NONPARAMETRIC REGRESSION WITH SPLINES

Let

$$Y_i = m(X_i) + \varepsilon_i \quad , \quad i = 1, \dots, n,$$

$\varepsilon_1, \dots, \varepsilon_n$ iid with $\mathbb{E}\,\varepsilon_1 = 0$, $\mathbb{E}\,\varepsilon_1^2 = \sigma^2$, independent of $X_1, \dots, X_n \in [0,1]$.

Idea: let $b_1, \dots, b_d : [0,1] \to \mathbb{R}$ be a set of $\underbrace{\text{basis functions}}$ and suppose

building block functions that we can add together to make more complicated functions

$$m(x) \approx \sum_{k=1}^{d} \alpha_k b_k(x).$$

Then we build the a least-squares estimator

$$\hat{m}_n(x) = \sum_{k=1}^{d} \hat{\alpha}_k b_k(x),$$

where

$$(\hat{\alpha}_1, \dots, \hat{\alpha}_d) = \underset{\alpha_1, \dots, \alpha_d}{\operatorname{argmin}} \sum_{i=1}^{n} \left(Y_i - \sum_{k=1}^{d} \alpha_k b_k(X_i) \right)^2.$$

Exercise: Show that by defining the matrix

$$B = \left(b_k(X_i) \right)_{1 \le k \le d,\, 1 \le i \le n},$$

we may write

(i) $\quad \hat{\underset{\sim}{\alpha}} = (\hat{\alpha}_1, \dots, \hat{\alpha}_d)^T = (B^T B)^{-1} B^T Y$

(ii) $\quad \hat{m}_n(x) = \underset{\sim}{b}_x^T \hat{\underset{\sim}{\alpha}}$, with $\underset{\sim}{b}_x = (b_1(x), \dots, b_d(x))^T$.

Concepts: (i) As $n \to \infty$, the number of basis functions $d \to \infty$, so the approximation

$$m(x) \approx \sum_{k=1}^{d} \alpha_k b_k(x)$$

improves.

(ii) The quality of this approximation depends on

- the choice of basis functions
- the smoothness of the true function m.

(iii) The approximation is never perfect, so we will always have some approximation bias.

Definition of splines: From Stone (1985):

Let K_n be a positive integer and let

$$I_{nk} = \left[\frac{k-1}{K_n}, \frac{k}{K_n}\right), \quad k = 1, \dots, K_n - 1$$

$$I_{nK_n} = \left[\frac{K_n - 1}{K_n}\ 1\right].$$

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

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

The functions in $\mathcal{M}_{n,1}$, $\mathcal{M}_{n,2}$, and $\mathcal{M}_{n,3}$ are called linear, quadratic, and cubic splines, respectively.

Define also

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

Result of DeBoor (1968):

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

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

Remark: K_n is the # intervals into which we subdivide $[0,1]$.

Example: For $m \in \mathcal{H}(1, L)$, i.e. $m \in \text{Lipschitz}(L)$, $\exists\, m_{n,0}^{spl} \in \underbrace{\mathcal{M}_{n,0}}_{\text{piecewise constant}}$ s.t.

$$\| m - m_{n,0}^{spl} \|_\infty \leq C \cdot K_n^{-1}.$$

We can see directly that for $m \in \text{Lipschitz}(L)$,

$$\| m - m_{n,0}^{spl} \|_\infty = \max_{1 \leq k \leq K_n} \sup_{x \in I_{nk}} \left| m(x) - m_{n,0}^{spl}(x) \right| \leq L \cdot \frac{1}{K_n}$$

if $m_{n,0}^{spl}$ a piecewise constant such that, e.g. $m_{n,0}^{spl}(x) = m\left(\frac{k-1}{K_n}\right)$ for $x \in I_{nk}$, $k = 1, \dots, K_n$.

B-spline basis for $\mathcal{M}_{n,r}$:

Let $0 = u_0 < u_1 < \dots < u_K = 1$ be knots.

