Chapter 3. Spline smoothing and semi-parametric Models (II) Part 1

March 9, 2007

1 Polynomial Spline

1.1 Basic idea

A polynomial of kth order is

$$P(x) = a_0 + a_1 x + ... + a_k x^k,$$

where $a_0, ..., a_k$ are constant parameters. If $k = 3$, it is called the cubic polynomial.

Recall that we are interested in the conditional expectation function of Y given X,

$$m(x) = E(Y|X = x)$$

or model

$$Y = m(x) + \varepsilon \tag{1.1}$$

The function $m(.)$ can be very complicated. Assuming it is smooth enough, it can be approximated by polynomial functions **locally**.

Note that the kernel smoothing approximate the function pointwise. i.e., for each point, we use a polynomial to approximate the underline function.

In contrast, polynomial splines approximate the function piecewise. That is we partition the region of x into several subintervals such that (1) in each interval a polynomial is used to approximate the true function (2) at the ends of interval, the function is smooth (has kth order derivatives).

The most popular order is 3, that is the cublic spline. For concreteness, we focus on the cubic splines, but the idea can also be applied to splines of any order k. The idea can be stated as follows. Let $t_1, ..., t_J$ be a fixed knot sequence such that

$$-\infty < t_1 < t_2 < ... < t_J < \infty$$

The cubic spline functions are twice continuously differentiable function of x and in each interval $(-\infty, t_1], [t_1, t_2], ..., [t_{J-1}, t_J], [t_J, \infty)$ is a cubic polynomial.

$t_j, j = 1, 2, ..., J$ are called knots. J is the number of knots.

The collection of all cubic spline functions forms a $(J + 4)$-dimensional linear space. There are two popular cubic spline bases

- power basis: $1, x, x^2, x^3, (x - t_j)^3_+,\ (j = 1, ..., J)$, where
$$(x - t_j)^3_+ = \begin{cases} (x - t_j)^3_+, & \text{if } x - t_j \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$
- B-spline basis (Boor (1978)).

1.2 Estimation

Let $B_1(x), B_2(x), ..., B_{J+4}(x)$ be the basis. The a cubic spline function can be expressed as

$$s(x) = \sum_{j=1}^{J+4} \theta_j B_j(x).$$

It is proved that for any function $m(x)$ with continuous derivatives on [a, b]. If the knots $t_{j+1} - t_j \to 0$ then,

$$s(x) \to m(x)$$

for any $x \in [a, b]$.

To simplify the statistical theory, people change to investigate a cubic spline model

$$Y = \sum_{j=1}^{J+4} \theta_j B_j(X) + \varepsilon \tag{1.2}$$

instead of (1.1). That is assume $m(x) = \sum_{j=1}^{J+4} \theta_j B_j(x)$.

Suppose $(X_1, Y_1), ..., (X_n, Y_n)$ is a random sample. The the least squares estimation is to minimize

$$\min_{\theta_1, ..., \theta_{J+4}} \sum_{i=1}^{n} \{Y_i - \sum_{j=1}^{J+4} \theta_j B_j(X_i)\}^2.$$

Let

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} B_1(X_1) & B_2(X_1) & \dots & B_{J+4}(X_1) \\ B_1(X_2) & B_2(X_2) & \dots & B_{J+4}(X_2) \\ \dots \\ B_1(X_n) & B_2(X_n) & \dots & B_{J+4}(X_n) \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \dots \\ \theta_{J+4} \end{pmatrix}$$

Then, the estimator for θ is

$$\hat{\theta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$

The estimated funciton is then

$$\hat{m}(x) = (B_1(x), ..., B_{J+4}(x))\hat{\theta} = \sum_{j=1}^{J+4} \hat{\theta}_j B_j(x).$$

1.3 The choice of knots and the number of knots

Theoretically, we can choose them by the AIC method or CV method (this procedure might be extremely complicated and time consuming). In practice, we can order the observations of X:

$$X_{[1]} \leq X_{[2]} \leq ... \leq X_{[n]},$$

Then the knots are choosen as

$$X_{[k]}, X_{[2k]}, X_{[3k]}, ...$$

where k is usually greater than 5. Then the number of knots is around n/k, i.e. $J \approx n/k$. In practice, the J is usually not very large.

Example 1.1 (simulation) *Suppose the model is*

$$Y = \sin(2\pi X) + 0.2\varepsilon$$

where $X \sim Uniform(0, 1)$ and $\varepsilon \sim N(0, 1)$ are independent. in this model

$$m(x) = \sin(2\pi x).$$

50 observations are sampled and plotted below. with different number of knots $k = 3, 8, 15, 25$, the estiamted functions are shown in figure 1

Comparing with kernel smoothing, it seems that polynomial spline can give a better estimation. Polynomial splines are not so sensitive to the tuning parameters (number of knots and the position of the knots) than kernel smoothing.

