Higher order spline HAL

Milos Mathias Koch - milos.koch@sund.ku.dk

September 3, 2025

	Distributional notation
$X \sim P_0$	X is distributed according to P_0.
$X \mid Y(=y) \sim P_0$	X is distributed according to P_0 given $Y(=y)$.
$X \sim Y$	X and Y are equal in distribution.
$X(P)$	The distribution of X.
$(X \mid Y(=y))(P)$	The distribution of X given $Y(=y)$.
$X \perp\!\!\!\perp Y$	X and Y are independent.
$X \perp\!\!\!\perp Y \mid Z$	X and Y are independent given Z.
$\mathcal{N}(\mu, \Sigma)$	The Gaussian distribution with mean μ and (co)variance Σ.
	Convergence notation
$\xrightarrow{P_0}$	Convergence in probability under P_0.
$\xrightarrow{P_n}$	Convergence in probability under the sequence $(P_n)_{n\in\mathbb{N}}$.
$\xrightarrow{\theta}$	Convergence in probability under P_θ.
$\overset{P_0}{\rightsquigarrow}$	Convergence in distribution under P_0.
$\overset{P_n}{\rightsquigarrow}$	Convergence in distribution under the sequence $(P_n)_{n\in\mathbb{N}}$.
$\overset{\theta}{\rightsquigarrow}$	Convergence in distribution under P_θ.

---

1 General results and notation

1.1 Knot points and index sets

The approximation results in this note depend on the L^2-approximation of cumulative functions in d dimensions of bounded sectional variation. It is shown in [van der Laan 2023] that such a function can be approximated with J knot points at a $O\big(r(d, M, J)\big)$ rate, where

$$r(d, M, J) \sim M \cdot \frac{\log(J)^{2(d-1)}}{J}.$$

We let $R \equiv R(d, J)$ be any set of J (deterministic) knot points that satisfy this approximation bound in $L^2(m)$ with m. the Lebesgue measure.

1.2 Motivation

It is well known that any càdlàg function can represented with basis zero-order splines

$$f(x) = f(0) + \sum_{s \subseteq \{1,\dots,d\}} \int 1(u_s \leq x_s) df(u_s, 0_{-s}),$$

where for subset s and vector x, x_s is the subvector of x with only the entries with indices in s and x_{-s} is complementary subvector. For convenience, we will write $f_s(x) = f(x_s, 0_{-s})$ for the s-section of f.

Example 1.1. Consider the càdlàg function of variation $f : [0, 1]^2 \to \mathbb{R}$ given by

$$f(x, y) = 1 + 2x + 3y + 4xy.$$

Then we can represent f with integrals with respect to the Radon-Nikodym derivatives of the sections $f_s, s \subseteq \{1, 2\}$. As f is smooth the Radon-Nikodym derivatives coincide with the classical derivatives, and so

$$\begin{aligned}
\frac{df_{\{1,2\}}}{dm}(x, y) &= \frac{df}{dxdy}(x, y) = 4 \\
\frac{df_{\{1\}}}{dm}(x, y) &= \frac{df}{dx}(x, 0) = 2 + 4 \cdot 0 = 2 \\
\frac{df_{\{1\}}}{dm}(x, y) &= \frac{df}{dy}(0, y) = 3 + 4 \cdot 0 = 3,
\end{aligned}$$

and so

$$\begin{aligned}
&f(0) + \int 1(u_1 \leq x) df_{\{1\}}(u_1, 0) + \int 1(u_2 \leq y) df_{\{2\}}(0, u_2) + \int 1(u \leq (x, y)) df_{\{1,2\}}(u_1, u_2) \\
=&f(0) + 2 \int 1(u_1 \leq x) du1 + 3 \int 1(u_2 \leq y) du_2 + 4 \int (u \leq (x, y)) d(u_1, u_2) \\
=&1 + 2x + 3y + 4xy,
\end{aligned}$$

where we have equipped $[0, 1]^2$ with the coordinatewise partial order.

