STAT 824 sp 2023 Lec 05 slides

Nonparametric regression: Least-squares splines

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Table of Contents

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be indep. realizations of $(X, Y) \in [0, 1] \times \mathbb{R}$, where

$$Y = m(X) + \varepsilon, \quad \text{for some } m : [0, 1] \to \mathbb{R},$$

where ε is independent of X with $\mathbb{E}\varepsilon = 0$ and $\mathbb{E}\varepsilon^2 = \sigma^2$.

<u>Idea</u>: Given a set of basis functions $b_1, \dots, b_{d_n} : [0, 1] \to \mathbb{R}$, suppose

$$m(x) \approx \sum_{k=1}^{d_n} \alpha_k b_k(x) \quad \text{for some} \quad \alpha_1, \dots, \alpha_{d_n}.$$

Then estimate $\alpha_1, \dots, \alpha_{d_n}$ with least squares to get $\hat{m}_n(x) = \sum_{k=1}^{d_n} \hat{\alpha}_k b_k(x)$.

- As $n \to \infty$, let $d_n \to \infty$ so that the approximation improves.
- Quality of approximation depends on
  1. the type and number of basis functions.
  2. the smoothness of the true function m.
- There will always be some approximation bias.

A non-parametric least squares estimator

For a set of basis functions $b_1, \dots, b_{d_n} : [0,1] \to \mathbb{R}$, let

$$\mathcal{B}_n = \{m : m(x) = \sum_{k=1}^{d_n} \alpha_k b_k(x), \alpha_1, \dots, \alpha_{d_n} \in \mathbb{R}\}.$$

Given $(X_1, Y_1), \dots, (X_n, Y_n)$, the least squares estimator of m in $\mathcal{B}_n$ is given by

$$\hat{m}_n = \underset{g \in \mathcal{B}_n}{\operatorname{argmin}} \sum_{i=1}^{n} [Y_i - g(X_i)]^2.$$

Exercise: Let $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ and define the matrix

$$\mathbf{B} = (b_k(X_i))_{1 \le i \le n, 1 \le k \le d_n}.$$

Show that $\hat{m}_n(x) = \mathbf{b}_x^T \hat{\alpha}$, where

$$\hat{\alpha} = \left(\mathbf{B}^T \mathbf{B}\right)^{-1} \mathbf{B}^T \mathbf{Y} \quad \text{and} \quad \mathbf{b}_x = (b_1(x), \dots, b_{d_n}(x))^T.$$

Splines: see Stone (1985) [2]

For a positive integer K_n let

$$I_{nk} = \left[\frac{k-1}{K_n}, \frac{k}{K_n}\right), \quad k = 1, \dots, K_n - 1, \quad \text{and} \quad I_{nK_n} = \left[\frac{K_n-1}{K_n}, 1\right].$$

For $r \geq 1$, define the set of functions

$$\mathcal{M}_{n,r} = \left\{ m : [0,1] \to \mathbb{R} : \begin{array}{l} m \text{ is a polynomial of degree } r \text{ or less on} \\ \text{each interval } I_1, \dots, I_{nK_n}, \text{ and } m \text{ is } r-1 \text{ times} \\ \text{continuously differentiable on } [0,1] \end{array} \right\}.$$

Moreover, let

$$\mathcal{M}_{n,0} = \{m : [0,1] \to \mathbb{R} : m \text{ is piecewise constant on } I_1, \dots, I_{nK_n}\}$$

- Fns in $\mathcal{M}_{n,1}$, $\mathcal{M}_{n,2}$, and $\mathcal{M}_{n,3}$ are called *linear, quadratic, and cubic splines*.
- Values j/K_n, $j = 0, \dots, K_n$ are called *knots*. Can choose knots differently.
- Functions in these spaces can nicely approximate functions in Hölder classes.

For a function $g : \mathcal{T} \to \mathbb{R}$, we write $\|g\|_\infty = \sup_{x \in \mathcal{T}} |g(x)|$.

