

Application of the Scaling Functions to Nonparametric Regression

Sorin MANOLE

Constantin Brâncoveanu University Pitești

For estimating regression function we can use many proceedings. In this paper, we have chosen to apply scaling functions to the estimation of regression functions.

When one knows many bivariate data with the values of two variables, in the goal to express a correlation between the two variables we use the regression function. The raw estimator of this function must be "smoothed out" in some way to get a final estimator. For this, we use the scaling functions, examples of such function being the Battle-Lemarié family and Daubechies family. After introducing several notions (multiresolution analysis, filter and projection of function onto approximation spaces), these are applied to obtain the estimators. In the last part, we present the algorithm for estimating nonparametric regression function through the scaling functions.

Keywords: nonparametric regression, scaling functions, filter, multiresolution analysis, approximation space, estimator.

Introduction

The typical case is that one when we have bivariate data (x_i, y_i) , $i = 1, 2, \dots, n$ and we want to use the data to quantify the relationship between the two variables. The standard regression model is $y_i = f(x_i) + \varepsilon_i$, $i = 1, 2, \dots, n$ (1), where the ε_i 's are independent and identically distributed $N(0, \sigma^2)$ random variables. The primary interest is in estimating the regression function f from the data. In parametric regression, we assume a particular structure on the regression function f (polynomial, logarithmic, exponential etc.). Statistical inference involves determining the parameter. Although parametric regression is used widely in practice, in many situations, one might be reluctant to choose a specific form of a model to fit a particular set of data. The field of nonparametric regression has developed to fit a curve to data without assuming any particular structure on the regression function f . We assume that the values x_i 's are increasingly ordered: $x_1 < x_2 < \dots < x_n$ and let $d_i = x_i - x_{i-1}$, $i = 2, 3, \dots, n$. We define d_1

as mean of other values d_i $d_1 = \frac{\sum_{i=2}^n d_i}{n-1} = \frac{x_n - x_1}{n-1}$.

If we note $a = x_1 - d_1 = x_0$, $b = x_n$, we obtain

$a = x_0 < x_1 < x_2 < \dots < x_n = b$, which means that the values of variable x lie on the interval $[a, b]$. Because we wish as these values to be placed in interval $[0, 1]$, we achieve a re-scaling of these through linear transformation $\varphi: [a, b] \rightarrow [0, 1]$, $\varphi(x) = \frac{x-a}{b-a}$. Using notation $x'_i = \varphi(x_i)$, $i = 1, 2, \dots, n$, we have $0 = x'_0 < x'_1 < x'_2 < \dots < x'_n = 1$. For this reasons, it will be assumed that the variables x_i lie on the unit interval $[0, 1]$. Frequently, x_i are equally spaced, which means that $d_1 = d_2 = \dots = d_n$. In this situation we obtain $x'_i = \frac{i}{n}$, $i = 1, 2, \dots, n$ (2). The raw estimator

of the regression function f is the piecewise constant function

$$\tilde{f}(x) = \begin{cases} y_i, & x_{i-1} \leq x < x_i, i = 1, 2, \dots, n \\ 1, & x = 1 \end{cases} \quad (3).$$

If the values x_i are equally spaced, we have

$$\tilde{f}(x) = \begin{cases} y_i, & \frac{i-1}{n} \leq x < \frac{i}{n}, i = 1, 2, \dots, n \\ 1, & x = 1 \end{cases} \quad (4).$$

Farther we introduce some theoretical elements necessary about scaling functions.

Scaling functions

A function $f: \mathbf{R} \rightarrow \mathbf{R}$ ($f: [a, b] \rightarrow \mathbf{R}$) belongs to the square-integrable function space

$L^2(\mathbf{R})$ ($L^2[a, b]$), if both f and f^2 are integrable on \mathbf{R} (on $[a, b]$). Therefore, we have $\int_{-\infty}^{+\infty} f^2(x)dx < \infty$ ($\int_a^b f^2(x)dx < \infty$). For two functions $f, g \in L^2(\mathbf{R})$ ($f, g \in L^2[a, b]$) we define inner product $\langle f, g \rangle = \int_{-\infty}^{+\infty} f(x)g(x)dx$ ($\langle f, g \rangle = \int_a^b f(x)g(x)dx$), by means him we obtain the norm of a function $f \in L^2(\mathbf{R})$ ($f \in L^2[a, b]$), $\|f\| = \sqrt{\langle f, f \rangle}$. A basis $\{f_j\}_{j \in \mathbf{Z}}$ for $L^2(\mathbf{R})$ ($L^2[a, b]$) is a orthonormal basis if $\|f_j\| = 1, \forall j \in \mathbf{Z} \langle f_j, f_k \rangle = 0, \forall j, k \in \mathbf{Z}, j \neq k$. The vectors of a set $A \subset L^2(\mathbf{R})$ ($A \subset L^2[a, b]$) span a subspace V if $V = \left\{ \sum_{a \in A} \alpha_a a / \alpha_a \in \mathbf{R}, \forall a \in A \right\}$ and we write $V = Sp(A)$. A sequence of subspaces $(V_j)_{j \in \mathbf{Z}}$ of $L^2(\mathbf{R})$ ($L^2[a, b]$) form a *multiresolution analysis* if

