in view of STONE (1982), one cannot really complain about the smoothness condition, but the boundary condition (5.2) makes (5.3) quite problematic. By way of example, if $f^{(m)}(0) \neq 0$, then one does not get any decrease in the global error, and so the bound (5.3) is achievable only for smoothing splines of polynomial order $4m$. It would be nice if the smoothing spline estimator of order $2m$ could be suitably modified such that (5.3) would apply under the sole condition (5.1). This would provide a measure of *adaptation*: One may underestimate (guess) the smoothness of f_o by a factor 2 if we may characterize the distinction $f_o \in W^{m,2}(0,1)$ vs. $f_o \in W^{2m,2}(0,1)$ in this way.

There is essentially only one boundary correction method, viz. the application of the Bias Correction Principle of Eubank and Speckman (1990b) by Huang (2001) as discussed in this section. The relaxed boundary splines of Oehlert (1992) avoid the problem rather than correcting it; see § 6.

(5.4) The Bias Reduction Principle (Eubank and Speckman, 1990b). Suppose one wishes to estimate a parameter $\theta_o \in \mathbb{R}^n$ and has available two estimators, each one flawed in its own way. One estimator, $\widetilde{\theta}$, is unbiased,

$$\mathbb{E}[\,\widetilde{\theta}\,] = \theta_o \ , \tag{5.5}$$

and each component has finite variance but otherwise has no known good properties. The other estimator, $\widehat{\theta}$, is biased but nice,

$$\mathbb{E}[\,\widehat{\theta}\,] = \theta_o + Ga + b \ , \tag{5.6}$$

for some $G \in \mathbb{R}^{n\times m}$, and $a \in \mathbb{R}^m$, $b \in \mathbb{R}^n$. It is assumed that G is known but that a and b are not. Let Π_G be the orthogonal projector onto the range of G. (If G has full column rank, then $\Pi_G = G(G^{T}G)^{-1}G^{T}$.)

Then, the estimator

$$\theta^{\#} = \widehat{\theta} + \Pi_G\,\big(\,\widetilde{\theta} - \widehat{\theta}\,\big) \tag{5.7}$$

satisfies

$$\mathbb{E}[\,\theta^{\#}\,] = \theta_o + \gamma \tag{5.8}$$

with
$$\|\,\gamma\,\| \leqslant \min\big\{\,\|\,Ga+b\,\|\,,\,\|\,b\,\|\,\big\}$$

and
$$\mathbb{E}[\,\|\,\theta^{\#} - \mathbb{E}[\,\theta^{\#}\,]\,\|^2\,] \leqslant \mathbb{E}[\,\|\,\widehat{\theta} - \mathbb{E}[\,\widehat{\theta}\,]\,\|^2\,] + \lambda\, m \ .$$

Here, $\lambda = \lambda_{\max}(\mathbb{V}\mathrm{ar}[\,\widetilde{\theta}\,]\,)$ is the largest eigenvalue of $\mathbb{V}\mathrm{ar}[\,\widetilde{\theta}\,]$.

Proof of the Bias Reduction Principle. One verifies that

$$c^{\#} \stackrel{\text{def}}{=} \mathbb{E}[\,\theta^{\#} - \theta_o\,] = (\,I - \Pi_G\,)\,(\,Ga + b\,)\ .$$

Now, since Π_G is an orthogonal projector, so is $I - \Pi_G$, and therefore $\|\,c^{\#}\,\| \leqslant \|\,Ga{+}b\,\|$. On the other hand, $(\,I{-}\Pi_G)G = O$, so $c^{\#} = (\,I{-}\Pi_G\,)\,b$, and $\|\,c^{\#}\,\| \leqslant \|\,b\,\|$.

For the variance part, it is useful to rewrite $\theta^{\#}$ as

$$\theta^{\#} = (\,I - \Pi_G)\,\widehat{\theta} + \Pi_G\,\widetilde{\theta}\ ,$$

so that by Pythagoras' theorem

$$\| \theta^{\#} - \mathbb{E}[\theta^{\#}] \|^2 = \| (I - \Pi_G)(\widehat{\theta} - \mathbb{E}[\widehat{\theta}]) \|^2 + \| \Pi_G(\widetilde{\theta} - \mathbb{E}[\widetilde{\theta}]) \|^2 .$$

The first term on the right is bounded by $\| \widetilde{\theta} - \mathbb{E}[\widehat{\theta}] \|^2$. For the second term, we have

$$\mathbb{E}\Big[\| \Pi_G(\widetilde{\theta} - \mathbb{E}[\widetilde{\theta}]) \|^2 \Big] = \mathrm{trace}\Big(\Pi_G \, \mathbb{V}\mathrm{ar}[\widetilde{\theta}] \, \Pi_G^T \Big) .$$

Let $\Lambda = \lambda I$. Then, $\Lambda - \mathbb{V}\mathrm{ar}[\widetilde{\theta}]$ is semi-positive-definite, so that

$$\mathrm{trace}(\Lambda - \mathbb{V}\mathrm{ar}[\widetilde{\theta}]) \geqslant 0 .$$

It follows that

$$\begin{aligned}
&\mathbb{E}[\| \Pi_G(\widetilde{\theta} - \mathbb{E}[\widetilde{\theta}]) \|^2] \\
&\qquad = \mathrm{trace}\big(\Pi_G \, \Lambda \, \Pi_G^T \big) - \mathrm{trace}\big(\Pi_G \, (\Lambda - \mathbb{V}\mathrm{ar}[\widetilde{\theta}]) \, \Pi_G^T \big) \\
&\qquad \leqslant \mathrm{trace}\big(\Pi_G \, \Lambda \, \Pi_G^T \big) = \lambda \, \mathrm{trace}\big(\Pi_G \, \Pi_G^T \big) = \lambda \, m .
\end{aligned}$$

The bound on the variance of $\theta^{\#}$ follows. Q.e.d.

The Bias Reduction Principle is useful when Ga is much larger than b and m is small. Under these circumstances, the bias is reduced dramatically, whereas the variance is increased by only a little. Note that $\Pi_G(\widetilde{\theta} - \widehat{\theta})$ is a "correction" to the estimator $\widehat{\theta}$.

We now wish to apply the Bias Reduction Principle to compute boundary corrections to the spline estimator of § 3. Actually, corrections to the values $f^{nh}(x_{in})$, $i = 1, 2, \cdots, n$, will be computed. For corrections to the spline function, see Exercise (5.21).

For the implementation of this scheme, the boundary behavior of the smoothing spline estimator must be described in the form (5.6). Thus, the boundary behavior must be "low-dimensional".

(5.9) The asymptotic behavior of the bias of the smoothing spline estimator near the boundary. Let $f_o \in W^{2m,2}(0,1)$. Then, $f_o^{(k)}$ is continuous for $k = 0, 1, \cdots, 2m-1$. Now, for $k = m, \cdots, 2m-1$, let L_k and R_k be polynomials (yet to be specified), and consider

$$p_o(x) = \sum_{\ell=m}^{2m-1} f_o^{(\ell)}(0) \, L_k(x) + f_o^{(\ell)}(1) \, R_k(x) . \tag{5.10}$$

We now wish to choose the L_k and R_k such that $g_o \stackrel{\text{def}}{=} f_o - p_o \in W^{2m,2}(0,1)$ satisfies

$$g_o^{(k)}(0) = g_o^{(k)}(1) = 0 , \quad k = m, \cdots, 2m-1 . \tag{5.11}$$

One verifies that it is sufficient that, for all k,

$$(5.12)\qquad \begin{aligned} & L_k^{(\ell)}(0) = L_k^{(\ell)}(1) = 0 \quad \text{for } \ell = m, \cdots, 2m-1 \;, \\ & \text{except that} \quad L_k^{(k)}(0) = 1 \;, \\ & \text{and} \quad R_k(x) = (-1)^k \, L_k(1-x) \;. \end{aligned}$$

The construction of the L_k is an exercise in Hermite-Birkhoff interpolation; see Kincaid and Cheney (1991). For the case $m = 2$, see Exercise (5.20).

Now, let g_h be the solution to

$$\begin{aligned} & \text{minimize} \quad \| f - g_o \|^2 + h^{2m} \| f^{(m)} \|^2 \\ & \text{subject to} \quad f \in W^{m,2}(0,1) \;, \end{aligned}$$

and construct the functions $L_{k,h}$ similarly, based on the $L_{k,o} = L_k$. Let

$$(5.13)\qquad \eta_{k,h} \stackrel{\text{def}}{=} h^{-k} \left\{ L_{k,h} - L_k \right\} \;, \quad k = m, \cdots, 2m-1 \;.$$

Then, g_h satisfies $\| g_h - g_o \|_{m,h}^2 = \mathcal{O}\big(h^{4m} \big)$, and by Exercise (4.15) applied to noiseless data,

$$(5.14)\qquad \| \eta_{k,h} \|_{m,h}^2 = \mathcal{O}\big(1 \big) \;, \quad k = m, \cdots, 2m-1 \;.$$

By linearity, it follows that

$$(5.15)\qquad f_h = f_o + \sum_{k=m}^{2m-1} h^k \left\{ f_o^{(k)}(0) \, \eta_{k,h} + f_o^{(k)}(1) \, \zeta_{k,h} \right\} + \varepsilon_h \;,$$

with $\| \varepsilon_h \|_{m,h} = \mathcal{O}\big(h^{2m} \big)$, and $\zeta_{k,h} = (-1)^k \, \eta_{k,h}$ for all k. Of course, by the Quadrature Lemma (2.27), the corresponding bounds hold for the sums:

$$(5.16)\qquad \tfrac{1}{n} \sum_{i=1}^{n} | \, \varepsilon_h(x_{in}) \, |^2 = \mathcal{O}\big(h^{4m} \big),$$

$$(5.17)\qquad \tfrac{1}{n} \sum_{i=1}^{n} | \, \eta_{k,h}(x_{in}) \, |^2 = \mathcal{O}\big(1 \big) \;.$$

(5.18) Computing boundary corrections (Huang, 2001). The Bias Reduction Principle may now be applied to compute boundary corrections. In the notation of the Bias Reduction Principle (5.4), take

$$\theta_o = \big(f_o(x_{1,n}), f_o(x_{2,n}), \cdots, f_o(x_{n,n}) \big)^T$$

and consider the estimators

$$\widehat{\theta} = \big(f^{nh}(x_{1,n}), f^{nh}(x_{2,n}), \cdots, f^{nh}(x_{n,n}) \big)^T$$

and

$$\widetilde{\theta} = \big(y_{1,n}, y_{2,n}, \cdots, y_{n,n} \big)^T \;.$$

Then, $\widetilde{\theta}$ is an unbiased estimator of θ_o. The asymptotic behavior of $\widehat{\theta}$ is described by

$$\mathbb{E}[\, \widehat{\theta} \,] = \theta_o + F \, a_0 + G \, a_1 + \varepsilon_h \;,$$

with ε_h as in (5.15) and $F,\, G \in \mathbb{R}^{n\times m}$, given by

$$F_{i,k} = \eta_{k,h}(x_{in}) \ , \quad G_{i,k} = \zeta_{k,h}(x_{in})$$

for $i = 1, 2, \cdots, n$ and $k = m, \cdots, 2m-1$. The vectors a_0 and a_1 contain the (unknown) derivatives of f_o at the endpoints. The estimator $\theta^{\#}$ may now be computed as per (5.7) and satisfies

$$\begin{aligned} \mathbb{E}\Big[\, \tfrac{1}{n} \sum_{i=}^{n} |\, \theta_i^{\#} - f_o(x_{in})\,|^2 \,\Big] &\leqslant \mathbb{E}\Big[\, \tfrac{1}{n} \sum_{i=}^{n} |\, (\varepsilon_{k,h})_i \,|^2 \,\Big] + 2\, m\, n^{-1}\, \sigma^2 \\ &= \mathcal{O}\big(\, h^{4m} + (nh)^{-1} \,\big) \ . \end{aligned} \tag{5.19}$$

The boundary behavior (5.15) is due to Rice and Rosenblatt (1983). We consider an analogous result for trigonometric sieves.