Key result from deBoor (1968) [1]

For each $m \in \mathcal{H}(\beta, L)$ on $[0, 1]$, there exists a function $m^{\text{spl}}_{n,r} \in \mathcal{M}_{n,r}$, where $r \geq \beta - 1$ such that

$$\|m - m^{\text{spl}}_{n,r}\|_\infty \leq C \cdot K_n^{-\beta}$$

for some constant $C > 0$.

Idea is to let $K_n \to \infty$ as $n \to \infty$, so that this approximation error goes to zero.

Exercise: For $m \in \text{Lipschitz}(L)$ on $[0, 1]$, show that $\exists\ m^{\text{spl}}_{n,0} \in \mathcal{M}_{n,0}$ such that

$$\sup_{x \in [0,1]} |m(x) - m^{\text{spl}}_{n,0}(x)| \leq \frac{L}{K_n}.$$

We now define the order r least-squares splines estimator of m as

$$\hat{m}_{n,r}^{\text{spl}} = \underset{g \in \mathcal{M}_{n,r}}{\operatorname{argmin}} \sum_{i=1}^{n} [Y_i - g(X_i)]^2.$$

Bound on MSE $\hat{m}_{n,r}^{\text{spl}}(x_0)$

If $m \in \mathcal{H}(\beta, L)$ on $[0,1]$, then for $r \geq \beta - 1$, we have

$$\text{MSE}\, \hat{m}_{n,r}^{\text{spl}}(x_0) \leq C \cdot \left(K_n^{-2\beta} + \frac{K_n}{n} \right)$$

for all $x_0 \in [0,1]$ for large enough n, provided (C1), (C2), and (C3) hold.

We will study the conditions (C1), (C2), and (C3) later on.

Exercise:

1. Find the value of K_n which minimizes the MSE bound.
2. Give the minimum bound over choices of K_n.
3. Anything interesting about this?

For our spline spaces $\mathcal{M}_{n,r}$, we need sets of basis functions $b_1, \dots, b_{d_n}$ such that

$$\mathcal{M}_{n,r} = \left\{ m : [0,1] \to \mathbb{R} : m = \textstyle\sum_{k=1}^{d_n} \alpha_k b_k, \alpha_1, \dots, \alpha_{d_n} \in \mathbb{R} \right\}.$$

B-splines: Cox-deBoor recursion formula

For a non-decreasing set of knots $0 = u_0 \leq u_1 \leq \cdots \leq u_K = 1$, let

$$N_{k,0}(u) = \begin{cases} 1, & u_k \leq u < u_{k+1} \\ 0, & \text{otherwise} \end{cases} \qquad \text{for } k = 0, \dots, K-1,$$

and

$$N_{k,r}(u) = \frac{u - u_k}{u_{k+r} - u_k} N_{k,r-1}(u) + \frac{u_{k+r+1} - u}{u_{k+r+1} - u_{k+1}} N_{k+1,r-1}(u)$$

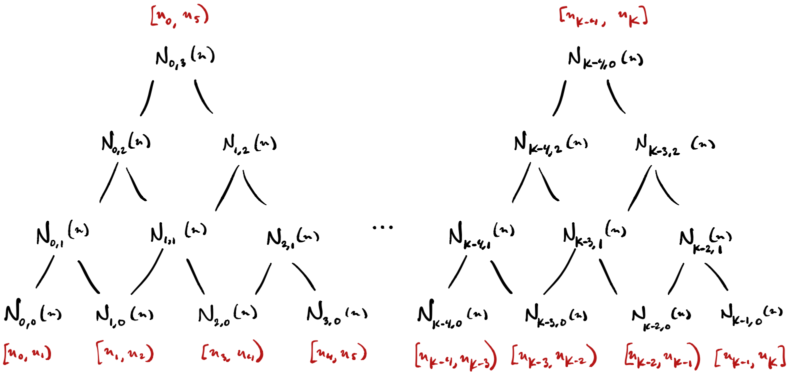
for $k = 0, \dots, K - r - 1$. These functions are called *B-splines*.

