S-shaped function estimation

Oliver Feng

Statistical Laboratory, University of Cambridge

Institut Henri Poincaré, 19 July 2022

Joint work with Yining Chen (LSE), Qiyang Han (Rutgers), Ray Carroll (Texas A&M) and Richard J. Samworth (Cambridge)

Outline

1. **Introduction to S-shaped regression functions**

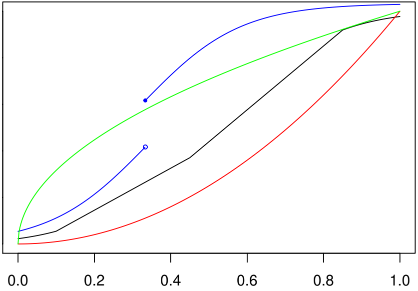
   Applications and examples

2. **Definitions and theory**

3. **Computation**

S-shaped functions

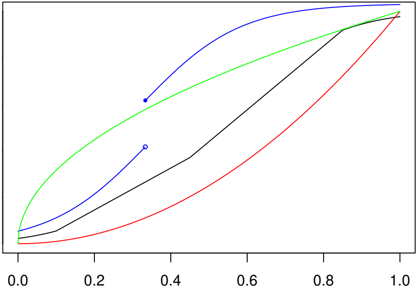
- We say that $f : [0, 1] \to \mathbb{R}$ is S-shaped if it is increasing, and if there exists an inflection point $m_0 \in [0, 1]$ such that f is convex on $[0, m_0]$ and concave on $[m_0, 1]$.
- f is not required to be continuous at m_0 or Lipschitz on $[0, 1]$.



S-shaped functions

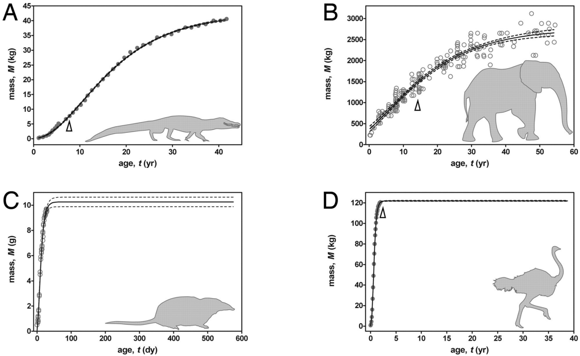
- We say that $f: [0,1] \to \mathbb{R}$ is S-shaped if it is increasing, and if there exists an inflection point $m_0 \in [0,1]$ such that f is convex on $[0, m_0]$ and concave on $[m_0, 1]$.

- f is not required to be continuous at m_0 or Lipschitz on $[0,1]$.
- Aim: Estimate an unknown S-shaped regression function and its inflection point(s).



S-shaped regression functions

- Modelling the dependence of a response variable on a covariate as an S-shaped function: many examples in applied science, such as **growth or development curves** for individuals or populations, and **learning curves** for skill proficiency

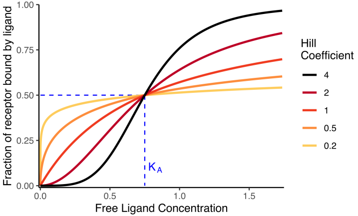


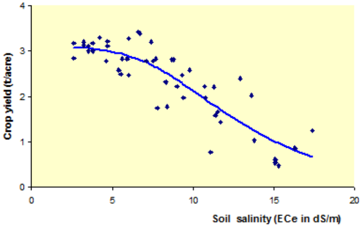
Source: Lee and Werner (2008)

S-shaped regression functions

Further examples:

- **Production or utility curves** in economics (e.g. output vs resource levels, or sales revenue vs advertising)
- **Dose-response curves** in biochemistry and medicine
- Dependence of crop yields on soil salinity (inverted S-shaped curve) in agronomy (van Genuchten and Gupta, 1993)



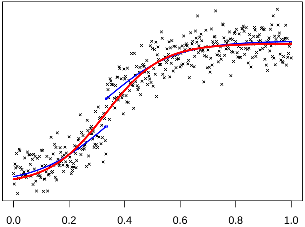


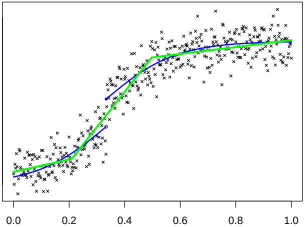
Parametric methods

- Restricting to parametric subclasses of sigmoidal curves, e.g. generalised **logistic functions**

$$x \mapsto C + \frac{B}{(1 + e^{-bx+c})^{\kappa}} \quad \text{with } B, b, \kappa > 0,\ C, c \in \mathbb{R};$$

- **Piecewise linear regression** with a fixed number of kinks, and bent cable (linear–quadratic–linear) models (Chiu et al., 2006)

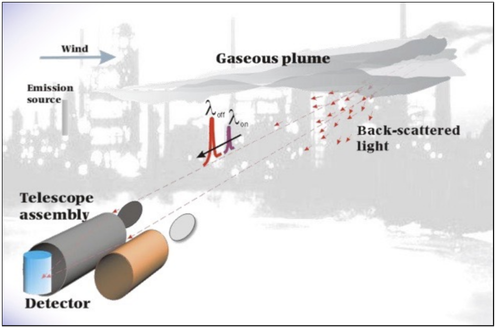




Nonparametric methods

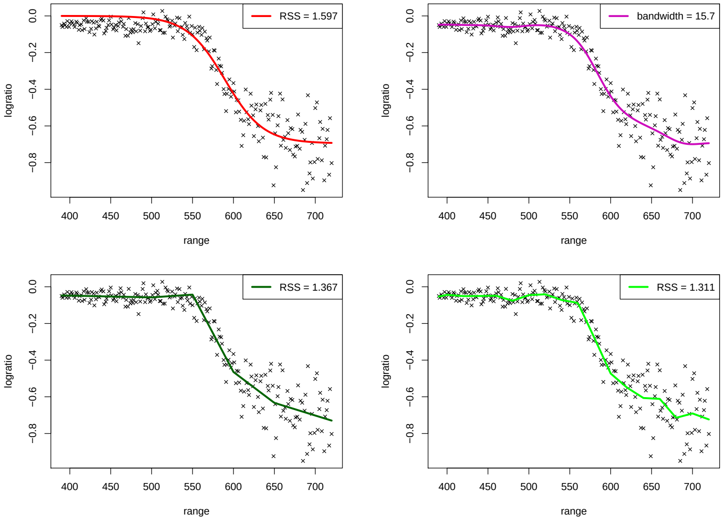
- Kernel-based methods:
  - Identification of the inflection point of a smooth signal via **local polynomial regression** and constrained bandwidth selection (Kachouie and Schwartzman, 2013)
  - Estimation of S-shaped production functions via **shape-constrained kernel least squares** and bandwidth selection by cross-validation (Yagi et al., 2019, 2020)
- (Penalised) least squares based on **cubic B-splines** defined with respect to user-specified knots (Liao and Meyer, 2017)
- Geometric / numerical analysis approach to identifying inflection points (Christopoulos, 2016)

Example: LIDAR air pollution data



- Model: $\log \frac{P(r_i;\lambda_{\mathrm{on}})}{P(r_i;\lambda_{\mathrm{off}})} = f_0(r_i) + \xi_i, \quad i = 1, \dots, n$
- $f_0(r) = -C \int_0^r g_0(s)\, ds$ for $r \geq 0$, where $g_0(s)$ is the concentration of mercury at distance s metres away from the detector, and $C = 1.6 \times 10^{-5}\,\mathrm{ng}^{-1}\,\mathrm{m}^2$

Example: LIDAR air pollution data



Outline

1. Introduction to S-shaped regression functions

2. Definitions and theory

- Consistency and robustness
- Finite-sample risk bounds
- Inflection point estimation

3. Computation

S-shaped least squares estimators (LSEs)

- For $m \in [0, 1]$, denote by $\mathcal{F}^m$ the class of all S-shaped functions on $[0, 1]$ with an inflection point at m: this is a convex cone

- Denote by $\mathcal{F} = \bigcup_{m\in[0,1]} \mathcal{F}^m$ the class of all S-shaped functions on $[0, 1]$: this is **not convex**

S-shaped least squares estimators (LSEs)

- For $m \in [0,1]$, denote by $\mathcal{F}^m$ the class of all S-shaped functions on $[0,1]$ with an inflection point at m: this is a convex cone
- Denote by $\mathcal{F} = \bigcup_{m\in[0,1]} \mathcal{F}^m$ the class of all S-shaped functions on $[0,1]$: this is **not convex**
- Observe $(x_1, Y_1), \ldots, (x_n, Y_n) \in [0,1] \times \mathbb{R}$ with $x_1 < \cdots < x_n$
- For a class $\tilde{\mathcal{F}}$ of functions on $[0,1]$, we say that $\tilde{f}_n \colon [0,1] \to \mathbb{R}$ is an LSE over $\tilde{\mathcal{F}}$ based on $\{(x_i, Y_i) : 1 \leq i \leq n\}$ if it minimises $f \mapsto \sum_{i=1}^{n} \bigl(Y_i - f(x_i)\bigr)^2 =: \mathrm{RSS}_n(f)$ over $\tilde{\mathcal{F}}$

Existence of S-shaped LSEs

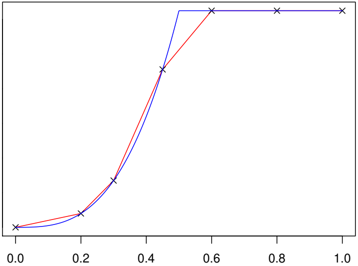
Proposition. For each $m \in [0,1]$, there exists an LSE $\tilde{f}_n^m$ over $\mathcal{F}^m$ that is uniquely determined at $x_1, \ldots, x_n$, and there exists an LSE $\tilde{f}_n$ over $\mathcal{F}$ with an inflection point in $\{x_2, \ldots, x_{n-1}\}$.