- 1) $\dots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots;$
- 2) $\bigcap_{j \in \mathbf{Z}} V_j = \{0\}, \quad \overline{\bigcup_{j \in \mathbf{Z}} V_j} = L^2(\mathbf{R})$ ($L^2[a, b]$);
- 3) $f \in V_j \Leftrightarrow f(2 \cdot) \in V_{j+1}, \quad \forall j \in \mathbf{Z};$
- 4) $f \in V_0 \Rightarrow f(\cdot - k) \in V_0, \quad \forall k \in \mathbf{Z};$
- 5) There exists a function $\Phi \in V_0$ such that the set $\{\Phi_{j,k} / k \in \mathbf{Z}\}$, where $\Phi_{j,k}(x) = 2^{\frac{j}{2}} \Phi(2^j x - k), \forall k \in \mathbf{Z}$ (5) constitutes an orthonormal basis for $V_j, \forall j \in \mathbf{Z}$. The fifth property show that $V_j = Sp(\{\Phi_{j,k} / k \in \mathbf{Z}\})$. The set $\{\Phi_{0,k} = \Phi(\cdot - k) / k \in \mathbf{Z}\}$, obtained by applying integer translations, form a basis for V_0 . The function Φ is called the *scaling function* since its dilates and translates constitute orthonormal bases for all V_j subspaces, which are simply scaled versions of V_0 . For any function $f \in L^2(\mathbf{R})$ ($f \in L^2[a, b]$), we note $P^j f$ the projection of a function f onto the subspace V_j . By virtue of the second property, approximation $P^j f$ of function f

converges of this function $P^j f \rightarrow f$, as $j \rightarrow \infty$ in the L^2 sense $\|P^j f - f\| \rightarrow 0$, as $j \rightarrow \infty$. Because $\{\Phi_{j,k} / k \in \mathbf{Z}\}$ is a basis for V_j , we get $(P^j f)(x) = \sum_{k \in \mathbf{Z}} c_{j,k} \Phi_{j,k}(x)$ (6). For functions f with support compact ($\text{supp}(f) = \overline{\{x \in \mathbf{R} / f(x) \neq 0\}}$), it is easily seen that only finitely many sequence elements $\{c_{j,k}\}_{k \in \mathbf{Z}}$ will be non-zero. Since the set $\{\Phi_{j,k} / k \in \mathbf{Z}\}$ is an orthonormal basis for V_j , the scaling function coefficients can be computed $c_{j,k} = \langle f, \Phi_{j,k} \rangle = \int_{-\infty}^{+\infty} f(u) \Phi_{j,k}(u) du$ (7). Far away, we propose to express the coefficients $c_{j,k}$ of the approximation of f in space V_j in terms of the coefficients $c_{j+1,k}$ of the approximation of f in space V_{j+1} . Note that $\Phi \in V_0$ and therefore also $\Phi \in V_1$ since $V_0 \subset V_1$. Because $\{\Phi_{1,k} / k \in \mathbf{Z}\}$ is an orthonormal basis for V_1 , there exists a sequence $\{h_k\}_{k \in \mathbf{Z}}$, which is called *filter*, such that $\Phi(x) = \sum_{k \in \mathbf{Z}} h_k \Phi_{1,k}(x)$ (8) and that the sequence elements may be written $h_k = \langle \Phi, \Phi_{1,k} \rangle = \int_{-\infty}^{+\infty} \Phi(x) \Phi_{1,k}(x) dx$ (9). This sequence is a square-summable sequence ($\{h_k\}_{k \in \mathbf{Z}} \in l^2(\mathbf{Z})$), that is $\sum_{k \in \mathbf{Z}} h_k^2 < \infty$. Using (5) and (8), we may show that $\Phi_{j,k}(x) = \sum_{l \in \mathbf{Z}} h_{l-2k} \Phi_{j+1,l}(x)$ (10). Substituting this result into (7) we obtain $c_{j,k} = \sum_{l \in \mathbf{Z}} h_{l-2k} c_{j+1,l}$ (11) relation which allow to write coefficients on $P^j f$ in terms of coefficients on $P^{j+1} f$. The simplest scaling function is $\Phi(x) = \begin{cases} 1, & x \in [0, 1) \\ 0, & \text{otherwise} \end{cases}$. Applying a usual dilation and usual translations, we obtain $\Phi_{j,k}(x) = \begin{cases} 2^{\frac{j}{2}}, & x \in [2^{-j} k, 2^{-j}(k+1)) \\ 0, & \text{otherwise} \end{cases}$