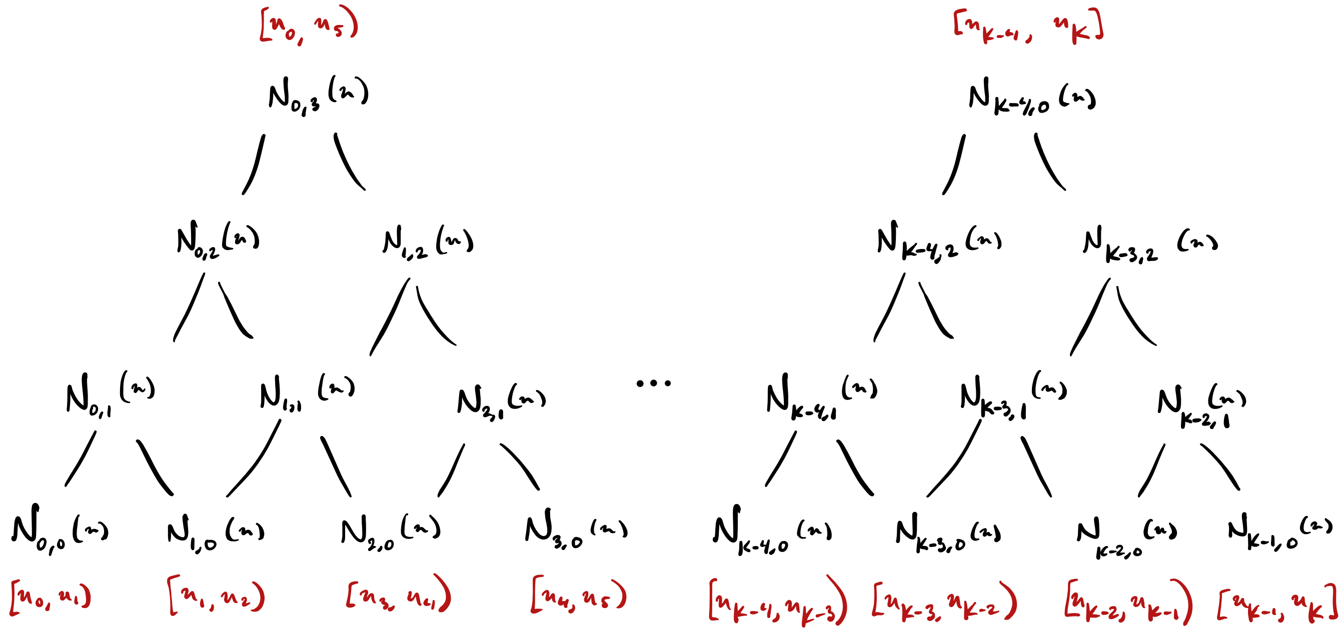
Define the $r=0$ order B-splines based on these knots as

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

Higher-order B-splines are constructed recursively (Cox-deBoor recursion formula):

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

The recursion has a structure like this:



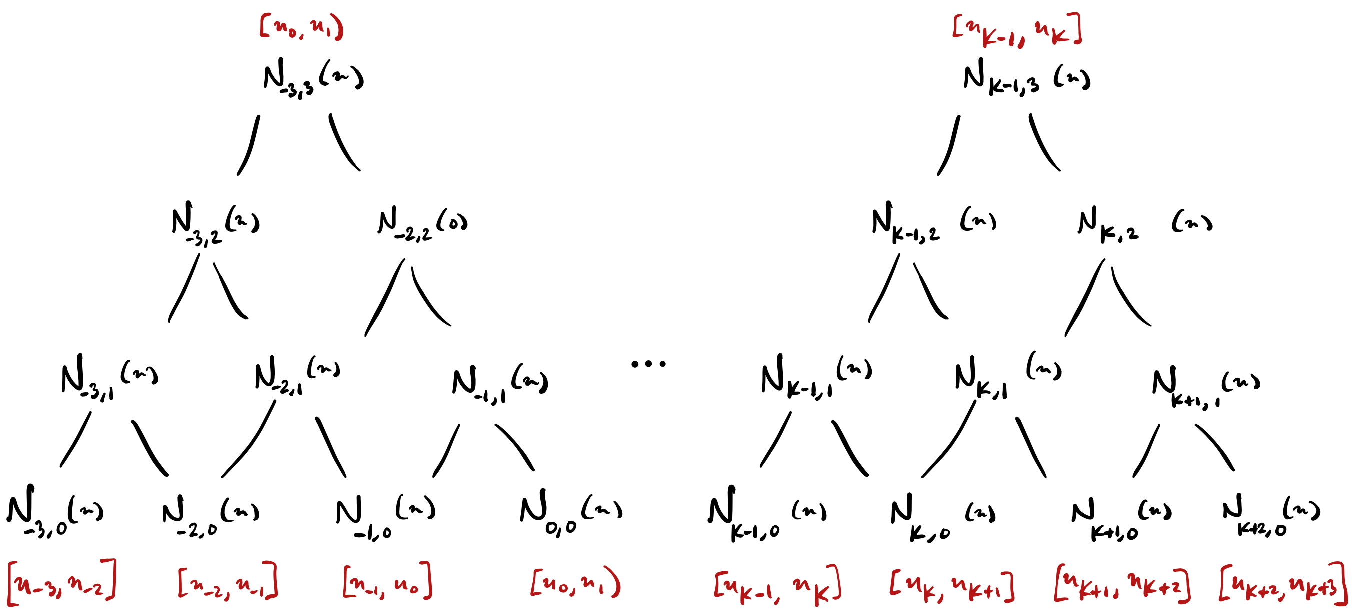
Typically the B-spline basis for splines of order r is constructed by including the end knots $r+1$ times:

