Shape-constrained splines

Applications and examples in R

Antonio Gasparrini and Zaid Chalabi

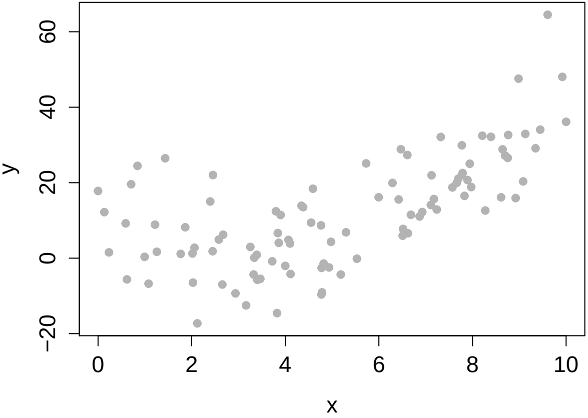
Department of Social and Environmental Health Research
London School of Hygiene and Tropical Medicine (LSHTM)

Centre for Statistical Methodology – LSHTM
30 October 2015

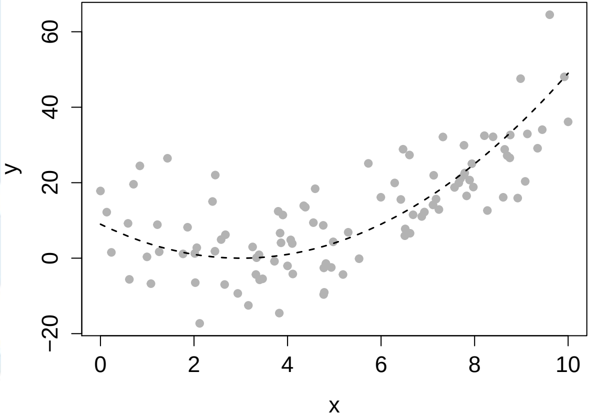
Outline

Outline

Scatterplot of x and y



True relationship



B-spline regression

The relationship between the predictor x_i and the response y_i, with $i = 1, \ldots, n$, can be defined by a function f:

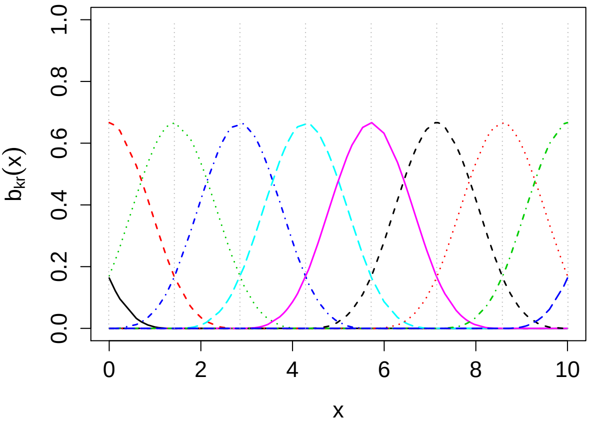
$$y_i = f(x_i)$$

The function can be approximated by $m = p + r$ **B-splines** of degree r, by setting $p + 1$ knots in the range $x_1 \leq x \leq x_n$:

$$y_i = f(x_i) \approx \sum_{k=1}^{m} \gamma_k b_{k,r}(x_i)$$

where $b_{k,r}$ is the (non-negative) k^{th} B-spline and γ_k its coefficient

Spline basis terms



Estimation

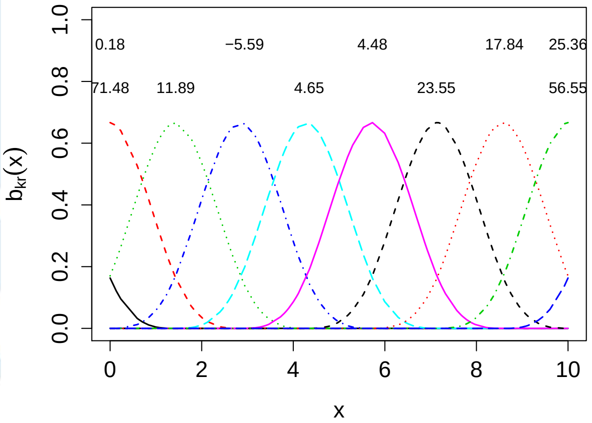
By defining:

$$\begin{aligned}\boldsymbol{\gamma} &= [\gamma_1, \ldots, \gamma_k, \ldots, \gamma_m]^{\mathsf{T}} \\ \mathbf{x}_i &= [b_{1,r}(x_i), \ldots, b_{k,r}(x_i), \ldots, b_{m,r}(x_i)]^{\mathsf{T}}\end{aligned}$$

