Lecture 4: Natural Splines

MATH5824 Generalised Linear and Additive Models

Richard P Mann

MATH5824 Generalised Linear and Additive Models

Reading

Course notes: Chapter 3, Sections 3.1–3.3

`www.richardpmann.com/MATH5824`

The Interpolation Problem

Given: n observations $\{(t_i, y_i)\}_{i=1}^n$ at distinct locations $t_1 < \cdots < t_n$.

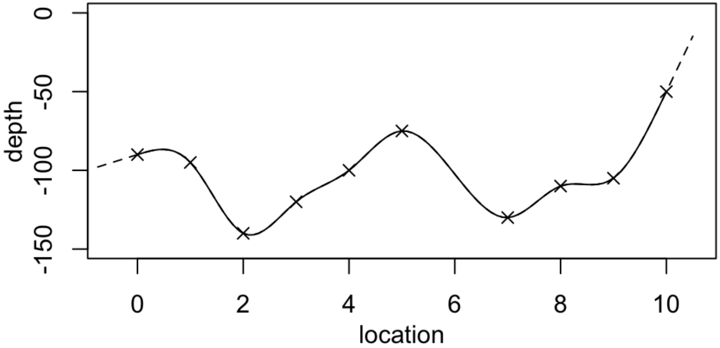
Goal: Find a smooth function f such that $f(t_i) = y_i$ for all i.

A cubic spline with knots at the data locations has $n + 4$ degrees of freedom for n interpolation conditions.

$\Rightarrow$ Infinitely many cubic splines interpolate the data.

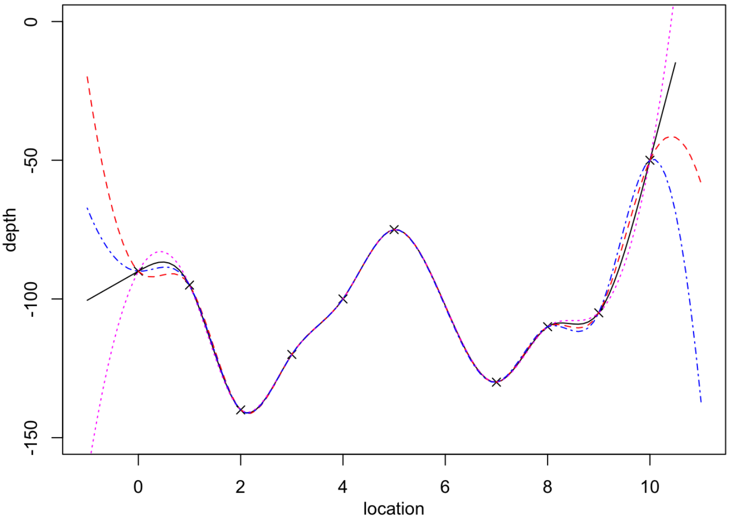
Question: How do we choose among them?

Cubic Interpolating Spline: Coal Seam Data



A cubic interpolating spline fitted to the coal seam data. The dashed line shows extrapolation beyond the data range.

Non-uniqueness of Interpolation



Multiple cubic interpolating splines can pass through the same data. The solid black

Natural Splines: The Key Idea

Definition

A spline is **natural** if, beyond the boundary knots t_1 and t_n, its higher-order derivatives are zero:

$$f^{(j)}(t) = 0 \quad \text{for } j = \frac{p+1}{2}, \ldots, p$$

whenever $t \leq t_1$ or $t \geq t_n$.

Effect: The spline is forced to be a low-order polynomial in the boundary regions.