Example 1.2 (motorcycle) **(data)** *See figure 2*

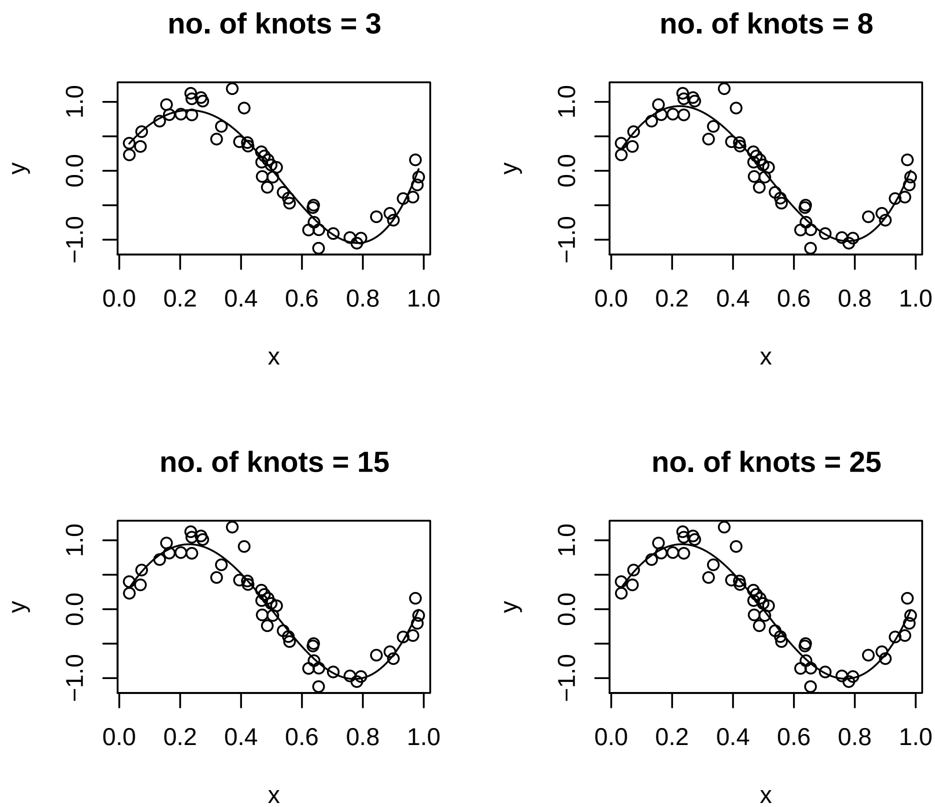

Figure 1: **(pspline.R) (c3a1.R)**

1.4 Statistical inference

If we further assume that

$$\varepsilon \sim N(0, \sigma^2)$$

Then it is easy to see that the estimator of θ follows normal.

$$(\hat{\theta} - \theta) \sim N(0, (\mathbf{X}^\top \mathbf{X})^{-1} \sigma^2)$$

Therefore, the distribution of $\hat{m}(x) = (B_1(x), ..., B_{J+4}(x))\hat{\theta}$ is

$$\hat{m}(x) - m(x) \sim N\{0, (B_1(x), ..., B_{J+4}(x))(\mathbf{X}^\top \mathbf{X})^{-1}(B_1(x), ..., B_{J+4}(x))^\top \sigma^2\}$$

Remark 1.3 *why the estimator $\hat{m}(x)$ is "unbiased"? because we assume the approximation by spline function is accurate [recall if we assume the polynomial approximaiton is accurate, we can also prove that kernel smoothing estimator is "unbiased"]*