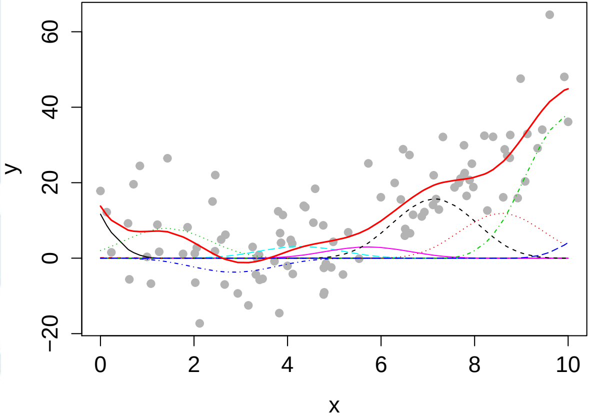
it is possible to rely on **standard estimation methods** by minimizing the least square objective:

$$\sum_{i=1}^{n} \big(y_i - f(x_i; \gamma)\big)^2 = ||\mathbf{y} - \mathbf{X}\gamma||^2$$

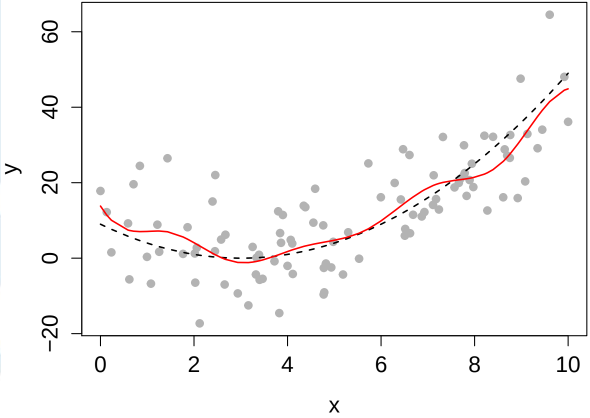
Splines and coefficients



Fitted curve



Comparison



Issues and motivation

Two issues:

1. 'Wiggly' shape: need to define **optimal smoothness** of the curve [NB: discussed in a previous CSM seminar]
2. Possibility of imposing **shape-constraints** on the curve [NB: today's topic]

The latter can be based on a priori assumptions: in many biological or epidemiological phenomena, we can assume for instance **monotonic increasing/decreasing** and/or **convex/concave** associations

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Outline

Derivatives of a spline function

For a B-spline function:

$$f(x;\gamma) = \sum_{k=1}^{m} \gamma_k b_{k,r}(x)$$

assuming equi-spaced knots at distance z, the derivatives can be computed as:

$$f'(x;\gamma) = z^{-1} \sum_{k=2}^{m} (\gamma_k - \gamma_{k-1}) b_{k,r-1}(x)$$

$$f''(x;\gamma) = z^{-2} \sum_{k=3}^{m} (\gamma_k - 2\gamma_{k-1} + \gamma_{k-2}) b_{k,r-2}(x)$$

Imposing constraints

Constraints on the first derivative

A sufficient condition for $f'(x;\gamma) > 0$ is $\gamma_k - \gamma_{k-1} > 0$ for $k = 2, \ldots, m$

Constraints on the second derivative

A sufficient condition for $f''(x;\gamma) > 0$ is $\gamma_k - 2\gamma_{k-1} + \gamma_{k-2} > 0$ for $k = 3, \ldots, m$

Constraints on **linear combinations of the coefficients** map into constraints on the **shape of the relationship**

In matrix terms

Defining the two difference matrices:

$$\mathbf{D}_1 = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix} ; \quad \mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & \cdots & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -2 & 1 \end{pmatrix} .$$

