Smoothing splines

(notes by Pavel Chigansky)

A common method of constructing estimators in parametric models is through optimising a suitable contrast function. One example is the Maximum Likelihood Estimator for which likelihood function plays the role of the contrast. Another example is the Least Squares Estimator in the regression problem, which minimizes squared deviations of the measurements from a curve in a postulated parametric subspace. The analogous approach works in the nonparametric estimation, albeit with appropriate adjustments.

1. The estimator

Consider the regression problem of estimating unknown function $f : [0,1] \mapsto \mathbb{R}$ from the data $(X_1, Y_1), \dots, (X_n, Y_n)$ generated by the equation

$$Y_j = f(X_j) + \varepsilon_j, \tag{1.1}$$

where ε_j's are i.i.d. random variables with zero mean and unit variance. For simplicity all design points X_j are assumed to be distinct.

If f is known to have m-derivatives, then it makes sense to fit the data by a similarly smooth function, e.g., from the Sobolev space $W_2^m([0,1])$ of m-times weakly differentiable functions with the finite norm

$$\|f\|_{m,2} = \big\|f^{(m)}\big\|_2^2 < \infty.$$

This can be achieved by minimizing the contrast functional

$$J_m(f) := \frac{1}{n}\sum_{j=1}^n \big(Y_j - f(X_j)\big)^2 + \lambda \big\|f^{(m)}\big\|_2^2 \tag{1.2}$$

over the above Sobolev space

$$\widehat{f}_n = \operatorname*{argmin}_{f \in W_2^m([0,1])} J_m(f). \tag{1.3}$$

The *penalty* term in (1.2) is multiplied by a tuning parameter $\lambda > 0$, which can be expected to play the role analogous to the bandwidth in kernel estimators. Indeed, for $\lambda = 0$, the minimum $J_m(f) = 0$ is attained by any function, which passes through all the data points. Any such estimator obviously overfits. Taking $\lambda \to \infty$ leads to the usual polynomial regression of degree m, which can hardly

lecture notes for "Advanced Statistical Models B" course.

be expected to approximate well a generic function in the Sobolev space. This is the under-fitting (or over-smoothing) extremity. Hence the choice of λ will have crucial affect on the estimation quality and must be tuned with care.

1.1. Splines and natural splines. Quite remarkably, the optimizer of $J_m(f)$ turns out to be a smooth *piecewise* polynomial function, which reduces the infinite dimensional optimization problem to a finite dimensional one. To construct the solution let us start with a few general notions.

DEFINITION 1.1. *A spline of order r with the knots at $x_1 < \dots < x_k$ is the piecewise polynomial function of the form*

$$s(x) = \sum_{j=0}^{r-1} \theta_j x^j + \sum_{j=1}^{k} \eta_j (x - x_j)_+^{r-1}, \quad x \in \mathbb{R}, \tag{1.4}$$

where [1] *θ_j's and η_j's are real constants.*

Obviously, a spline is a polynomial of degree $r-1$ on each interval $[x_{j-1}, x_j]$, has $r-2$ everywhere continuous derivatives and its $r-1$ derivative has jumps at the knots x_j's. Thus it consists of polynomials smoothly pieced together at the knot points. The set $S^r(x_1, \dots, x_k)$ of all splines of order r with knots $x_1, \dots, x_k$ is a linear space of dimension $r+k$, spanned by the functions (Problem 1)

$$1, x, \dots, x^{r-1}, (x - x_1)_+^{r-1}, \dots, (x - x_k)_+^{r-1}. \tag{1.5}$$

The solution to minimization problem (1.3) turns out to belong to an n - dimensional subspace of splines of order $r := 2m$ with $k := n$ knots at the design points, obtained through the additional requirement that (1.4) is a polynomial of degree $m-1$ outside the interval $[x_1, x_n]$. This subset of $S^r(x_1, \dots, x_k)$ is called *natural splines* and is denoted by $NS^{2m}(x_1, \dots, x_n)$. Obviously the additional condition implies

$$\theta_m = \dots = \theta_{2m-1} = 0. \tag{1.6}$$

A calculation shows that, in fact, $NS^{2m}(x_1, \dots, x_n)$ is a linear subspace of dimension n (see Problem 2).