Can compute row vector $\mathbf{N}_r(x) = (N_{0,r}(x), \dots, N_{K-r-1,r}(x))$, $x \in [0,1]$, with

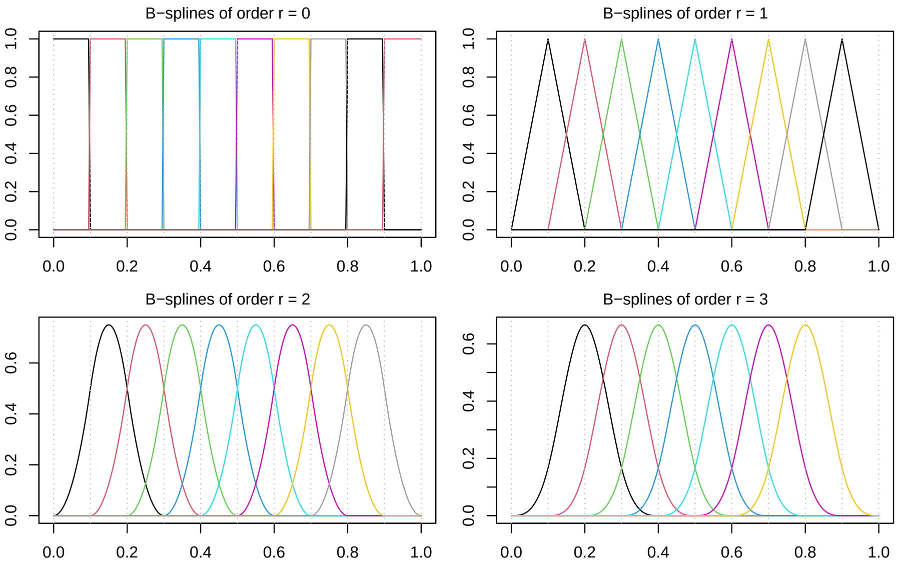
```
splineDesign(knots = u, x = x, ord = r + 1)
```

Require `splines` package.

The Cox-deBoor recursion has a structure like this:



Exercise: Show construction of $N_{0,1}$ based on knots $(u_0, u_1, u_2) = (0, 1/2, 1)$.


B−splines of order r = 0
B−splines of order r = 1
B−splines of order r = 2
B−splines of order r = 3
0.0
0.2
0.4
0.6
0.8
1.0

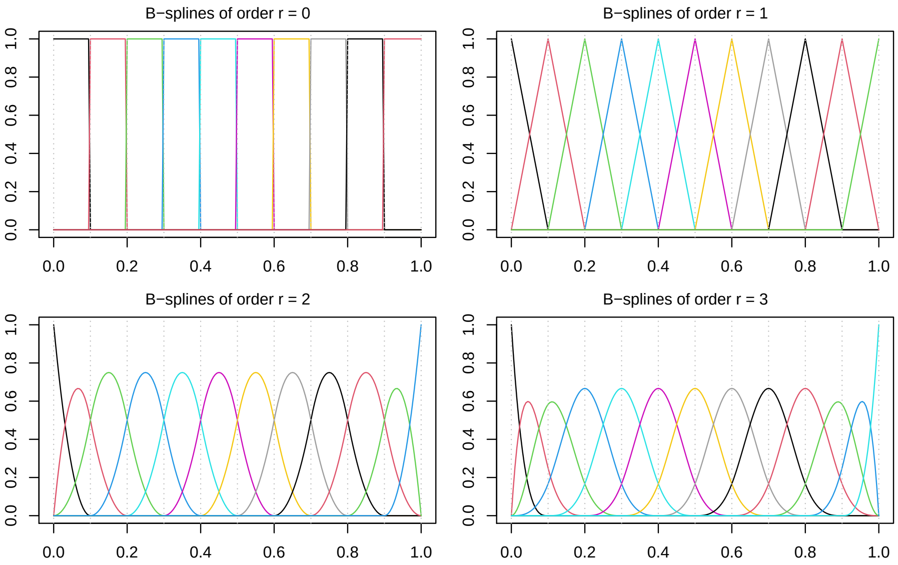
To handle boundary issues, a convention is to include the end knots $r+1$ times:

$$0 = u_{-r} = \cdots = u_0 < u_1 < \cdots < u_K = \cdots = u_{K+r}$$

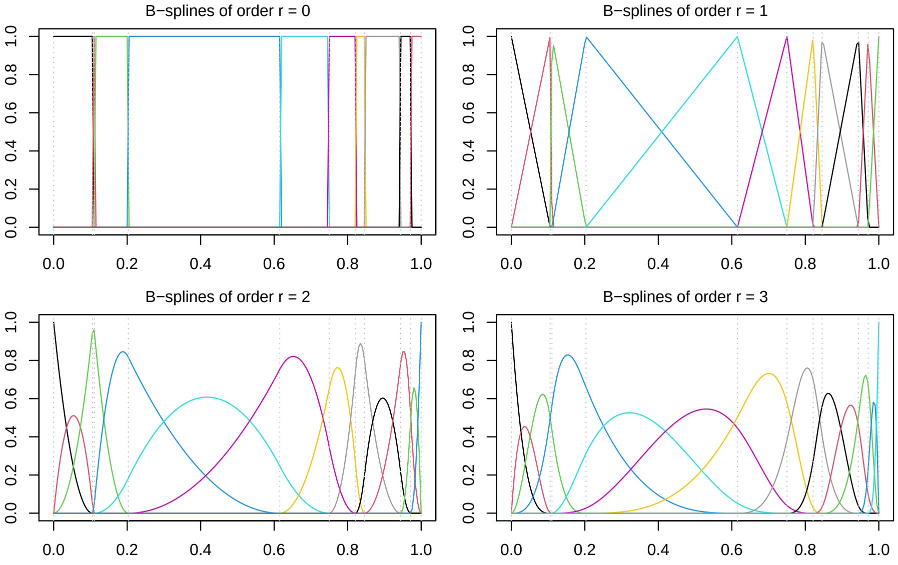
This results in $K+r$ basis functions when $[0,1]$ is subdivided into K intervals.

Exercise: Make beautiful plots of B-spline functions of order $r = 0, 1, 2, 3$ in R

1. with equally spaced knots.
2. with unequally spaced knots.



B−splines of order r = 0
B−splines of order r = 1
B−splines of order r = 2
B−splines of order r = 3
0.0
0.2
0.4
0.6
0.8
1.0


B−splines of order r = 0
B−splines of order r = 1
B−splines of order r = 2
B−splines of order r = 3
0.0
0.2
0.4
0.6
0.8
1.0

Replicating boundary knots r times results in $d_n = K_n + r$ basis functions.

For $X_1, \dots, X_n$, we can obtain the $n \times d_n$ design matrix $\mathbf{B}$ with

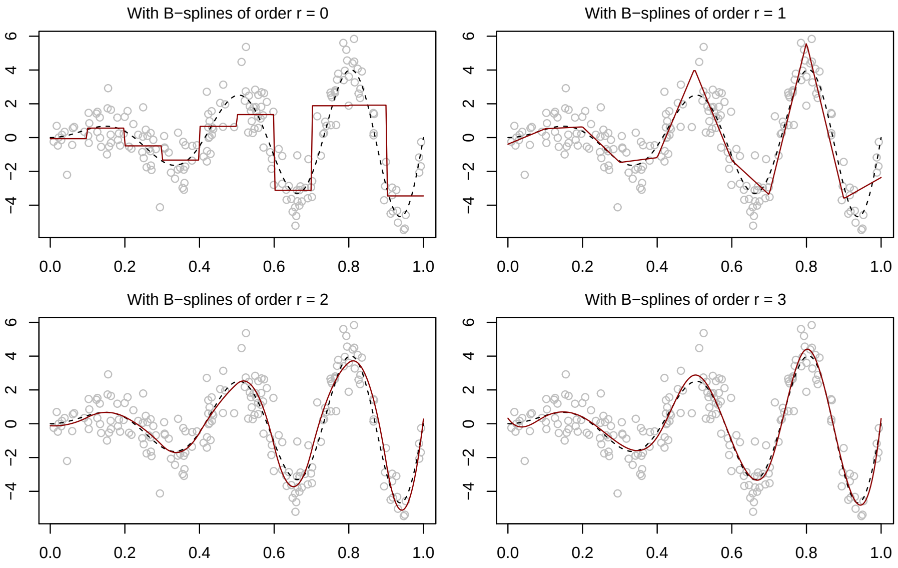
```
splineDesign(knots = u, x = X, ord = r + 1),
```

