Ordered monotone regression

Kaspar Rufibach

Biostatistics Unit
Institute of Social and Preventive Medicine
University of Zurich

9th German Open Conference on Probability and Statistics
March 2, 2010

Joint work with Fadoua Balabdaoui and Filippo Santambrogio
(both Paris-Dauphiné)

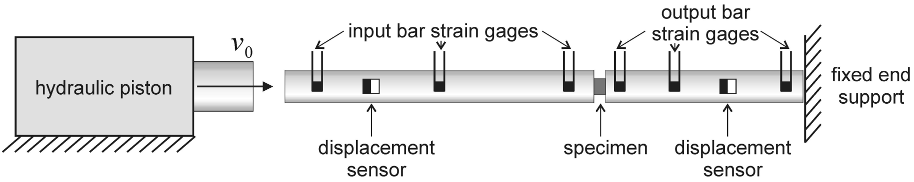
Agenda

1. Motivating example
2. Isotonic regression
3. Ordered monotone regression
4. Algorithms
5. Smooth estimates
6. Back to stress-strain example

Motivating example from mechanical engineering

Dynamic material tests, Shim and Mohr (2009).

Determine deformation resistance (strength of material) from uniaxial compression tests at different loading velocities.



Motivating example from mechanical engineering

Yields stress-strain curves.

x-axis: strain,

- force applied to experimental unit via piston,
- negative logarithm of ratio of current over initial piston stroke,
- maximum shortening of piston stroke: 63% $\Rightarrow$ x-values lie in $[0, 1]$.
  $-\log(1) = 0$, $-\log(1 - 0.63) = 1$.
- x-values in fact measured with error $\Rightarrow$ ignored.

y-axis: stress,

- measured compression of the sample,
- computed from cross-sectional area of specimen.

Motivating example from mechanical engineering

Substantial noise due to

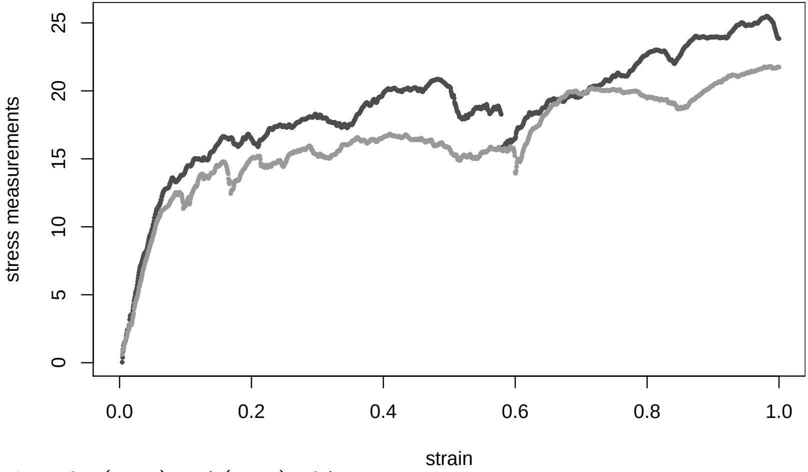
- variability in adjusting loading velocity and
- electrical noise in data acquisition.

Stress-strain curves at different loading velocities.

Regression curves are expected to be

1. monotone increasing $\Rightarrow$ stress $=$ increasing function of strain,
2. ordered $\Rightarrow$ deformation resistance is higher for lower loading velocity.

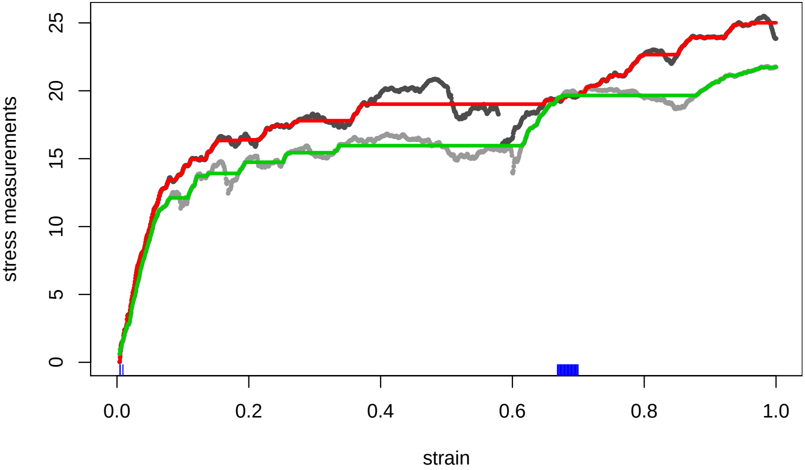
Motivating example from mechanical engineering



1495 pairs (x_i, y_i) and (x_i, z_i) with

- x_i: invoked strain,
- y_i, z_i: stress measurements for two loading velocities.