which show that approximation spaces V_j consists of functions that are piecewise constant, more exactly $V_j = \{f \in L^2(\mathbf{R}) / f \text{ is piecewise constant on } [2^{-j}k, 2^{-j}(k+1)), k \in \mathbf{Z}\}$. At the same time, according to (7), the coefficients $c_{j,k}$ are computed from the formula

$$c_{j,k} = \int_{2^{-j}k}^{2^{-j}(k+1)} 2^{\frac{j}{2}} f(u) du.$$

Since for j fixed, any $x \in \mathbf{R}$ lie on the only one interval of type $[2^{-j}k, 2^{-j}(k+1))$, using (6) we obtain $(P^j f)(x) = \frac{1}{2^{-j}} \int_{2^{-j}k}^{2^{-j}(k+1)} f(u) du$ which show

that this approximation of f in x is “mean” of all function values on the interval $[2^{-j}k, 2^{-j}(k+1))$. Another example of scaling function is the Battle-Lemarié family. The first member of the Battle-Lemarié family is identical to the scaling function defined before $N_1(x) = \begin{cases} 1, & x \in [0,1) \\ 0, & \text{otherwise} \end{cases}$. In constructing of

the members of this family, the cardinal B -spline functions are useful, which are piecewise polynomial functions, defined through convolution operator $*$ (for $f, g \in L^2(\mathbf{R})$) we have:

$$(f * g)(x) = \int_{-\infty}^{+\infty} f(x-v)g(v)dv.$$

The first cardinal B -spline function is $N_1(x)$ introduced before, and cardinal B -splines of higher order are defined recursively in terms of lower order B -splines by means of the convolution operator

$$N_m(x) = (N_{m-1} * N_1)(x) = \int_0^1 N_{m-1}(x-u)du,$$

$$n = 2, 3, \dots$$

It is interesting to note also that the function $N_m(x)$ is also the probability density function for the sum of m independent uniform $(0,1)$ random variables. Therefore, by the Central Limit Theorem, as m gets large, the spline function $N_m(x)$ approaches the probability density function of a normal random variable.

The second member of the Battle-Lemarié family begins with the piecewise linear spline $N_2(x)$. But this function is not or-

thogonal to its integer translates and therefore the function $N_2(x)$ must be “orthogonalized” to give an appropriate scaling function $N_2(x)$. Unfortunately, there is no closed form expression for the resulting scaling function, but it is given only in terms of its

Fourier transform
$$\hat{\Phi}(\omega) = \sqrt{\frac{3}{2\pi}} \frac{4 \sin^2 \frac{\omega}{2}}{\omega^2 \sqrt{1 + 2 \cos^2 \frac{\omega}{2}}}.$$

The function Φ is piecewise linear, it is not compactly supported, but it does decay exponentially. Mallat gives general formulas for constructing scaling functions for the Battle-Lemarié family, as well as for computing the associated filter $\{h_k\}_{k \in \mathbf{Z}}$. The resulting scaling functions will be polynomial splines. Another important family of scaling functions is the Daubechies family. They are compactly supported scaling functions. The derivation of this family represents quite a different approach. Thus, we begin with the filters, and then we derive from the filter coefficients the scaling function. Obvious, for compactly supported Φ , only finitely many of the h_k 's will be non-zero. Daubechies approach to constructing scaling functions begins with defining the trigonometric polynomial

$$m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{k \in \mathbf{Z}} h_k e^{-ik\omega} = \frac{1}{\sqrt{2}} \sum_{k \in \mathbf{Z}} h_k (\cos k\omega - i \sin k\omega)$$

The scaling functions are written in terms of this $m_0(\omega)$ function, and, again, they can only be expressed in terms of their Fourier

transforms
$$\hat{\Phi}(\omega) = \frac{1}{\sqrt{2\pi}} \prod_{j=1}^{\infty} m_0(2^{-j}\omega).$$
 There

are several possibilities for the function $m_0(\omega)$, depending on the number of non-zero elements in the filter. Clearly, the more non-zero filter elements make the smoother the resulting scaling functions.

Estimation of nonparametric regression

For nonparametric estimation of regression function it can us kernel smoothing, orthogonal series estimation and scaling functions estimation. This paper will focus on only the scaling function version of these methods as applied to nonparametric regression, that is recovery of a regression function f as in (1).