1.2. The spline estimator. Let $v \in \mathbb{R}^n$ be a fixed vector, $x_1, \dots, x_n$ be a given set of knots and $m \le n$ be a positive integer. Consider the minimization problem of finding a function which

$$\begin{aligned} &\text{minimizes } \|f^{(m)}\|_2^2 \ \text{ over } \ W_2^m([0,1]) \\ &\text{subject to } f(x_j) = v_j, \ j = 1, \dots, n. \end{aligned} \tag{P}$$

The solution to this seemingly infinite dimensional optimization problem is, in fact, finite dimensional as asserted by the following result to be proved in the next section.

[1] here $x_+ = \max(0, x)$ and $x_+^p = (x_+)^p$, $p \in \mathbb{Z}_+$

THEOREM 1.2. *Let $\phi_1,\dots,\phi_n$ be any basis of $NS^{2m}(x_1,\dots,x_n)$. Then the matrix $\Phi\in\mathbb{R}^{n\times n}$ with the entries*

$$\Phi_{ij}=\phi_j(x_i),\quad i,j\in\{1,\dots,n\},$$

is nonsingular and the problem (P) *has the unique solution*

$$f(x)=\sum_{j=1}^n b_j\phi_j(x),$$

where $b=\Phi^{-1}v$.

COROLLARY 1.3. *The minimizer of the functional $J_m(f)$ in* (1.2) *over the Sobolev space $W_2^m([0,1])$ is a natural spline in $NS^{2m}(X_1,\dots,X_n)$.*

PROOF. Let $\widetilde{f}\in W_2^m([0,1])$ be a minimizer of $J_m(f)$ and let $\widehat{f}\in NS^{2m}(x_1,\dots,x_n)$ be the solution to (P) with $v_j=\widetilde{f}(X_j)$. Then

$$\inf_{f\in W_2^m([0,1])} J_m(f)=J_m(\widetilde{f})=\frac{1}{n}\sum_{j=1}^n\big(Y_j-\widetilde{f}(X_j)\big)^2+\lambda\big\|\widetilde{f}^{(m)}\big\|_2^2\ge$$
$$\frac{1}{n}\sum_{j=1}^n\big(Y_j-\widehat{f}(X_j)\big)^2+\lambda\big\|\widehat{f}^{(m)}\big\|_2^2=J_m(\widehat{f}),$$

where the inequality holds due to Theorem 1.2. □

In view of Corollary 1.3 minimization of (1.2) over $f\in W_2^m([0,1])$ reduces to a simple finite dimensional optimization problem.

THEOREM 1.4. *Let $\phi_1,\dots,\phi_n$ be any basis of $NS^{2m}(X_1,\dots,X_n)$. Define the $n\times n$ matrices Φ and Ω with the entries $\Phi_{ij}=\phi_j(X_i)$ and $\Omega_{ij}=\langle\phi_i^{(m)},\phi_j^{(m)}\rangle$. Assume $n\ge m$, then the unique minimizer of* (1.2) *is the natural spline*

$$\widehat{f}(x)=\sum_{j=1}^n b_j\phi_j(x),\tag{1.7}$$

where the coefficients b_j's are the entries of the vector

$$b=\Big(\Phi^\top\Phi+n\lambda\Omega\Big)^{-1}\Phi^\top Y.\tag{1.8}$$

PROOF. For a natural spline as above

$$J_m(f)=\frac{1}{n}\sum_{k=1}^n\Big(Y_k-\sum_{j=1}^n b_j\phi_j(X_k)\Big)^2+\lambda\Big\|\sum_{j=1}^n b_j\phi_j^{(m)}\Big\|_2^2,$$

and the claimed formula for the unique minimizer of this expression over $b\in\mathbb{R}^n$ is obtained by equating the gradient to zero (Problem 3). □

1.3. Proof of Theorem 1.2. The key element of the proof of Theorem 1.2 is the following lemma.

LEMMA 1.5. *Let $\phi_1, \dots, \phi_n$ be a basis in $NS^{2m}(x_1, \dots, x_n)$. Then there are coefficients $\theta_{0,j}, \dots, \theta_{m-1,j}$ and $\eta_{1,j}, \dots, \eta_{n,j}$ such that*

$$\phi_j(x) = \sum_{i=0}^{m-1} \theta_{i,j} x^i + \sum_{i=1}^{n} \eta_{i,j} (x - x_i)_+^{2m-1}. \tag{1.9}$$