(5.20) Exercise. (a) Let $m = 2$. Verify that

$$L_2(x) = \tfrac{1}{4}\,(\,1-x\,)^2 - \tfrac{1}{10}\,(\,1-x\,)^5 \ , \ L_3(x) = -\tfrac{1}{12}\,(\,1-x)^4 + \tfrac{1}{20}\,(\,1-x\,)^5$$

satisfy (5.12), and verify (5.11).
(b) Verify (5.16).
(c) Prove that the bounds (5.17) are sharp.
(d) Prove (5.19).

(5.21) Exercise. Suppose we are not interested in $f^{nh}(x_{in})$, $i = 1, \cdots, n$, but in the actual spline $f^{nh}(x)$, $x \in [\,0\,,1\,]$. Formulate an algorithm to compute the boundary correction to the spline function. [Hint: One may think of the spline estimator as being given by its coefficients; in other words, it is still a finite-dimensional object. Unbiased estimators of f_o do not exist, but we do have an unbiased estimator of the spline interpolant of f_o using the data $f_o(x_{in})$, $i = 1, 2, \cdots, n$, which is a very accurate approximation to f_o. See Chapter 19 for the details on spline interpolation.]

Exercises: (5.20), (5.21).

6. Relaxed boundary splines

In this section, we discuss the solution of Oehlert (1992) to the boundary correction problem for smoothing splines. His approach is to avoid the problem altogether by modifying the roughness penalization in the smoothing spline problem (4.4). The choice of penalization by Oehlert (1992) is actually quite fortuitous: It is easy to analyze the resulting estimator, much in the style of §§ 2 and 3, but the choice itself is *magic*.

We operate again under the Gauss-Markov model (1.1)–(1.2) with asymptotically uniform designs; see Definition (2.22). For now, suppose that

$$f_o \in W^{2m,2}(0,1) \ . \tag{6.1}$$

In general, under these circumstances, the smoothing spline estimator of polynomial order $2m$, defined as the solution to (2.1), has mean integrated squared error $\mathcal{O}\bigl(n^{-2m/(2m+1)}\bigr)$, whereas the smoothness assumption (6.1) should allow for an error $\mathcal{O}\bigl(n^{-4m/(4m+1)}\bigr)$.

It is worthwhile to repeat the motivation of Oehlert (1992) for his suggested modification of (4.4). He observes that the global variance of the estimator is $\mathcal{O}\bigl((nh)^{-1}\bigr)$ and that the squared bias is $\mathcal{O}\bigl(h^{4m}\bigr)$ away from the boundary points but, in general, is only $\mathcal{O}\bigl(h^{2m}\bigr)$ near the boundary. Thus, it would be a good idea to reduce the bias near the boundary if this could be done without dramatically increasing the variance. His way of doing this is to downweight the standard roughness penalization near the endpoints. There would appear to be many ways of doing this, until one has to do it. Indeed, the analysis of Oehlert (1992) and the analysis below show that quite a few "things" need to happen.

The particular suggestion of Oehlert (1992) is as follows. Let $m \geqslant 1$ be an integer, and consider the vector space of functions defined on $(0,1)$

$$\mathcal{W}_m = \left\{ f \;\middle|\; \begin{array}{l} \forall\,\delta : 0 < \delta < \tfrac{1}{2} \implies f \in W^{m,2}(\delta, 1-\delta) \\ \quad |\, f\,|_{\mathcal{W}_m} < \infty \end{array} \right\}, \tag{6.2}$$

where the semi-norm $|\cdot|_{\mathcal{W}_m}$ is defined by way of

$$|\, f\,|^2_{\mathcal{W}_m} \stackrel{\text{def}}{=} \int_0^1 \{\, x(1-x)\,\}^m \,|\, f^{(m)}(x)\,|^2 \, dx \,. \tag{6.3}$$

The relaxed boundary spline estimator of the regression function is then defined as the solution $f = \psi^{nh}$ of the problem

$$\begin{array}{ll} \text{minimize} & RLS(f) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} |\, f(x_{in}) - y_{in}\,|^2 + h^{2m} \,|\, f\,|^2_{\mathcal{W}_m} \\ \text{subject to} & |\, f\,|_{\mathcal{W}_m} < \infty \,. \end{array} \tag{6.4}$$

Of course, the existence and uniqueness of the solution must be established, and the objective function $RLS(f)$ must be well-defined on $\mathcal{W}_m$. There are some difficulties in the case $m = 1$ that require some extra conditions on the design (asymptotic uniformity does not suffice, it appears). So, at the crucial moment, we assume that $m \geqslant 2$. Also, the assumption (6.1) may be replaced by the condition

$$f_o \in \mathcal{W}_{2m} \,. \tag{6.5}$$

The difficulties for $m = 1$ are illustrated in the following exercise.

(6.6) Exercise. Show that the function

$$f(x) = \bigl|\log\{\, x(1-x)\,\}\bigr|^{\alpha}\,, \quad x \in (0,1)\,,$$

belongs to $\mathcal{W}_1$ for $\alpha < \frac{1}{2}$ but not for $\alpha = \frac{1}{2}$.

The two “final” results are as follows. Note that there are almost no conditions in the existence and uniqueness theorem.

(6.7) THEOREM. *Let* $m \geqslant 1$. *The solution of* (6.4) *exists. If the design contains at least* m *distinct points, then the solution is unique.*

(6.8) THEOREM. *Let* $m \geqslant 2$. *Assume the Gauss-Markov model* (1.1)–(1.2), *and that the design is asymptotically uniform. Assuming* $f_o \in \mathcal{W}_{2m}$, *the solution* ψ^{nh} *of* (6.4) *satisfies*

$$\mathbb{E}[\, \| \, \psi^{nh} - f_o \, \|^2 \,] = \mathcal{O}\big(\, n^{-4m/(4m+1)} \, \big) \ ,$$

provided $h \asymp n^{-1/(4m+1)}$ *(deterministically).*

We now set out to prove Theorems (6.7) and (6.8). The proof essentially follows the development in §§ 2, 3, and 4: The relevant lemmas all have useful analogues, but some of the proofs are simple computations in terms of an orthogonal basis for the Hilbert spaces in question. This orthogonal basis (the Legendre polynomials, suitably scaled) was already featured prominently in OEHLERT (1992), and we are not above using it. In fact, this constitutes the *magic* of the particular choice of penalization. This section is devoted to preliminaries, analogous to § 2. In the next section, the existence, uniqueness, and convergence rates are established. We note that, for $m = 1$, Theorem (6.7) holds for a modified design, not including the endpoints of the interval; see Exercise (6.41).

Reproducing kernel Hilbert spaces. For $h > 0$, define the inner products on $\mathcal{W}_m$,

$$\langle\!\langle \, f \, , \, g \, \rangle\!\rangle_{h,\mathcal{W}_m} = \langle \, f \, , \, g \, \rangle + h^{2m} \, \langle \, f \, , \, g \, \rangle_{\mathcal{W}_m} \ , \tag{6.9}$$

where

$$\langle \, f \, , \, g \, \rangle_{\mathcal{W}_m} = \int_0^1 [\, x(1-x) \,]^m \, f^{(m)}(x) \, g^{(m)}(x) \, dx \ , \tag{6.10}$$

and the associated norms $\| \cdot \|_{h,\mathcal{W}_m}$ by way of

$$\| \, f \, \|_{h,\mathcal{W}_m}^2 = \langle\!\langle \, f \, , \, f \, \rangle\!\rangle_{h,\mathcal{W}_m} \ . \tag{6.11}$$

It is obvious that, with all these norms, $\mathcal{W}_m$ is a Hilbert space. Moreover, these norms are equivalent, but not uniformly in h; see Definition (2.7) and Exercise (2.8).

At this point, we introduce the shifted Legendre polynomials, which behave very nicely in all of the $\mathcal{W}_m$. As mentioned before, OEHLERT (1992) already made extensive use of this.

First, we summarize the relevant properties of the Legendre polynomials. One way to define the standard Legendre polynomials is through the recurrence relations

$$\begin{aligned} &P_{-1}(x) = 0 \;, \quad P_0(x) = 1 \;, \\ &(k+1)\, P_{k+1}(x) = (2k+1)\, x\, P_k(x) - k\, P_{k-1}(x) \;, \quad k \geqslant 0 \;. \end{aligned} \tag{6.12}$$

The shifted, normalized Legendre polynomials are here defined as

$$Q_k(x) = (2k+1)^{1/2}\, P_k(2x-1) \;, \quad k \geqslant 0 \;. \tag{6.13}$$

They satisfy the following orthogonality relations:

$$\big\langle\, Q_k \,,\, Q_\ell \,\big\rangle_{L^2(0,1)} = \begin{cases} 1 \,, & \text{if} \quad k = \ell \;, \\ 0 \,, & \text{otherwise} \;, \end{cases} \tag{6.14}$$

$$\big\langle\, Q_k \,,\, Q_\ell \,\big\rangle_{\mathcal{W}_m} = \begin{cases} 0 \,, & \text{if} \quad k \neq \ell \;, \\ \dfrac{4^m\,(k+m)!}{(k-m)!} \,, & \text{if} \quad k = \ell \;. \end{cases} \tag{6.15}$$

Note that the last inner product vanishes (also) for $k = \ell < m$. We also have the pointwise bounds

$$\begin{aligned} &|\, Q_k(x) \,| \leqslant (2k+1)^{1/2} && \text{for all } 0 \leqslant x \leqslant 1 \text{ and } k \geqslant 0 \;, \\ &|\, Q_k(x) \,| \leqslant c\, \{\, x(1-x) \,\}^{-1/4} && \text{for all } 0 < x < 1 \text{ and } k \geqslant 1 \end{aligned} \tag{6.16}$$

for a suitable constant c independent of k and x. A handy reference for all of this is Sansone (1959). Note that (6.14)–(6.15) prove the following lemma.

(6.17) Lemma. *For all* $h > 0$ *and all* $m \geqslant 1$,

$$\big\langle\!\big\langle\, Q_k \,,\, Q_\ell \,\big\rangle\!\big\rangle_{h,\mathcal{W}_m} = \begin{cases} 1 + (2h)^{2m}\, \lambda_{k,m} \,, & \textit{if} \quad k = \ell \;, \\ 0 \,, & \textit{otherwise} \;, \end{cases}$$

where $\lambda_{k,m} = (k+m)!/(k-m)!$. *Moreover, there exist constants* $c_m > 1$ *such that, for all* k, $k \geqslant m$, *we have* $(c_m)^{-1} \leqslant k^{-2m}\, \lambda_{k,m} \leqslant c_m$.

It follows that Q_k, $k \geqslant 0$, is an orthonormal basis for $L^2(0,1)$ and an orthogonal basis for $\mathcal{W}_m$. Also, it gives us a handy expression for the norms on $\mathcal{W}_m$, but we shall make them handier yet. For $f \in L^2(0,1)$, define

$$\widehat{f}_k = \big\langle\, f \,,\, Q_k \,\big\rangle \;, \quad k \geqslant 0 \;. \tag{6.18}$$

The following lemma is immediate.