Naive approach



Two naive approaches to "estimation":

- "Trial and error" with "parametric" functions ⇒ original approach!
- Two independent monotone functions ⇒ ordering not guaranteed.

Problem

For data (x_i, y_i) and (x_i, z_i) find the minimizer of

$$L_2(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{n} w_{1,i}(y_i - a_i)^2 + \sum_{i=1}^{n} w_{2,i}(z_i - b_i)^2$$

over $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^n$ such that

- $\mathbf{a}$ and $\mathbf{b}$ are increasing and
- $\mathbf{a} \leq \mathbf{b}$, with $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}_+^n$ given vectors of weights.

We provide:

- Existence, Uniqueness.
- Characterization.
- Computation.

In addition (see paper): Formal proof of PAVA validity for functionals satisfying Cauchy mean value property in one-curve problem.

A simpler problem

Minimize

$$L_1(\mathbf{a}) \quad = \quad \sum_{i=1}^{n} w_i(y_i - a_i)^2$$

over $\mathbf{a} \in \mathbb{R}^n$ s.t. $a_1 \leq a_2 \leq \cdots \leq a_n$ where $w_i > 0$.

Identify any vector $\mathbf{v} \in \mathbb{R}^n$ with cumsum function $Hy : [0, n] \rightarrow \mathbb{R}$:

- $Hy(0) = 0$,
- $Hy(k) = \sum_{i=1}^{k} y_i, \;\; k = 1, \ldots, n$,
- and linear on the intervals $[0, 1], \ldots, [n-1, n]$.

Theorem (GCM characterization)

A vector $\widehat{\mathbf{a}}^*$ minimizes L_1 if and only if

$$\begin{aligned} H\widehat{\mathbf{a}}^*(n) &= H\mathbf{y}(n), \\ H\widehat{\mathbf{a}}^* &\leq H\mathbf{y} \text{ on } [0, n], \\ H\widehat{\mathbf{a}}^*(t) &= H\mathbf{y}(t) \text{ at knot points } t \in]0, n[\text{ of } H\widehat{\mathbf{a}}^*. \end{aligned}$$

min-max representation and generalizations

Equivalent representation for least-squares problem:

$$\widehat{a}_i^* \quad = \quad \max_{s \leq i} \min_{t \geq i} M(\{s, \ldots, t\})$$

where $M(\{s, \ldots, t\}) = \sum_{i=s}^{t} y_i w_i / \sum_{i=s}^{t} w_i$.

- Used to derive theoretical results, not for computation.
- Valid for functionals M that satisfy Cauchy mean value property.
- Computation via PAVA.

Two-curve problem: obvious questions

- Geometrical characterizations using GCMs?
- PAVA-like algorithm?
- min-max representation (that only depends on data)?

Two-curve problem: obvious questions

- Geometrical characterizations using GCMs? Not found.
- PAVA-like algorithm? Not found.
- min-max representation (that only depends on data)? Partially.

Two-curve problem: obvious questions

- Geometrical characterizations using GCMs? Not found.
- PAVA-like algorithm? Not found.
- min-max representation (that only depends on data)? Partially.

Alternatives:

- Analytical characterizations.
- Projected subgradient algorithm (general), Dykstra's algorithm (least squares).
- First value of lower and last of upper curve.

$$L_2(a,b) \;=\; \sum_{i=1}^n w_{1,i}(y_i - a_i)^2 + \sum_{i=1}^n w_{2,i}(z_i - b_i)^2$$

Theorem (Balabdaoui, R., and Santambrogio, 2010)

- *Minimizer* $(\widehat{\mathbf{a}}^*, \widehat{\mathbf{b}}^*)$ *exists and is unique.*
- *Necessary and sufficient characterization can be given (using directional derivatives).*
- *We have the following representations:*

$$\begin{aligned}\widehat{a}_i^* &= \max_{s \le i} \min_{t \ge i} \left(M_1(\{s,\ldots,t\}) \wedge \widehat{b}_s^*\right)\\ \widehat{b}_i^* &= \max_{s \le i} \min_{t \ge i} \left(M_2(\{s,\ldots,t\}) \vee \widehat{a}_t^*\right)\end{aligned}$$

for $i = 1, \ldots, n$, *where* M_1 *and* M_2 *are the weighted mean functionals.*

Problem with min-max representation:

- Expression for $\widehat{a}_i^*$ depends on $\widehat{\mathbf{b}}^*$ and vice versa.
- Idea for algorithm: iterate above min-max representations. Does not work!