For $r=0$, use knots $0 = u_0 < u_1 < \cdots < u_K = 1$

For $r=1$, use knots $0 = u_{-1} = u_0 < u_1 < \cdots < u_K = u_{K+1} = 1$

For $r \geq 1$ use knots $0 = u_{-r} = \cdots = u_0 < u_1 < \cdots < u_K = \cdots = u_{K+r} = 1$

For $r=3$, the recursion formula under this choice of knots looks like:



Replicating boundary knots in this way, we end up with

$$d_n = K_n + r$$

basis functions when we subdivide $[0,1]$ into K_n intervals and use splines of order r.

__Exercise:__ Make pictures of basis functions using spline.des from the splines package in R.

__Remark:__ Cubic B-splines $(r=3)$ are the most often used in practice.

Equipped with some number d_n of B-spline basis functions, which we now denote by $b_1, \ldots, b_{d_n}$, we have

$$\mathcal{M}_{n,r} = \left\{ m : [0,1] \to \mathbb{R} : m(x) = \sum_{k=1}^{d_n} \alpha_k b_k(x), \quad \alpha_1, \ldots, \alpha_d \in \mathbb{R} \right\}.$$

That is, the functions $b_1, \ldots, b_{d_n}$ form a basis for $\mathcal{M}_{n,r}$.

We now introduce the least-squares spline estimator

$$\begin{aligned} \hat{m}^{spl}_{n,r} &= \underset{g \in \mathcal{M}_{n,r}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(Y_i - g(X_i) \right)^2, \\ &= \sum_{k=1}^{d_n} \hat{\alpha}_k b_k, \end{aligned}$$

where

$$(\hat{\alpha}_1, \ldots, \hat{\alpha}_{d_n})^T = \underset{\underset{\sim}{\alpha} \in \mathbb{R}^{d_n}}{\operatorname{argmin}} \| \underset{\sim}{Y} - B \underset{\sim}{\alpha} \|_2^2, \qquad B = \left(b_k(X_i) \right)_{1 \le k \le d_n, 1 \le i \le n}.$$

__Exercise:__ Generate some data and plot $\hat{m}^{spl}_{n,r}$ based on splines of order $r = 0, 1, 2, 3$.

MSE of $\hat{m}^{spl}_{n,r}(x_0)$:

Now consider the bias and variance of $\hat{m}^{spl}_{n,r}(x_0)$ for some $x_0 \in [0,1]$.

Conditions: Let $m \in \mathcal{H}(\beta, L)$ and let $m^{spl}_{n,r} \in \mathcal{M}_{n,r}$ satisfy $\|m - m^{spl}_{n,r}\|_\infty \leq C \cdot K_n^{-\beta}$.
Let $X_1, \ldots, X_n \in [0,1]$ be deterministic such that for some $n_0 > 0$,

$$(C.1) \quad K_n^{-1} c_1 \leq \lambda_{min}\left(\tfrac{1}{n} B^T B\right) \leq \lambda_{max}\left(\tfrac{1}{n} B^T B\right) \leq C_1 \cdot K_n^{-1}$$