The estimator of regression function f will be the projection of raw estimator \tilde{f} , defined in (3), onto the approximation space V_J . Therefore, according to (6), we have $\hat{f}_J(x) = (P^J \tilde{f})(x) = \sum_{k \in \mathbf{Z}} \hat{c}_{J,k} \Phi_{J,k}(x)$ (12). If in previous section we discussed the function $f \in L^2(\mathbf{R})$, possibly with infinite support, farther we have the function \tilde{f} , evi-

$$\hat{f}_J(x) = \sum_{k \in \mathbf{Z}} \left(\sum_{i=1}^n \int_{x_{i-1}}^{x_i} y_i \Phi_{J,k}(u) du \right) \Phi_{J,k}(x) = \sum_{i=1}^n y_i \int_{x_{i-1}}^{x_i} \left(\sum_{k \in \mathbf{Z}} \Phi_{J,k}(u) \Phi_{J,k}(x) \right) du = \sum_{i=1}^n y_i \int_{x_{i-1}}^{x_i} E_J(x, u) du \quad (14)$$

where $E_J(x, u) = \sum_{k \in \mathbf{Z}} \Phi_{J,k}(u) \Phi_{J,k}(x)$ (15)

The sum in (15) is finite because $\text{supp}(\tilde{f}) = [0, 1]$ is compact. E_J defined above is a variable kernel. It is not apparent that the estimator \hat{f}_J depends explicitly on the choice of parameter J . A relatively large value of parameter J gives a narrow smoothing window, so the estimate of the function at x depends primarily on data near x . Thus, estimators for J large tend to be rather wiggly. Conversely, a smaller choice of parameter J gives a wider smoothing window, resulting in more averaging and a smoother estimate. Large values of J give estimators with smaller bias, but larger variance. Selecting a smaller parameter J will decrease the variance, but increase the bias. The developments above show that this estimator is at the same time both an orthogonal series and a kernel estimator. For better emphasize of the correlation between the two variable x and y , we need more estimators \hat{f}_J , both for larger values of J , and for smaller values of J . Therefore, in the beginning choose a scaling function Φ , a larger value J_1 of the parameter J and a smaller value J_0 of the same parameter, and afterwards we shall use the following algorithm:

1) We compute the estimated coefficients $\hat{c}_{J_1,k}$, which appear in writing of estimator \hat{f}_{J_1} , using formula

$$\hat{c}_{J_1,k} = \sum_{i=1}^n y_i \int_{x_{i-1}}^{x_i} \Phi_{J_1,k}(u) du \quad (16), \text{ according to (13). If}$$

it is necessary, the integral given in (16) is

cently square-integrable function on the interval $[0, 1]$. The coefficients $\hat{c}_{J,k}$, which are estimation of "true" coefficients, according to (7) are computed by

$$\hat{c}_{J,k} = \langle \tilde{f}, \Phi_{J,k} \rangle = \int_0^1 \tilde{f}(u) \Phi_{J,k}(u) du \quad (13).$$

Substituting (13) into (12), using (3) and rearranging terms, we obtain

computed on the basis of a numerical method (Bessel, Hermite, Simpson, Romberg etc.).

2) Using (9) we compute the filter coefficients $h_k = \int_0^1 \Phi(x) \sqrt{2} \Phi(2x - k) dx$ (17) possibly on the basis of a numerical method.

3) For M large we compute values of the regression function estimator \hat{f}_{J_1} in the points $\frac{l}{M}$, according to (12)

$$\hat{f}_{J_1}\left(\frac{l}{M}\right) = \sum_k \hat{c}_{J_1,k} \Phi_{J_1,k}\left(\frac{l}{M}\right), \quad l = 0, 1, \dots, M \quad (18) \text{ and}$$

we represent it graphically.

4) Using (11) we determine the coefficients which appear in writing of estimator \hat{f}_{J_1-1} , $\hat{c}_{J_1-1,k} = \sum_l h_{l-2k} \hat{c}_{J_1,l}$ (19)

5) With $J_1 - 1$ in place of J_1 we take again the proceeding from step 3), until the value of this parameter gets equal to J_0 .

Bibliography

1. Antoniadis, A., Gregoire, G., McKeague, I. W. – *Wavelet Methods for Curve Estimation*, Journal of the American Statistical Association 89, pp. 1340-1353, 1994
2. Daubechies, I. – *Orthonormal Bases of Compactly Supported Wavelets*, SIAM Journal on Mathematical Analysis 24, pp. 499-519, 1993
3. Eubank, R. L. – *Spline Smoothing and Nonparametric regression*, Marcel Dekker, New York, 1988
4. Ogden, R. T. – *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhäuser, Boston, 1997
5. Tătaru, G. – *Serii Fourier. Transformata Laplace*, Editura Economică București, 2003
6. Walter, G. G. – *Wavelets and other Orthogonal Systems with Applications*, CRC Press, Boca Raton, Florida, 1994