Dykstra's algorithm

Problem of minimizing

$$L_2(a,b) \quad = \quad \sum_{i=1}^{n} w_{1,i}(y_i - a_i)^2 + \sum_{i=1}^{n} w_{2,i}(z_i - b_i)^2$$

over $(a,b) \in \mathbb{R}^n \times \mathbb{R}^n$ s.t. a and b are increasing and $a \leq b$

is equivalent

to least squares projection of $(\mathbf{y}, \mathbf{z})$ on intersection of convex cones

$$\{(a,b) : a \text{ is increasing}\}, \ \{(a,b) : b \text{ is increasing}\}, \ \text{and } \{(a,b) : a \leq b\}.$$

What the doctor ordered: algorithm by Dykstra (1983).
Simply iterate projections.

Dykstra's algorithm

$$
\begin{aligned}
C_1 &:= \{(a,b) : a \text{ is increasing}\}, \\
C_2 &:= \{(a,b) : b \text{ is increasing}\}, \\
C_3 &:= \{(a,b) : a \le b\}.
\end{aligned}
$$

Computation of projections:

- C_1, C_2: PAVA,
- C_3: replace each pair (a_i, b_i) violating constraint (i.e. $a_i > b_i$) by weighted average $(w_{1,i}a_i + w_{2,i}b_i)/(w_{1,i} + w_{2,i})$.

Dykstra's algorithm:

- Often converges slow $\Rightarrow$ here very fast.
- Only three projections involved (that are easy to compute).
- Implemented in `R` package `OrdMonReg`.

Projected subgradient algorithm

Dykstra's algorithm: only applicable to functionals of type

$$L_2(a, b) \quad = \quad \sum_{i=1}^n w_{1,i}(y_i - a_i)^2 + \sum_{i=1}^n w_{2,i}(z_i - b_i)^2.$$

Projected subgradient algorithm: can handle

$$L_3(a, b) \quad = \quad F_{\mathbf{y},\mathbf{w}_1}(\mathbf{a}) + \sum_{i=1}^n w_{2,i}(z_i - b_i)^2$$

where F is convex and differentiable (as a function of $\mathbf{a}$).

Second term can likely be generalized as well.

Derivation and implementation more involved.
Computes solution in least squares case slower.

A smoothed version of the estimates

In fact: estimates only determined at x_i's.

One-curve problem: estimate step function $\Rightarrow$ derivative of GCM of cumsum. We define estimates in two-curve as step functions as well $\Rightarrow$ non-smooth.

Define smooth estimates for some kernel K_h:

$$\tilde{a}_h^*(x) = \frac{\sum_{i=1}^n K_h(x - x_i)\widehat{a}_i^*}{\sum_{i=1}^n K_h(x - x_i)} \qquad \tilde{b}_h^*(x) = \frac{\sum_{i=1}^n K_h(x - x_i)\widehat{b}_i^*}{\sum_{i=1}^n K_h(x - x_i)}.$$

Theorem (Mukerjee, 1988)

A monotone function remains monotone after kernel smoothing if the kernel is log-concave.

Proof: Log-concave densities have monotone likelihood ratio.

Ordering: maintained if we use same kernel and bandwidth for both functions $\Rightarrow$ smoothing is simply taking weighted means.

Back to our motivating example

1495 pairs (x_i, y_i), (x_i, z_i).

y_i, z_i: stress results for two different loading velocities.

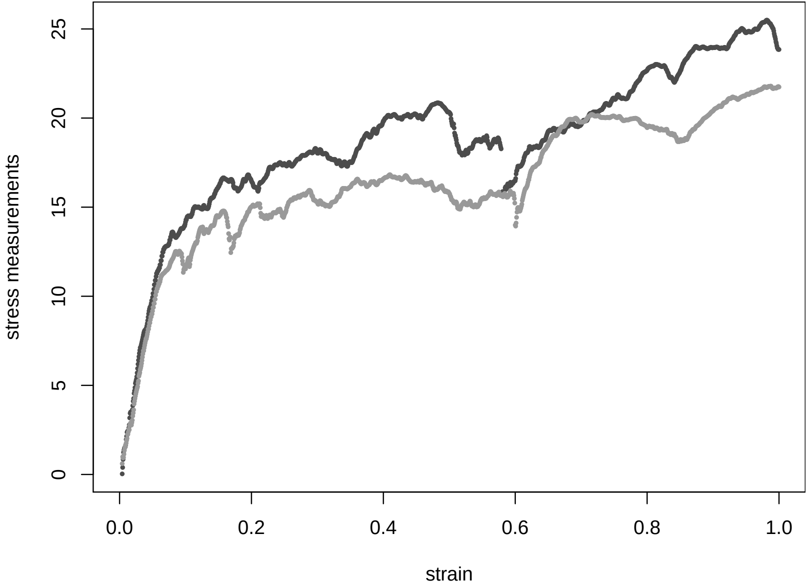
Estimated curves have to be isotonic, ordered, and ideally smooth.