$$(C.2) \quad \left\| \left(\tfrac{1}{n} B^T B\right)^{-1} \right\|_\infty \leq C_2 \cdot K_n$$

$$(C.3) \quad \left\| \tfrac{1}{n} B^T \left(\underset{\sim}{m} - \underset{\sim}{m}^{spl}_{n,r} \right) \right\|_\infty \leq C_3 \cdot K_n^{-1-\beta},$$

where

$$\underset{\sim}{m} = (m(X_1), \ldots, m(X_n))^T \quad \text{and} \quad \underset{\sim}{m}^{spl}_{n,r} = (m^{spl}_{n,r}(X_1), \ldots, m^{spl}_{n,r}(X_n))^T$$

for all $n \geq n_0$, where K_n is the number of subintervals $[0,1]$ is divided into, and C, c_1, C_1, C_2, and C_3 are positive constants.

We will need the following:

- For a $n \times m$ matrix A, $\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^m |A_{ij}|$ is the max abs. row sum.

- From the construction of the B-spline basis for $\mathcal{M}_{n,r}$, we have

$$\|\underset{\sim}{b}_x\|_1 = 1 \quad \text{as well as} \quad \|\underset{\sim}{b}_x\|_2 \leq 1 \quad \text{for all } x \in [0,1].$$

- Moreover, since $m^{spl}_{n,r}(x_0) = b_{x_0}^T \underset{\sim}{\alpha}$ for some $\underset{\sim}{\alpha} \in \mathbb{R}^d$, $\underset{\sim}{m}^{spl}_{n,r} = B \underset{\sim}{\alpha}$, so that

$$\underset{\sim}{b}_{x_0}^T (B^T B)^{-1} B^T \underset{\sim}{m}^{spl}_n = \underset{\sim}{b}_{x_0}^T (B^T B)^{-1} B^T B \underset{\sim}{\alpha} = \underset{\sim}{b}_{x_0}^T \underset{\sim}{\alpha} = m^{spl}_n(x_0)$$

Result: Bound for MSE $\hat{m}_n^{spl}(x_0)$

If $m \in \mathcal{H}(\beta, L)$ on $[0,1]$, then under (C.1), (C.2), and (C.3), for $r \geq \beta - 1$, we have

$$\text{MSE}\, \hat{m}_{n,r}^{spl}(x_0) \leq C \cdot \left(K_n^{-2\beta} + \frac{K_n}{n} \right) \quad \text{for some } C > 0$$

for all $n \geq n_0$.

Exercise: (i) Find the optimal K_n. (ii) Plug in to get optimal MSE.

Solution: Ignoring the constant in front, we have

$$\text{(i)} \quad \frac{\partial}{\partial K_n}\left(K_n^{-2\beta} + \frac{K_n}{n} \right) = (2\beta) K_n^{-2\beta-1} + \frac{1}{n} = 0$$

$$\Leftrightarrow \qquad K_n = n^{-\frac{1}{2\beta+1}} \left(\frac{1}{2\beta}\right)^{\frac{1}{2\beta+1}}$$

So choose $K_n = c \cdot n^{-\frac{1}{2\beta+1}}$ for some constant $c > 0$.

(ii) For $K_n = c \cdot n^{-\frac{1}{2\beta+1}}$, we have

$$C \cdot \left(K_n^{-2\beta} + \frac{K_n}{n} \right) = C \cdot \left(c \cdot n^{-\frac{1}{2\beta+1}} \right)^{2\beta} + \frac{c \cdot n^{-\frac{1}{2\beta+1}}}{n}$$

$$= C \cdot \left(c^{2\beta}\, n^{-\frac{2\beta}{2\beta+1}} + c \cdot n^{-\frac{2\beta}{2\beta+1}} \right)$$

$$= \tilde{C} \cdot n^{-\frac{2\beta}{2\beta+1}}$$

for some constant $\tilde{C} > 0$.