If $s = \sum_{j=1}^n b_j \phi_j$ and $f \in W_2^m([0,1])$ then

$$\big\langle f^{(m)}, s^{(m)} \big\rangle = (-1)^m (2m-1)! \sum_{i=1}^n f(x_i) \sum_{j=1}^n b_j \eta_{i,j}. \tag{1.10}$$

PROOF. The representation (1.9) is obvious since (1.5) is a basis and (1.6) holds. Since $s^{m+j}(x)$ vanish outside the interval $[x_1, x_n]$ for all $j = 0, \dots, m-1$, integrating by parts we obtain

$$\begin{aligned}
&\big\langle f^{(m)}, s^{(m)} \big\rangle = \int_0^1 f^{(m)}(x) s^{(m)}(x) dx = \\
&(-1)^{m-1} \int_0^1 f'(x) s^{(2m-1)}(x) dx = \\
&(-1)^{m-1} \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} f'(x) s^{(2m-1)}(x) dx = \\
&(-1)^{m-1} \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} f'(x) \sum_{j=1}^n b_j \phi_j^{(2m-1)}(x) dx = \\
&(-1)^{m-1} \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} f'(x) \sum_{j=1}^n b_j (2m-1)! \sum_{i=1}^n \eta_{i,j} \mathbf{1}_{\{x \ge x_i\}} dx = \\
&(-1)^{m-1} (2m-1)! \sum_{k=1}^{n-1} \big(f(x_{k+1}) - f(x_k) \big) \sum_{j=1}^n b_j \sum_{i=1}^k \eta_{i,j} = \\
&(-1)^{m-1} (2m-1)! \sum_{j=1}^n b_j \bigg(\sum_{k=2}^n f(x_k) \sum_{i=1}^{k-1} \eta_{i,j} - \sum_{k=1}^{n-1} f(x_k) \sum_{i=1}^k \eta_{i,j} \bigg) = \\
&(-1)^{m-1} (2m-1)! \sum_{j=1}^n b_j \bigg(\sum_{k=1}^n f(x_k) \Big(\sum_{i=1}^{k-1} \eta_{i,j} - \sum_{i=1}^k \eta_{i,j} \Big) + f(x_n) \sum_{i=1}^n \eta_{i,j} \bigg) = \\
&(-1)^m (2m-1)! \sum_{j=1}^n b_j \sum_{k=1}^n f(x_k) \eta_{k,j}.
\end{aligned}$$

The last equality holds since

$$\phi_j^{(2m-1)}(x) = (2m-1)! \sum_{j=1}^n \eta_{i,j} \stackrel{\dagger}{=} 0, \quad x \ge x_n,$$

where † holds since a natural spline must equal a polynomial of degree m on $[x_n, \infty)$. □

PROOF OF THEOREM 1.2. Let us first check that the matrix Φ is nonsingular, that is, $\Phi u = 0$ implies $u = 0$. To this end, let $s(x) = \sum_{j=1}^n u_j \phi_j(x)$ and note that $s(x_i) = (\Phi u)_i = 0$, $i = 1, \dots, n$. Applying Lemma 1.5 with $f := s$ in (1.10) we obtain $\|s^{(m)}\|_2^2 = 0$. Thus s must be polynomial of degree $m-1$, which vanishes at $n \geq m$ points. Since a nonzero polynomial cannot have more zeros than its degree, s can only be the zero function, that is, $u_j = 0$.

Now for any two functions $f \in W_2^m([0,1])$ and a spline s,

$$\left\|f^{(m)}\right\|^2 = \left\|s^{(m)} + f^{(m)} - s^{(m)}\right\|^2 = \left\|s^{(m)}\right\|^2 + 2\langle s^{(m)}, f^{(m)} - s^{(m)}\rangle + \left\|f^{(m)} - s^{(m)}\right\|^2.$$

If $f(x_j) = s(x_j) = v_j$, then $f - s$ vanishes at x_j's, and the cross term vanishes due to Lemma 1.5. Thus we arrive at the bound $\left\|f^{(m)}\right\|^2 \geq \left\|s^{(m)}\right\|^2$ and s is a minimizer and, since Φ is nonsingular, its coordinates in the basis $\phi_1, \dots, \phi_n$ are given by $\Phi^{-1} v$.