- Form $\hat{f}_n^m$ by linear interpolation: this may not lie in $\mathcal{F}^m$ if $m \notin \{x_1, \ldots, x_n\}$; it is the LSE over a modified class $\mathcal{H}^m$

Existence of S-shaped LSEs

Proposition. For each $m \in [0, 1]$, there exists an LSE $\tilde{f}_n^m$ over $\mathcal{F}^m$ that is uniquely determined at $x_1, \ldots, x_n$, and there exists an LSE $\tilde{f}_n$ over $\mathcal{F}$ with an inflection point in $\{x_2, \ldots, x_{n-1}\}$.

- Form $\hat{f}_n^m$ by linear interpolation: this may not lie in $\mathcal{F}^m$ if $m \notin \{x_1, \ldots, x_n\}$; it is the LSE over a modified class $\mathcal{H}^m$



Existence of S-shaped LSEs

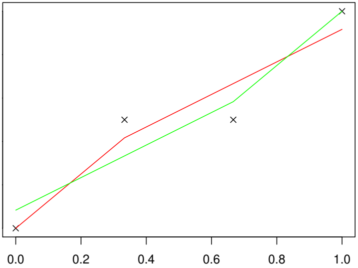
Proposition. For each $m \in [0, 1]$, there exists an LSE $\tilde{f}_n^m$ over $\mathcal{F}^m$ that is uniquely determined at $x_1, \ldots, x_n$, and there exists an LSE $\tilde{f}_n$ over $\mathcal{F}$ with an inflection point in $\{x_2, \ldots, x_{n-1}\}$.

- $\hat{f}_n^m$ is unique because $\mathcal{H}^m$ is convex, but $\tilde{f}_n$ may not even be uniquely defined on $\{x_1, \ldots, x_n\}$

Existence of S-shaped LSEs

Proposition. For each $m \in [0, 1]$, there exists an LSE $\tilde{f}_n^m$ over $\mathcal{F}^m$ that is uniquely determined at $x_1, \ldots, x_n$, and there exists an LSE $\tilde{f}_n$ over $\mathcal{F}$ with an inflection point in $\{x_2, \ldots, x_{n-1}\}$.

- $\hat{f}_n^m$ is unique because $\mathcal{H}^m$ is convex, but $\tilde{f}_n$ may not even be uniquely defined on $\{x_1, \ldots, x_n\}$



L^2 projection framework

- For a general probability distribution P on $[0,1] \times \mathbb{R}$ with a finite second moment, consider minimising

$$f \mapsto L(f,P) = \int_{[0,1]\times\mathbb{R}} \bigl(y - f(x)\bigr)^2 \, dP(x,y)$$

over $\mathcal{F}^m$ for a fixed $m \in [0,1]$, or over $\mathcal{F}$.

L^2 projection framework

- For a general probability distribution P on $[0,1] \times \mathbb{R}$ with a finite second moment, consider minimising

$$f \mapsto L(f, P) = \int_{[0,1]\times\mathbb{R}} \big(y - f(x)\big)^2 \, dP(x, y)$$

over $\mathcal{F}^m$ for a fixed $m \in [0,1]$, or over $\mathcal{F}$.

- $\psi_m^0(P) = \text{argmin}_{f \in \mathcal{F}^m} L(f, P)$ and $\psi^0(P) = \text{argmin}_{f \in \mathcal{F}} L(f, P)$ are well-defined under mild conditions on P, but uniqueness (P^X almost everywhere) is not guaranteed for the latter.

L^2 projection framework

- For a general probability distribution P on $[0,1] \times \mathbb{R}$ with a finite second moment, consider minimising

$$f \mapsto L(f,P) = \int_{[0,1]\times\mathbb{R}} \big(y - f(x)\big)^2 \, dP(x,y)$$

over $\mathcal{F}^m$ for a fixed $m \in [0,1]$, or over $\mathcal{F}$.

- $\psi_m^0(P) = \operatorname{argmin}_{f \in \mathcal{F}^m} L(f,P)$ and $\psi^0(P) = \operatorname{argmin}_{f \in \mathcal{F}} L(f,P)$ are well-defined under mild conditions on P, but uniqueness (P^X almost everywhere) is not guaranteed for the latter.

- Continuity results for $(m,P) \mapsto \psi_m^0(P)$, $P \mapsto \psi^0(P)$, $(m,P) \mapsto \inf_{f \in \mathcal{F}^m} L(f,P)$ and $P \mapsto \inf_{f \in \mathcal{F}} L(f,P)$ with respect to the 2-Wasserstein distance W_2 and notions of set convergence

Consistency and robustness

- Regression framework (triangular array scheme):
  $Y_{ni} = f_0(x_{ni}) + \xi_{ni}$ for $n \in \mathbb{N}$ and $1 \leq i \leq n$, where $\xi_{n1}, \ldots, \xi_{nn}$ are i.i.d. with mean zero and finite variance

Consistency and robustness

- Regression framework (triangular array scheme): $Y_{ni} = f_0(x_{ni}) + \xi_{ni}$ for $n \in \mathbb{N}$ and $1 \leq i \leq n$, where $\xi_{n1}, \ldots, \xi_{nn}$ are i.i.d. with mean zero and finite variance

- **Consistency** when $f_0 \in \mathcal{F}$ has a unique inflection point m_0:

Theorem. Suppose $\mathbb{P}_n^X = n^{-1} \sum_{i=1}^n \delta_{x_{ni}}$ converges weakly to a distribution P_0^X with supp $P_0^X = [0,1]$ and $P_0^X(\{m\}) = 0$ for all $m \in [0,1]$. For $n \in \mathbb{N}$, let $\tilde{g}_n \in \{\hat{f}_n^{m_0}, \tilde{f}_n\}$ and let $\tilde{m}_n$ denote any inflection point of $\tilde{g}_n$. Then for any closed $A \subseteq [0,1] \setminus \{m_0\}$,

$$\tilde{m}_n \xrightarrow{p} m_0, \qquad \sup_{x \in A} |(\tilde{g}_n - f_0)(x)| \xrightarrow{p} 0.$$

Consistency and robustness

- Regression framework (triangular array scheme): $Y_{ni} = f_0(x_{ni}) + \xi_{ni}$ for $n \in \mathbb{N}$ and $1 \leq i \leq n$, where $\xi_{n1}, \ldots, \xi_{nn}$ are i.i.d. with mean zero and finite variance

- **Consistency** when $f_0 \in \mathcal{F}$ has a unique inflection point m_0:

Theorem. Suppose $\mathbb{P}_n^X = n^{-1} \sum_{i=1}^n \delta_{x_{ni}}$ converges weakly to a distribution P_0^X with $\operatorname{supp} P_0^X = [0,1]$ and $P_0^X(\{m\}) = 0$ for all $m \in [0,1]$. For $n \in \mathbb{N}$, let $\tilde{g}_n \in \{\hat{f}_n^{m_0}, \tilde{f}_n\}$ and let $\tilde{m}_n$ denote any inflection point of $\tilde{g}_n$. Then for any closed $A \subseteq [0,1] \setminus \{m_0\}$,

$$\tilde{m}_n \xrightarrow{p} m_0, \qquad \sup_{x \in A} |(\tilde{g}_n - f_0)(x)| \xrightarrow{p} 0.$$

- **Robustness to misspecification** ($f_0 \notin \mathcal{F}$): "$\hat{f}_n^{m_0}, \tilde{f}_n$ converge to the projections of f_0 onto $\mathcal{F}^{m_0}, \mathcal{F}$ respectively"

Finite-sample risk bounds

- For fixed $n \in \mathbb{N}$, consider a fixed design regression model $Y_i = f_0(x_i) + \xi_i,\ i = 1, \ldots, n$ in which $\xi_1, \ldots, \xi_n$ are independent and sub-Gaussian with parameter 1.
- Global loss function: for $g\colon [0,1] \to \mathbb{R}$, let $\|g\|_n \equiv \|g\|_{L^2(\mathbb{P}_n^X)} = \bigl(\sum_{i=1}^n g^2(x_i)/n\bigr)^{1/2}$.
- We prove sharp oracle inequalities to quantify the worst-case and adaptive performance of S-shaped LSEs $\tilde{f}_n$: these apply to any regression function $f_0\colon [0,1] \to \mathbb{R}$ and take the form

$$\mathbb{E}\bigl(\|\tilde{f}_n - f_0\|_n\bigr) \leq \inf_{f \in \mathcal{F}} \bigl\{\|f_0 - f\|_n + r_n(f)\bigr\}, \quad \text{where}$$

  - $\|f_0 - f\|_n$ is an approximation error term with leading constant 1;
  - $r_n(f)$ is an estimation error term.