Proof of bound on MSE $\hat{m}^{spl}_{n,r}(x_0)$:

Variance: We have

$$\mathrm{Var}\ \hat{m}^{spl}_{n,r}(x_0) = \mathrm{Var}\left[\underset{\sim}{b}_{x_0}^{T}(B^TB)^{-1}B^TY\right]$$

$$= \underset{\sim}{b}_{x_0}^{T}(B^TB)^{-1}b_{x_0}\ \sigma^2$$

$$= \frac{1}{n}\underset{\sim}{b}_{x_0}^{T}\left(\frac{1}{n}B^TB\right)^{-1}b_{x_0}\cdot\sigma^2$$

Fact: For A $d\times d$, sym, pos. def.,

$$\sup_{x\in\mathbb{R}^d}\frac{xAx}{\|x\|_2^2} = \lambda_{max}(A)$$

$$\le \frac{K_n}{n}\underbrace{\|\underset{\sim}{b}_{x_0}\|_2^2}_{\le 1}\ \lambda_{max}\left(\left(\frac{1}{n}B^TB\right)^{-1}\right)\cdot\sigma^2$$

$$\le \frac{K_n}{n}\frac{\sigma^2}{c_1}$$

Bias: We have, for some $m^{spl}_{n,r}\in\mathcal{M}_{n,r}$

$$\mathbb{E}\,\hat{m}^{spl}_{n,r}(x_0) - m(x_0) = \underset{\sim}{b}_{x_0}^{T}(B^TB)^{-1}B^T\underset{\sim}{m} - m(x_0)$$

$\underset{\sim}{b}_{x_0}^{T}(B^TB)^{-1}B^T\underset{\sim}{m}^{spl}_{n,r} = m^{spl}_{n,r}(x_0)$,

$$= \underset{\sim}{b}_{x_0}^{T}(B^TB)^{-1}B^T\underset{\sim}{m} - m^{spl}_{n,r}(x_0) + m^{spl}_{n,r}(x_0) - m(x_0)$$

since $\underset{\sim}{m}^{spl}_{n,r} = B\underset{\sim}{\alpha}$ for some $\underset{\sim}{\alpha}\in\mathbb{R}^d$,
so
$\underset{\sim}{b}_{x_0}^{T}(B^TB)^{-1}B^TB\underset{\sim}{\alpha} = \underset{\sim}{b}_{x_0}^{T}\underset{\sim}{\alpha} = m^{spl}_{n,r}(x_0)$

$$= \underset{\sim}{b}_{x_0}^{T}(B^TB)^{-1}B^T\left(\underset{\sim}{m} - \underset{\sim}{m}^{spl}_{n,r}\right) + m^{spl}_{n,r}(x_0) - m(x_0)$$

So the absolute value of the bias is bounded by

$$\left|\mathbb{E}\,\hat{m}^{spl}_{n,r}(x_0) - m(x_0)\right| \le \left|\underset{\sim}{b}_{x_0}^{T}(B^TB)^{-1}B^T\left(\underset{\sim}{m} - \underset{\sim}{m}^{spl}_{n,r}\right)\right| + \left\|m^{spl}_{n,r} - m\right\|_\infty,$$

where $\|m^{spl}_{n} - m\|_\infty \le C\cdot K_n^{-\beta}$, from de Boor (1968), and

$$\left|\underset{\sim}{b}_{x_0}^{T}(B^TB)^{-1}B^T\left(\underset{\sim}{m} - \underset{\sim}{m}^{spl}_{n,r}\right)\right| = \left|\underset{\sim}{b}_{x_0}^{T}\left(\frac{1}{n}B^TB\right)^{-1}\left(\frac{1}{n}B^T\left(\underset{\sim}{m} - \underset{\sim}{m}^{spl}_{n,r}\right)\right)\right|$$

$$\|Ax\|_1 = \left\| \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{d1} & \cdots & a_{dd} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \right\|_1$$

$$= \left\| x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{d1} \end{pmatrix} + \cdots + x_d \begin{pmatrix} a_{1d} \\ \vdots \\ a_{dd} \end{pmatrix} \right\|_1$$

$$\le \|\underset{\sim}{x}\|_1 \max_{1 \le k \le d} \left\| \begin{pmatrix} a_{1k} \\ \vdots \\ a_{dk} \end{pmatrix} \right\|_1 .$$

$$\le \left\| \left(\tfrac{1}{n} B^T B \right)^{-1} \underset{\sim}{b}_{x_0} \right\|_1 \left\| \tfrac{1}{n} B^T \left(\underset{\sim}{m} - \underset{\sim}{m}^{spl}_{n,r} \right) \right\|_\infty$$