To argue that this minimizer is unique, note that the bound is saturated if and only if $\left\|f^{(m)} - s^{(m)}\right\|^2 = 0$, which implies that $f - s$ is a polynomial of degree $m-1$. Since it has $n \geq m$ zeros at $x_1, \dots, x_n$, as we already argued above, the only such polynomial is the zero function and hence the bound is in fact saturated only at $f = s$. $\square$

2. Computation

Computation of the smoothing splines by means of Theorem 1.4 requires finding a basis for natural splines. It is typically easier to approach the construction of the estimator (1.7) otherwise, by exploiting the piecewise polynomial structure of splines. Let us demonstrate the ideas for the natural cubic spline $m = 2$, which is the most commonly used spline in practice. The key to the computation is the following result.

LEMMA 2.1. *Let $s \in NS^4(x_1, \dots, x_n)$ and $h_i = x_{i+1} - x_i$. Define the $(n-2) \times n$ matrix*

$$\Delta = \begin{pmatrix} \frac{1}{h_1} & -\left(\frac{1}{h_1} + \frac{1}{h_2}\right) & \frac{1}{h_2} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{h_2} & -\left(\frac{1}{h_2} + \frac{1}{h_3}\right) & \frac{1}{h_3} & 0 & \cdots & 0 \\ \cdots & & & & & & \cdots \\ 0 & \cdots & & 0 & \frac{1}{h_{n-2}} & -\left(\frac{1}{h_{n-2}} + \frac{1}{h_{n-1}}\right) & \frac{1}{h_{n-1}} \end{pmatrix}$$

and the symmetric tridiagonal $(n-2) \times (n-2)$ matrix

$$W = \begin{pmatrix} \frac{h_1+h_2}{3} & \frac{h_2}{6} & 0 & 0 & 0 & \cdots & 0 \\ \frac{h_2}{6} & \frac{h_2+h_3}{3} & \frac{h_3}{6} & 0 & 0 & \cdots & 0 \\ 0 & \frac{h_3}{6} & \frac{h_3+h_4}{3} & \frac{h_4}{6} & 0 & \cdots & 0 \\ \cdots & & & & & & \cdots \\ 0 & \cdots & & & 0 & \frac{h_{n-2}}{6} & \frac{h_{n-2}+h_{n-1}}{3} \end{pmatrix}.$$

Let $K = \Delta^\top W^{-1} \Delta$ and S be the vector with entries $s(x_i)$, $i = 1, \dots, n$, then

$$\|s''\|^2 = S^\top K S. \tag{2.1}$$

PROOF. (Problem 4) □

Note that the functional (1.2) evaluated at a cubic spline as in the lemma, can be written as

$$J_2(s) = \frac{1}{n}\sum_{j=1}^{n}\big(Y_j - S_j\big)^2 + \lambda S^\top K S,$$

where S_j's are the values of the spline at X_j's. Minimizing this over S gives

$$S = (I + n\lambda K)^{-1} Y. \tag{2.2}$$

The whole spline can now be restored using the formulas, which relate coefficients of the spline $s(x)$ on each interval $[X_i, X_{i+1}]$ to its values $s(X_j) = S_j$ at the knots (see Problem 4). Using the matrix inversion lemma (Woodbury identity) the two matrix inversions in (2.2) can be reduced to a single inversion of a banded matrix, which is why the above approach is computationally advantageous over the formulas in Theorem 1.4.

3. MSE risk bound

One way to approach the risk analysis of the smoothing splines estimator is to find a basis, which diagonalizes simultaneously the matrices Φ and Ω in Theorem 1.4. Below we consider the case of linear splines $m = 1$ with uniform design, for which this basis can be found explicitly. This special case sheds much light on the structure of the splines estimators, which is preserved in a greater generality.

3.1. Demmler-Reinsch basis (m=1). Consider the uniform grid of points

$$x_i = \frac{i - 1/2}{n}, \quad i = 1, \dots, n.$$

A linear natural spline is piecewise linear between the knots and is constant outside $[x_1, x_n]$. It turns out that the required basis in this case can be constructed by linear interpolation of the constant and the cosine functions

$$\sqrt{2}\cos(k\pi x), \quad k = 1, \dots, n-1.$$

Namely, define $\phi_1(x) \equiv 1$ and for $k = 2, \dots, n-1$,

$$\phi_{k+1}(x) = \sqrt{2}\begin{cases} \cos(k\pi x_1), & x \in [0, x_1], \\ \cos(k\pi x_{i+1})\dfrac{x - x_i}{x_{i+1} - x_i} + \cos(k\pi x_i)\dfrac{x_{i+1} - x}{x_{i+1} - x_i}, & x \in [x_i, x_{i+1}],\ 1 \le i < n \\ \cos(k\pi x_n), & x \in [x_n, 1] \end{cases}$$