Finite-sample risk bounds: worst-case

- Let $R := n^{-1}(x_n - x_1)/\min_{2 \le i \le n}(x_i - x_{i-1})$. For $f \in \mathcal{F}$, define

$$V(f) := f(x_n) - f(x_1) = \max_{1 \le i \le n} f(x_i) - \min_{1 \le i \le n} f(x_i).$$

Finite-sample risk bounds: worst-case

- Let $R := n^{-1}(x_n - x_1)/\min_{2\leq i\leq n}(x_i - x_{i-1})$. For $f \in \mathcal{F}$, define

$$V(f) := f(x_n) - f(x_1) = \max_{1\leq i\leq n} f(x_i) - \min_{1\leq i\leq n} f(x_i).$$

Theorem. There exists a universal $C > 0$ such that for every $f_0 \colon [0,1] \to \mathbb{R}$, $n \geq 2$ and LSE $\tilde{f}_n$ over $\mathcal{F}$, we have

$$\|\tilde{f}_n - f_0\|_n \leq$$

$$\inf_{f\in\mathcal{F}} \left\{ \|f - f_0\|_n + \frac{C\bigl(1+V(f)\bigr)^{1/3}}{n^{1/3}} \wedge \frac{CR^{1/10}\bigl(1+V(f)\bigr)^{1/5}}{n^{2/5}} \right\} + \sqrt{\frac{8t}{n}}$$

with probability at least $1 - e^{-t}$, for every $t > 0$.

Finite-sample risk bounds: worst-case

- Let $R := n^{-1}(x_n - x_1)/\min_{2\leq i\leq n}(x_i - x_{i-1})$. For $f \in \mathcal{F}$, define

$$V(f) := f(x_n) - f(x_1) = \max_{1\leq i\leq n} f(x_i) - \min_{1\leq i\leq n} f(x_i).$$

Theorem. There exists a universal $C > 0$ such that for every $f_0 \colon [0,1] \to \mathbb{R}$, $n \geq 2$ and LSE $\tilde{f}_n$ over $\mathcal{F}$, we have

$$\mathbb{E}_{f_0}(\|\tilde{f}_n - f_0\|_n) \leq$$

$$\inf_{f\in\mathcal{F}} \left\{ \|f - f_0\|_n + \frac{C\big(1 + V(f)\big)^{1/3}}{n^{1/3}} \wedge \frac{CR^{1/10}\big(1 + V(f)\big)^{1/5}}{n^{2/5}} \right\} + \sqrt{\frac{2\pi}{n}}.$$

Finite-sample risk bounds: worst-case

- Let $R := n^{-1}(x_n - x_1)/\min_{2\leq i\leq n}(x_i - x_{i-1})$. For $f \in \mathcal{F}$, define

$$V(f) := f(x_n) - f(x_1) = \max_{1\leq i\leq n} f(x_i) - \min_{1\leq i\leq n} f(x_i).$$

Conclusion: when $f_0 \in \mathcal{F}$, or f_0 is close to $\mathcal{F}$ in an $L^2(\mathbb{P}_n^X)$ sense:

- $\tilde{f}_n$ attains the optimal worst-case risk of order $n^{-2/5}$ with respect to $L^2(\mathbb{P}_n^X)$-loss when the design points are 'near-equispaced'.

Finite-sample risk bounds: worst-case

- Let $R := n^{-1}(x_n - x_1)/\min_{2\le i\le n}(x_i - x_{i-1})$. For $f \in \mathcal{F}$, define

$$V(f) := f(x_n) - f(x_1) = \max_{1\le i\le n} f(x_i) - \min_{1\le i\le n} f(x_i).$$

Conclusion: when $f_0 \in \mathcal{F}$, or f_0 is close to $\mathcal{F}$ in an $L^2(\mathbb{P}_n^X)$ sense:

- $\tilde{f}_n$ attains the optimal worst-case risk of order $n^{-2/5}$ with respect to $L^2(\mathbb{P}_n^X)$-loss when the design points are 'near-equispaced'.

- For adversarially chosen design configurations, the worst-case risk of $\tilde{f}_n$ can be of order $n^{-1/3}$.

Finite-sample risk bounds: adaptation

- Let $\mathcal{H}$ be the set of piecewise affine $f \in \mathcal{F}$ with kinks in $\{x_2, \ldots, x_{n-1}\}$.
- For $f \in \mathcal{H}$, denote by $k(f)$ the number of affine pieces of f, i.e. the smallest $k \in [n]$ such that f is affine on each of k subintervals $I_1, \ldots, I_k$ that partition $[0, 1]$.

Finite-sample risk bounds: adaptation

- Let $\mathcal{H}$ be the set of piecewise affine $f \in \mathcal{F}$ with kinks in $\{x_2, \dots, x_{n-1}\}$.
- For $f \in \mathcal{H}$, denote by $k(f)$ the number of affine pieces of f, i.e. the smallest $k \in [n]$ such that f is affine on each of k subintervals $I_1, \dots, I_k$ that partition $[0,1]$.

Theorem. For every $f_0 \colon [0,1] \to \mathbb{R}$, $n \geq 2$ and LSE $\tilde{f}_n$ over $\mathcal{F}$, we have

$$\|\tilde{f}_n - f_0\|_n \leq \inf_{f \in \mathcal{H}} \left\{ \|f - f_0\|_n + \sqrt{\frac{32\big(k(f)+1\big)}{n} \log\left(\frac{en}{k(f)+1}\right)} \right\} + \sqrt{\frac{2(t + \log n)}{n}}$$

with probability at least $1 - e^{-t}$, for every $t > 0$.

Finite-sample risk bounds: adaptation

- Let $\mathcal{H}$ be the set of piecewise affine $f \in \mathcal{F}$ with kinks in $\{x_2, \ldots, x_{n-1}\}$.
- For $f \in \mathcal{H}$, denote by $k(f)$ the number of affine pieces of f, i.e. the smallest $k \in [n]$ such that f is affine on each of k subintervals $I_1, \ldots, I_k$ that partition $[0, 1]$.

Theorem. For every $f_0 \colon [0,1] \to \mathbb{R}$, $n \geq 2$ and LSE $\tilde{f}_n$ over $\mathcal{F}$, we have

$$\mathbb{E}_{f_0}(\|\tilde{f}_n - f_0\|_n) \leq \inf_{f \in \mathcal{H}} \left\{ \|f - f_0\|_n + 8\sqrt{\frac{k(f)+1}{n} \log\left(\frac{en}{k(f)+1}\right)} \right\}.$$

Finite-sample risk bounds: adaptation

- Let $\mathcal{H}$ be the set of piecewise affine $f \in \mathcal{F}$ with kinks in $\{x_2, \ldots, x_{n-1}\}$.
- For $f \in \mathcal{H}$, denote by $k(f)$ the number of affine pieces of f, i.e. the smallest $k \in [n]$ such that f is affine on each of k subintervals $I_1, \ldots, I_k$ that partition $[0, 1]$.

Conclusion: when $f_0 \in \mathcal{F}$, or f_0 is close to $\mathcal{F}$ in an $L^2(\mathbb{P}_n^X)$ sense,

- $\tilde{f}_n$ adaptively attains the parametric rate of order $n^{-1/2}$ (up to a logarithmic factor) when the approximating function $f \in \mathcal{F}$ is piecewise affine with a small number of affine pieces.

