Proposals for a cyclic (restricted cubic spline)

Lara Lusa, Crt Ahlin

October 2016

1 Definition of restricted cubic splines

We start with the definition of cubic splines:

$$s(x) = \gamma_0 + \gamma_1 X + \gamma_2 X^2 + \gamma_3 X^3 + \beta_1 (X - t_1)^3_+ + \beta_2 (X - t_2)^3_+ + ... + \beta_k (X - t_k)^3_+ + \tag{1}$$

$$s(x) = \gamma_0 + \gamma_1 X + \gamma_2 X^2 + \gamma_3 X^3 + \sum_{j=1}^{k} \beta_j (X - t_j)^3_+ \tag{2}$$

Restricted cubic splines have to be linear in the first and last interval. Linearity for $x \leq t_1$ implies that $\gamma_2 = \gamma_3 = 0$

Linearity for $x \geq t_k$ implies that the spline can be defined using k-2 coefficients, re-parametrizing the function

$$s(x) = \gamma_0 + \gamma_1 x + \sum_{j=1}^{k-2} \beta_j \left((x - t_j)^3_+ + (x - t_{k-1})^3_+ \frac{t_k - t_j}{t_k - t_{k-1}} + (x - t_k)^3_+ \frac{t_{k-1} - t_j}{t_k - t_{k-1}} \right)$$

Let us define $b(x, t_j) = \left((x - t_j)^3_+ + (x - t_{k-1})^3_+ \frac{t_k - t_j}{t_k - t_{k-1}} + (x - t_k)^3_+ \frac{t_{k-1} - t_j}{t_k - t_{k-1}} \right)$

$$s(x) = \gamma_0 + \gamma_1 x + \sum_{j=1}^{k-2} \beta_j b(x, t_j)$$

For $x \leq t_1$

$$b(x, t_j) = b'(x, t_j) = b''(x, t_j) = 0 \forall t_j : j = 1, ..., k$$

and

$$s(x_{min}) = \gamma_0 + \gamma_1 x_{min}$$

For $x \geq t_k$

$$b(x, t_j) = (x - t_j)^3 + (x - t_{k-1})^3 \frac{t_k - t_j}{t_k - t_{k-1}} + (x - t_k)^3 \frac{t_{k-1} - t_j}{t_k - t_{k-1}}$$

$$b'(x,t_j) = 3(x-t_j)^2 + 3(x-t_{k-1})^2\frac{t_k-t_j}{t_k-t_{k-1}} + 3(x-t_k)^2\frac{t_{k-1}-t_j}{t_k-t_{k-1}}$$

$$b''(x,t_j) = 6(x-t_j) + 6(x-t_{k-1})\frac{t_k-t_j}{t_k-t_{k-1}} + 6(x-t_k)\frac{t_{k-1}-t_j}{t_k-t_{k-1}}$$

and

$$s(x_{max}) = \gamma_0 + \gamma_1 x_{max} + \sum_{j=1}^{k-2}\beta_j b(x_{max},t_j)$$

2 Constraining RCS to achieve the same estimated value at the beginning and end of the cycle

Suppose that the covariate that we want to model is a cyclic variable and that data are organized so that the minimum and maximum value correspond to the beginning and the end of the cycle.

2.1 Constraint on the estimated value at the beginning and end of the cycle

It would be desirable that the estimated value at the beginning and the end of the cycle are equal. We can achieve this goal by imposing an additional constrain, s(x_{min})=s(x_{max}).

$$\gamma_0 + \gamma_1 x_{min} = \gamma_0 + \gamma_1 x_{max} + \sum_{j=1}^{k-2}\beta_j b(x_{max},t_j)$$

$$\gamma_1 = \frac{\sum_{j=1}^{k-2}\beta_j b(x_{max},t_j)}{(x_{min}-x_{max})}$$

$$s(x) = \gamma_0 + \frac{\sum_{j=1}^{k-2}\beta_j b(x_{max},t_j)}{(x_{min}-x_{max})}x + \sum_{j=1}^{k-2}\beta_j b(x,t_j) = \gamma_0 + \sum_{j=1}^{k-2}\beta_j\left(b(x,t_j) + \frac{b(x_{max},t_j)}{(x_{min}-x_{max})}x\right)$$

Hence, with a simple modification of the design matrix we can easily introduce a constraint on the equality of the estimated spline function at the beginning and at the end of the cycle.