(6.19) Lemma. *Let* $m \geqslant 1$. *For all* $h > 0$ *and* $f \in \mathcal{W}_m$,

$$f = \sum_{k \geqslant 0} \widehat{f}_k\, Q_k \;,$$

with convergence in the $\mathcal{W}_m$-topology, and for all f, $g \in \mathcal{W}_m$,

$$\langle\!\langle\, f\,,\, g\,\rangle\!\rangle_{h,\mathcal{W}_m} = \sum_{k\geqslant 0} \bigl(1 + (2h)^{2m}\, \lambda_{k,m} \bigr)\ \widehat{f}_k\ \widehat{g}_k\ .$$

Finally, for all $f \in \mathcal{W}_m$,

$$\|\, f\,\|^2_{h,\mathcal{W}_m} = \sum_{k\geqslant 0} \bigl(1 + (2h)^{2m}\, \lambda_{k,m} \bigr)\, |\, \widehat{f}_k\, |^2 < \infty\ .$$

The representation above for the norms is nice, but the behavior of the $\lambda_{k,m}$ is a bit of a bummer. So, let us define the equivalent norms

$$\text{(6.20)} \qquad |\!|\!|\, f\,|\!|\!|_{h,\mathcal{W}_m} = \Bigl\{ \sum_{k\geqslant 0} \bigl(1 + (2hk)^{2m} \bigr)\, |\, \widehat{f}_k\, |^2 \Bigr\}^{1/2}\ .$$

(6.21) Lemma. *Let $m \geqslant 1$. The norms $|\!|\!|\,\cdot\,|\!|\!|_{h,\mathcal{W}_m}$ and $\|\,\cdot\,\|_{h,\mathcal{W}_m}$ are equivalent, uniformly in h, $0 < h \leqslant 1$; i.e., there exists a constant γ_m such that, for all h, $0 < h \leqslant 1$, and all $f \in \mathcal{W}_m$,*

$$(\gamma_m)^{-1}\, \|\, f\,\|_{h,\mathcal{W}_m} \leqslant |\!|\!|\, f\,|\!|\!|_{h,\mathcal{W}_m} \leqslant \gamma_m\, \|\, f\,\|_{h,\mathcal{W}_m}\ .$$

Proof. By Lemma (6.17), we obviously have

$$1 + (2h)^{2m}\, \lambda_{k,m} \leqslant c_m \bigl(1 + (2hk)^{2m} \bigr)\ ,$$

with the same c_m as in Lemma (6.17). Thus, $c_m^{-1}\, \|\, f\,\|_{h,\mathcal{W}_m} \leqslant |\!|\!|\, f\,|\!|\!|_{h,\mathcal{W}_m}$.
Also, for $k \geqslant m$, the lower bound of Lemma (6.17) on $\lambda_{k,m}$ is useful. For $0 < h \leqslant 1$ and $0 \leqslant k < m$, we have

$$\frac{1 + (2hk)^{2m}}{1 + (2m)^{2m}} \leqslant 1 = 1 + (2h)^{2m}\, \lambda_{k,m}\ ,$$

so that with $\gamma_m = \max\{\, c_m\,,\, 1 + (2m)^{2m}\,\}$,

$$|\!|\!|\, f\,|\!|\!|_{h,\mathcal{W}_m} \leqslant \gamma_m\, \|\, f\,\|_{h,\mathcal{W}_m}\ .$$

The lemma follows. Q.e.d.

We are now ready to show that the point evaluations $x \mapsto f(x)$ are bounded linear functionals on $\mathcal{W}_m$; in other words, that the $\mathcal{W}_m$ are reproducing kernel Hilbert spaces. First, define the functions

$$\text{(6.22)} \qquad \Phi_h(x) = \min\bigl(h^{-1/2}\,,\, \{\, x(1-x)\,\}^{-1/4} \bigr)\ .$$

(6.23) Lemma (The case $m = 2$). *There exists a constant c such that, for all h, $0 < h \leqslant 1$, and all $f \in \mathcal{W}_2$,*

$$|\, f(x)\,| \leqslant c_m\, \Phi_h(x)\, h^{-1/2}\, \|\, f\,\|_{h,\mathcal{W}_2} \quad \textit{for all}\ \ x \in (0,1)\ .$$

PROOF. Using the representation of Lemma (6.19) for $f \in \mathcal{W}_m$, we get

$$| f(x) | \leqslant \sum_{k \geqslant 0} | \widehat{f}_k | \, | Q_k(x) | \, .$$

Now, with the second inequality of (6.16),

$$| f(x) | \leqslant c \left\{ x(1-x) \right\}^{-1/4} \sum_{k \geqslant 0} | \widehat{f}_k | \, ,$$

and with Cauchy-Schwarz, the last series is bounded by

$$||| f |||_{h,\mathcal{W}_2} \left\{ \sum_{k \geqslant 0} \left(1 + (2hk)^4 \right)^{-1} \right\}^{1/2} .$$

Now, the infinite series is dominated by

$$\int_0^\infty \left(1 + (2hx)^4 \right)^{-1} dx = h^{-1} \int_0^\infty \left(1 + (2\,t\,)^4 \right)^{-1} d\,t \, = c\,h^{-1} \, ,$$

for a suitable constant. Thus,

$$| f(x) | \leqslant c\,h^{-1/2} \left\{ x(1-x) \right\}^{-1/4} ||| f |||_{h,\mathcal{W}_2} \, . \tag{6.24}$$

For all x, we use the first bound of (6.16). With Cauchy-Schwarz, this gives the bound

$$| f(x) | \leqslant ||| f |||_{h,\mathcal{W}_2} \left\{ \sum_{k \geqslant 0} \frac{2k+1}{1+(2hk)^4} \right\}^{1/2} .$$

Now, we may drop the $+1$ in the numerator, and then the infinite series behaves like

$$\int_0^\infty \frac{2x}{1+(2hx)^4} \, dx = h^{-2} \int_0^\infty \frac{2\,t}{1+(2\,t\,)^4} \, d\,t \, = c\,h^{-2}$$

for (another) constant c. Thus,

$$| f(x) | \leqslant c\,h^{-1} ||| f |||_{h,\mathcal{W}_2} \, . \tag{6.25}$$

By the equivalence of the norms, uniformly in h, $0 < h \leqslant 1$, the lemma follows from (6.24) and (6.25). Q.e.d.

(6.26) LEMMA. *For all $m \geqslant 1$, there exists a constant c_m such that, for all $f \in \mathcal{W}_{m+1}$ and all h, $0 < h \leqslant 1$,*

$$\| f \|_{h,\mathcal{W}_m} \leqslant c_m \| f \|_{h,\mathcal{W}_{m+1}} \, .$$

PROOF. With the representation of Lemma (6.19) and (6.20),

$$||| f |||^2_{h,\mathcal{W}_m} \leqslant c\, ||| f |||^2_{h,\mathcal{W}_{m+1}} \, ,$$

where

$$c = \sup \left\{ \frac{1+(2hk)^{2m}}{1+(2hk)^{2m+2}} \;\middle|\; k \geqslant 0, \ 0 < h \leqslant 1 \right\} \leqslant \sup_{t>0} \frac{1+\,t^{2m}}{1+\,t^{2m+2}} < \infty \, .$$

Together with the equivalence of the norms, that is all that there is to it. Q.e.d.

The final result involving the Legendre polynomials or, more to the point, the equivalent norms, is an integration-by-parts formula.

(6.27) LEMMA. *Let* $m \geqslant 1$. *For all* $f \in \mathcal{W}_m$ *and all* $g \in \mathcal{W}_{2m}$,

$$\langle f , g \rangle_{\mathcal{W}_m} \leqslant \| f \| \, \| g \|_{1,\mathcal{W}_{2m}} \ .$$

PROOF. Using the representation of Lemma (6.19), and Lemma (6.17), the inner product may be written as, and then bounded by,

$$\sum_{k \geqslant m} \left(1 + 2^{2m} \lambda_{k,m} \right) \widehat{f}_k \, \widehat{g}_k \leqslant c_m \sum_{k \geqslant m} \left(1 + (2k)^{2m} \right) | \widehat{f}_k | \, | \widehat{g}_k | \ .$$

Now, with Cauchy-Schwarz, the right-hand side may be bounded by

$$\| f \| \left\{ \sum_{k \geqslant 0} \left(1 + (2k)^{2m} \right)^2 | \widehat{g}_k |^2 \right\}^{1/2} ,$$

and, in turn, the infinite series may be bounded by

$$2 \sum_{k \geqslant 0} \left(1 + (2k)^{4m} \right) | \widehat{g}_k |^2 = 2 \, ||| \, g \, |||^2_{1,\mathcal{W}_{2m}} \ .$$ Q.e.d.

(6.28) REMARK. The reason we called Lemma (6.27) an integration-by-parts formula is because it is. Recall that

$$\langle f , g \rangle_{\mathcal{W}_m} = \int_0^1 \{ x(1-x) \}^m f^{(m)}(x) \, g^{(m)}(x) \, dx \ ,$$

so that integrating by parts m times gives

$$\int_0^1 g^{(2m)}(x) \, (-D)^m \Big\{ \{ x(1-x) \}^m f^{(m)}(x) \Big\} \, dx \ ,$$

where D denotes differentiation with respect to x, *provided* the boundary terms vanish. Showing that they do is harder than it looks (e.g., are the boundary values actually defined?), but the expansion in Legendre polynomials avoids the issue.

Quadrature. The last technical result deals with quadrature. The only hard part is an embedding result where apparently, the Legendre polynomials are of no use. We must slug it out; cf. the proof of Lemma (2.10).

(6.29) EMBEDDING LEMMA. *There exists a constant* c *such that*

$$\| f \|_{W^{1,2}(0,1)} \leqslant c \, \| f \|_{1,\mathcal{W}_2} \quad \textit{for all } f \in \mathcal{W}_2 \, .$$

PROOF. For $x,\, y \in (0,1)$, with $y < x$,

$$| f'(x) - f'(y) | \leqslant \int_y^x | f''(t) | \, dt \leqslant c(x,y) \, | f |_{\mathcal{W}_2}$$

with
$$c^2(x,y) = \int_y^x [\, t(1-t)\,]^{-2} \, dt \ .$$

It follows that, for any closed subinterval $[\, a\,,\, b\,] \subset (0,1)$, we have

$$c(x,y) \leqslant C(a,b) \, | x - y |^{1/2} \quad \text{for } x,\, y \in [\, a\,,\, b\,] \ ,$$

for a suitable constant $C(a,b)$. Thus, f' is continuous in $(0,1)$.

Now, let $M = \left(\frac{1}{4}, \frac{3}{4}\right)$, and choose $y \in M$ such that

$$| f'(y) |^2 = 2 \, \| f' \|_M^2$$

in the notation of (2.10). This is possible by the Mean Value Theorem. Then,

$$\int_0^y | f'(t) |^2 \, dt \leqslant 2 \, | f'(y) |^2 + 2 \int_0^y | f'(t) - f'(y) |^2 \, dt \ .$$

Now, by Hardy's inequality (see Lemma (6.31) and Exercise (6.32) below),

$$\begin{aligned} \int_0^y | f'(t) - f'(y) |^2 \, dt &\leqslant 4 \int_0^y t^2 \, | f''(t) |^2 \, dt \\ &\leqslant 64 \int_0^y [\, t(1-t)\,]^2 \, | f''(t) |^2 \, dt \, . \end{aligned}$$

Also, $| f'(y) |^2 = 2 \, \| f' \|_M^2$. By the Interpolation Lemma (2.12) with $h = 1$, we have the bound

$$\| f' \|_M^2 \leqslant c \, \| f \|_M^2 + c_1 \, \| f'' \|_M^2 \tag{6.30}$$

for constants c and c_1 independent of y (since y is bounded away from 0). Since, on the interval M, the weight function $\{\, x(1-x)\,\}^2$ is bounded from below by $\frac{9}{256}$, then

$$\int_{\frac{1}{4}}^{\frac{3}{4}} | f''(t) |^2 \, dt \leqslant c \int_{\frac{1}{4}}^{\frac{3}{4}} \{\, x(1-x)\,\}^2 \, | f''(t) |^2 \, dt$$

with $c = 256/9$, so that $\| f'' \|_M^2 \leqslant c \, | f |_{\mathcal{W}_2}^2$ and we obtain

$$\| f' \|_{(0,y)}^2 \leqslant c \, \| f \|_{1,\mathcal{W}_2}^2 \ .$$

The same bound applies to $\| f' \|_{(y,1)}^2$. The lemma follows. Q.e.d.