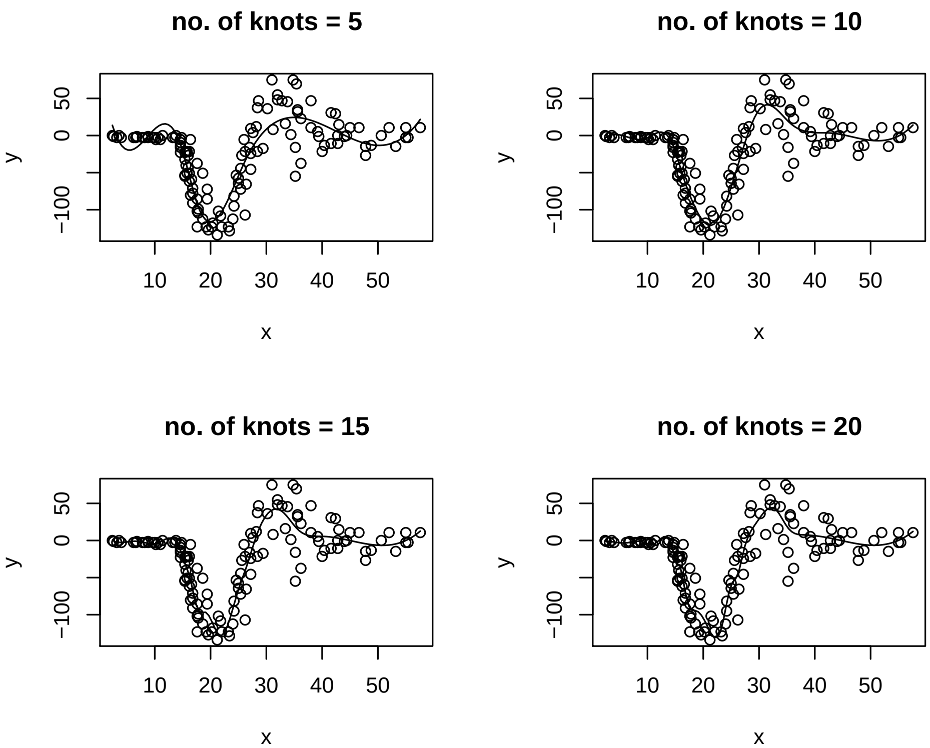

Figure 2: **(pspline.R) (c3a2.R)**

Therefore the 95% confidence band is

$$\hat{m}(x) \pm 1.96\{(B_1(x), ..., B_{J+4}(x))(\mathbf{X}^\top \mathbf{X})^{-1}(B_1(x), ..., B_{J+4}(x))^\top \sigma^2\}^{1/2}$$

In practice, σ^2 can be estimated by

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} \{Y_i - \hat{m}(X_i)\}^2$$

Example 1.4 (motorcycle) **(data)** *See figure 3*

1.5 Codes in R

There are different kinds of splines. Examples are the polynomial splines, smoothing splines and penalized splines. For smoothing spline, the codes are smooth.spline() and predict()

Example 1.5 (motorcycle) **(data)** *See figure 4*

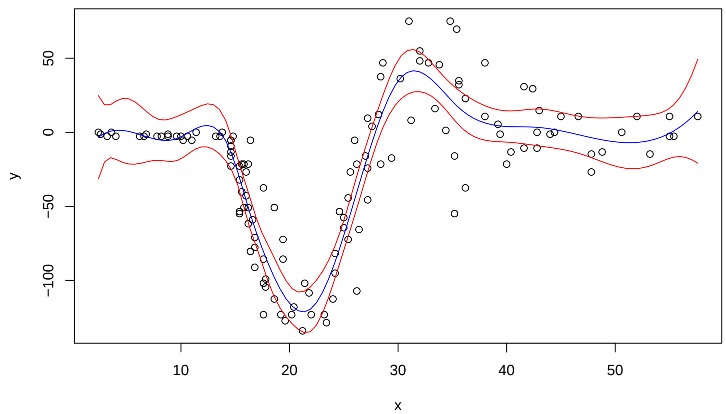

Figure 3: **(pspline.R) (c3a3.R)**

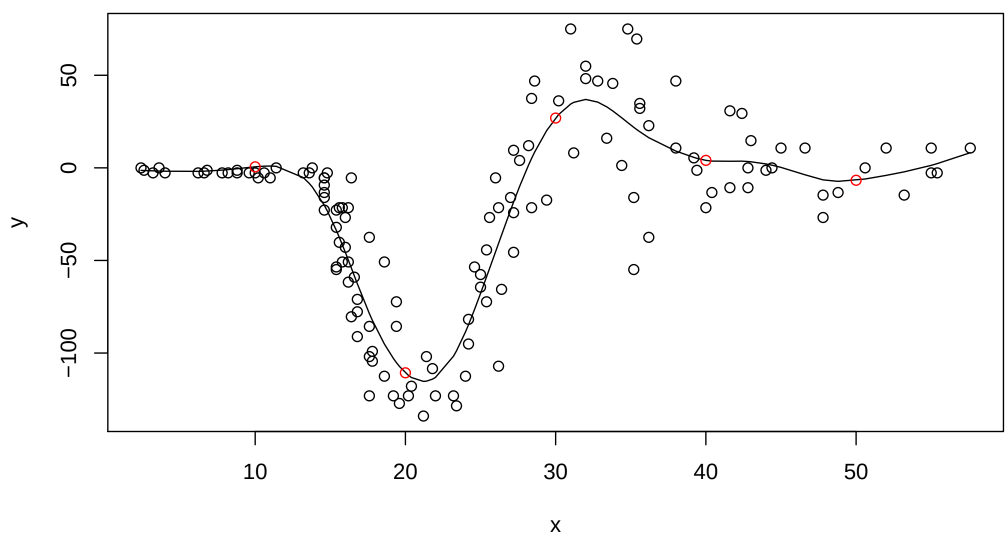

Figure 4: **(pspline.R) (c3a4.R)**