The crux of the article [van der Laan 2023] is that estimating these sectional derivatives leads to a faster convergence rate when estimating the entire function.

We consider the function space of k-order smooth càdlàg functions, $D^{(k)}[0, 1]^d$ on the d-dimensional unit cube, which are defined as càdlàg functions, f, for any nested[1] list of subset, $S(k+1) = (s_1, \dots, s_{k+1})$ of $\{1, \dots, d\}$ the $S(k + 1)$-derivative of f exists in the following sense:

[1] In the sense of $s_1 \supset s_2 \supset \dots \supset s_{k+1}$.

1. If at any point along the list, $s_{j+1} = \varnothing$, we define any subsequent derivative as

$$f^{(l)}_{S(l)} = f^{(j)}_{S(l)}(x(s_l), 0(-s_l)), \text{ for } l = j+1, \dots, k$$

2. For non-empty s_{k+1}, we define $f^{(k+1)}$ as the Radon-Nikodym derivative of $f_{S(k)}$ with respect to $m_{s_{k+1}}$, the Lebesgue measure on $[0,1]^{s_{k+1}} = \{x_{s_{k+1}} \mid x \in [0,1]^d\}$.

We let $\mathcal{S}(k)$ denote the set of all such nested lists.

Lemma 1.2. $\#\mathcal{S}(k) = (k+1)^d$

Proof. For each $x \in \{1, \dots, d\}$, then the inclusion of x in each entry of a list $S(k) \in \mathcal{S}(k)$ can be represented with a non-decreasing binary sequence, as such

s_k,	$s_{k-1}, \dots,$	s_i,s_{i-1},	$\dots$,s_1
0,	0, $\dots$,	0,1,	$\dots$,1.

So for each x, there is $k+2$ possible lists[2]. In our example above, x is identified with $i-1$.

There are d such x's, whence we see the result. □

To model these derivatives, we will need higher order splines.

Definition 1.3. For a knot point $u \in [0,1]$, let ϕ^0_u be the zero-order spline $\phi^0_u(x) = 1(u \leq x)$.

Recursively define the j-order spline as

$$\phi^j_u(x) = \int_{(u,x]} \phi^{j-1}_{u'} dm(u')$$

for $j = 1, \dots k$.

For $u \in [0,1]^d$, we can then define the d-dimensional k-order spline as

$$\phi^k_u(x) = \prod_{j=1}^{d} \phi^k_{u_j}(x_j),$$

and the d-dimensional k-order $S(k)$ spline at 0 as

$$\Phi_{S(k)} = \prod_{j=1,\#s_j>0}^{k} \phi^j_0(x_{s_k \setminus s_{k+1}})$$

[2]One for each entry in the list and one for the list that contains x in none of its entries

Example 1.4.

$$\phi^0_{0.3}(x) = 1(0.3 \le x) = \begin{cases} 1, & x \ge 0.3, \\ 0, & x < 0.3. \end{cases}$$

$$\phi^1_{0.3}(x) = \max(x - 0.3, 0).$$

Let $u = (0.2,\, 0.5)$, $x = (x_1, x_2)$. Then

$$\phi^2_{(0.2,\,0.5)}(x) = \phi^2_{0.2}(x_1) \cdot \phi^2_{0.5}(x_2),$$

where, under Lebesgue measure,

$$\phi^2_{0.2}(x_1) = \frac{1}{2}\max(x_1 - 0.2, 0)^2, \quad \phi^2_{0.5}(x_2) = \frac{1}{2}\max(x_2 - 0.5, 0)^2.$$

Hence,

$$\phi^2_{(0.2,\,0.5)}(x) = \frac{1}{4}\max(x_1 - 0.2, 0)^2 \cdot \max(x_2 - 0.5, 0)^2.$$

In general, the single-dimensional splines are very much alike Taylor polynomials, and we realize

$$\phi^k_u(x) \;=\; \frac{(x-u)_+^k}{k!} \qquad \text{where} \quad (x-u)_+ := \max(x-u, 0).$$

$$\phi^k_u(x) = \begin{cases} \dfrac{(x-u)^k}{k!}, & x \ge u, \\ 0, & x < u. \end{cases}$$

$$\phi^k_u(x) \;=\; \prod_{j=1}^{d} \frac{\left(x_j - u_j\right)_+^k}{k!}, \qquad u = (u_1, \ldots, u_d) \in [0,1]^d,\; x = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

1.3 Main theorem

Similarly to a Taylor expansion, we want to represent a function in terms of its derivatives. To this end we define primitives, or antiderivatives, in terms of these derivatives.