To prove the version of Hardy's inequality alluded to above, we quote the following result from HARDY, LITTLEWOOD, and POLYA (1951).

(6.31) LEMMA. *Let* $K : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ *be homogeneous of degree* -1*; i.e.,* $K(tx, ty) = t^{-1} K(x,y)$ *for all nonnegative* t, x, *and* y. *Then, for all* $f \in L^2(\mathbb{R}^+)$,

$$\int_{\mathbb{R}^+} \Bigl| \int_{\mathbb{R}^+} K(x,y)\, f(y)\, dy \Bigr|^2 dx \leqslant k \, \| f \|^2_{L^2(\mathbb{R}^+)} ,$$

where $k = \displaystyle\int_{\mathbb{R}^+} y^{-1/2} K(1,y)\, dy$.

PROOF. For nonnegative $f, g \in L^2(\mathbb{R}^*)$,

$$\begin{aligned}
&\int_{\mathbb{R}^+} f(x) \int_{\mathbb{R}^+} K(x,y)\, g(y)\, dy\, dx \\
&\qquad = \int_{\mathbb{R}^+} f(x) \int_{\mathbb{R}^+} x\, K(x,xy)\, g(xy)\, dy\, dx \quad \text{(change of variable)} \\
&\qquad = \int_{\mathbb{R}^+} f(x) \int_{\mathbb{R}^+} K(1,y)\, g(xy)\, dy\, dx \quad \text{(homogeneity)} \\
&\qquad = \int_{\mathbb{R}^+} K(1,y) \int_{\mathbb{R}^+} f(x)\, g(xy)\, dx\, dy \quad \text{(Fatou)} .
\end{aligned}$$

Now, with Cauchy-Schwarz,

$$\begin{aligned}
\int_{\mathbb{R}^+} f(x)\, g(xy)\, dx\, dy &\leqslant \| f \|_{L^2(\mathbb{R}^+)} \Bigl\{ \int_{\mathbb{R}^+} | g(xy) |^2 dx \Bigr\}^{1/2} \\
&\leqslant y^{-1/2} \| f \|_{L^2(\mathbb{R}^+)} \| g \|_{L^2(\mathbb{R}^+)} ,
\end{aligned}$$

the last equality by a change of variable. Thus, for all nonnegative f, $g \in L^2(\mathbb{R}^+)$,

$$\int_{\mathbb{R}^+} f(x) \int_{\mathbb{R}^+} K(x,y)\, g(y)\, dy\, dx \leqslant k \, \| f \|_{L^2(\mathbb{R}^+)} \| g \|_{L^2(\mathbb{R}^+)} ,$$

with the constant k as advertised. Obviously, then this holds also for *all* $f, g \in L^2(\mathbb{R}^+)$. Finally, take

$$f(x) = \int_{\mathbb{R}^+} K(x,y)\, g(y)\, dy , \quad x \in \mathbb{R}^+ ,$$

and we are in business. Q.e.d.

(6.32) EXERCISE. (a) Show that the function

$$K(x,y) = y^{-1}\, \mathbb{1}(\, x < y\,) , \quad x, y > 0 ,$$

is homogeneous of degree -1.

(b) Show the following consequence of Lemma (6.31): For all integrable functions f on $\mathbb{R}^+$,

$$\int_{\mathbb{R}^+} \Bigl| \int_x^\infty y^{-1} f(y)\, dy \Bigr|^2 dx \leqslant 4 \int_{\mathbb{R}^+} x^2 \, | f(x) |^2 dx .$$

(c) Use (b) to show that, for all functions f with a measurable derivative,

$$\int_0^\infty |\, f(t)\,|^2\, dt \leqslant 4 \int_0^\infty \ t^2\, |\, f'(t)\,|^2\, dt \ .$$

(d) Use (c) to show that, for all functions f with a measurable derivative,

$$\int_0^T |\, f(t) - f(T)\,|^2\, dt \leqslant 4 \int_0^T \ t^2\, |\, f'(t)\,|^2\, dt \ .$$

(6.33) Lemma. *Let $m \geqslant 2$. For asymptotically uniform designs, there exists a constant c_m such that, for all $f \in \mathcal{W}_m$ and all h, $0 < h \leqslant 1$,*

$$\Bigl|\, \tfrac{1}{n} \sum_{i=1}^n |\, f(x_{in})\,|^2 - \|\, f\,\|^2 \,\Bigr| \leqslant c_m\, (\, nh^2\,)^{-1}\, \|\, f\,\|_{h,\mathcal{W}_m} \ .$$

Proof. By Definition (2.22), the left-hand side is bounded by

$$c\, n^{-1}\, \|\, (\, f^2\,)'\,\|_1 \ .$$

Now,

$$\|\, (\, f^2\,)'\,\|_1 \leqslant 2\,\|\, f\,\|\,\|\, f'\,\| \leqslant 2\,\|\, f\,\|\,\|\, f\,\|_{W^{1,2}(0,1)} \leqslant c\,\|\, f\,\|\,\|\, f\,\|_{1,\mathcal{W}_2} \ ,$$

the last inequality by Lemma (6.29). Finally for $0 < h \leqslant 1$,

$$\|\, f\,\|_{1,\mathcal{W}_2} \leqslant h^{-2}\, \|\, f\,\|_{h,\mathcal{W}_2} \ ,$$

and of course $\|\, f\,\| \leqslant \|\, f\,\|_{h,\mathcal{W}_2}$. Thus,

$$\|\, (\, f^2\,)'\,\|_1 \leqslant c\, h^{-2}\, \|\, f\,\|_{h,\mathcal{W}_2}^2 \ . \tag{6.34}$$

The lemma then follows from Lemma (6.26). Q.e.d.

(6.35) Remark. The inequality (6.34) does not appear to be sharp as far as the rate h^{-2} is concerned. The example $f(x) = (\, x - \lambda\,)_+^2$ (for appropriate λ) shows that the rate $h^{-3/2}$ may apply. Verify this. What is the best possible rate? (The authors do not know.)

Reproducing kernels. We finally come to the existence of the reproducing kernels, implied by Lemmas (6.23) and (6.26), and its consequences for random sums.

(6.36) Theorem [Reproducing kernel Hilbert spaces]. *Let $m \geqslant 2$. Then, $\mathcal{W}_m$ is a reproducing kernel Hilbert space with reproducing kernels $R_{m,h}(x, y)$, $x, y \in [\,0\,,\,1\,]$, so that, for all $f \in \mathcal{W}_m$,*

$$f(x) = \bigl\langle\!\bigl\langle\, R_{m,h}(x, \cdot\,)\,,\, f\,\bigr\rangle\!\bigr\rangle_{h,\mathcal{W}_m} \ , \quad \textit{for all } x \in [\,0\,,\,1\,]\ ,$$

and, for a suitable constant c_m not depending on h,

$$\|\, R_{m,h}(x, \cdot\,)\,\|_{h,\mathcal{W}_m} \leqslant c_m\, \Phi_h(x)\, h^{-1/2} \ .$$

The reproducing kernel Hilbert space setting has consequences for random sums. For the noise vector d_n and design x_{in}, $i = 1, 2, \cdots, n$, let

$$(6.37) \qquad S^{nh}(t) = \tfrac{1}{n} \sum_{i=1}^{n} d_{in}\, R_{mh}(x_{in}, t)\,, \quad t \in [0, 1]\,.$$

(6.38) Lemma. *Let* $m \geqslant 2$. *Then, for all* $f \in \mathcal{W}_m$ *and* h, $0 < h \leqslant 1$, *and for all designs,*

$$\Big| \tfrac{1}{n} \sum_{i=1}^{n} d_{in}\, f(x_{in}) \Big| \leqslant \| f \|_{h,\mathcal{W}_m} \, \| S^{nh} \|_{h,\mathcal{W}_m} \;.$$

If, moreover, the noise vector d_n *satisfies* (1.2) *and the design is asymptotically uniform, then there exists a constant* c *such that, for all* n *and all* h, $0 < h \leqslant 1$, *with* $nh^2 \to \infty$,

$$\mathbb{E}[\, \| S^{nh} \|^2_{h,\mathcal{W}_m} \,] \leqslant c\,(nh)^{-1} \;.$$

Proof. For the first inequality, use the representation

$$f(x_{in}) = \langle\!\langle\, f\,,\, R_{mh}(x_{in}, \cdot\,)\,\rangle\!\rangle_{h.\mathcal{W}_m}$$

to see that the sum equals $\langle\!\langle\, f\,,\, S^{nh}\,\rangle\!\rangle_{h.\mathcal{W}_m}$. Then Cauchy-Schwarz implies the inequality.

For the second inequality, note that the expectation equals

$$n^{-2} \sum_{i,j=1}^{n} \mathbb{E}[\, d_{in}\, d_{jn}\,]\, \langle\!\langle\, R_{mh}(x_{in}, \cdot\,)\,,\, R_{mh}(x_{jn}, \cdot\,)\,\rangle\!\rangle_{h.\mathcal{W}_m} \;.$$

By the assumption (1.2) and Theorem (6.36), this equals and may be bounded as

$$(6.39) \quad \sigma^2\, n^{-2} \sum_{i=1}^{n} \|\, R_{mh}(x_{in}, \cdot\,) \,\|^2_{h,\mathcal{W}_m} \leqslant c\,(nh)^{-1} \cdot \tfrac{1}{n} \sum_{i=1}^{n} |\, \Phi_h(x_{in}) \,|^2$$

for a suitable constant c. By the asymptotic uniformity of the design, see Definition (2.22), we have

$$(6.40) \qquad \tfrac{1}{n} \sum_{i=1}^{n} |\, \Phi_h(x_{in}) \,|^2 \leqslant \|\, \Phi_h \,\|^2 + c\, n^{-1/2}\, \|\, \Phi_h^2 \,\|_{W^{1,1}(0,1)} \;.$$

Now, $\|\, \Phi_h^2 \,\|_1 = \|\, \Phi_h \,\|^2$ and

$$\|\, \Phi_h \,\|^2 \leqslant \int_0^1 \{\, x(1-x) \,\}^{-1/2}\, dx = \pi \;.$$

Also,

$$\|\, \{\, \Phi_h^2 \,\}' \,\|_1 = \int_a^{1-a} \Big|\, \frac{d}{dx} \{\, x(1-x) \,\}^{-1/2} \Big|\, dx \;,$$

where a is the smallest solution of

$$h^{-1/2} = \{\, x(1-x)\,\}^{-1/4} \ .$$

So, $a \asymp h^2$. Now, $\{\, x(1-x)\,\}^{-1/4}$ is decreasing on $(\, a\,,\, \frac{1}{2}\,)$, and so

$$\int_a^{\frac{1}{2}} \left| \frac{d}{dx} \{\, x(1-x)\,\}^{-1/2} \right| dx = \{\, a(1-a)\,\}^{-1/2} - 2 \asymp h^{-1} \ ,$$

and the same bound applies to the integral over $(\, \frac{1}{2}\,,\, 1-a\,)$. To summarize, all of this shows that, for a suitable constant c,

$$\|\, \Phi_h \,\|^2 \leqslant \pi \quad , \quad \|\, \{\, \Phi_h^2 \,\}' \,\|_1 \leqslant c\, h^{-1} \ , \quad \text{and so} \quad \|\, \Phi_h^2 \,\|_{W^{1,1}(0,1)} \leqslant c\, h^{-1} \ ,$$

and then (6.40) shows that

$$\tfrac{1}{n} \sum_{i=1}^{n} |\, \Phi_h(x_{in}) \,|^2 \leqslant \pi + c\, (nh^2)^{-1/2} \ .$$

This implies the advertised bound on $\mathbb{E}[\, \|\, S^{nh} \,\|^2_{h,\mathcal{W}_m}\,]$. Q.e.d.