A direct calculation (Problem 7) shows that ϕ_k's are linearly independent and therefore form a basis of $NS^2(x_1, \dots, x_n)$. Moreover the matrices, introduced in Theorem 1.4, satisfy

$$\Phi^\top \Phi = \Phi\Phi^\top = nI \tag{3.1}$$

and

$$\Omega = \operatorname{diag}\big(0, \gamma_1, \dots, \gamma_{n-1}\big), \qquad \gamma_j = \left(2n\sin\left(\frac{j\pi}{2n}\right)\right)^2. \tag{3.2}$$

The formula (1.8) now simplifies to

$$b_1 = \frac{1}{n}\sum_{j=1}^{n} Y_j$$

$$b_{j+1} = \frac{\sqrt{2}}{1+\lambda\gamma_j}\frac{1}{n}\sum_{i=1}^{n}\cos(j\pi x_i)Y_i, \quad j = 1, \dots, n-1$$

and the corresponding estimator reads

$$\widehat{f}(x) = \frac{1}{n}\sum_{m=1}^{n} Y_m + \sum_{j=1}^{n-1}\frac{2}{1+\lambda\gamma_j}\frac{1}{n}\sum_{m=1}^{n}\cos(j\pi x_m)Y_m\cos(j\pi x) =$$
$$\frac{1}{n}\sum_{m=1}^{n} K_n(x, x_m; \lambda)Y_m,$$

where we defined the kernel

$$K_n(x,y;\lambda) := 1 + 2\sum_{j=1}^{n-1}\frac{\cos(j\pi y)\cos(j\pi x)}{1+\lambda\gamma_j}. \tag{3.3}$$

3.2. The upper MSE bound. The large sample asymptotic analysis of the MSE risk is based on the following approximation of kernel (3.3).

LEMMA 3.1. *Let $\lambda_n := n^{-a}$ with $a \in (0,1)$, then*

$$\sup_{x,y\in[0,1]}\Big|K_n(x,y;\lambda_n) - K(x,y;\lambda_n)\Big| \le \frac{C}{n\lambda_n}, \tag{3.4}$$

where C is some constant and

$$K(x,y;\lambda) = \frac{1}{2\sqrt{\lambda}}\Big(e^{-|x-y|/\sqrt{\lambda}} + e^{-(x+y)/\sqrt{\lambda}} + e^{-(1-x+1-y)/\sqrt{\lambda}}\Big). \tag{3.5}$$

PROOF. When n is large, in view of (3.2), it makes sense to approximate the sum in (3.3) by the infinite series, where γ_j is replaced with $(j\pi)^2$,

$$\widetilde{K}(x,y;\lambda) = 1 + 2\sum_{j=1}^{\infty}\frac{\cos(j\pi y)\cos(j\pi x)}{1+\lambda(j\pi)^2} =$$
$$\frac{e^{-|x-y|/\sqrt{\lambda}} + e^{-2/\sqrt{\lambda}}e^{|x-y|/\sqrt{\lambda}} + e^{-(x+y)/\sqrt{\lambda}} + e^{(x+y-2)/\sqrt{\lambda}}}{2\sqrt{\lambda}(1-e^{-2/\sqrt{\lambda}})}.$$

The closed form expression is obtained by writing the product of cosines as the sum, thus splitting the summation into two explicitly computable series (through counter integration on the complex plane).

A simple calculation shows that the obtained kernel approximates kernel (3.5):

$$\sup_{0\le x,y\le 1}\big|\widetilde{K}(x,y;\lambda_n) - K(x,y;\lambda_n)\big| \le \frac{2}{\sqrt{\lambda_n}}e^{-1/\sqrt{\lambda_n}} \le 2\lambda_n^{(p-1)/2} \le \frac{2}{n\lambda_n},$$

where the second inequality holds[2] for any $p>0$ and the last one for $p>2/a-1$. Hence it remains to check that

$$\big|K_n(x,y;\lambda_n)-\widetilde{K}(x,y;\lambda_n)\big|\le \frac{C}{\lambda_n n}, \tag{3.6}$$

with a constant C. To this end, we have

$$\big|K_n(x,y;\lambda_n)-\widetilde{K}(x,y;\lambda_n)\big|\le 2\sum_{j=1}^{n-1}\left|\frac{1}{1+\lambda_n\gamma_j}-\frac{1}{1+\lambda_n(j\pi)^2}\right|+\sum_{j=n}^{\infty}\frac{2}{1+\lambda_n(j\pi)^2}.$$