Inflection point estimation

- Assumptions on a sequence of regression models $Y_{ni} = f_0(x_{ni}) + \xi_{ni}, \ i = 1, \ldots, n$ (triangular array scheme):

Inflection point estimation

- Assumptions on a sequence of regression models $Y_{ni} = f_0(x_{ni}) + \xi_{ni},\ i = 1, \ldots, n$ (triangular array scheme):
- $x_{ni} = i/n$ and $\xi_{n1}, \ldots, \xi_{nn} \overset{\text{iid}}{\sim} P_\xi$ for all $n \in \mathbb{N}$ and $1 \leq i \leq n$, where P_ξ is sub-Gaussian with parameter 1.

Inflection point estimation

- Assumptions on a sequence of regression models $Y_{ni} = f_0(x_{ni}) + \xi_{ni},\ i = 1, \ldots, n$ (triangular array scheme):
- $x_{ni} = i/n$ and $\xi_{n1}, \ldots, \xi_{nn} \overset{\text{iid}}{\sim} P_\xi$ for all $n \in \mathbb{N}$ and $1 \le i \le n$, where P_ξ is sub-Gaussian with parameter 1.

Local smoothness condition: $f_0 \in \mathcal{F}$ has a unique inflection point $m_0 \in (0,1)$, and there exists $B > 0$ such that as $x \to m_0$, either

$$f_0(x) = f_0(m_0) - B\big(1 + o(1)\big)\,\mathrm{sgn}(x - m_0)|x - m_0|^\alpha$$

for some $\alpha \in (0,1)$, or

$$f_0(x) = f_0(m_0) + f_0'(m_0)(x - m_0) + B\big(1 + o(1)\big)\,\mathrm{sgn}(x - m_0)|x - m_0|^\alpha$$

for some $\alpha \in (1, \infty)$.

Inflection point estimation

- Assumptions on a sequence of regression models $Y_{ni} = f_0(x_{ni}) + \xi_{ni},\ i = 1, \ldots, n$ (triangular array scheme):

Local smoothness condition: $f_0 \in \mathcal{F}$ has a unique inflection point $m_0 \in (0, 1)$, and there exists $B > 0$ such that as $x \to m_0$, either

$$f_0(x) = f_0(m_0) - B\big(1 + o(1)\big)\,\mathrm{sgn}(x - m_0)|x - m_0|^\alpha$$

for some $\alpha \in (0, 1)$, or

$$f_0(x) = f_0(m_0) + f_0'(m_0)(x - m_0) + B\big(1 + o(1)\big)\,\mathrm{sgn}(x - m_0)|x - m_0|^\alpha$$

for some $\alpha \in (1, \infty)$.

- When $\alpha \geq 3$ is an odd integer, this holds if f_0 is locally C^α at m_0 and $f_0^{(k)}(m_0) = 0 \neq f_0^{(\alpha)}(m_0)$ for $2 \leq k \leq \alpha - 1$.

Inflection point estimation

Theorem. Under the assumptions on the previous slide, let $(\tilde{f}_n)$ be any sequence of LSEs over $\mathcal{F}$, and for each n, let $\tilde{m}_n$ be an inflection point of $\tilde{f}_n$. Then $\tilde{m}_n - m_0 = O_p\big((n/\log n)^{-1/(2\alpha+1)}\big)$.

Inflection point estimation

Theorem. Under the assumptions on the previous slide, let $(\tilde{f}_n)$ be any sequence of LSEs over $\mathcal{F}$, and for each n, let $\tilde{m}_n$ be an inflection point of $\tilde{f}_n$. Then $\tilde{m}_n - m_0 = O_p\big((n/\log n)^{-1/(2\alpha+1)}\big)$.

- For example, for $f_0 \colon x \mapsto \Phi(x - m_0)$, we have $\alpha = 3$ and therefore $\tilde{m}_n - m_0 = O_p\big((n/\log n)^{-1/7}\big)$.

Inflection point estimation

Theorem. Under the assumptions on the previous slide, let $(\tilde{f}_n)$ be any sequence of LSEs over $\mathcal{F}$, and for each n, let $\tilde{m}_n$ be an inflection point of $\tilde{f}_n$. Then $\tilde{m}_n - m_0 = O_p\big((n/\log n)^{-1/(2\alpha+1)}\big)$.

- For example, for $f_0 \colon x \mapsto \Phi(x - m_0)$, we have $\alpha = 3$ and therefore $\tilde{m}_n - m_0 = O_p\big((n/\log n)^{-1/7}\big)$.

- Proof idea: If m_n is a long way from m_0, then on the interval between them, $\mathrm{RSS}_n(\hat{f}_n^{m_n}) > \mathrm{RSS}_n(\hat{f}_n^{m_0})$ due to misspecification: there is a long subinterval between m_0 and m_n on which one of $f_0, \hat{f}_n^{m_n}$ is convex and the other is concave.

Inflection point estimation

Theorem. Under the assumptions on the previous slide, let $(\tilde{f}_n)$ be any sequence of LSEs over $\mathcal{F}$, and for each n, let $\tilde{m}_n$ be an inflection point of $\tilde{f}_n$. Then $\tilde{m}_n - m_0 = O_p\big((n/\log n)^{-1/(2\alpha+1)}\big)$.

- For example, for $f_0\colon x \mapsto \Phi(x - m_0)$, we have $\alpha = 3$ and therefore $\tilde{m}_n - m_0 = O_p\big((n/\log n)^{-1/7}\big)$.

- Proof idea: If m_n is a long way from m_0, then on the interval between them, $\mathrm{RSS}_n(\hat{f}_n^{m_n}) > \mathrm{RSS}_n(\hat{f}_n^{m_0})$ due to misspecification: there is a long subinterval between m_0 and m_n on which one of $f_0, \hat{f}_n^{m_n}$ is convex and the other is concave.

- However, $\mathrm{RSS}_n(\tilde{f}_n) - \mathrm{RSS}_n(\hat{f}_n^{m_0}) \leq 0$ by definition of $\tilde{f}_n = \hat{f}_n^{\tilde{m}_n}$.

Inflection point estimation

Theorem. Under the assumptions on the previous slide, let $(\tilde{f}_n)$ be any sequence of LSEs over $\mathcal{F}$, and for each n, let $\tilde{m}_n$ be an inflection point of $\tilde{f}_n$. Then $\tilde{m}_n - m_0 = O_p\big((n/\log n)^{-1/(2\alpha+1)}\big)$.

Complementary **local asymptotic minimax lower bound**:

- For $r > 0$, let $\mathcal{F}(f_0, r) := \{f \in \mathcal{F} : \int_0^1 (f - f_0)^2 < r^2\}$.
- For $f \in \mathcal{F}$, write $\mathcal{I}_f$ for the set of inflection points of f and let $d(x, \mathcal{I}_f) := \inf_{z \in \mathcal{I}_f} |x - z|$ for $x \in [0, 1]$.

Inflection point estimation

Theorem. Under the assumptions on the previous slide, let $(\tilde{f}_n)$ be any sequence of LSEs over $\mathcal{F}$, and for each n, let $\tilde{m}_n$ be an inflection point of $\tilde{f}_n$. Then $\tilde{m}_n - m_0 = O_p\big((n/\log n)^{-1/(2\alpha+1)}\big)$.

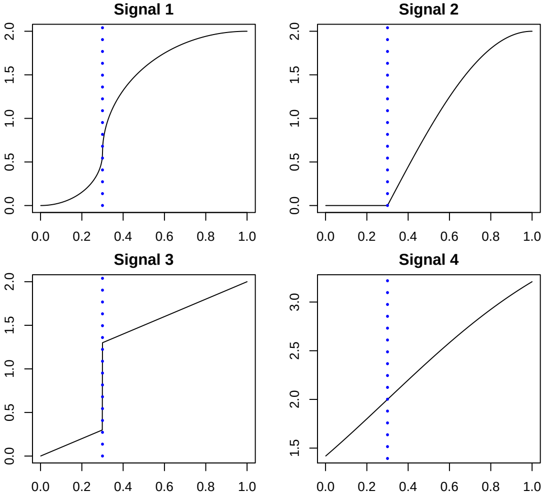
Complementary **local asymptotic minimax lower bound**:

- For $r > 0$, let $\mathcal{F}(f_0, r) := \{f \in \mathcal{F} : \int_0^1 (f - f_0)^2 < r^2\}$.
- For $f \in \mathcal{F}$, write $\mathcal{I}_f$ for the set of inflection points of f and let $d(x, \mathcal{I}_f) := \inf_{z \in \mathcal{I}_f} |x - z|$ for $x \in [0, 1]$.