(6.41) Exercise. Some of the results in this section also hold for $m = 1$. (a) Show that, (6.24) holds for $m = 1$ and that instead of the uniform bound we have

$$|\, f(x) \,| \leqslant c\, h^{-1} \left| \log\{\, x(1-x)\,\} \right|^{1/2} |||\, f \,|||_{h,\mathcal{W}_1} \ .$$

(b) Show that, for $a = 1/n$,

$$\int_a^{1-a} |\, \log\{\, x(1-x)\,\} \,|^{1/2} \,|\, f\,'(x) \,|\, dx \leqslant c\, (\, \log n\,)^2 \,|\, f \,|_{\mathcal{W}_1} \ .$$

(c) Prove the case $m = 1$ of Lemma (6.38) for the designs

$$x_{in} = i/(n-1) \quad \text{and} \quad x_{in} = (i - \tfrac{1}{2})/(n-1) \ , \quad i = 1, 2, \cdots, n \, .$$

Indeed, for $m = 1$, the requirement on the designs is the asymptotic uniformity of Definition (2.22) together with the assumption that

$$\text{(d)} \qquad \sup_{n \geqslant 1} \ \tfrac{1}{n} \sum_{i=1}^{n} \{\, x_{in}\,(\, 1 - x_{in}\,)\,\}^{-1/2} < \infty \ .$$

[Hint: For (a), proceed analogously to the proof of Lemma (2.10). For (b), Cauchy-Schwarz does it. For (c), use (6.39), but with Φ_h replaced by $\Psi_h(x) = \{\, x\,(\, 1-x\,)\,\}^{-1/4}$.]

Exercises: (6.6), (6.32), (6.40).

7. Existence, uniqueness, and rates

In this section, we actually prove Theorems (6.7) and (6.8). This pretty much goes along the lines of §§ 3 and 4. We start out with the quadratic behavior.

(7.1) LEMMA. *Let* $m \geqslant 1$, *and let* f^{nh} *be a solution of* (6.4). *Then, for all* $f \in \mathcal{W}_m$ *and all* $h > 0$,

$$\frac{1}{n} \sum_{i=1}^{n} |\, f(x_{in}) - f^{nh}(x_{in})\,|^2 + h^{2m}\,|\, f - f^{nh}\,|^2_{\mathcal{W}_m} =$$
$$\frac{1}{n} \sum_{i=1}^{n} \bigl(\, f(x_{in}) - y_{in}\,\bigr)\bigl(\, f(x_{in}) - f^{nh}(x_{in})\,\bigr) + h^{2m} \bigl\langle\, f\,,\, f - f^{nh}\,\bigr\rangle_{\mathcal{W}_m} \;.$$

(7.2) UNIQUENESS LEMMA. *Let* $m \geqslant 1$, *and suppose that the design contains at least* m *distinct points. Then the solution of* (6.4) *is unique.*

(7.3) EXERCISE. Prove it. [Hint: Copy the proofs of Lemmas (3.1) and (3.6) with some cosmetic changes.]

We go on to prove the convergence rates of Theorem (6.8).

PROOF OF THEOREM (6.8). The starting point is the quadratic behavior of Lemma (7.1). After the usual manipulations with

$$\varepsilon^{nh} = f^{nh} - f_o \;, \tag{7.4}$$

this gives the equality

$$\frac{1}{n} \sum_{i=1}^{n} |\,\varepsilon^{nh}(x_{in})\,|^2 + h^{2m}\,|\,\varepsilon^{nh}\,|^2_{\mathcal{W}_m} = \frac{1}{n} \sum_{i=1}^{n} d_{in}\,\varepsilon^{nh}(x_{in}) - h^{2m} \bigl\langle\, f_o\,,\,\varepsilon^{nh}\,\bigr\rangle_{\mathcal{W}_m} \;. \tag{7.5}$$

Now, we just need to apply the appropriate results.

For the bias part, note that by Lemma (6.27)

$$- h^{2m} \bigl\langle\, f_o\,,\,\varepsilon^{nh}\,\bigr\rangle_{\mathcal{W}_m} \leqslant h^{2m}\,\|\,\varepsilon^{nh}\,\|\,\|\, f_o\,\|_{1,\mathcal{W}_{2m}} \;. \tag{7.6}$$

For the random sum on the right of (7.5), we use Lemma (6.38), so

$$\frac{1}{n} \sum_{i=1}^{n} d_{in}\,\varepsilon^{nh}(x_{in}) \leqslant \|\,\varepsilon^{nh}\,\|_{h,\mathcal{W}_m}\,\|\,S^{nh}\,\|_{h,\mathcal{W}_m} \;. \tag{7.7}$$

For the sum on the left-hand side of (7.5), Lemma (6.33) provides the lower bound

$$\|\,\varepsilon^{nh}\,\|^2 - c\,(nh^2)^{-1}\,\|\,\varepsilon^{nh}\,\|_{h,\mathcal{W}_m} \;,$$

so that

$$\zeta^{nh}\,\|\,\varepsilon^{nh}\,\|^2_{h,\mathcal{W}_m} \leqslant \frac{1}{n} \sum_{i=1}^{n} |\,\varepsilon^{nh}(x_{in})\,|^2 + h^{2m}\,|\,\varepsilon^{nh}\,|^2_{\mathcal{W}_m} \;, \tag{7.8}$$

with $\zeta^{nh} \to 1$, provided $nh^2 \to \infty$.

Substituting (7.6), (7.7), and (7.6) into (7.5) results in

$$\zeta^{nh} \, \| \, \varepsilon^{nh} \, \|^2_{h,\mathcal{W}_m} \leqslant \| \, S^{nh} \, \|_{h,\mathcal{W}_m} \, \| \, \varepsilon^{nh} \, \|_{h,\mathcal{W}_m} + c \, h^{2m} \, \| \, \varepsilon^{nh} \, \| \, \| \, f_o \, \|_{1,\mathcal{W}_{2m}} \ ,$$

and the right-hand side may be bounded by

$$\| \, \varepsilon^{nh} \, \|_{h,\mathcal{W}_m} \, \big\{ \, \| \, S^{nh} \, \|_{h,\mathcal{W}_m} + c \, h^{2m} \, \| \, f_o \, \|_{1,\mathcal{W}_{2m}} \, \big\} \ ,$$

so that, for a different c,

$$\| \, \varepsilon^{nh} \, \|^2_{h,\mathcal{W}_m} \leqslant c \, \| \, S^{nh} \, \|^2_{h,\mathcal{W}_m} + \ c \, h^{4m} \, \| \, f_o \, \|^2_{1,\mathcal{W}_{2m}} \ .$$

Finally, Lemma (6.38) gives that

$$\mathbb{E}\big[\, \| \, \varepsilon^{nh} \, \|^2_{h,\mathcal{W}_m} \, \big] = \mathcal{O}\big(\, (nh)^{-1} + h^{4m} \, \big) \ ,$$

provided, again, that $nh^2 \to \infty$. For the optimal choice $h \asymp n^{-1/(4m+1)}$, this is indeed the case, and then

$$\mathbb{E}\big[\, \| \, \varepsilon^{nh} \, \|^2_{h,\mathcal{W}_m} \, \big] = \mathcal{O}\big(\, n^{-4m/(4m+1)} \, \big) \ .$$

This completes the proof. Q.e.d.

Finally, we prove the existence of the solution of (6.4). The following (compactness) result is useful. Define the mapping $T : \mathcal{W}_m \to C[\,0\,,\,1\,]$ by

$$\text{(7.9)} \qquad Tf(x) = \int_{\frac{1}{2}}^{x} (\, x - \, t \,)^{m-1} \, f^{(m)}(\, t \,) \, dt \ , \quad x \in [\,0\,,\,1\,] \ .$$

(7.10) Lemma. *Let $m \geqslant 2$. There exists a constant c such that, for all $f \in \mathcal{W}_m$ and all $x, \, y \in [\,0\,,\,1\,]$,*

$$\big| \, Tf(x) - Tf(y) \, \big| \leqslant c \, | \, y - x \, |^{1/2} \, | \, f \, |_{\mathcal{W}_m} \ .$$

Proof. First we show that T is bounded. Let $0 < x \leqslant \frac{1}{2}$. Note that

$$\big| \, Tf(x) \, \big|^2 \leqslant c(x) \left| \, \int_{\frac{1}{2}}^{x} \big\{ \, t \, (1 - t \,) \, \big\}^m \, | \, f^{(m)}(\, t \,) \, |^2 \, dt \, \right|$$

with

$$c(x) = \left| \, \int_{\frac{1}{2}}^{x} \frac{(\, x - \, t \,)^{2m-2}}{\big\{ \, t \, (1 - t \,) \, \big\}^m} \, dt \, \right| \ .$$

Now, for $0 < x < \, t \, \leqslant \frac{1}{2}$, we have

$$0 \leqslant \frac{(\, t \, - x \,)^{2m-2}}{\big\{ \, t \, (1 - t \,) \, \big\}^m} \leqslant 2^m \, (\, t \, - x \,)^{m-2} \leqslant 4$$

since $m \geqslant 2$. It follows that $c(x) \leqslant 2$ on the interval $0 \leqslant x \leqslant \frac{1}{2}$. The same argument applies to the case $x \geqslant \frac{1}{2}$, so that

$$|\,Tf(x)\,| \leqslant 2\,|\,f\,|_{\mathcal{W}_m} \ . \tag{7.11}$$

Thus, T is a bounded linear mapping from $\mathcal{W}_m$ into $L^\infty(0,1)$ in the $|\cdot|_{\mathcal{W}_m}$ topology on $\mathcal{W}_m$.

Let $f \in \mathcal{W}_m$, and set $g = Tf$. Then, from (7.11), the function g is bounded, so surely $\|\,g\,\| \leqslant 2\,|\,f\,|_{\mathcal{W}_m}$. Of course, $g^{(m)} = f^{(m)}$ (almost everywhere), so that $|\,g\,|_{\mathcal{W}_m} = |\,f\,|_{\mathcal{W}_m}$. It follows that

$$\|\,g\,\|_{1,\mathcal{W}_m} \leqslant 3\,|\,f\,|_{\mathcal{W}_m} \ .$$

Now, by the Embedding Lemma (6.29), for a suitable constant c,

$$\|\,g'\| \leqslant c\,\|\,g\,\|_{1,\mathcal{W}_2} \leqslant c_m\,\|\,g\,\|_{1,\mathcal{W}_m} \leqslant \widetilde{c}\,|\,f\,|_{\mathcal{W}_m} \ .$$

It follows that, for all $x,\ y \in [\,0\,,\,1\,]$,

$$|\,g(x) - g(y)\,| = \left|\int_y^x g'(\,t\,)\,dt\right| \leqslant |\,x-y\,|^{1/2}\,\|\,g'\,\| \leqslant \widetilde{c}\,|\,x-y\,|^{1/2}\,|\,f\,|_{\mathcal{W}_m} \ ,$$

as was to be shown. Q.e.d.