A second parametrization, with which we can achieve the same result, while maintaining the linear term in the model, is the following:

$$\gamma_0 + \gamma_1 x_{min} = \gamma_0 + \gamma_1 x_{max} + \sum_{j=1}^{k-2}\beta_j b(x_{max},t_j)$$

$$\beta_{k-2}b(x_{max},t_{k-2}) = \gamma_1(x_{min}-x_{max}) - \sum_{j=1}^{k-3}\beta_j b(x_{max},t_j)$$

$$\beta_{k-2} = \frac{\gamma_1(x_{min}-x_{max}) - \sum_{j=1}^{k-3}\beta_j b(x_{max},t_j)}{b(x_{max},t_{k-2})}$$

$$s(x)=\gamma_0+\gamma_1\left(X+\frac{x_{min}-x_{max}}{b(x_{max},t_{k-2})}b(x,t_{k-2})\right)+\sum_{j=1}^{k-3}\beta_j\left(b(x,t_j)-\frac{b(x_{max},t_j)}{b(x_{max},t_{k-2})}b(x,t_{k-2})\right)$$

Using the first or second parametrization, which are equivalent, we achieve that the estimated value in the extreme points is the same.

2.2 Constraint on the equality of the first derivatives of the spline at the beginning and end of the cycle

The first constrain guarantees the equality of the estimated values of the function at the beginning and end of the cycles, but the slopes of the linear functions estimated for the values smaller than the first node and larger than the last node do not necessarily coincide.

Adding the additional constraint
$s'(x_{min})=s'(x_{max})$
we can achieve a smoother estimate at the end of the cycle, where

$$s(x)=\gamma_0+\gamma_1\left(X+\frac{x_{min}-x_{max}}{b(x_{max},t_{k-2})}b(x,t_{k-2})\right)+\sum_{j=1}^{k-3}\beta_j\left(b(x,t_j)-\frac{b(x_{max},t_j)}{b(x_{max},t_{k-2})}b(x,t_{k-2})\right). \quad (3)$$

$$s'(x_{min})=\gamma_1$$

$$s'(x_{max})=\gamma_1\left(1+\frac{x_{min}-x_{max}}{b(x_{max},t_{k-2})}b'(x_{max},t_{k-2})\right)+\sum_{j=1}^{k-3}\beta_j\left(b'(x_{max},t_j)-\frac{b(x_{max},t_j)}{b(x_{max},t_{k-2})}b'(x,t_{k-2})\right).$$

$$\beta_{k-3}\left(b'(x_{max},t_{k-3})-\frac{b(x_{max},t_{k-3})}{b(x_{max},t_{k-2})}b'(x_{max},t_{k-2})\right)=$$

$$-\gamma_1\left(\frac{x_{min}-x_{max}}{b(x_{max},t_{k-2})}b'(x_{max},t_{k-2})\right)-\sum_{j=1}^{k-4}\beta_j\left(b'(x_{max},t_j)-\frac{b(x_{max},t_j)}{b(x_{max},t_{k-2})}b'(x_{max},t_{k-2})\right)$$

$$\beta_{k-3}=\frac{-\gamma_1\left(\frac{x_{min}-x_{max}}{b(x_{max},t_{k-2})}b'(x_{max},t_{k-2})\right)-\sum_{j=1}^{k-4}\beta_j\left(b'(x_{max},t_j)-\frac{b(x_{max},t_j)}{b(x_{max},t_{k-2})}b'(x_{max},t_{k-2})\right)}{b'(x_{max},t_{k-3})-\frac{b(x_{max},t_{k-3})}{b(x_{max},t_{k-2})}b'(x_{max},t_{k-2})}$$