The last term satisfies the bound

$$\sum_{j=n}^{\infty}\frac{1}{1+\lambda_n(j\pi)^2}\le\frac{1}{\lambda_n}\sum_{j=n}^{\infty}j^{-2}\le\frac{1}{\lambda_n}\int_{n-1}^{\infty}\frac{1}{x^2}dx=\frac{1}{\lambda_n(n-1)}.$$

To bound the first term, we can use the elementary inequalities

$$\tfrac12 x\le\sin x\le x\quad\text{and}\quad|\sin x-x|\le\frac16 x^3,\quad\forall x\in[0,\tfrac{\pi}{2}],$$

which imply

$$\sum_{j=1}^{n-1}\left|\frac{1}{1+\lambda_n\gamma_j}-\frac{1}{1+\lambda_n(j\pi)^2}\right|=\lambda_n\sum_{j=1}^{n-1}\frac{|(j\pi)^2-\lambda_n\gamma_j|}{(1+\lambda_n\gamma_j)(1+\lambda_n(j\pi)^2)}\le$$

$$\frac{1}{\lambda_n}\sum_{j=1}^{n-1}\frac{|(j\pi)^2-\gamma_j|}{\gamma_j(j\pi)^2}=\frac{1}{\lambda_n}\sum_{j=1}^{n-1}\frac{|j\pi-\sqrt{\gamma_j}|(j\pi+\sqrt{\gamma_j})}{\gamma_j(j\pi)^2}\le$$

$$\frac{1}{\lambda_n}\sum_{j=1}^{n-1}\frac{2n\frac16\big(\frac{j\pi}{2n}\big)^3(j\pi+j\pi)}{\frac14(j\pi)^2(j\pi)^2}\le\frac{1}{n\lambda_n}.$$

This proves (3.6) and, in turn, the claimed bound. □

This lemma shows that at the interior points the linear splines estimator behaves as the kernel estimator with the Laplacian kernel $\frac12 e^{-|u|}$. Using this approximation we can estimate the variance and the bias and optimize over the bandwidth. Let us sketch the derivation for the MSE at a fixed *interior* point $x_0\in(0,1)$.

More specifically (3.4) implies that as $n\to\infty$,

$$\mathrm{Var}_f\big(\widehat{f}_n(x_0)\big)=\frac{\sigma^2}{n^2}\sum_{m=1}^{n}K_n(x_0,x_m;\lambda_n)^2=\frac{\sigma^2}{n}\int_0^1 K(x_0,x;\lambda_n)^2dx\big(1+o(1)\big),$$

where the integral can be estimated using the explicit (3.5)

$$\int_0^1 K(x_0,x;\lambda_n)^2dx\le$$

$$\frac{3}{4\lambda_n}\left(\int_0^1 e^{-2|x-x_0|/\sqrt{\lambda_n}}dx+\int_0^1 e^{-2(x+x_0)/\sqrt{\lambda_n}}dx+\int_0^1 e^{-2(1-x+1-x_0)/\sqrt{\lambda_n}}dx\right)\le$$

$$\frac{3}{4\lambda_n}\left(\sqrt{\lambda_n}2\int_0^\infty e^{-2s}ds+e^{-2x_0/\sqrt{\lambda_n}}+e^{-2(1-x_0)/\sqrt{\lambda_n}}\right)\le\frac{C}{\sqrt{\lambda_n}}$$

[2] $e^{-x}\le 1/x^p$ for all $x>0$ and $p>1$

with a constant C (which depends on x_0). Hence

$$\operatorname{Var}_f(\widehat{f}_n(x_0)) \le \frac{C}{n\sqrt{\lambda_n}}, \quad \text{as } n \to \infty,$$

for some constant C independent of f. Similarly,

$$\mathbb{E}_f \widehat{f}_n(x_0) = \frac{1}{n}\sum_{m=1}^{n} K_n(x_0, x_m; \lambda) f(x_m) = \int_0^1 K(x_0, x; \lambda_n) f(x) dx (1 + o(1)).$$