(7.12) Corollary. *Let $m \geqslant 2$. Then the mapping $T : \mathcal{W}_m \to C[\,0\,,\,1\,]$ is compact in the $|\cdot|_{\mathcal{W}_m}$ topology on $\mathcal{W}_m$.*

Proof. This follows from the Arzelà-Ascoli theorem. Q.e.d.

(7.13) Lemma. *Let $m \geqslant 2$. Then the relaxed boundary smoothing problem* (6.4) *has a solution.*

Proof. Obviously, the objective function $RLS(f)$ of (6.4) is bounded from below (by 0), so there exists a minimizing sequence, denoted by $\{\,f_k\,\}_k$. Then, obviously,

$$h^{2m}\,|\,f_k\,|^2_{\mathcal{W}_m} \leqslant RLS(f_k) \leqslant RLS(f_1) \ ,$$

the last inequality without loss of generality. Thus, there exists a subsequence, again denoted by $\{\,f_k\,\}_k$, for which $\{\,f_k^{(m)}\,\}_k$ converges, in the weak topology on $\mathcal{W}_m$ induced by the $|\cdot|_{\mathcal{W}_m}$ semi-norm, to some element φ_o. Then,

$$|\,\varphi_o\,|^2_{\mathcal{W}_m} \leqslant \liminf_{k\to\infty}\,|\,f_k\,|^2_{\mathcal{W}_m} \ ,$$

and by the compactness of T in this setting, then

$$\lim_{k\to\infty}\,\|\,Tf_k - T\varphi_o\,\|_\infty = 0 \ .$$

Finally, use Taylor's theorem with exact remainder to write

$$f_k(x) = p_k(x) + Tf_k(x)$$

for suitable polynomials p_k. Now, proceed as in the proof of the Existence Lemma (3.7) for smoothing splines. Consider the restrictions of the f_k to the design points,

$$r_n f_k \stackrel{\text{def}}{=} \big(f_k(x_{1,n}), f_k(x_{2,n}), \cdots , f_k(x_{n,n}) \big) , \quad k = 1, 2, \cdots .$$

We may extract a subsequence from $\{ f_k \}_k$ for which $\{ r_n f_k \}_k$ converges in $\mathbb{R}^n$ to some vector v_o. Then, for the corresponding polynomials,

$$\lim_{k\to\infty} p_k(x_{in}) = [\, v_o \,]_i - T\varphi_o(x_{in}) , \quad i = 1, 2, \cdots , n ,$$

and there exists a polynomial p_o of order m such that

(7.14) $$p_o(x_{in}) = [\, v_o \,]_i - T\varphi_o(x_{in}) , \quad i = 1, 2, \cdots , n .$$

Finally, define $\psi_o = p_o + T\varphi_o$, and then, for the (subsub) sequence in question,

$$\lim_{k\to\infty} \mathit{RLS}(f_k) \geqslant \mathit{RLS}(\psi_o) ,$$

so that ψ_o minimizes $\mathit{RLS}(f)$ over $f \in \mathcal{W}_m$. Q.e.d.

(7.15) Exercise. Prove Theorem (6.8) for the case $m = 1$ when the design is asymptotically uniform in the sense of Definition (2.22) and satisfies condition (d) of Exercise (6.41).

Exercises: (7.3), (7.15).

8. Partially linear models

The gravy train of the statistical profession is undoubtedly data analysis by means of the linear model

(8.1) $$y_{in} = x_{in}^T \beta_o + d_{in} , \quad i = 1, 2, \cdots , n ,$$

in which the vectors $x_{in} \in \mathbb{R}^d$ embody the design of the experiment (d is some fixed integer $\geqslant 1$), $\beta_o \in \mathbb{R}^d$ are the unknown parameters to be estimated, $y_n = (y_{1,n}, y_{2,n}, \cdots , y_{n,n})^T$ are the observed response variables, and the collective noise is $d_n = (d_{1,n}, d_{2,n}, \cdots , d_{n,n})^T$, with independent components, assumed to be normally distributed,

(8.2) $$d_n \sim \text{Normal}(0, \sigma^2 I) ,$$

with σ^2 unknown. The model (8.1) may be succinctly described as

(8.3) $$y_n = X_n \beta_o + d_n ,$$

with the design matrix $X_n = (x_{1,n} \,|\, x_{2,n} \,|\, \cdots \,|\, x_{n,n})^T \in \mathbb{R}^{n\times d}$.

If X_n has full column rank, then the maximum likelihood estimator of β_o is given by

(8.4) $$\widehat{\beta}_n = (\, X_n^T X_n)^{-1} X_n^T y_n$$

and is normally distributed,

(8.5) $$\sqrt{n}\,(\,\widehat{\beta}_n - \beta_o\,) \sim N\big(0, \sigma^2\,(\,X_n^T X_n\,)^{-1}\,\big) \ ,$$

and the train is rolling.

In this section, we consider the *partially* linear model

(8.6) $$y_{in} = z_{in}^T\,\beta_o + f_o(x_{in}) + d_{in} \ , \quad i = 1, 2, \cdots, n \ ,$$

where $z_{in} \in \mathbb{R}^d$, $x_{in} \in [\,0\,,\,1\,]$ (as always), the function f_o belongs to $W^{m,2}(0,1)$ for some integer $m \geqslant 1$, and

(8.7) the d_{in} are iid, zero-mean random variables with a finite fourth moment.

In analogy with (8.3), this model may be described as

(8.8) $$y_n = Z_n\,\beta_o + r_n f_o + d_n \ ,$$

with $r_n f_o = \big(\,f_o(x_{1,n}), f_o(x_{2,n}), \cdots, f_o(x_{n,n})\,\big)^T$. Thus, r_n is the *restriction operator* from $[\,0\,,\,1\,]$ to the x_{in}.

Such models arise in the standard regression context, where interest is really in the model $y_n = Z_n\,\beta_o + d_n$ but the additional covariates x_{in} cannot be ignored. However, one does not wish to assume that these covariates contribute linearly or even parametrically to the response variable. See, e.g., Engle, Granger, Rice, and Weiss (1986), Green, Jennison, and Seheult (1985), or the introductory example in Heckman (1988). Amazingly, under reasonable (?) conditions, one still gets best asymptotically normal estimators of β_o; that is, asymptotically, the contribution of the nuisance parameter f_o vanishes.

In this section, we exhibit asymptotically normal estimators of β_o and also pay attention to the challenge of estimating f_o at the optimal rate of convergence.

The assumptions needed are as follows. We assume that the x_{in} are deterministic and form a uniformly asymptotic design; e.g., equally spaced as in

(8.9) $$x_{in} = \frac{i-1}{n-1} \ , \quad i = 1, 2, \cdots, n \ .$$

The z_{in} are assumed to be random, according to the model

(8.10) $$z_{in} = g_o(x_{in}) + \varepsilon_{in} \ , \quad i = 1, 2, \cdots, n \ ,$$

in which

(8.11) the ε_{in} are mutually independent, zero-mean random variables, with finite fourth moment, and
$$\mathbb{E}[\,\varepsilon_{in}\,\varepsilon_{in}^T\,] = V \in \mathbb{R}^{d\times d} \ ,$$
with V positive-definite. Moreover, the ε_{in} are independent of d_n in the model (8.6).

In (8.10), $g_o(x) = \mathbb{E}[\, z \,|\, x \,]$ is the conditional expectation of z and is assumed to be a smooth function of x; in particular,

$$g_o \in W^{1,2}(0,1) \ . \tag{8.12}$$

(Precisely, *each component* of g_o belongs to $W^{1,2}(0,1)$.) Regarding f_o, we assume that, for some integer $m \geqslant 1$,

$$f_o \in W^{m,2}(0,1) \ . \tag{8.13}$$

Below, we study two estimators of β_o, both related to smoothing spline estimation. However, since the model (8.3) and the normality result (8.5) constitute the guiding light, the methods and notations used appear somewhat different from those in the previous sections.

The simplest case. To get our feet wet, we begin with the case in which $g_o(x) = 0$ for all x, so that

(8.14)
the z_{in} are mutually independent, zero-mean random variables, with finite fourth moment, independent of the d_{in}, and satisfy
$$\mathbb{E}[\, z_{in}\, z_{in}^T \,] = V \ ,$$
with V positive-definite.

Under these circumstances, by the strong law of large numbers,

$$\tfrac{1}{n}\, Z_n^T Z_n \longrightarrow_{\text{as}} V \ . \tag{8.15}$$

The estimator under consideration goes by the name of the partial spline estimator, the solution to

$$\begin{array}{ll} \text{minimize} & \tfrac{1}{n} \,\|\, Z_n\,\beta + r_n f - y_n \,\|^2 + h^{2m} \,\|\, f^{(m)} \,\|^2 \\ \text{subject to} & \beta \in \mathbb{R}^d \ , \ f \in W^{m,2}(0,1) \ . \end{array} \tag{8.16}$$

One verifies that the solution (β^{nh}, f^{nh}) exists and is unique *almost surely*, and that f^{nh} is an ordinary ("natural") spline function of polynomial order $2m$ with the x_{in} as knots. With (8.4) in mind, we wish to express the objective function in (8.16) in linear algebra terms. For fixed β, the Euler equations (3.18) applied to (8.15) imply that the natural spline function f is completely determined in terms of its function values at the knots, encoded in the vector $r_n f$. Then, there exists a symmetric, semi-positive-definite matrix $M \in \mathbb{R}^{n \times n}$, depending on the knots x_{in} only, such that

$$\|\, f^{(m)} \,\|^2 = (r_n f)^T M\, r_n f \quad \text{for all natural splines } f \ . \tag{8.17}$$

So, the problem (8.16) may be written as

$$\begin{array}{ll} \text{minimize} & \tfrac{1}{n} \,\|\, Z_n\,\beta + r_n f - y_n \,\|^2 + h^{2m}\, (r_n f)^T M\, r_n f \\ \text{subject to} & \beta \in \mathbb{R}^d \ , \ r_n f \in \mathbb{R}^n \ . \end{array} \tag{8.18}$$

Here, the notation $r_n f$ is suggestive but otherwise denotes an arbitrary vector in $\mathbb{R}^n$.

The solution (β, f) to (8.18) is uniquely determined by the normal equations

$$\begin{aligned} Z_n^T\,(\,Z_n\,\beta + r_n f - y_n\,) &= 0\ , \\ (\,I + n\,h^{2m}\,M)\,r_n f + Z_n\,\beta - y_n &= 0\ . \end{aligned} \tag{8.19}$$

Eliminating $r_n f$, we get the explicit form of the partial spline estimator

$$\beta^{nh} = \bigl(\,Z_n^T(\,I - S_h\,)\,Z_n\,\bigr)^{-1}\,Z_n^T(\,I - S_h\,)\,y_n\ , \tag{8.20}$$

in which

$$S_h = (\,I + n\,h^{2m}\,M\,)^{-1} \tag{8.21}$$

is the natural smoothing spline operator. Note that S_h is symmetric and positive-definite. The following exercise is useful.

(8.22) Exercise. Let $\delta_n = (\,\delta_{1,n}, \delta_{2,n}, \cdots, \delta_{n,n})^T \in \mathbb{R}^n$ and let $f = \varphi$ be the solution to

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{n}\sum_{i=1}^{n} |\,f(x_{in}) - \delta_{in}\,|^2 + h^{2m}\,\|\,f^{(m)}\,\|^2 \\ &\text{subject to} \quad f \in W^{m,2}(0,1)\ . \end{aligned}$$

Show that $r_n\varphi = S_h\,\delta_n$.

