Multiscale Change Point Segmentation

Axel Munk\textsuperscript{1,2}

\textsuperscript{1} Faculty for Mathematics and Computer Science, Georg-August University Göttingen, Germany
\textsuperscript{2} Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany

Merle Behr, UGöttingen
Klaus Frick, Buchs NTB, CH
Thomas Hotz, TU Illmenau
Housen Li, MPI bpc
Florian Pein, UGöttingen
Hannes Sieling, Blue Yonder Big Data Analytics, Hamburg

www.stochastik.math.uni-goettingen.de/munk
I. The Regressogram/Change Point-Problem

Applications:

quality control, electrical engineering, signal processing, quantum optics, financial econometrics, genetics, ...
I. The Regressogram/Change Point-Problem

Membrane biophysics:

(A) 0.3ms of current recordings of a phospholipid bilayer containing recombined protein Tim23 excited at 160mV, sampled at 50kHz (15K data), Meinecke lab Med.Dep. Göttingen

(B) Time trace of one coordinate of an atom of MD T4 lysozyme, de Groot lab, MPIbpc
I. The Regressogram/Change Point-Problem

Statistical Methods:

Wavelet based (multiscale): Antoniadis/Gijbels’02 (JNPS), Donoho/Johnstone’04 (Biometrika), Fryzlewicz/Nason/von Sachs’04,07,08, Kolazły/Knowak’05 (Biometrika), Killick et al’13 (EJS), ...

Kernel based: Müller’92 (AoS), …, Arlot et al.’12 (arXiv), ...

Aggregation: Rigollet/Tsybakov’12 (Stat. Science)

Bayesian approaches: Yao’84 (AoS), Chib’98 (JoE), Barry/Hartigan’93 (JASA), Green’95 (Biom.), Ghoasl et al.’99 (AISM), Fearnhead’06 (SC), Luong et al’12 (arXiv), Du/Kou’15 (JASA), ...

(Penalized) maximum likelihood: Hinkley’70 (Biometrika), Braun/Braun/Müller’00 (Biometrika), Au, Yao’89 (Sankhya), Birge/Massart’01 (JEMS), Zhang/Siegund’07 (Biometrics), …, Boysen et al.’09 (AoS), Harachou/Levy-Leduc’10 (JASA), …

Time series: Bai/Perron’98 (Econometrika), Yao’93 (Biometrika), Lavielle/Moulines’00 (JTSA), Huskova/Antoch’03 (TMMP), Mercurio/Spokoiny’04 (AoS), Preuß et al.’14 (JASA), ...

HMM/State space: Fearnhead/Clifford’03 (JRSS-B), Fuh’04 (AoS), Cappe et al.’05, …

Distributional changes, Online prediction, sequentially, optimal stopping, ...

Monographs: Ibragimov/Kashinskii’81, Baseville/Nikivorov’93, Carlstein et al.’94, Csorgő/Horvath’97, Chen/Gupta’00, Korostelev/Korosteleva’11, …
I. The Regressogram/Change Point-Problem

"segmentation of time series": ~36 millions hits

Statistically vs Computational efficiency

fast (local) search/segmentation methods:
CBS, Ohlsen et al.’04, (Biostatistics), Venkatraman/Ohlsen’07 (Bioinformatics), PELT, Killick et al. 12/14 (JASA, JSS)
I. The Regressogram/Change Point-Problem

Hybrid approach:

Statistical Multiscale Change Point Estimation (SMUCE):

- make inference on number of segments
  
  We aim for statements like:
  "With 90% prob. the number of selected segments is correct"

- multiscale estimation/detection

- simultaneous (honest/uniform) confidence statements on jump locations/size/signal

- computationally fast
I. A Gentle Introduction to Statistical Multiscale Changepoint Estimation (SMUCE)
Candidate function (MLE, \#jumps=4)

Example: Gaussian white noise, variance = 1
Marginal residuals fit well...