- **Natural linear spline** $(p = 1)$: constant outside $[t_1, t_n]$
- **Natural cubic spline** $(p = 3)$: linear outside $[t_1, t_n]$

Natural Cubic Spline: Boundary Conditions

For a natural **cubic** spline, the boundary conditions are:

$$f''(t_1) = f''(t_n) = 0 \quad \text{and} \quad f'''(t_1) = f'''(t_n) = 0$$

More precisely:

$$\lim_{\varepsilon \to 0} f^{(2)}(t_1 + \varepsilon) = \lim_{\varepsilon \to 0} f^{(2)}(t_n - \varepsilon) = 0$$

$$\lim_{\varepsilon \to 0} f^{(3)}(t_1 + \varepsilon) = \lim_{\varepsilon \to 0} f^{(3)}(t_n - \varepsilon) = 0$$

These impose $p + 1 = 4$ additional constraints.

Degrees of Freedom for Natural Splines

Starting from: $\mathrm{df}_{\text{spline}} = m + p + 1$

Natural boundary conditions: $p + 1$ additional constraints

Therefore:

$$\mathrm{df}_{\text{natural spline}} = (m + p + 1) - (p + 1) = m$$

The number of degrees of freedom equals the number of knots m, **regardless of the spline order** p.

For interpolation: With knots at data locations ($m = n$), we have exactly n degrees of freedom matching n interpolation conditions.

$\Rightarrow$ The natural cubic interpolating spline is **unique**.

Alternative Representations

Proposition (3.1)

Natural splines have convenient closed-form representations.

Linear natural spline ($p = 1$):

$$f(t) = a_0 + \sum_{i=1}^{m} b_i |t - t_i|, \qquad \sum_{i=1}^{m} b_i = 0$$

Cubic natural spline ($p = 3$):

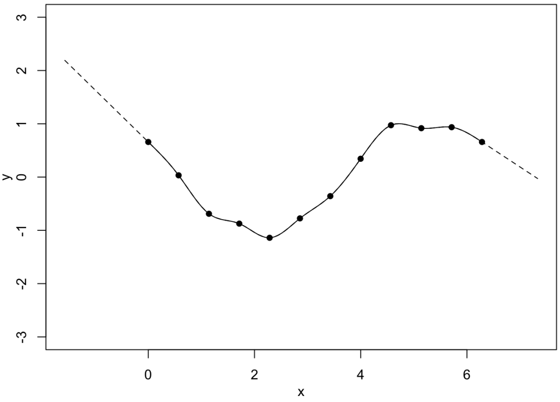
$$f(t) = a_0 + a_1 t + \sum_{i=1}^{m} b_i |t - t_i|^3, \qquad \sum_{i=1}^{m} b_i = 0, \quad \sum_{i=1}^{m} b_i t_i = 0$$

These representations build in the natural boundary conditions automatically via the side constraints on $\{b_i\}$.

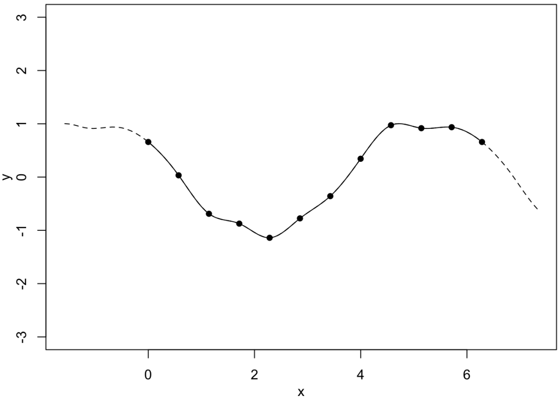
Periodic Splines

For periodic data (e.g., seasonal patterns), an alternative boundary condition:

$$f^{(\ell)}(t_1) = f^{(\ell)}(t_m) \quad \text{for } \ell = 0, 1, \ldots, p-1$$



Natural spline



Periodic spline

Summary

Key points:

- Without boundary conditions, infinitely many splines interpolate the data
- Natural splines impose linearity (cubic) or constancy (linear) beyond the boundary knots
- A natural spline with m knots has exactly m degrees of freedom
- Natural cubic interpolating splines are unique when knots are at data locations
- Alternative representations express natural splines using $|t - t_i|^p$ basis functions
- Periodic splines are an alternative for cyclic data

Next lecture: Roughness penalties and fitting interpolating splines in R.