Proposition. Under the same assumptions, with $\xi_{n1}, \ldots, \xi_{nn} \overset{\text{iid}}{\sim} N(0, 1)$ for all n, we have

$$\sup_{\tau > 0} \liminf_{n \to \infty} \inf_{\check{m}_n} \sup_{f \in \mathcal{F}(f_0, \tau/\sqrt{n})} n^{1/(2\alpha+1)} \, \mathbb{E}_f\big(d(\check{m}_n, \mathcal{I}_f)\big) > 0.$$

Inflection point result: illustration



Inflection point result: illustration

- S-shaped signals with $m_0 = 0.3$:

$$f_1(x) = \begin{cases} 2(0.3 - \sqrt{0.09 - x^2}) & \text{for } x \in [0, 0.3) \\ 2\{0.3 + \sqrt{0.49 - (1-x)^2}\} & \text{for } x \in [0.3, 1]; \end{cases}$$

$$f_2(x) = \sin\big((x - 0.3)\pi/1.4\big)\mathbb{1}_{\{x \geq 0.3\}};$$

$$f_3(x) = x + \mathbb{1}_{\{x \geq 0.3\}};$$

$$f_4(x) = 4/\big(1 + e^{-2(x-0.3)}\big).$$

- f_1: $\alpha = 1/2 \Rightarrow$ rate $O_p(n^{-1/2})$
- f_2: does not satisfy the assumption for any $\alpha > 0$; rate $O_p(n^{-1/3})$ for kink estimation
- f_3: '$\alpha = 0$'; rate $O_p(n^{-1})$ for changepoint estimation
- f_4: $\alpha = 3 \Rightarrow$ rate $O_p(n^{-1/7})$

Subinterval localisation

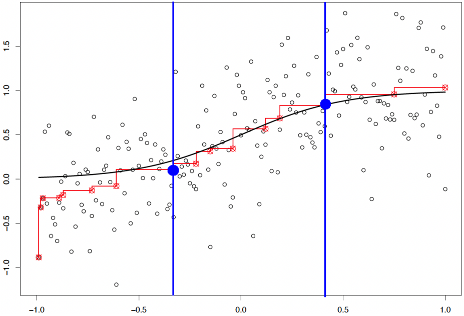
- To derive the rate of convergence of $\tilde{m}_n$, we localise the analysis of $\tilde{f}_n$ to a neighbourhood of m_0.

Subinterval localisation

- To derive the rate of convergence of $\tilde{m}_n$, we localise the analysis of $\tilde{f}_n$ to a neighbourhood of m_0.
- There is a subinterval localisation result for the piecewise constant isotonic LSE $\bar{f}_n$ over the class of increasing functions on $[0, 1]$, based on observations $(x_1, Y_1), \ldots, (x_n, Y_n)$.

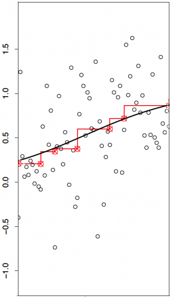
Subinterval localisation

- To derive the rate of convergence of $\tilde{m}_n$, we localise the analysis of $\tilde{f}_n$ to a neighbourhood of m_0.
- There is a subinterval localisation result for the piecewise constant isotonic LSE $\bar{f}_n$ over the class of increasing functions on $[0, 1]$, based on observations $(x_1, Y_1), \ldots, (x_n, Y_n)$.



Subinterval localisation

- To derive the rate of convergence of $\tilde{m}_n$, we localise the analysis of $\tilde{f}_n$ to a neighbourhood of m_0.
- There is a subinterval localisation result for the piecewise constant isotonic LSE $\bar{f}_n$ over the class of increasing functions on $[0, 1]$, based on observations $(x_1, Y_1), \ldots, (x_n, Y_n)$.



Subinterval localisation

- To derive the rate of convergence of $\tilde{m}_n$, we localise the analysis of $\tilde{f}_n$ to a neighbourhood of m_0.
- There is a subinterval localisation result for the piecewise constant isotonic LSE $\bar{f}_n$ over the class of increasing functions on $[0, 1]$, based on observations $(x_1, Y_1), \ldots, (x_n, Y_n)$.

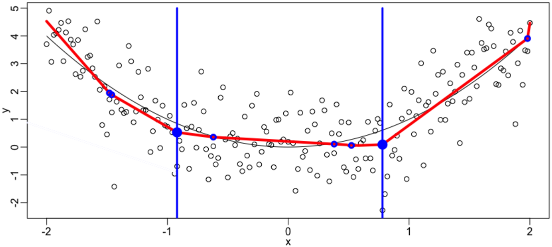
Proposition. Let $x_k < x_\ell$ be knots of $\bar{f}_n$, so that $\bar{f}_n(x_i) < \bar{f}_n(x_{i+1})$ for $i \in \{k, \ell\}$. Denote by $\bar{f}_{(k:\ell]}$ the isotonic LSE based on $\{(x_i, Y_i) : k + 1 \leq i \leq \ell\}$. Then $\bar{f}_n = \bar{f}_{(k:\ell]}$ on $[x_{k+1}, x_\ell]$.

Subinterval localisation

- To derive the rate of convergence of $\tilde{m}_n$, we localise the analysis of $\tilde{f}_n$ to a neighbourhood of m_0.
- There is a subinterval localisation result for the piecewise constant isotonic LSE $\bar{f}_n$ over the class of increasing functions on $[0, 1]$, based on observations $(x_1, Y_1), \ldots, (x_n, Y_n)$.
- However, the analogous property does not hold in general for an (increasing) convex LSE $\breve{f}_n$ if $x_k < x_\ell$ are taken to be kinks of $\breve{f}_n$.

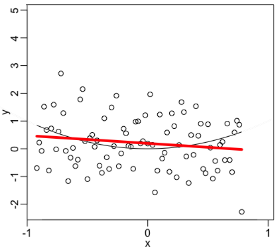
Subinterval localisation

- To derive the rate of convergence of $\tilde{m}_n$, we localise the analysis of $\tilde{f}_n$ to a neighbourhood of m_0.
- There is a subinterval localisation result for the piecewise constant isotonic LSE $\bar{f}_n$ over the class of increasing functions on $[0, 1]$, based on observations $(x_1, Y_1), \ldots, (x_n, Y_n)$.
- However, the analogous property does not hold in general for an (increasing) convex LSE $\check{f}_n$ if $x_k < x_\ell$ are taken to be kinks of $\check{f}_n$.



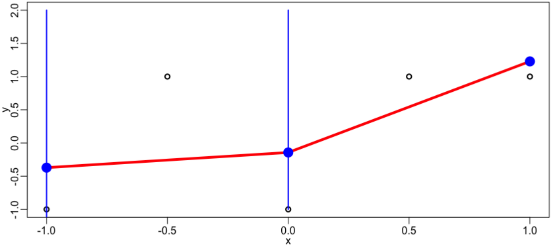
Subinterval localisation

- To derive the rate of convergence of $\tilde{m}_n$, we localise the analysis of $\tilde{f}_n$ to a neighbourhood of m_0.
- There is a subinterval localisation result for the piecewise constant isotonic LSE $\bar{f}_n$ over the class of increasing functions on $[0, 1]$, based on observations $(x_1, Y_1), \ldots, (x_n, Y_n)$.
- However, the analogous property does not hold in general for an (increasing) convex LSE $\check{f}_n$ if $x_k < x_\ell$ are taken to be kinks of $\check{f}_n$.



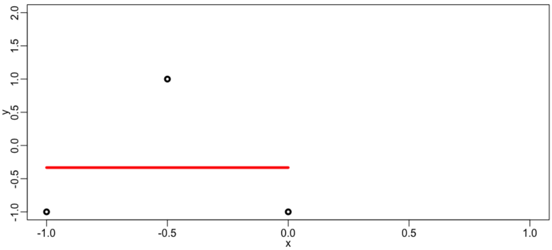
Subinterval localisation

- To derive the rate of convergence of $\tilde{m}_n$, we localise the analysis of $\tilde{f}_n$ to a neighbourhood of m_0.
- There is a subinterval localisation result for the piecewise constant isotonic LSE $\bar{f}_n$ over the class of increasing functions on $[0,1]$, based on observations $(x_1, Y_1), \ldots, (x_n, Y_n)$.
- However, the analogous property does not hold in general for an (increasing) convex LSE $\check{f}_n$ if $x_k < x_\ell$ are taken to be kinks of $\check{f}_n$.