Practitioners: fit "parametric" models using trial-and-error until above properties are met $\Rightarrow$ ad-hoc, time-consuming.

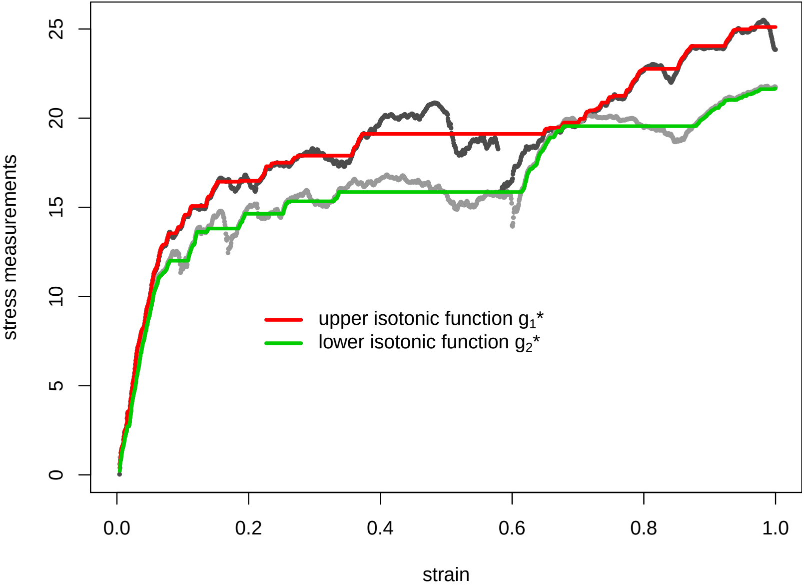
We use normal kernel $K_h(x) = \phi(x/h) \Rightarrow$ log-concave.

Bandwidth: $h = 0.1n^{-1/5} = 0.023$.

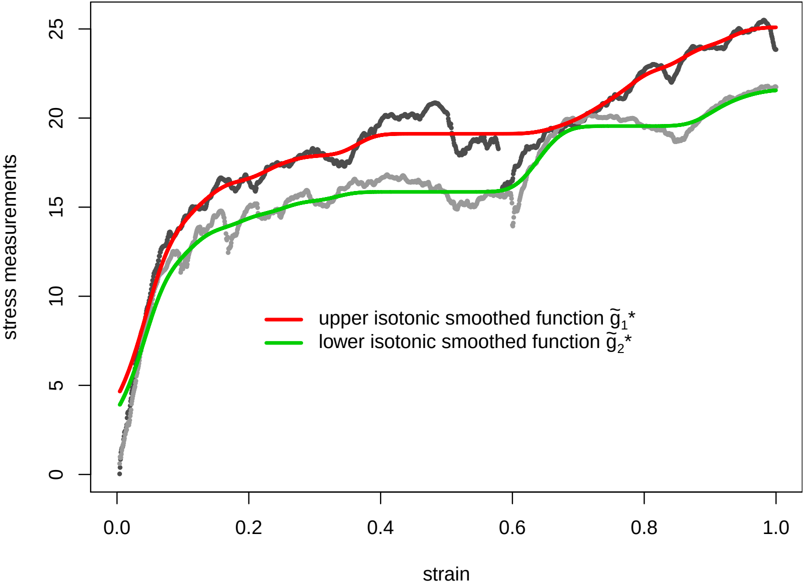
Example: data



Example: ordered monotone estimates



Example: ordered monotone smoothed estimates



Summary and outlook

Summary:

- Existence, uniqueness, and characterization of ordered monotone regression.
- Dykstra's and projected subgradient algorithm.
- Smoothed estimates that maintain ordering and monotonicity.
- Algorithms for some generalized one- and two-curve problem implemented in `R` package `OrdMonReg`.

Outlook:

- Initial problem of practitioners was solved.
- Now: estimation, then smoothing. Penalize (non-)smoothness as well?
- Asymptotics: presumably even more difficult than in one-curve problem.
- Further applications?
- More than two curves algorithm by Beran & Dümbgen (2010)?

Thank you for your attention.