Although **global residuals** look normal, **local residual patterns** indicate that the candidate is not a reasonable solution.
Small scale scanning
scale size = 10

local t-test:
H: residual signal = 0
on scale 10
Small scale scanning
scale size = 10

Violators: local t-test rejections
residual signal on scale 10 not zero
Local t-test on scale 100

Large scale scanning
scale size = 100
Violators: local t-test rejections
residual signal on scale 100 not zero

Large scale scanning
scale size = 100
Candidate function (MLE, #jumps = 4)

Statistical **Multiscale Testing**
Scan residuals at **all** scales (= interval lengths) **simultaneously** (= multiplicity issue)
Candidate function (MLE, #jumps = 5)
Candidate function (MLE, #jumps = 6)
Candidate function (MLE, #jumps = 7)
Candidate function
(MLE, \#jumps = 8)

Statistical **Multiscale** Scanning
selects final candidate which does not violate
any local t-test on each scale
Candidate function (MLE, \#jumps = 8)

Statistical **Multiscale** Scanning selects final candidate which does not violate any local t-test on each scale

**Issues:**
- Multiplicity? (Finite sample) error control?
- How to calibrate scales?
- How to get candidate functions?
II. SMUCE: Statistical Multiscale Change Point Estimator

Combine two different routes

**Estimation:** Modify LSE according to SMSC constraint

Statistical **multiscale shape constraint** for model selection/detection: **Testing** and **confidence set**
II. Multiscale Testing in Change Point Regression
Some Terminology

Data model: \( Y_i \sim EF(\vartheta(i/n)) \)

here: \( Y_i = \vartheta(i/n) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2) \) i.i.d.

Regression function

\[
\vartheta \in S := \{ f : [0, 1] \to \mathbb{R} : f \text{ right cont., locally constant, } k \text{ jumps } , k \in \mathbb{N} \} \\
\text{Jump Space}
\]

Set of discontinuities (jumps): \( J(\vartheta) := \{ t \in [0, 1) : \vartheta(t_-) \neq \vartheta(t_+) \} \).

Number of jumps: \( K = \#J(\vartheta) \)
The Local Likelihood Multiscale Constraint: *Gauss*

For a given parameter $\theta_0 \in \Theta$ and an interval $I = \{i, \ldots, j\}$ with length *(scale)* $j - i + 1$ let the *local likelihood-ratio statistic* (Siegmund/Yakir’00)

$$T_i^j(Y, \theta_0) = (j - i + 1)^{-1} \left( \sum_{l=i}^{j} Y_l - \theta_0 \right)^2$$
From Local to **Multiscale** Constraint

For a given parameter $\theta_0 \in \Theta$ and an interval $I = \{i, \ldots, j\}$ with length (scale) $j - i + 1$ let the *local likelihood-ratio statistic* (Siegmund/Yakir’00)

$$T^j_i (Y, \theta_0) = (j - i + 1)^{-1} \left( \sum_{l=i}^{j} Y_l - \theta_0 \right)^2$$

As a goodness of fit measure for a given candidate $\vartheta \in S$ we employ the *scale calibrated log-likelihood-ratio multiscale statistic* on the system of intervals where $\vartheta$ is constant

$$T_n (Y, \vartheta) = \max_{1 \leq i < j \leq n \atop \vartheta \equiv \theta_{[i,j]} \text{ on } [i/n, j/n]} \sqrt{2T^j_i (Y, \theta_{[i,j]})} - \sqrt{2\log \frac{en}{j - i + 1}}.$$
From Local to **Multiscale** Constraint

For a given parameter $\theta_0 \in \Theta$ and an interval $I = \{i, \ldots, j\}$ with length (scale) $j - i + 1$ let the *local likelihood-ratio statistic* (Siegmund/Yakir’00)

$$T_i^j(Y, \theta_0) = (j - i + 1)^{-1} \left( \sum_{l=i}^{j} Y_l - \theta_0 \right)^2$$

As a goodness of fit measure for a given candidate $\vartheta \in S$ we employ the *scale calibrated log-likelihood-ratio multiscale statistic* on the system of intervals where $\vartheta$ is constant

$$T_n(Y, \vartheta) = \max_{1 \leq i < j \leq n, \vartheta \equiv \theta_{[i,j]} \text{ on } [i/n, j/n]} \sqrt{2T_i^j(Y, \theta_{[i,j]})} - \sqrt{2 \log \frac{en}{j - i + 1}}.$$
$T_n = \max_{1 \leq i \leq j \leq n} \left\{ \frac{|\sum_{l=i}^{j} \epsilon_l|}{\sqrt{j-i+1}} - \sqrt{2 \log \frac{e n}{j-i+1}} \right\}$

$\ln 1000 \sim 7$ (Kabluchko, M.’09, ESAIMProb.Stat.)