Definition 1.5. A kth order primitive of a kth order smooth function, $f : [0,1]^d \to \mathbb{R}$ can be represented as

$$f(x) = \int_{(0,y_1]} \cdots \int_{(0,y_{k-1}]} f^{(k)}(y_k) \prod_{j=k}^{1} dm(y_j),$$

where $\prod_{j=k}^{1}$ is a descending product. We define the linear operator, $m^k : D[0,1]^d \to D^{(k)}$, as

$$m^k(f) = \int_{(0,y_1]} \cdots \int_{(0,y_{k-1}]} f(y_k) \prod_{j=k}^{1} dm(y_j),$$

from which we see $m^k(f^{(k)}) = f$. We can then define f to be càdlàg k order smooth if $f^{(k)}$ is in $D[0,1]^d$. This should be compared to the ordinary definition of C^k.

This leads us directly to our main representation result

Theorem 1.6. Let $f \in D^{(k)}[0,1]^d$, then

$$\begin{aligned} f(x) &= \sum_{S(k+1) \in \mathcal{S}(k+1)} \Phi_{S(k+1)}(x) m^k\left(f^{(k)}\right) \\ &= \sum_{S(k+1) | \#s_{k+1}=0} \Phi_{S(k+1)}(x) f_{t(S(k+1))}^{(t(S(k+1)))}(0_{s_{t(S(k+1))}}) \\ &+ \sum_{S(k+1) | \#s_{k+1}>0} \Phi_{S(k+1)}(x) m^k\left(f^{(k)}\right). \end{aligned}$$

Note that as the lists are nested, we are not counting any terms more than once in the second representation.

Case 1: $d = 1, k = 3$

$S(4)$	$\Phi_{S(4)}(x)$	$t(S(4))$	f term
$(\emptyset, \emptyset, \emptyset, \emptyset)$	1	0	$f_0^{(0)}(0)$
$(\{1\}, \emptyset, \emptyset, \emptyset)$	$\phi_0^1(x_1)$	1	$\phi_0^1(x_1) f_1^{(1)}(0)$
$(\{1\}, \{1\}, \emptyset, \emptyset)$	$\phi_0^1(x_1)\phi_0^2(x_1)$	2	$\phi_0^1(x_1)\phi_0^2(x_1) f_2^{(2)}(0)$
$(\{1\}, \{1\}, \{1\}, \emptyset)$	$\phi_0^1(x_1)\phi_0^2(x_1)\phi_0^3(x_1)$	3	$\phi_0^1(x_1)\phi_0^2(x_1)\phi_0^3(x_1) f_3^{(3)}(0)$
$(\{1\}, \{1\}, \{1\}, \{1\})$	$\phi_0^1(x_1)\phi_0^2(x_1)\phi_0^3(x_1)$	4	$\phi_0^1(x_1)\phi_0^2(x_1)\phi_0^3(x_1) m^3(f^{(3)}(x_1))$