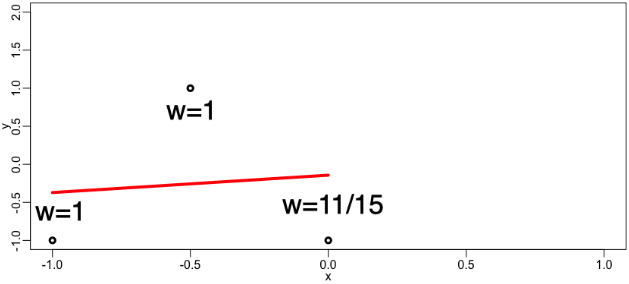


Subinterval localisation

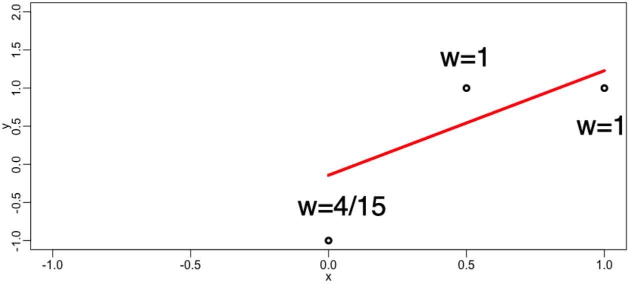
- To derive the rate of convergence of $\tilde{m}_n$, we localise the analysis of $\tilde{f}_n$ to a neighbourhood of m_0.
- There is a subinterval localisation result for the piecewise constant isotonic LSE $\bar{f}_n$ over the class of increasing functions on $[0, 1]$, based on observations $(x_1, Y_1), \ldots, (x_n, Y_n)$.
- However, the analogous property does not hold in general for an (increasing) convex LSE $\check{f}_n$ if $x_k < x_\ell$ are taken to be kinks of $\check{f}_n$.



Remedy: boundary adjustments



Remedy: boundary adjustments



Remedy: boundary adjustments

Proposition. For $m \in [0,1]$, let $\hat{f}_n^m$ be the LSE over $\mathcal{H}^m$ based on $\{(x_i, Y_i) : i \in [n]\}$, and define the residuals $\hat{\varepsilon}_i = Y_i - \hat{f}_n^m(x_i)$ for $i \in [n]$. Then for all $j \in [n]$, the boundary weights

$$\underline{w}_j := -\frac{\sum_{i=1}^{j-1} \hat{\varepsilon}_i}{\hat{\varepsilon}_j} \mathbb{1}_{\{\hat{\varepsilon}_j \neq 0\}}, \quad \overline{w}_j := -\frac{\sum_{i=j+1}^{n} \hat{\varepsilon}_i}{\hat{\varepsilon}_j} \mathbb{1}_{\{\hat{\varepsilon}_j \neq 0\}}$$

lie in $[0,1]$, and $\underline{w}_j + \overline{w}_j = \mathbb{1}_{\{\hat{\varepsilon}_j \neq 0\}}$.

Remedy: boundary adjustments

Proposition. For $m \in [0,1]$, let $\hat{f}_n^m$ be the LSE over $\mathcal{H}^m$ based on $\{(x_i, Y_i) : i \in [n]\}$, and define the residuals $\hat{\varepsilon}_i = Y_i - \hat{f}_n^m(x_i)$ for $i \in [n]$. Then for all $j \in [n]$, the boundary weights

$$\underline{w}_j := -\frac{\sum_{i=1}^{j-1} \hat{\varepsilon}_i}{\hat{\varepsilon}_j} \mathbb{1}_{\{\hat{\varepsilon}_j \neq 0\}}, \quad \overline{w}_j := -\frac{\sum_{i=j+1}^{n} \hat{\varepsilon}_i}{\hat{\varepsilon}_j} \mathbb{1}_{\{\hat{\varepsilon}_j \neq 0\}}$$

lie in $[0,1]$, and $\underline{w}_j + \overline{w}_j = \mathbb{1}_{\{\hat{\varepsilon}_j \neq 0\}}$. Let $x_k < x_\ell$ be knots of $\hat{f}_n^m$ such that $x_{k+1} \leq m \leq x_{\ell-1}$. Then $\hat{f}_n^m$ minimises

$$f \mapsto \sum_{i=1}^{k-1} \big(Y_i - f(x_i)\big)^2 + \underline{w}_k \big(Y_k - f(x_k)\big)^2$$

over all increasing convex $f : [0,1] \to \mathbb{R}$.

Remedy: boundary adjustments

Proposition. For $m \in [0,1]$, let $\hat{f}_n^m$ be the LSE over $\mathcal{H}^m$ based on $\{(x_i, Y_i) : i \in [n]\}$, and define the residuals $\hat{\varepsilon}_i = Y_i - \hat{f}_n^m(x_i)$ for $i \in [n]$. Then for all $j \in [n]$, the boundary weights

$$\underline{w}_j := -\frac{\sum_{i=1}^{j-1} \hat{\varepsilon}_i}{\hat{\varepsilon}_j} \mathbb{1}_{\{\hat{\varepsilon}_j \neq 0\}}, \quad \overline{w}_j := -\frac{\sum_{i=j+1}^{n} \hat{\varepsilon}_i}{\hat{\varepsilon}_j} \mathbb{1}_{\{\hat{\varepsilon}_j \neq 0\}}$$

lie in $[0,1]$, and $\underline{w}_j + \overline{w}_j = \mathbb{1}_{\{\hat{\varepsilon}_j \neq 0\}}$. Let $x_k < x_\ell$ be knots of $\hat{f}_n^m$ such that $x_{k+1} \leq m \leq x_{\ell-1}$. Then $\hat{f}_n^m$ minimises

$$f \mapsto \overline{w}_\ell \big(Y_\ell - f(x_\ell)\big)^2 + \sum_{i=\ell+1}^{n} \big(Y_i - f(x_i)\big)^2$$

over all increasing concave $f \colon [0,1] \to \mathbb{R}$.

Remedy: boundary adjustments

Proposition. For $m \in [0,1]$, let $\hat{f}_n^m$ be the LSE over $\mathcal{H}^m$ based on $\{(x_i, Y_i) : i \in [n]\}$, and define the residuals $\hat{\varepsilon}_i = Y_i - \hat{f}_n^m(x_i)$ for $i \in [n]$. Then for all $j \in [n]$, the boundary weights

$$\underline{w}_j := -\frac{\sum_{i=1}^{j-1} \hat{\varepsilon}_i}{\hat{\varepsilon}_j} \mathbb{1}_{\{\hat{\varepsilon}_j \neq 0\}}, \quad \overline{w}_j := -\frac{\sum_{i=j+1}^{n} \hat{\varepsilon}_i}{\hat{\varepsilon}_j} \mathbb{1}_{\{\hat{\varepsilon}_j \neq 0\}}$$

lie in $[0,1]$, and $\underline{w}_j + \overline{w}_j = \mathbb{1}_{\{\hat{\varepsilon}_j \neq 0\}}$. Let $x_k < x_\ell$ be knots of $\hat{f}_n^m$ such that $x_{k+1} \leq m \leq x_{\ell-1}$. Then $\hat{f}_n^m$ minimises

$$f \mapsto \overline{w}_k \big(Y_k - f(x_k)\big)^2 + \sum_{i=k+1}^{\ell-1} \big(Y_i - f(x_i)\big)^2 + \underline{w}_\ell \big(Y_\ell - f(x_\ell)\big)^2$$

over all $f \in \mathcal{H}^m$.

Outline

1. Introduction to S-shaped regression functions

2. Definitions and theory

3. Computation

`ScanAll`: brute-force method

`ScanSelected`: refined search

`SeqConReg`: sequential procedure

Computation of S-shaped LSEs: `ScanAll`

- Aim: given $(x_1, Y_1), \ldots, (x_n, Y_n) \in [0,1] \times \mathbb{R}$ with $x_1 < \cdots < x_n$, compute the S-shaped LSE $\hat{f}_n$ with minimal inflection point, i.e.

$$\hat{f}_n = \hat{f}_n^{\hat{m}_n}, \quad \text{where} \quad \hat{m}_n = x_{\hat{\jmath}_n} \quad \text{and} \quad \hat{\jmath}_n = \operatorname*{sargmin}_{1 \le j \le n} \mathrm{RSS}_n(\hat{f}_n^{x_j})$$

Computation of S-shaped LSEs: `ScanAll`

- Aim: given $(x_1, Y_1), \ldots, (x_n, Y_n) \in [0,1] \times \mathbb{R}$ with $x_1 < \cdots < x_n$, compute the S-shaped LSE $\hat{f}_n$ with minimal inflection point, i.e.

$$\hat{f}_n = \hat{f}_n^{\hat{m}_n}, \quad \text{where} \quad \hat{m}_n = x_{\hat{\jmath}_n} \quad \text{and} \quad \hat{\jmath}_n = \operatorname*{sargmin}_{1 \le j \le n} \mathrm{RSS}_n(\hat{f}_n^{x_j})$$