$$\le \underbrace{\|\underset{\sim}{b}_{x_0}\|_1}_{\le 1} \left\| \left(\tfrac{1}{n} B^T B \right)^{-1} \right\|_\infty \left\| \tfrac{1}{n} B^T \left(\underset{\sim}{m} - \underset{\sim}{m}^{spl}_{n,r} \right) \right\|_\infty$$

$$\le C_2 \cdot K_n \cdot C_3 \cdot K_n^{-1-\beta} .$$

Putting the bias and variance bounds together completes the proof. $\square$

Exercise : Let $m \in \text{Lipschitz}(L)$, with $\hat{m}^{spl}_{n,0} \in \mathcal{M}_{n,0}$, i.e. piecewise constant.

(i) Give $\lambda_{min}\left(\frac{1}{n} B^T B\right)$ and $\lambda_{max}\left(\frac{1}{n} B^T B\right)$

(ii) Give $\left\| \left(\frac{1}{n} B^T B \right)^{-1} \right\|_\infty$.

(iii) Give $\left\| \frac{1}{n} B^T \left(\underset{\sim}{m} - \underset{\sim}{m}^{spl}_{n,0} \right) \right\|_\infty$.

(iv) Discuss reasonable bounds for these quantities

Then discuss assumptions about $X_1, \dots, X_n$ needed to bound these. Talk about $MSE\ \hat{m}^{spl}_n(x_0)$ in this setting.

Solution: First: $B = \left(\mathbb{1}\left(X_i \in I_{nk} \right) \right)_{1 \le k \le K_n,\, 1 \le i \le n}$, so

$$\left(\tfrac{1}{n} B^T B \right) = \operatorname{diag}\left(\frac{\#\{X_i \in I_{n1}\}}{n}, \dots, \frac{\#\{X_i \in I_{nK_n}\}}{n} \right).$$

This gives

$$\lambda_{min}\left(\tfrac{1}{n} B^T B \right) = \min_{1 \le k \le K_n} \frac{\#\{X_i \in I_{nk}\}}{n}$$

$$\lambda_{max}\left(\tfrac{1}{n} B^T B \right) = \max_{1 \le k \le K_n} \frac{\#\{X_i \in I_{nk}\}}{n} .$$

Also

$$\left\| \left(\frac{1}{n} B^T B \right)^{-1} \right\|_\infty = \max_{1 \le k \le K_n} \frac{n}{\#\{X_i \in I_{nk}\}}$$

and

$$\left\| \frac{1}{n} B^T (\underset{\sim}{m} - \underset{\sim}{m}_n^{spl}) \right\|_\infty = \max_{1 \le k \le K_n} \left| \frac{1}{n} \sum_{i=1}^n (m(X_i) - m_n^{spl}(X_i)) \mathbb{1}(X_i \in I_{nk}) \right|$$

$$\le \max_{1 \le k \le K_n} \| m - m_n^{spl} \|_\infty \frac{\#\{X_i \in I_{nk}\}}{n}$$

$$\le C \cdot K_n^{-1} \cdot \max_{1 \le k \le K_n} \frac{\#\{X_i \in I_{nk}\}}{n} .$$

If we can assume

$$\frac{1}{K_n} \frac{1}{c} \le \frac{\#\{X_i \in I_{nk}\}}{n} \le c \frac{1}{K_n} \qquad (\star)$$

for some $c > 0$, then we would have

(i) $\frac{1}{K_n} \frac{1}{c} \le \lambda_{min}\left(\frac{1}{n} B^T B\right) \le \lambda_{max}\left(\frac{1}{n} B^T B\right) \le c \frac{1}{K_n}$

(ii) $\left\| \left(\frac{1}{n} B^T B \right)^{-1} \right\|_\infty \le c \cdot K_n$

(iii) $\left\| \frac{1}{n} B^T (\underset{\sim}{m} - \underset{\sim}{m}_n^{spl}) \right\|_\infty \le C \cdot K_n^{-1} \cdot c \cdot K_n^{-1}$

We usually get some version of $(\star)$ by assuming $X_1, \dots, X_n \in [0,1]$ are generated according to some density f which is bounded away from zero and bounded above.

Our work would lead to

$$\text{MSE } \hat{m}_n^{spl}(x_0) \le C \cdot \left(K_n^{-2} + \frac{K_n}{n} \right).$$