p=1 Piece-wise Polynomials and Regression Splines

Stephen Vardeman
Analytics Iowa LLC
ISU Statistics and IMSE

Set-up with *K* knots

Continue consideration of the case of a one-dimensional input variable x, and now K "knots"

$$\xi_1 < \xi_2 < \cdots < \xi_K$$

and forms for $f(x)$ that are

1. polynomials of order M (or less) on all intervals (ξ_{j-1}, ξ_j), and (potentially, at least)
2. have derivatives of some specified order at the knots, and (potentially, at least)
3. are linear outside (ξ_1, ξ_K).

Obvious basis functions

Let $I_1(x) = I[x < \xi_1]$, for $j = 2, \ldots, K$ set $I_j(x) = I[\xi_{j-1} \le x < \xi_j]$, and define $I_{K+1}(x) = I[\xi_K \le x]$. One can have piecewise polynomials using basis functions

$$
\begin{aligned}
& I_1(x), I_2(x), \ldots, I_{K+1}(x) \\
& xI_1(x), xI_2(x), \ldots, xI_{K+1}(x) \\
& \vdots \\
& x^M I_1(x), x^M I_2(x), \ldots, x^M I_{K+1}(x)
\end{aligned}
$$

Further, continuity and differentiability (at the knots) conditions for $f(x) = \sum_{m=1}^{(M+1)(K+1)} \beta_m h_m(x)$ follow from enforcing some linear relations between appropriate β_ms. This is conceptually simple, but messy. It is much cleaner to simply begin with a set of basis functions that are tailored to have desired continuity/differentiability properties.

Another basis and "natural cubic splines"

A set of $M+1+K$ basis functions for piecewise polynomials of degree M with derivatives of order $M-1$ at all knots is easily seen to be

$$1, x, x^2, \ldots, x^M, (x-\xi_1)_+^M, (x-\xi_2)_+^M, \ldots, (x-\xi_K)_+^M$$

(since the value and first $M-1$ derivatives of $(x-\xi_j)_+^M$ at ξ_j are all 0). The choice of $M=3$ is fairly standard.

Since extrapolation with polynomials typically gets worse with order, it is common to impose a restriction that outside (ξ_1, ξ_K) a form $f(x)$ be linear. For the case of $M=3$ this can be accomplished by beginning with basis functions $1, x, (x-\xi_1)_+^3, (x-\xi_2)_+^3, \ldots, (x-\xi_K)_+^3$ and imposing restrictions necessary to force 2nd and 3rd derivatives to the right of ξ_K to be 0. This produces so-called "natural" (linear outside (ξ_1, ξ_K)) cubic regression splines.

Development of natural cubic splines

Notice that (considering $x > \xi_K$)

$$\frac{d^2}{dx^2}\left(\alpha_0 + \alpha_1 x + \sum_{j=1}^{K} \beta_j \left(x - \xi_j\right)_+^3\right) = 6\sum_{j=1}^{K} \beta_j \left(x - \xi_j\right) \qquad (1)$$

and

$$\frac{d^3}{dx^3}\left(\alpha_0 + \alpha_1 x + \sum_{j=1}^{K} \beta_j \left(x - \xi_j\right)_+^3\right) = 6\sum_{j=1}^{K} \beta_j \qquad (2)$$

So, linearity for large x requires (from (2)) that $\sum_{j=1}^{K} \beta_j = 0$. Further, substituting this into (1) means that linearity also requires that $\sum_{j=1}^{K} \beta_j \xi_j = 0$.

Development of natural cubic splines

Using then $\sum_{j=1}^{K} \beta_j = 0$ to conclude that $\beta_K = -\sum_{j=1}^{K-1} \beta_j$ and substituting into $\sum_{j=1}^{K} \beta_j \xi_j = 0$ yields

$$\beta_{K-1} = -\sum_{j=1}^{K-2} \beta_j \left(\frac{\xi_K - \xi_j}{\xi_K - \xi_{K-1}} \right)$$

and then

$$\beta_K = \sum_{j=1}^{K-2} \beta_j \left(\frac{\xi_K - \xi_j}{\xi_K - \xi_{K-1}} \right) - \sum_{j=1}^{K-2} \beta_j$$

These then suggest the set of basis functions consisting of $1, x$ and for $j = 1, 2, \ldots, K-2$ the functions

$$(x - \xi_j)_+^3 - \left(\frac{\xi_K - \xi_j}{\xi_K - \xi_{K-1}} \right) (x - \xi_{K-1})_+^3 + \left(\frac{\xi_K - \xi_j}{\xi_K - \xi_{K-1}} \right) (x - \xi_K)_+^3 - (x - \xi_K)_+^3$$

Bases

These are the functions $1, x$ and for $j = 1, 2, \ldots, K-2$

$$(x-\xi_j)_+^3 - \left(\frac{\xi_K - \xi_j}{\xi_K - \xi_{K-1}}\right)(x-\xi_{K-1})_+^3 + \left(\frac{\xi_{K-1} - \xi_j}{\xi_K - \xi_{K-1}}\right)(x-\xi_K)_+^3$$

These are essentially the basis functions that HTF call their N_j.

There are other (harder to motivate, but in the end more pleasing and computationally more attractive) sets of basis functions for natural polynomial splines. See the B-spline material at the end of HTF Chapter 5.