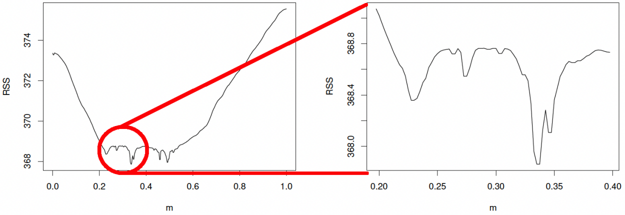
- Naive brute-force approach (`ScanAll`): compute $\hat{f}_n^m$ for *every* $m \in \{x_1, \ldots, x_n\}$, e.g. by an active set or support reduction algorithm (Dümbgen et al., 2007, Groeneboom et al., 2008)

Computation of S-shaped LSEs: `ScanAll`

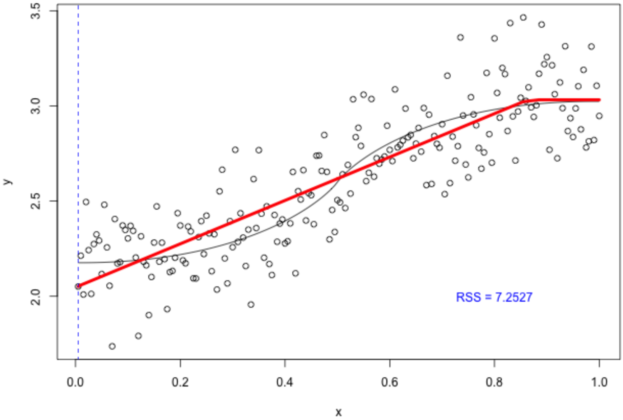
- Aim: given $(x_1, Y_1), \ldots, (x_n, Y_n) \in [0,1] \times \mathbb{R}$ with $x_1 < \cdots < x_n$, compute the S-shaped LSE $\hat{f}_n$ with minimal inflection point, i.e.

$$\hat{f}_n = \hat{f}_n^{\hat{m}_n}, \quad \text{where} \quad \hat{m}_n = x_{\hat{\jmath}_n} \quad \text{and} \quad \hat{\jmath}_n = \operatorname*{sargmin}_{1 \le j \le n} \mathrm{RSS}_n(\hat{f}_n^{x_j})$$

- Naive brute-force approach (`ScanAll`): compute $\hat{f}_n^m$ for *every* $m \in \{x_1, \ldots, x_n\}$, e.g. by an active set or support reduction algorithm (Dümbgen et al., 2007, Groeneboom et al., 2008)
  - $j \mapsto \mathrm{RSS}_n(f_n^{x_j})$ may have multiple local minima



ScanAll: **example**



Subinterval localisation

- For $j \in \{1, \ldots, n\}$, form $\hat{h}_n^j$ by concatenating the increasing convex LSE $\hat{f}_{1,j}$ on $\{(x_i, Y_i) : 1 \leq i \leq j\}$ and the increasing concave LSE $\hat{f}_{n,j+1}$ on $\{(x_i, Y_i) : j+1 \leq i \leq n\}$

Subinterval localisation

- For $j \in \{1, \ldots, n\}$, form $\hat{h}_n^j$ by concatenating the increasing convex LSE $\hat{f}_{1,j}$ on $\{(x_i, Y_i) : 1 \leq i \leq j\}$ and the increasing concave LSE $\hat{f}_{n,j+1}$ on $\{(x_i, Y_i) : j+1 \leq i \leq n\}$

- $\hat{h}_n^j$ may not lie in $\mathcal{F}^{x_j}$, in which case $\hat{f}_n^{x_j} \neq \hat{h}_n^j$; nevertheless, we have the following two important facts:

Subinterval localisation

- For $j \in \{1, \ldots, n\}$, form $\hat{h}_n^j$ by concatenating the increasing convex LSE $\hat{f}_{1,j}$ on $\{(x_i, Y_i) : 1 \leq i \leq j\}$ and the increasing concave LSE $\hat{f}_{n,j+1}$ on $\{(x_i, Y_i) : j+1 \leq i \leq n\}$

- $\hat{h}_n^j$ may not lie in $\mathcal{F}^{x_j}$, in which case $\hat{f}_n^{x_j} \neq \hat{h}_n^j$; nevertheless, we have the following two important facts:

A. If $\hat{h}_n^j \in \mathcal{F}^{x_j}$, then $\hat{f}_n^{x_j} = \hat{h}_n^j$ (LSE over a larger set)

B. Subinterval localisation: Given any LSE $\tilde{f}_n$ over $\mathcal{F}$, if either x_j is its smallest inflection point or x_{j+1} is its largest inflection point, then $\tilde{f}_n = \hat{h}_n^j$ (with weights $\overline{w}_j = 0 = 1 - \underline{w}_j$) and $Y_j \leq \tilde{f}_n(x_j) \leq \tilde{f}_n(x_{j+1}) \leq Y_{j+1}$

Refined strategy: `ScanSelected`

- Computational gains based on subinterval localisation: to locate $\hat{\jmath}_n = \text{sargmin}_{1 \leq j \leq n} \mathrm{RSS}_n(\hat{f}_n^{x_j})$,

1. We can refine the search by discarding those j for which $Y_j > Y_{j+1}$: only $n - 1$ pairwise comparisons required
2. For the remaining indices j, compute $\hat{h}_n^j$ by fitting $\hat{f}_{1,j}$ and $\hat{f}_{n,j+1}$ separately on disjoint subsets of the original data

Refined strategy: `ScanSelected`

- Computational gains based on subinterval localisation: to locate $\hat{\jmath}_n = \mathrm{sargmin}_{1 \le j \le n} \mathrm{RSS}_n(\hat{f}_n^{x_j})$,

1. We can refine the search by discarding those j for which $Y_j > Y_{j+1}$: only $n-1$ pairwise comparisons required
2. For the remaining indices j, compute $\hat{h}_n^j$ by fitting $\hat{f}_{1,j}$ and $\hat{f}_{n,j+1}$ separately on disjoint subsets of the original data

- Let $\mathcal{J}$ be the set of j for which $\hat{h}_n^j$ is an S-shaped function in $\mathcal{F}^{x_j}$. Then $\hat{f}_n^{x_j} = \hat{h}_n^j$ for all $j \in \mathcal{J}$ by Fact **A** and $\hat{\jmath}_n \in \mathcal{J}$ by Fact **B**, so

$$\hat{\jmath}_n = \operatorname*{sargmin}_{j \in \mathcal{J}} \mathrm{RSS}_n(\hat{h}_n^j)$$

Refined strategy: `ScanSelected`

- Computational gains based on subinterval localisation: to locate $\hat{\jmath}_n = \text{sargmin}_{1 \le j \le n} \mathrm{RSS}_n(\hat{f}_n^{x_j})$,

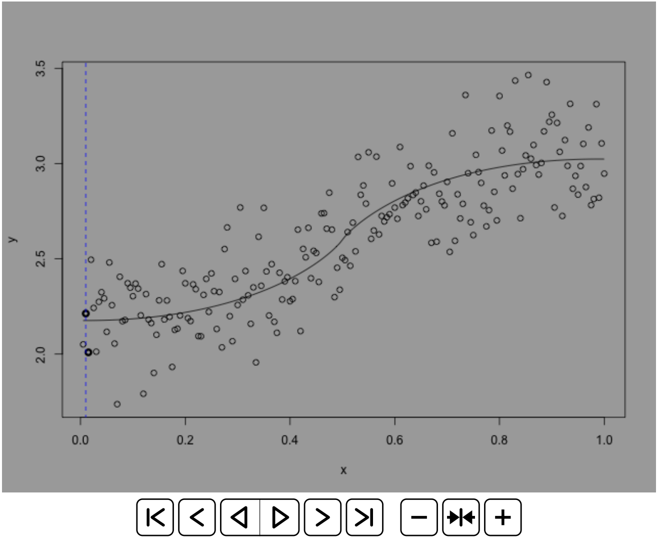
  1. We can refine the search by discarding those j for which $Y_j > Y_{j+1}$: only $n-1$ pairwise comparisons required
  2. For the remaining indices j, compute $\hat{h}_n^j$ by fitting $\hat{f}_{1,j}$ and $\hat{f}_{n,j+1}$ separately on disjoint subsets of the original data

- Let $\mathcal{J}$ be the set of j for which $\hat{h}_n^j$ is an S-shaped function in $\mathcal{F}^{x_j}$. Then $\hat{f}_n^{x_j} = \hat{h}_n^j$ for all $j \in \mathcal{J}$ by Fact **A** and $\hat{\jmath}_n \in \mathcal{J}$ by Fact **B**, so

$$\hat{\jmath}_n = \operatorname*{sargmin}_{j \in \mathcal{J}} \mathrm{RSS}_n(\hat{h}_n^j)$$

- Can we use $\hat{f}_{1,j-1}$ as a warm start for computing $\hat{f}_{1,j}$?