If f is a bounded function and x_0 is an interior point, the contribution of the second and the third terms in (3.5) to this integral is again exponentially negligible. Let us now assume that f has bounded second derivative $\|f''\|_\infty \le L$. Then Taylor's approximation gives

$$\int_0^1 \frac{1}{2\sqrt{\lambda_n}} e^{-|x_0-x|/\sqrt{\lambda_n}} f(x) dx = \int_{-(1-x_0)/\sqrt{\lambda_n}}^{x_0/\sqrt{\lambda_n}} \tfrac{1}{2} e^{-|u|} f(x_0 - \sqrt{\lambda_n} u) du =$$
$$\int_{-(1-x_0)/\sqrt{\lambda_n}}^{x_0/\sqrt{\lambda_n}} \tfrac{1}{2} e^{-|u|} \left(f(x_0) - \sqrt{\lambda_n} u f'(x_0) + \int_0^{-\sqrt{\lambda_n} u} \int_0^s f''(x_0 + t) dt ds \right) du =$$
$$f(x_0) \int_{\mathbb{R}} \tfrac{1}{2} e^{-|u|} du - \sqrt{\lambda_n} f'(x_0) \int_{\mathbb{R}} u \tfrac{1}{2} e^{-|u|} du + R_n = f(x_0) + R_n,$$

where the residual satisfies $|R_n| \le CL\lambda_n$ with a constant C. Thus we obtain the following bound for the bias term

$$|\mathbb{E}_f \widehat{f}_n(x_0) - f(x_0)| \le CL\lambda_n,$$

which holds for all n large enough, uniformly over $f \in \Sigma(2, L)$. The squared bias and the variance terms are balanced by the choice $\lambda_n = n^{-2/5}$ and we obtain the familiar MSE bound

$$\sup_{f \in \Sigma(2,L)} \mathbb{E}_f \Big(n^{2/5} \big(\widehat{f}_n(x_0) - f(x_0) \big) \Big)^2 \le C(L).$$

Computer experiment

Implement the cubic spline regressor, using the formulas from Section 2 and Problem 4.

(1) Generate a sample for the regression problem (1.1) with the uniform design

$$X_j = j/n, \quad j = 1, \dots, n, \quad n = 300,$$

normal $N(0, \sigma^2)$ noise with $\sigma^2 = 1$ and some smooth function f with explicitly computable squared norm $\|f\|^2$. Apply your estimator with the smoothing parameter $\lambda := c\sigma^2 n^{-2/5}/\|f\|^2$ where $c = 1$ and plot the obtained estimate at the design points. Repeat with $c = 0.1$ and $c = 2$ for the same data and comment on the visible effects you observe.

(2) Choose the optimal "oracle" value of c by approximating the discrete MSE

$$R_c(f,\widehat{f}_n) = \mathbb{E}_f \frac{1}{n}\sum_{j=1}^n \big(f(X_j)-\widehat{f}_n(X_j)\big)^2$$

on the grid of values of c around $c=1$, by means of MC averaging over $M=10,000$ trials. Plot the MSE values for all c's and choose the minimizer. Plot the estimate generated by the "oracle" cubic spline estimator for the data in question (1).

Exercises

PROBLEM 1. Show that the functions in (1.5) form a basis in $S^r(x_1,\dots,x_k)$.

PROBLEM 2. Argue that $NS^{2m}(x_1,\dots,x_n)$ is a linear subspace and show that its dimension equals n.

PROBLEM 3. Fill in the details in the proof of Theorem 1.4.

PROBLEM 4. This problem is a guided proof of Lemma 2.1.

(1) Argue that the second derivative of the cubic spline $s\in S^4(x_1,\dots,x_n)$ is a continuous piecewise linear function

$$s''(x) = z_{i+1}\frac{x-x_i}{h_i} + z_i\frac{x_{i+1}-x}{h_i}, \quad x\in[x_i,x_{i+1}],$$

where $z_i = s''(x_i)$ and $h_i = x_{i+1}-x_i$.

(2) Argue that if $s\in NS^4(x_1,\dots,x_n)$, then $z_1=z_n=0$.

(3) Show that $\|s''\|^2 = z^\top W z$ with $z=(z_2,\dots,z_{n-1})$.

(4) Argue that for $x\in[x_i,x_{i+1}]$ the spline $s(x)$ from (1) can be written as

$$s(x) = z_{i+1}\frac{(x-x_i)^3}{6h_i} + z_i\frac{(x_{i+1}-x)^3}{6h_i} + c_i(x-x_i) + d_i(x_{i+1}-x),$$

where c_i and d_i are some constants.