Substituting in the equation 3

$$
\begin{aligned}
s(x) = & \gamma_0 + \gamma_1 \left(X + \frac{x_{min} - x_{max}}{b(x_{max}, t_{k-2})} b(x, t_{k-2}) \right) + \\
& \sum_{j=1}^{k-4} \beta_j \left(b(x, t_j) - \frac{b(x_{max}, t_j)}{b(x_{max}, t_{k-2})} b(x, t_{k-2}) \right) + \\
& \beta_{k-3} \left(b(x, t_{k-3}) - \frac{b(x_{max}, t_{k-3})}{b(x_{max}, t_{k-2})} b(x, t_{k-2}) \right) = \\
& \gamma_0 + \gamma_1 \left(X + \frac{x_{min} - x_{max}}{b(x_{max}, t_{k-2})} b(x, t_{k-2}) \right) + \\
& \sum_{j=1}^{k-4} \beta_j \left(b(x, t_j) - \frac{b(x_{max}, t_j)}{b(x_{max}, t_{k-2})} b(x, t_{k-2}) \right) + \\
& - \gamma_1 \frac{\frac{x_{min} - x_{max}}{b(x_{max}, t_{k-2})} b'(x_{max}, t_{k-2})}{b'(x_{max}, t_{k-3}) - \frac{b(x_{max}, t_{k-3})}{b(x_{max}, t_{k-2})} b'(x_{max}, t_{k-2})} \left(b(x, t_{k-3}) - \frac{b(x_{max}, t_{k-3})}{b(x_{max}, t_{k-2})} b(x, t_{k-2}) \right) + \\
& - \frac{\sum_{j=1}^{k-4} \beta_j \left(b'(x_{max}, t_j) - \frac{b(x_{max}, t_j)}{b(x_{max}, t_{k-2})} b'(x_{max}, t_{k-2}) \right)}{b'(x_{max}, t_{k-3}) - \frac{b(x_{max}, t_{k-3})}{b(x_{max}, t_{k-2})} b'(x_{max}, t_{k-2})} \left(b(x, t_{k-3}) - \frac{b(x_{max}, t_{k-3})}{b(x_{max}, t_{k-2})} b(x, t_{k-2}) \right) = \\
& \gamma_0 + \\
& \gamma_1 \Bigg(X + \frac{x_{min} - x_{max}}{b(x_{max}, t_{k-2})} b(x, t_{k-2}) \\
& - \frac{\frac{x_{min} - x_{max}}{b(x_{max}, t_{k-2})} b'(x_{max}, t_{k-2})}{b'(x_{max}, t_{k-3}) - \frac{b(x_{max}, t_{k-3})}{b(x_{max}, t_{k-2})} b'(x_{max}, t_{k-2})} \left(b(x, t_{k-3}) - \frac{b(x_{max}, t_{k-3})}{b(x_{max}, t_{k-2})} b(x, t_{k-2}) \right) \Bigg) + \\
& \sum_{j=1}^{k-4} \beta_j \Bigg(b(x, t_j) - \frac{b(x_{max}, t_j)}{b(x_{max}, t_{k-2})} b(x, t_{k-2}) + \\
& - \frac{b'(x_{max}, t_j) - \frac{b(x_{max}, t_j)}{b(x_{max}, t_{k-2})} b'(x_{max}, t_{k-2})}{b'(x_{max}, t_{k-3}) - \frac{b(x_{max}, t_{k-3})}{b(x_{max}, t_{k-2})} b'(x_{max}, t_{k-2})} \left(b(x, t_{k-3}) - \frac{b(x_{max}, t_{k-3})}{b(x_{max}, t_{k-2})} b(x, t_{k-2}) \right) \Bigg)
\end{aligned} \tag{4}
$$

This derivation is used to construct the design matrix for the functions in the R package peRiodic.