In view of the model (8.8), we then get that

$$\beta^{nh} - \beta_o = \text{variation} + \text{bias}\ , \tag{8.23}$$

with

$$\begin{aligned} \text{variation} &= \bigl(\,Z_n^T\,(\,I - S_h\,)\,Z_n\,\bigr)^{-1}\,Z_n^T\,(\,I - S_h\,)\,d_n\ , \\ \text{bias} &= \bigl(\,Z_n^T\,(\,I - S_h\,)\,Z_n\,\bigr)^{-1}\,Z_n^T\,(\,I - S_h\,)\,r_n f_o\ . \end{aligned} \tag{8.24}$$

In the above, we tacitly assumed that $Z_n^T\,(\,I - S_h\,)\,Z_n$ is nonsingular. Asymptotically, this holds by (8.15) and the fact that

$$\tfrac{1}{n}\,Z_n^T\,S_h\,Z_n = \mathcal{O}_P\bigl(\,(nh)^{-1}\,\bigr)\ , \tag{8.25}$$

as we show below. The same type of argument shows that

$$\tfrac{1}{n}\,Z_n^T\,S_h\,d_n = \mathcal{O}_P\bigl(\,(nh)^{-1}\,\bigr)\ , \tag{8.26}$$

$$\tfrac{1}{n}\,Z_n^T\,(\,I - S_h\,)\,r_n f_o = \mathcal{O}_P\bigl(\,(nh)^{-1/2}\,h^m\,\bigr)\ . \tag{8.27}$$

This gives

$$\beta^{nh} - \beta_o = (\,Z_n^T\,Z_n)^{-1}\,Z_n^T\,d_n + \mathcal{O}_P\bigl(\,(nh)^{-1} + n^{-1/2}\,h^{m-1/2}\,\bigr)\ ,$$

and the asymptotic normality of $\beta^{nh} - \beta_o$ follows for the appropriate h (but (8.25)–(8.27) need proof).

(8.28) THEOREM. *Under the assumptions* (8.2), (8.7), *and* (8.9)–(8.14),

$$\sqrt{n}\,(\,\beta^{nh} - \beta_o) \longrightarrow_{\mathrm{d}} \Upsilon \sim \mathrm{Normal}(\,0\,,\,\sigma^2\,V\,)\;,$$

provided $h \to 0$ *and* $nh^2 \to \infty$.

Note that Theorem (8.28) says that β^{nh} is asymptotically a minimum variance unbiased estimator.

(8.29) EXERCISE. Complete the proof of the theorem by showing that

$$(\,Z_n^T\,Z_n)^{-1}\,Z_n^T\,d_n \longrightarrow_{\mathrm{d}} U \sim \mathrm{Normal}(\,0\,,\,\sigma^2\,V^{-1}\,)\;.$$

PROOF OF (8.26). Note that $d \geqslant 1$ and $Z_n^T \in \mathbb{R}^{d\times n}$. We actually pretend that $d = 1$.

In accordance with Exercise (8.22), let $\mathfrak{z}^{nh}$ be the natural spline of order $2m$ with the x_{in} as knots satisfying $r_n\mathfrak{z}^{nh} = S_h Z_n$. Then,

$$\|\,\mathfrak{z}^{nh}\,\|_{m,h} = \mathcal{O}_P\big(\,(nh)^{-1/2}\,\big)\;; \tag{8.30}$$

see the Random Sum Lemma (2.20). Now,

$$\tfrac{1}{n}\,Z_n^T S_h\,d_n = \tfrac{1}{n}\,d_n^T S_h\,Z_n = \tfrac{1}{n}\sum_{i=1}^{n} d_{in}\,\mathfrak{z}^{nh}(x_{in})\;,$$

so that in the style of the Random Sum Lemma (2.20),

$$\tfrac{1}{n}\,Z_n^T S_h\,d_n = \Big\langle\,\tfrac{1}{n}\sum_{i=1}^{n} d_{in}\,\mathcal{R}_{mh}(\,\cdot\,,x_{in})\,,\,\mathfrak{z}^{nh}\,\Big\rangle_{m,h}\;,$$

whence

$$\big|\,\tfrac{1}{n}\,Z_n^T S_h\,d_n\,\big| \leqslant \|\,\tfrac{1}{n}\sum_{i=1}^{n} d_{in}\,\mathcal{R}_{mh}(\,\cdot\,,x_{in})\,\|_{m,h}\,\|\,\mathfrak{z}^{nh}\,\|_{m,h}\;. \tag{8.31}$$

Finally, observe that

$$\mathbb{E}\Big[\,\|\,\tfrac{1}{n}\sum_{i=1}^{n} d_{in}\,\mathcal{R}_{mh}(\,\cdot\,,x_{in})\,\|_{m,h}^2\,\Big] = \mathcal{O}\big(\,(nh)^{-1}\,\big) \tag{8.32}$$

by assumption (8.14). Thus, (8.26) follows for $d = 1$. Q.e.d.

(8.33) EXERCISE. Clean up the proof for the case $d \geqslant 2$. Note that $Z_n^T\,S_h\,d_n \in \mathbb{R}^d$, so we need not worry about the choice of norms.

(8.34) EXERCISE. Prove (8.25) for $d = 1$ by showing that

$$\tfrac{1}{n}\,Z_n^T\,S_h\,Z_n = \Big\langle\,\tfrac{1}{n}\sum_{i=1}^{n} z_{in}\,\mathcal{R}_{mh}(\,\cdot\,,x_{in})\,,\,\mathfrak{z}^{nh}\,\Big\rangle_{m,h}\;,$$

with $\mathfrak{z}^{nh}$ as in the proof of (8.26) and properly bounding the expression on the right. Then, do the general case $d \geqslant 2$.

Proof of (8.27). Note that $S_h\, r_n f_o = r_n f_{hn}$, with f_{hn} the solution to the noise-free problem (4.18). Then, the results of Exercises (4.22) and (4.23) imply that $\| f_o - f_{hn} \|_{m,h} = \mathcal{O}(h^m)$. Thus,

$$\begin{aligned} \tfrac{1}{n}\, Z_n^T\, (\, I - S_h\,)\, r_n f_o &= \tfrac{1}{n} \sum_{i=1}^{n} z_{in}\, (\, f_o(x_{in}) - f_{hn}(x_{in})\,) \\ &= \Bigl\langle\, \tfrac{1}{n} \sum_{i=1}^{n} z_{in}\, \mathcal{R}_{mh}(\,\cdot\,, x_{in})\,,\, f_o - f_{hn}\, \Bigr\rangle_{m,h}\ , \end{aligned}$$

so that

$$\bigl|\, \tfrac{1}{n}\, Z_n^T\, (\, I - S_h\,)\, r_n f_o\, \bigr| \leqslant \Bigl\|\, \tfrac{1}{n} \sum_{i=1}^{n} z_{in}\, \mathcal{R}_{mh}(\,\cdot\,, x_{in})\, \Bigr\|_{m,h}\, \|\, f_o - f_{hn}\, \|_{m,h}\ ,$$

and the rest is old hat. Q.e.d.

(8.35) Exercise. Show that, under the conditions of Theorem (8.28),

$$\|\, f^{nh} - f_o\, \|_{m,h}^2 = \mathcal{O}_P\bigl(\, (nh)^{-1} + h^{2m}\,\bigr)\ .$$

Thus, for $h \asymp n^{-1/(2m+1)}$, we get the optimal convergence rate for f^{nh} as well as the asymptotic normality of β^{nh}.

Arbitrary designs. We now wish to see what happens when the z_{in} do not satisfy (8.14) but only (8.11). It will transpire that one can get asymptotic normality of β^{nh} but not the optimal rate of convergence for f^{nh}, at least not at the same time. Thus, the lucky circumstances of Exercise (8.35) fail to hold any longer. However, a fix is presented later.

Again, as estimators we take the solution (β^{nh}, f^{nh}) of (8.18), and we need to see in what form (8.25)–(8.27) hold. When all is said and done, it turns out that (8.27) is causing trouble, as we now illustrate.

It is useful to introduce the matrix G_n,

(8.36) $$G_n = r_n g_o = \bigl[\, g_o(x_{1,n})\, |\, g_o(x_{2,n})\, |\ \cdots\ |\, g_o(x_{n,n})\,\bigr]^T \in \mathbb{R}^{n \times d}\ ,$$

and define

(8.37) $$\widetilde{Z}_n = Z_n - G_n\ .$$

Thus, $\widetilde{Z}_n$ shares the properties of Z_n for the simplest case.

Trying to prove (8.27) for arbitrary designs. Write

$$\tfrac{1}{n}\, Z_n^T\, (\, I - S_h\,)\, r_n f_o = \tfrac{1}{n}\, \widetilde{Z}_n^T\, (\, I - S_h\,)\, r_n f_o + \tfrac{1}{n}\, G_n^T\, (\, I - S_h\,)\, r_n f_o\ .$$

For the first term on the right, we do indeed have

$$\tfrac{1}{n}\, \widetilde{Z}_n^T\, (\, I - S_h\,)\, r_n f_o = \mathcal{O}_P\bigl(\, (nh)^{-1/2}\, h^m\,\bigr)\ ,$$

see the proof of (8.27) for the simplest case.

For the second term, recall that $(I - S_h) r_n f_o = r_n(f_o - f_{hn})$, and this is $\mathcal{O}\bigl(h^m \bigr)$. Thus, we only get

$$\tfrac{1}{n} G_n^T (I - S_h) r_n f_o = \mathcal{O}\bigl(h^m \bigr) \;. \tag{8.38}$$

Moreover, it is easy to see that in general this is also an asymptotic lower bound. Thus (8.27) must be suitably rephrased. Q.e.d.

So, if all other bounds stay the same, asymptotic normality is achieved only if $h \ll n^{-1/(2m)}$, but then we do not get the optimal rate of convergence for the estimator of f_o since the required $h \asymp n^{-1/(2m+1)}$ is excluded.

(8.39) Exercise. Prove (suitable modifications of) (8.25) and (8.26) for the arbitrary designs under consideration. Also, verify the asymptotic normality of β^{nh} for $nh^2 \to \infty$ and $nh^{2m} \to 0$.

So, what is one to do? From a formal mathematical standpoint, it is clear that a slight modification of (8.38), and hence a slight modification of (8.27), would hold, viz.

$$\tfrac{1}{n} G_n^T (I - S_h)^2 r_n f_o = \mathcal{O}\bigl(h^{m+1} \bigr) \;, \tag{8.40}$$

but then everything else must be modified as well. All of this leads to two-stage estimators, in which the conditional expectation $g_o(x) = \mathbb{E}[\, z \,|\, x \,]$ is first estimated nonparametrically and then the estimation of β_o and f_o is considered.

(8.41) Exercise. Prove (8.40) taking

$$\tfrac{1}{n} G_n^T (I - S_h)^2 r_n f_o = \tfrac{1}{n} \sum_{i=1}^{n} \{\, g_o(x_{in}) - g_{hn}(x_{in}) \,\} \{\, f_o(x_{in}) - f_{hn}(x_{in}) \,\}$$

as the starting point. (g_{hn} is defined analogously to f_{hn}.)