(5) Check that the constants c_i and d_i are related to the values of the spline at x_i and x_{i+1} through the formulas

$$d_i = \frac{s(x_i)}{h_i} - z_i\frac{h_i}{6} \quad\text{and}\quad c_i = \frac{s(x_{i+1})}{h_i} - z_{i+1}\frac{h_i}{6}. \tag{3.7}$$

(6) Use continuity of the derivative s' at the knots to show that the vectors z and $S=(s(x_1),\dots,s(x_n))$ are related by the formula $Wz=\Delta S$. Deduce the expression (2.1).

(7) Specify the cubic smoothing spline estimator for the regression problem.

PROBLEM 5. Following the ansatz from the previous problem, derive the computational procedure for the linear smoothing natural spline.

PROBLEM 6. ([2, Exercise 1.11]) In this problem we will see that the cubic smoothing spline estimator for the regression model (1.1) is equivalent to a certain kernel estimator, under appropriate assumptions. We will consider the estimator, cf. (1.3),

$$\widehat{f}_n = \operatorname*{argmin}_{f \in W} \left(\frac{1}{n} \sum_{j=1}^{n} \big(Y_j - f(X_j)\big)^2 + \lambda \int \big(f''(x)\big)^2 dx \right),$$

where the function space W will be specified below.

(1) First suppose that W is the set of all the functions $f : [0,1] \mapsto \mathbb{R}$ such that f' is absolutely continuous. Prove that the estimator $\widehat{f}_n$ reproduces polynomials of degree less or equal 1 if $n \geq 2$.

(2) Hereafter the uniform design $X_k = k/n$ is assumed. Suppose next that W is the set of functions $f : [0,1] \mapsto \mathbb{R}$ such that (i) f' is absolutely continuous and (ii) the periodicity condition is satisfied: $f(0) = f(1)$ and $f'(0) = f'(1)$. Prove that the minimization problem which defines $\widehat{f}_n$ is equivalent to

$$\min_{\theta} \sum_{j=1}^{\infty} \Big(-2\widehat{\theta}_j \theta_j + \theta_j^2 \big(\lambda \pi^4 a_j^2 + 1\big)\big(1 + O(n^{-1})\big) \Big), \tag{3.8}$$

where $\theta = (\theta_j)_{j \in \mathbb{N}}$ is the sequence of Fourier coefficients of f, $\widehat{\theta}_j$'s are the estimates of θ_j's,

$$\widehat{\theta}_j = \frac{1}{n} \sum_{k=1}^{n} Y_k \phi_j(k/n),$$

a_j's are the sequence, introduced in the definition of periodic Sobolev ellipsoids $W^{\mathrm{per}}(\beta, L)$ (in the previous chapters), and term $O(n^{-1})$ is uniform in θ.

(3) Assume that the term $O(n^{-1})$ in (3.8) is negligible. Formally replacing it by zero, find the solution of (3.8) and conclude that the periodic spline estimator is approximately equal to a weighted projection estimator

$$\widehat{f}_n(x) \approx \sum_{j=1}^{\infty} \lambda_j^* \widehat{\theta}_j \phi_j(x),$$

with weights λ_j^* written explicitly.

(4) Use (3) to show that for small λ the spline estimator $\widehat{f}_n$ is approximated by the kernel estimator

$$\widetilde{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} Y_i K\left(\frac{X_i - x}{h}\right)$$

with with bandwidth $h = \frac{1}{2}\lambda^{1/4}$ and the Silverman kernel

$$K(u) = \int_{-\infty}^{\infty} \frac{\cos(2\pi t u)}{1 + (2\pi t)^4} dt = \frac{1}{2} \exp\big(-|u|/\sqrt{2}\big) \sin(|u|/\sqrt{2} + \pi/4).$$

PROBLEM 7. This problem is a guided proof of the diagonalizing property of the Demmler-Reinsch basis.

(1) Prove (3.1).

(2) Argue that ϕ_k's form a basis.

(3) Prove (3.2).

References

[1] Randall L. Eubank, Nonparametric Regression and Spline Smoothing, CRC press, (1999)
[2] AB Tsybakov, Introduction to Nonparametric Estimation, Springer Science & Business Media, 22 Oct 2008

DEPARTMENT OF STATISTICS, THE HEBREW UNIVERSITY, MOUNT SCOPUS, JERUSALEM 91905, ISRAEL
E-mail address: Pavel.Chigansky@mail.huji.ac.il