Case 2: $d = 2, k = 2$

$S(3)$	$\Phi_{S(3)}(x)$	$t(S(3))$	f term
$(\emptyset, \emptyset, \emptyset)$	1	0	$f_0^{(0)}(0)$
$(\{1\}, \emptyset, \emptyset)$	$\phi_0^1(x_1)$	1	$\phi_0^1(x_1) f_1^{(1)}(0)$
$(\{2\}, \emptyset, \emptyset)$	$\phi_0^1(x_2)$	1	$\phi_0^1(x_2) f_1^{(1)}(0)$
$(\{1,2\}, \emptyset, \emptyset)$	$\phi_0^1(x_1)\phi_0^1(x_2)$	1	$\phi_0^1(x_1)\phi_0^1(x_2) f_1^{(1)}(0)$
$(\{1\}, \{1\}, \emptyset)$	$\phi_0^1(x_1)\phi_0^2(x_1)$	2	$\phi_0^1(x_1)\phi_0^2(x_1) f_2^{(2)}(0)$
$(\{2\}, \{2\}, \emptyset)$	$\phi_0^1(x_2)\phi_0^2(x_2)$	2	$\phi_0^1(x_2)\phi_0^2(x_2) f_2^{(2)}(0)$
$(\{1,2\}, \{1\}, \emptyset)$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_1)$	2	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_1) f_2^{(2)}(0)$
$(\{1,2\}, \{2\}, \emptyset)$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_2)$	2	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_2) f_2^{(2)}(0)$
$(\{1,2\}, \{1,2\}, \emptyset)$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_1)\phi_0^2(x_2)$	2	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_1)\phi_0^2(x_2) f_2^{(2)}(0)$
$(\{1\}, \{1\}, \{1\})$	$\phi_0^1(x_1)\phi_0^2(x_1)$	3	$\phi_0^1(x_1)\phi_0^2(x_1) m^2(f^{(2)}(x_1))$
$(\{2\}, \{2\}, \{2\})$	$\phi_0^1(x_2)\phi_0^2(x_2)$	3	$\phi_0^1(x_2)\phi_0^2(x_2) m^2(f^{(2)}(x_2))$
$(\{1,2\}, \{1\}, \{1\})$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_1)$	3	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_1) m^2(f^{(2)}(x_1))$
$(\{1,2\}, \{2\}, \{2\})$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_2)$	3	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_2) m^2(f^{(2)}(x_2))$
$(\{1,2\}, \{1,2\}, \{1\})$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_1)\phi_0^2(x_2)$	3	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_1)\phi_0^2(x_2) m^2(f^{(2)}(x_1))$
$(\{1,2\}, \{1,2\}, \{2\})$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_1)\phi_0^2(x_2)$	3	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_1)\phi_0^2(x_2) m^2(f^{(2)}(x_2))$
$(\{1,2\}, \{1,2\}, \{1,2\})$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_1)\phi_0^2(x_2)$	3	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^2(x_1)\phi_0^2(x_2) m^2(f^{(2)}(x_1, x_2))$