Then:

$$\begin{aligned} f'(x;\gamma) > 0 \quad &\Rightarrow \quad \mathbf{D}_1\gamma > \mathbf{0} \\ f''(x;\gamma) > 0 \quad &\Rightarrow \quad \mathbf{D}_2\gamma > \mathbf{0} \end{aligned}$$

Type of constraints

- Monotonically increasing: $f'(x_i;\gamma) > 0 \quad \Rightarrow \quad \mathbf{D}_1\gamma > \mathbf{0}$
- Monotonically decreasing: $f'(x_i;\gamma) < 0 \quad \Rightarrow \quad -\mathbf{D}_1\gamma > \mathbf{0}$
- Convex: $f''(x_i;\gamma) > 0 \quad \Rightarrow \quad \mathbf{D}_2\gamma > \mathbf{0}$
- Concave: $f''(x_i;\gamma) < 0 \quad \Rightarrow \quad -\mathbf{D}_2\gamma > \mathbf{0}$
- Monotonically increasing and convex: $f'(x_i;\gamma) > 0$ and $f''(x_i;\gamma) > 0 \quad \Rightarrow \quad [\mathbf{D}_1^\mathsf{T}\mathbf{D}_2^\mathsf{T}]^\mathsf{T}\gamma > \mathbf{0}$
- ...

Estimation can be performed through **linear constrained optimization** using the log-likelihood function

Outline

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

An extension by Pya and Wood

Pya and Wood (2010, 2015) proposed **shape constrained additive models**

This approach simultaneously addresses two issues:

1. imposing constraints on the shape through a **re-parameterization** of the model
2. defining the optimal smoothness of the curve via additive models with **penalized splines**

Re-parameterization as an unconstrained model

In the case of monotonically increasing shapes, the constraints are enforced by setting $\gamma_k - \gamma_{k-1} > 0$ for $k = 2, \ldots, m$

This can also be obtained by re-parameterizing the B-splines with **unconstrained coefficients** ϕ as:

$$\gamma_1 = \phi_1 \ , \ \gamma_k = \phi_1 + \sum_{j=2}^{k} e^{\phi_j}, k = 2, \ldots, m$$

Matrix formulation

Setting $\gamma = \mathbf{P}\tilde{\phi}$, with $\tilde{\phi} = [\phi_1, e^{\phi_2}, \ldots, e^{\phi_m}]^T$, and:

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

the regression model can written in terms of unconstrained *working* parameters as:

$$f(x;\phi) = \mathbf{XP}\tilde{\phi}$$

Fitted through **non-linear unconstrained optimization**

Shape constrained penalized splines

The **optimal smoothness** of the relationship can be found by penalizing differences between adjacent coefficients (Wood 2006)

The least square objective can be modified to:

$$||\mathbf{y} - \mathbf{X}\mathbf{P}\tilde{\phi}||^2 + \lambda\phi^{\mathsf{T}}\mathbf{S}\phi$$

where $\mathbf{S} = \mathbf{D}_2^{\mathsf{T}}\mathbf{D}_2$ is a known **penalty matrix** (with $\mathbf{D}_2$ previously defined), and λ is a **smoothing parameter**

Maximizing the log-likelihood

$$
\begin{aligned}
\mathcal{L}_p(\phi, \lambda) =& \mathcal{L}(\phi) - \frac{1}{2}\lambda\phi^{\mathsf{T}}\mathbf{S}\phi \\
J(\phi) =& \frac{\partial \mathcal{L}_p}{\partial \phi} \\
H(\phi) =& \frac{\partial^2 \mathcal{L}_p}{\partial \phi^2}
\end{aligned}
$$

Starting from an initial estimate $\phi^{(0)}$, solve iteratively using the **Newton-Raphson** method, with:

$$\phi^{(i+1)} = \phi^{(i)} - H(\phi^{(i)})^{-1} J(\phi^{(i)})$$

This method is integrated with smoothing parameter (λ) selection

Outline

Outline

Discussion

Interesting method with several **potential applications**

(Relatively) complex estimation and computational techniques

Comparison of **linear constrained** and **non-linear unconstrained** optimization

Shape constrained additive models fully implemented in the R package `scam`

Extensions

Framework already extended to **bi-dimensional risk surfaces** through a tensor product basis functions

Not easy to address the original problem that motivated the research: **shape-constrained exposure-lag-response** functions