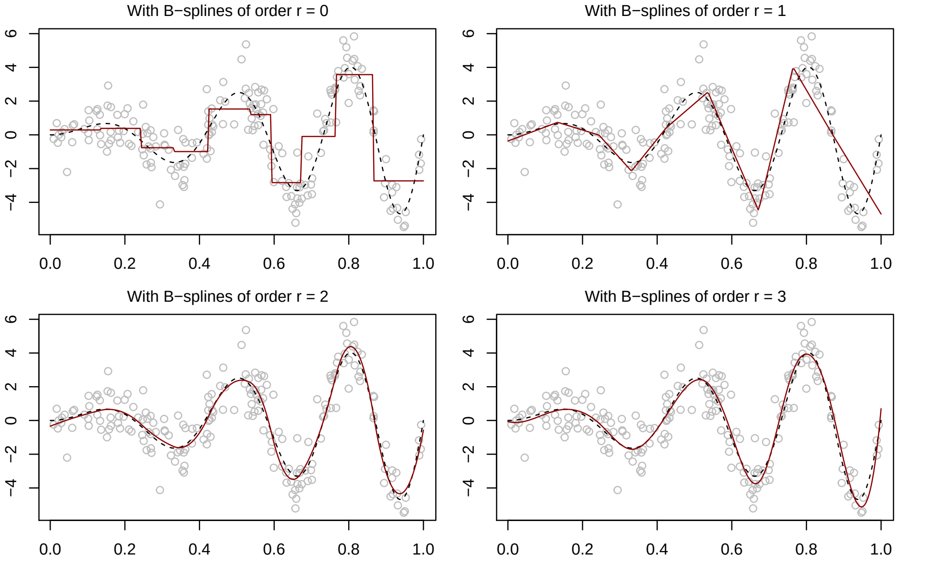
where `X` is a vector containing the values $X_1, \dots, X_n$.

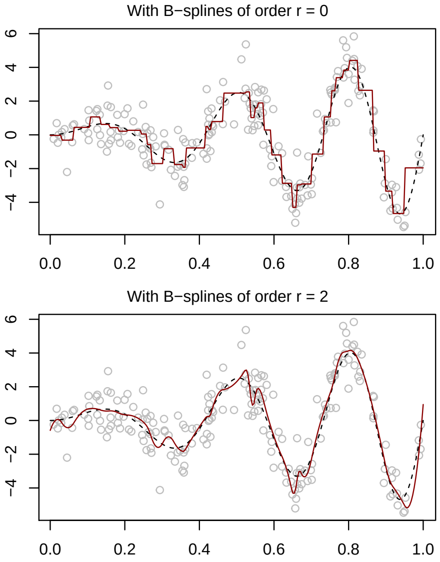
Note that (with the replicated boundary knots) the rows of $\mathbf{B}$ always sum to 1.

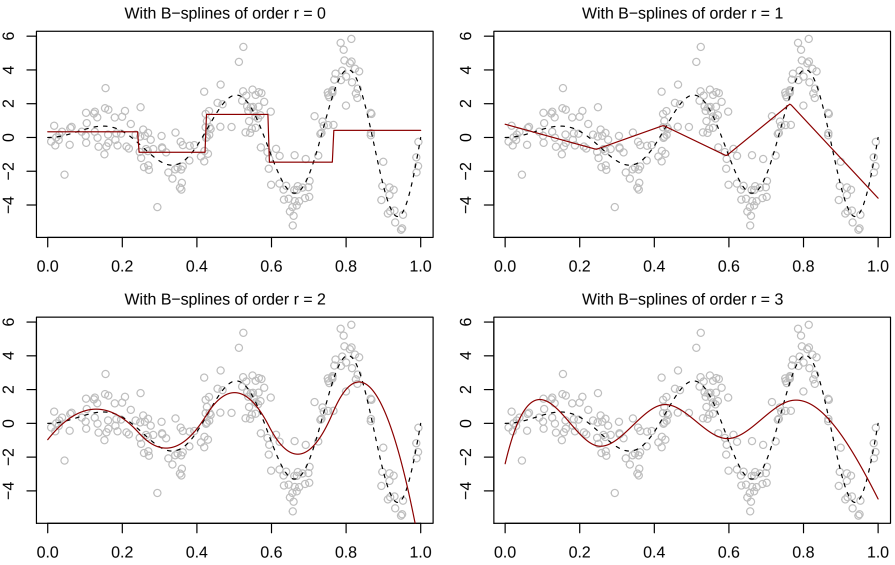
Exercise:

1. For $n = 200$, generate data $Y_i = m(X_i) + \varepsilon_i$ with
   - $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Uniform}(0, 1)$, indep. of $\varepsilon_1, \dots, \varepsilon_n \overset{\text{ind}}{\sim} \text{Normal}(0, 1)$
   - $m(x) = 5x \cdot \sin(2\pi(1 + x)^2)$
2. Plot $\hat{m}_{n,r}^{\text{spl}}$ under $K_n = 10$ for $r = 0, 1, 2, 3$ with
   - knots equally spaced in $[0, 1]$
   - knots at equally space quantiles of $X_1, \dots, X_n$
3. Try different values of K_n.


With B−splines of order r = 0
With B−splines of order r = 1
With B−splines of order r = 2
With B−splines of order r = 3
−4
−2
0
2
4
6
0.0
0.2
0.4
0.6
0.8
1.0


With B−splines of order r = 0
With B−splines of order r = 1
With B−splines of order r = 2
With B−splines of order r = 3
−4
−2
0
2
4
6
0.0
0.2
0.4
0.6
0.8
1.0


With B−splines of order r = 0
With B−splines of order r = 1
With B−splines of order r = 2
With B−splines of order r = 3
−4 −2 0 2 4 6
0.0 0.2 0.4 0.6 0.8 1.0


With B−splines of order r = 0
With B−splines of order r = 1
With B−splines of order r = 2
With B−splines of order r = 3
−4
−2
0
2
4
6
0.0
0.2
0.4
0.6
0.8
1.0

Conditions for bounding MSE $\hat{m}^{\text{spl}}_{n,r}(x_0)$; see Zhou (1998) [3]

Let $m \in \mathcal{H}(\beta, L)$ on $[0,1]$ and let $m^{\text{spl}}_{n,r} \in \mathcal{M}_{n,r}$ satisfy $\|m - m^{\text{spl}}_{n,r}\|_\infty \leq C \cdot K_n^{-\beta}$.

Let $X_1, \dots, X_n \in [0,1]$ be deterministic such that for large enough n,

(C1) $K_n^{-1} \cdot c_1 \leq \lambda_{\min}\left(n^{-1}\mathbf{B}^T\mathbf{B}\right) \leq \lambda_{\max}\left(n^{-1}\mathbf{B}^T\mathbf{B}\right) \leq C_1 \cdot K_n^{-1}$

(C2) $\left\| \left(n^{-1}\mathbf{B}^T\mathbf{B}\right)^{-1} \right\|_\infty \leq C_2 \cdot K_n$

(C3) $\left\|n^{-1}\mathbf{B}^T(\mathbf{m} - \mathbf{m}^{\text{spl}}_{n,r})\right\|_\infty \leq C_3 \cdot K_n^{-1-\beta}$,

where

$$\mathbf{m} = (m(X_1), \dots, m(X_n))^T \quad \text{and} \quad \mathbf{m}^{\text{spl}}_{n,r} = (m^{\text{spl}}_{n,r}(X_1), \dots, m^{\text{spl}}_{n,r}(X_n))^T.$$

Exercise:

1. Use above to get bounds on the bias and variance of $\hat{m}^{\text{spl}}_{n,r}(x_0)$.
2. Consider (C1), (C2), and (C3) in the case of $\beta = 1$, $r = 0$.

Carl De Boor.
On uniform approximation by splines.
J. Approx. Theory, 1(1):219–235, 1968.

Charles J Stone.
Additive regression and other nonparametric models.
The Annals of Statistics, pages 689–705, 1985.

S Zhou, X Shen, DA Wolfe, et al.
Local asymptotics for regression splines and confidence regions.
The Annals of Statistics, 26(5):1760–1782, 1998.