ScanSelected: **example**



Sequential procedure: `SeqConReg`

- More efficient approach: we reveal new observations one by one, and update the increasing convex and increasing concave least squares fits using a mixed primal-dual bases algorithm (Fraser and Massam, 1989, Meyer, 1999).

Sequential procedure: SeqConReg

- More efficient approach: we reveal new observations one by one, and update the increasing convex and increasing concave least squares fits using a mixed primal-dual bases algorithm (Fraser and Massam, 1989, Meyer, 1999).

- Given $\hat{f}_{1,j-1}$ and a new observation (x_j, Y_j), note that if $Y_j \geq \hat{f}_{1,j-1}(x_j)$, then $\hat{f}_{1,j} = \hat{f}_{1,j-1}$ on $\{x_1, \ldots, x_j\}$.

Sequential procedure: `SeqConReg`

- More efficient approach: we reveal new observations one by one, and update the increasing convex and increasing concave least squares fits using a mixed primal-dual bases algorithm (Fraser and Massam, 1989, Meyer, 1999).

- Given $\hat{f}_{1,j-1}$ and a new observation (x_j, Y_j), note that if $Y_j \geq \hat{f}_{1,j-1}(x_j)$, then $\hat{f}_{1,j} = \hat{f}_{1,j-1}$ on $\{x_1, \ldots, x_j\}$.

- Otherwise, if $Y_j < \hat{f}_{1,j-1}(x_j)$, then start with the LSE $\hat{f}_{1,j-1}$ based on $(x_1, Y_1), \ldots, (x_{j-1}, Y_{j-1}), (x_j, \tilde{Y}_j \equiv \hat{f}_{1,j-1}(x_j))$, and decrease the value of $\tilde{Y}_j$.

Sequential procedure: SeqConReg

- More efficient approach: we reveal new observations one by one, and update the increasing convex and increasing concave least squares fits using a mixed primal-dual bases algorithm (Fraser and Massam, 1989, Meyer, 1999).

- Given $\hat{f}_{1,j-1}$ and a new observation (x_j, Y_j), note that if $Y_j \geq \hat{f}_{1,j-1}(x_j)$, then $\hat{f}_{1,j} = \hat{f}_{1,j-1}$ on $\{x_1, \ldots, x_j\}$.

- Otherwise, if $Y_j < \hat{f}_{1,j-1}(x_j)$, then start with the LSE $\hat{f}_{1,j-1}$ based on $(x_1, Y_1), \ldots, (x_{j-1}, Y_{j-1}), (x_j, \tilde{Y}_j \equiv \hat{f}_{1,j-1}(x_j))$, and decrease the value of $\tilde{Y}_j$.

- The LSE fit is a piecewise linear function of $\tilde{Y}_j$; we need to track the changes in the 'active' set of kinks.

Summary of algorithm (R package Sshaped)

1. Discard all $j \in \{1, \ldots, n\}$ for which $Y_j > Y_{j+1}$.
2. For each of the remaining j, use SeqConReg to compute the increasing convex LSE $\hat{f}_{1,j}$ and increasing concave LSE $\hat{f}_{n,j+1}$. Concatenate these by linear interpolation to form $\hat{h}_n^j$. Discard j if $\hat{h}_n^j \notin \mathcal{F}^{x_j}$, i.e. if

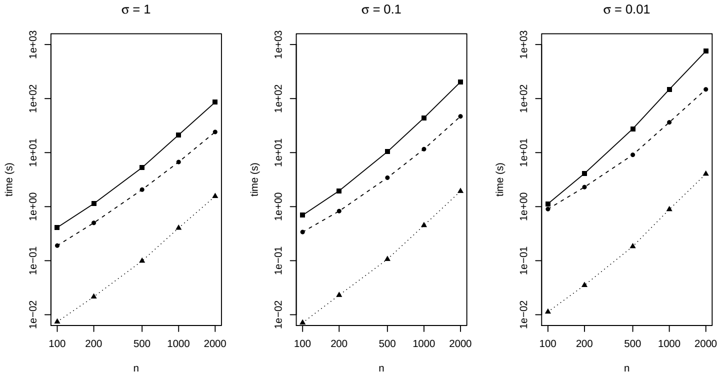
$$\frac{\hat{f}_{n,j+1}(x_{j+2}) - \hat{f}_{n,j+1}(x_{j+1})}{x_{j+2} - x_{j+1}} > \frac{\hat{f}_{n,j+1}(x_{j+1}) - \hat{f}_{1,j}(x_j)}{x_{j+1} - x_j}.$$

3. Let $\mathcal{J}$ be the set of indices j retained after Steps 1 and 2. Find

$$\tilde{\jmath} = \operatorname*{sargmin}_{j \in \mathcal{J}} \mathrm{RSS}_n(\hat{h}_n^j).$$

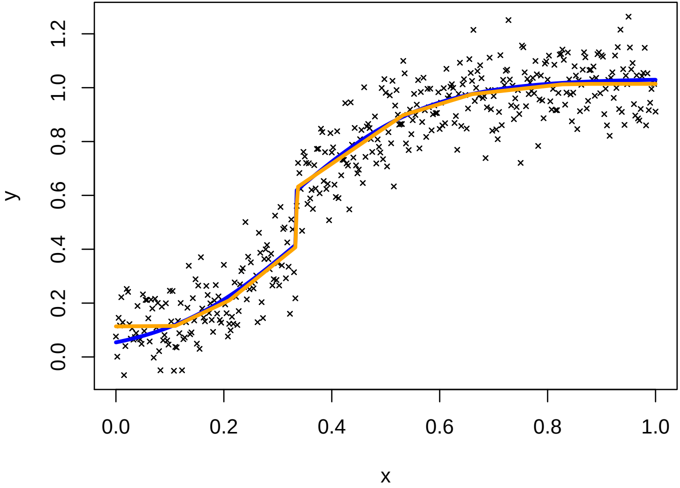
   Return $(x_{\tilde{\jmath}}, \hat{h}_n^{\tilde{\jmath}}) = (\hat{m}_n, \hat{f}_n)$.

Running time comparison

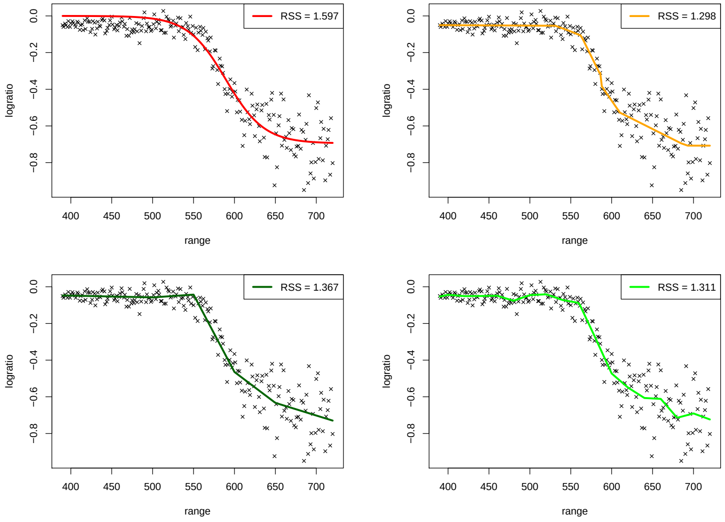


Log-log plots of the running time (in seconds) of the SeqConReg (▲), ScanSelected (●) and ScanAll (■) algorithms, for sample sizes $n \in \{100, 200, 500, 1000\}$ and noise levels $\sigma \in \{1, 0.1, 0.01\}$

Example: simulated data



Example: LIDAR air pollution data



Thank you for listening!

References

Dümbgen, L., Hüsler, A. and Rufibach, K. (2007). Active set and EM algorithms for log-concave densities based on complete and censored data. Available at `https://arxiv.org/abs/0707.4643v4`.

Feng, O. Y., Chen, Y., Han, Q., Carroll, R. J. and Samworth, R. J. (2022). Nonparametric, tuning-free estimation of S-shaped functions. *J. Roy. Statist. Soc., Ser. B, to appear*.

Feng, O. Y., Chen, Y., Han, Q., Carroll, R. J. and Samworth, R. J. (2021). `Sshaped`: Estimation of an S-shaped function. `R` package version 0.99. Available at `https://cran.r-project.org/web/packages/SShaped`.

Fraser, D. A. S. and Massam, H. (1989). A mixed primal-dual bases algorithm for regression under inequality constraints. Application to concave regression. *Scand. J. Statist.*, **16**, 65–74.

Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2008). The support reduction algorithm for computing non-parametric function estimates in mixture models. *Scand. J. Statist.*, **35**, 385–399.

Meyer, M. (1999). An extension of the mixed primal-dual bases algorithm to the case of more constraints than dimensions. *J. Statist. Plann. Inference*, **81**, 13–31.