Two-stage estimators for arbitrary designs. Suppose we estimate the conditional expectation $g_o(x)$ by a smoothing spline estimator g^{nh} (componentwise). In our present finite-dimensional context, then

$$r_n \, g^{nh} = S_h \, Z_n \;. \tag{8.42}$$

With this smoothing spline estimator of g_o, let

$$G^{nh} = r_n g^{nh} = \bigl[\; g^{nh}(x_{1,n}) \,|\, g^{nh}(x_{2,n}) \,|\, \cdots , g^{nh}(x_{n,n}) \,\bigr]^T \in \mathbb{R}^{n \times d} \;, \tag{8.43}$$

and define

$$\mathcal{Z}^{nh} = Z_n - G^{nh} \;. \tag{8.44}$$

Now, following Chen and Shiau (1991), consider the estimation problem

$$\begin{array}{ll} \text{minimize} & \tfrac{1}{n} \, \| \, \mathcal{Z}^{nh} \beta + r_n \, \varphi - y_n \, \|^2 + h^{2m} \, \| \, \varphi^{(m)} \, \|^2 \\ \text{subject to} & \beta \in \mathbb{R}^d \;,\; \varphi \in W^{m,2}(0,1) \;. \end{array} \tag{8.45}$$

The solution is denoted by $(\beta^{nh,1}, f^{nh,1})$. Note that, with $\varphi = f + (g^{nh})^{T}\beta$, the objective function may also be written as

$$\tfrac{1}{n}\|\, Z_n\, \beta + r_n f - y_n \,\| + h^{2m} \,\|\, \{\, f + (g^{nh})^{T}\, \beta \,\}^{(m)}\, \|^2 \ ;$$

in other words, the (estimated) conditional expectation is part of the roughness penalization, with the same smoothing parameter h.

It is a straightforward exercise to show that this two-stage estimator of β_o is given by

$$\beta^{nh,1} = \big(\, \mathcal{Z}_n^T (\, I - S_h\,)^3\, \mathcal{Z}_n \,\big)^{-1}\, \mathcal{Z}_n^T (\, I - S_h\,)^2\, y_n \ , \tag{8.46}$$

so that

$$\beta^{nh,1} - \beta_o = \text{variation} + \text{bias} \ , \tag{8.47}$$

with

$$\begin{aligned} \text{variation} &= \big(\, \mathcal{Z}_n^T (\, I - S_h\,)^3\, \mathcal{Z}_n \,\big)^{-1}\, \mathcal{Z}_n^T (\, I - S_h\,)^2\, d_n \ , \\ \text{bias} &= \big(\, \mathcal{Z}_n^T (\, I - S_h\,)^3\, \mathcal{Z}_n \,\big)^{-1}\, \mathcal{Z}_n^T (\, I - S_h\,)^2\, r_n f_o \ . \end{aligned} \tag{8.48}$$

The crucial results to be shown are

$$\tfrac{1}{n}\, \mathcal{Z}_n^T\, \mathcal{Z}_n \longrightarrow_{\text{as}} V \ , \tag{8.49}$$

$$\tfrac{1}{n}\, \mathcal{Z}_n^T (\, -3\, S_h + 3\, {S_h}^2 - {S_h}^3\,)\, \mathcal{Z}_n = \mathcal{O}_P\big(\, (nh)^{-1}\,\big) \ , \tag{8.50}$$

$$\tfrac{1}{n}\, Z_n^T\, (-2\, S_h + {S_h}^2\,)\, d_n = \mathcal{O}_P\big(\, (nh)^{-1}\,\big) \ , \tag{8.51}$$

$$\tfrac{1}{n}\, Z_n^T\, (\, I - S_h\,)^2\, r_n f_o = \mathcal{O}_P\big(\, n^{-1/2}\, h^{m-1/2} + h^{m+1}\,\big) \ , \tag{8.52}$$

with V as in (8.11). They are easy to prove by the previously used methods. All of this then leads to the following theorem.

(8.53) Theorem. *Under the assumptions* (8.2), (8.7), *and* (8.9)–(8.13),

$$\sqrt{n}\, (\, \beta^{nh,1} - \beta_o) \longrightarrow_{\text{d}} \Upsilon \sim \text{Normal}(\, 0\, , \sigma^2\, V\,) \ ,$$

provided $nh^2 \to \infty$ *and* $nh^{2m+2} \to 0$.

(8.54) Exercise. (a) Prove (8.49) through (8.52).
(b) Assume that $g_o \in W^{m,2}(0,1)$. Prove that, for $h \asymp n^{-1/(2m+1)}$, we get the asymptotic normality of $\beta^{nh,1}$ as advertised in Theorem (8.53) as well as the optimal rate of convergence for the estimator of f_o, viz. $\|\, f^{nh,1} - f_o \,\| = \mathcal{O}_P\big(\, n^{-m/(2m+1)}\,\big)$.

We finish this section by mentioning the estimator

$$\beta^{nh,2} = \big(\, \mathcal{Z}_n^T (\, I - S_h\,)^2\, \mathcal{Z}_n \,\big)^{-1}\, \mathcal{Z}_n^T (\, I - S_h\,)^2\, y_n \tag{8.55}$$

of Speckman (1988), who gives a piecewise regression interpretation. (To be precise, Speckman (1988) considers kernel estimators, not just smoothing splines.) The estimator for f_o is then given by

$$r_n f^{nh,2} = S_h \, (\, y_n - Z_n \, \beta^{nh,2} \,) \; . \tag{8.56}$$

The asymptotic normality of $\beta^{nh,2}$ may be shown similarly to that of $\beta^{nh,1}$.

(8.57) Exercise. State and prove the analogue of Theorem (8.53) for the estimator $\beta^{nh,2}$.

This completes our discussion of spline estimation in partially linear models. It is clear that it could be expanded considerably. By way of example, the smoothing spline $r_n g^{nh} = S_h Z_n$, see (8.42), applies to each component of Z_n separately, so it makes sense to have different smoothing parameters for each component so

$$r_n \, (g^{nh})_j = S(h_j) \, (Z_n)_j \; , \quad j = 1, 2, \cdots, d \; , \tag{8.58}$$

with the notation $S(h) \equiv S_h$. Here $(Z_n)_j$ denotes the j-th column of Z_n.

Exercises: (8.22), (8.29), (8.33), (8.34), (8.35), (8.39), (8.41), (8.54), (8.57).

9. Estimating derivatives

Estimating derivatives is an interesting problem with many applications. See, e.g., D'Amico and Ferrigno (1992) and Walker (1998), where cubic and quintic splines are considered. In this section, we briefly discuss how smoothing splines may be used for this purpose and how error bounds may be obtained.

The problem is to estimate $f_o'(x)$, $x \in [\,0\,,\,1\,]$, in the model

$$y_{in} = f_o(x_{in}) + d_{in} \; , \quad i = 1, 2, \cdots, n \; , \tag{9.1}$$

under the usual conditions (4.1)–(4.4). We emphasize the last condition,

$$f_o \in W^{m,2}(0,1) \; . \tag{9.2}$$

As the estimator of f_o', we take $(f^{nh})'$, the derivative of the spline estimator. We recall that, under the stated conditions,

$$\mathbb{E}[\, \| \, f^{nh} - f_o \, \|_{m,h}^2 \,] = \mathcal{O}\big(\, n^{-2m/(2m+1)} \, \big) \; , \tag{9.3}$$

provided $h \asymp n^{-1/(2m+1)}$ (deterministically); see Corollary (4.7). Now, recall Lemma (2.17),

$$\| \, \varphi \, \|_{k,h} \leqslant c_m \, \| \, \varphi \, \|_{m,h} \; ,$$

valid for all $\varphi \in W^{m,2}(0,1)$ and for all $k = 0, 1, \cdots, m$, with a constant c_m depending on m only. Applying this to the problem at hand with

$k = 1$ yields

$$\mathbb{E}[\, h^2 \, \| \, (f^{nh} - f_o)' \, \|^2 \,] = \mathcal{O}\big(\, n^{-2m/(2m+1)} \, \big) \ ,$$

so that

$$\mathbb{E}[\, \| \, (f^{nh} - f_o)' \, \|^2 \,] = \mathcal{O}\big(\, n^{-2(m-1)/(2m+1)} \, \big) \ , \tag{9.4}$$

provided $h \asymp n^{-1/(2m+1)}$. This argument applies to all derivatives of order $< m$. We state it as a theorem.

(9.5) THEOREM. *Assume the conditions* (4.1) *through* (4.4) *and that the design is asymptotically uniform. Then, for* $d = 1, 2, \cdots, m-1$,

$$\mathbb{E}[\, \| \, (f^{nh})^{(d)} - f_o^{(d)} \, \|^2 \,] = \mathcal{O}\big(\, n^{-2(m-d)/(2m+1)} \, \big) \ ,$$

provided $h \asymp n^{-1/(2m+1)}$.

(9.6) EXERCISE. Prove the remaining cases of the theorem.

Some final comments are in order. It is not surprising that we lose accuracy in differentiation compared with plain function estimation. However, it is surprising that the asymptotically optimal value of h does not change (other than through the constant multiplying $n^{-1/(2m+1)}$). We also mention that RICE and ROSENBLATT (1983) determine the optimal convergence rates as well as the constants. Inasmuch as we get the optimal rates, the proof above is impeccable. Of course, our proof does not give any indication *why* these are the correct rates. The connection with kernel estimators through the "equivalent" kernels might provide some insight; see Chapter 14.

EXERCISE: (9.6).

10. Additional notes and comments

Ad § 1: Nonparametric regression is a huge field of study, more or less (less !) evenly divided into smoothing splines, kernel estimators, and local polynomials, although wavelet estimators are currently in fashion. It is hard to do justice to the extant literature. We mention WAHBA (1990), EUBANK (1999), HÄRDLE (1990), GREEN and SILVERMAN (1990), FAN and GIJBELS (1996), ANTONIADIS (2007), and GYÖRFI, KOHLER, KRZYŻAK, and WALK (2002) as general references.

Ad § 2: For everything you might ever need to know about the Sobolev spaces $W^{m,2}(0,1)$, see ADAMS (1975), MAZ'JA (1985), and ZIEMER (1989). The statement and proof of the Interpolation Lemma (2.12) comes essentially from ADAMS (1975).

The reference on reproducing kernel Hilbert spaces is Aronszajn (1950). Meschkowski (1962) and Hille (1972) are also very informative. For a survey of the use of reproducing kernels in statistics and probability, see Berlinet and Thomas-Agnan (2004). For more on Green's functions, see, e.g., Stakgold (1967).

The definition (2.22) of asymptotically uniform designs is only the tip of a sizable iceberg; see Amstler and Zinterhof (2001) and references therein.

Ad § 3: Rice and Rosenblatt (1983) refer to the natural boundary conditions (3.18) as unnatural (boundary) conditions, which is wrong in the technical sense but accurate nevertheless.

Ad § 6: The Embedding Lemma (6.13) is the one-dimensional version of a result in Kufner (1980). Of course, the one-dimensional version is much easier than the multidimensional case.

Ad § 8: The "simplest" case of the partially linear model was analyzed by Heckman (1988). Rice (1986a) treated arbitrary designs and showed that asymptotic normality of β^{nh} requires undersmoothing of the spline estimator of f_o. The two-stage estimator $\beta^{nh,1}$ and the corresponding asymptotic normality and convergence rates are due to Chen and Shiau (1991). The estimator $\beta^{nh,2}$ was introduced and studied by Speckman (1988) for "arbitrary" kernels S_h. Chen (1988) considered piecewise polynomial estimators for f_o. Both of these authors showed the asymptotic normality of their estimators and the optimal rate of convergence of the estimator for f_o. Bayesian interpretations may be found in Eubank (1988) and Heckman (1988). Eubank, Hart, and Speckman (1990) discuss the use of the trigonometric sieve combined with boundary correction using the Bias Reduction Principle; see § 15.4. Finally, Bunea and Wegkamp (2004) study model selection in the partially linear model.

http://www.springer.com/978-0-387-40267-3

Maximum Penalized Likelihood Estimation
Volume II: Regression
Eggermont, P.P.; LaRiccia, V.N.
2009, XX, 572 p., Hardcover
ISBN: 978-0-387-40267-3