Case 3: $d = 3, k = 1$

$S(2)$	$\Phi_{S(2)}(x)$	$t(S(2))$	f term
$(\emptyset, \emptyset)$	1	0	$f_0^{(0)}(0)$
$(\{1\}, \emptyset)$	$\phi_0^1(x_1)$	1	$\phi_0^1(x_1)f_1^{(1)}(0)$
$(\{2\}, \emptyset)$	$\phi_0^1(x_2)$	1	$\phi_0^1(x_2)f_1^{(1)}(0)$
$(\{3\}, \emptyset)$	$\phi_0^1(x_3)$	1	$\phi_0^1(x_3)f_1^{(1)}(0)$
$(\{1,2\}, \emptyset)$	$\phi_0^1(x_1)\phi_0^1(x_2)$	1	$\phi_0^1(x_1)\phi_0^1(x_2)f_1^{(1)}(0)$
$(\{1,3\}, \emptyset)$	$\phi_0^1(x_1)\phi_0^1(x_3)$	1	$\phi_0^1(x_1)\phi_0^1(x_3)f_1^{(1)}(0)$
$(\{2,3\}, \emptyset)$	$\phi_0^1(x_2)\phi_0^1(x_3)$	1	$\phi_0^1(x_2)\phi_0^1(x_3)f_1^{(1)}(0)$
$(\{1,2,3\}, \emptyset)$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)$	1	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)f_1^{(1)}(0)$
$(\{1\}, \{1\})$	$\phi_0^1(x_1)$	2	$\phi_0^1(x_1)m^1(f^{(1)}(x_1))$
$(\{2\}, \{2\})$	$\phi_0^1(x_2)$	2	$\phi_0^1(x_2)m^1(f^{(1)}(x_2))$
$(\{3\}, \{3\})$	$\phi_0^1(x_3)$	2	$\phi_0^1(x_3)m^1(f^{(1)}(x_3))$
$(\{1,2\}, \{1\})$	$\phi_0^1(x_1)\phi_0^1(x_2)$	2	$\phi_0^1(x_1)\phi_0^1(x_2)m^1(f^{(1)}(x_1))$
$(\{1,2\}, \{2\})$	$\phi_0^1(x_1)\phi_0^1(x_2)$	2	$\phi_0^1(x_1)\phi_0^1(x_2)m^1(f^{(1)}(x_2))$
$(\{1,3\}, \{1\})$	$\phi_0^1(x_1)\phi_0^1(x_3)$	2	$\phi_0^1(x_1)\phi_0^1(x_3)m^1(f^{(1)}(x_1))$
$(\{1,3\}, \{3\})$	$\phi_0^1(x_1)\phi_0^1(x_3)$	2	$\phi_0^1(x_1)\phi_0^1(x_3)m^1(f^{(1)}(x_3))$
$(\{2,3\}, \{2\})$	$\phi_0^1(x_2)\phi_0^1(x_3)$	2	$\phi_0^1(x_2)\phi_0^1(x_3)m^1(f^{(1)}(x_2))$
$(\{2,3\}, \{3\})$	$\phi_0^1(x_2)\phi_0^1(x_3)$	2	$\phi_0^1(x_2)\phi_0^1(x_3)m^1(f^{(1)}(x_3))$
$(\{1,2,3\}, \{1\})$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)$	2	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)m^1(f^{(1)}(x_1))$
$(\{1,2,3\}, \{2\})$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)$	2	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)m^1(f^{(1)}(x_2))$
$(\{1,2,3\}, \{3\})$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)$	2	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)m^1(f^{(1)}(x_3))$
$(\{1,2,3\}, \{1,2\})$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)$	2	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)m^1(f^{(1)}(x_1,x_2))$
$(\{1,2,3\}, \{1,3\})$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)$	2	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)m^1(f^{(1)}(x_1,x_3))$
$(\{1,2,3\}, \{2,3\})$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)$	2	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)m^1(f^{(1)}(x_2,x_3))$
$(\{1,2,3\}, \{1,2,3\})$	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)$	2	$\phi_0^1(x_1)\phi_0^1(x_2)\phi_0^1(x_3)m^1(f^{(1)}(x_1,x_2,x_3))$

Assume that we, for each $S(k+1)$ have a set of knot points $R(S(k+1), J)$ such that we can estimate

$$f^{\tilde{(k)}}(x_{s_{k+1}}) = \sum_{u \in R(\# s_{k+1}, J)} \beta_u \phi_u^0.$$

This is the HAL-MLE in the case where $R(\# s_{k+1}, J)$ are the section of the covariates. The by linearity of m^k, we have

$$
\begin{aligned}
\tilde{f}(x) &= \sum_{S(k+1)\in\mathcal{S}(k+1)} \Phi_{S(k+1)}(x) m^k\left(f^{(k)}\right) \\
&= \sum_{S(k+1)\mid \# s_{k+1}=0} \Phi_{S(k+1)}(x) f_{t(S(k+1))}^{(t(S(k+1)))}(0_{s_{t(S(k+1))}}) \\
&+ \sum_{S(k+1)\mid \# s_{k+1}>0} \Phi_{S(k+1)}(x) \sum_{u\in R(\# s_{k+1},J)} \beta_u \phi_u^k.
\end{aligned}
$$

This is the higher order HAL-MLE.

References

[van der Laan 2023] Mark van der Laan, *Higher Order Spline Highly Adaptive Lasso Estimators of Functional Parameters: Pointwise Asymptotic Normality and Uniform Convergence Rates*, *arχiv*, 2023.