$n = 1000$, $\epsilon_i$ standard normal.; rel. frequency of scales (intervals) exceeding fixed threshold (90% quantile of limit distribution of $T_n$)
III. SMUCE

Statistical Multiscale Change Point Estimator

Shape Constraint Estimation and Confidence Sets
Step 1: Estimate "model dimension" $K$

Solve the (nonconvex) optimization problem

$$\inf_{\vartheta \in S} \# J(\vartheta) \quad \text{s.t.} \quad T_n(Y, \vartheta) \leq q \quad \text{(MJ)}$$

Minimizes number of jumps $K = \# J(\vartheta)$

Sparsity enforcing $K = \ell_0(\vartheta)$

Multiscale shape constraint:

Fluctuation control over local residuals (nonconvex)

Minimal number of jumps $\hat{K}(q)$ s.t.

multiresolution constraint (MJ) is valid

Related estimators: Boysen et al.’09 (AoS), Davies et al.’12 (CSDA)
Step II: Signal, Jump Locations and Confidence Set

Solve the (nonconvex) optimization problem

$$\inf_{\vartheta \in S} \# J(\vartheta) \quad \text{s.t.} \quad T_n(Y, \vartheta) \leq q \quad (MJ)$$

- Estimated number of change-points: Minimizer $\hat{K}(q)$ of (MJ)
Solve the (nonconvex) optimization problem

\[
\inf_{\vartheta \in S} \# J(\vartheta) \quad \text{s.t.} \quad T_n(Y, \vartheta) \leq q \quad \text{(MJ)}
\]

- Estimated number of change-points: Minimizer \( \hat{K}(q) \) of (MJ)
- Confidence Set for \( \vartheta \): All solutions of (MJ)

\[C(q) = \{ \vartheta \in S : \# J(\vartheta) = \hat{K}(q) \text{ and } T_n(Y, \vartheta) \leq q \}\]
Step II: Signal, Jump Locations and Confidence Set

Solve the (nonconvex) optimization problem

$$\inf_{\vartheta \in \mathcal{S}} \#J(\vartheta) \quad \text{s.t.} \quad T_n(Y, \vartheta) \leq q$$

(MJ)

- Estimated number of change-points: Minimizer $\hat{K}(q)$ of (MJ)
- Confidence Set for $\vartheta$: All solutions of (MJ)

$$C(q) = \{ \vartheta \in \mathcal{S} : \#J(\vartheta) = \hat{K}(q) \text{ and } T_n(Y, \vartheta) \leq q \}$$

- SMUCE: Constraint MLE $\hat{\vartheta}(q)$ within $C(q)$, i.e.

$$\hat{\vartheta}(q) = \arg\max_{\vartheta \in C(q)} \sum_{i=1}^{n} \log \left( f_{\vartheta(i/n)}(Y_i) \right).$$
Step II: Signal, Jump Locations and Confidence Set

Solve the (nonconvex) optimization problem

\[
\inf_{\vartheta \in S} \# J(\vartheta) \quad \text{s.t.} \quad T_n(Y, \vartheta) \leq q \quad \text{(MJ)}
\]

- **Estimated number of change-points**: Minimizer \( \hat{K}(q) \) of (MJ)
- **Confidence Set** for \( \vartheta \): All solutions of (MJ)
  \[
  C(q) = \{ \vartheta \in S : \# J(\vartheta) = \hat{K}(q) \text{ and } T_n(Y, \vartheta) \leq q \}
  \]
- **SMUCE**: Constraint MLE \( \hat{\vartheta}(q) \) whithin \( C(q) \), i.e.
  \[
  \hat{\vartheta}(q) = \arg\max_{\vartheta \in C(q)} \sum_{i=1}^{n} \log \left( f_{\vartheta(i/n)}(Y_i) \right).
  \]
- **Statistical error control**: Choice of \( q \).
Controlling the model selection error

Solve the (nonconvex) optimization problem

\[
\inf_{\vartheta \in \mathcal{S}} \# J(\vartheta) \quad \text{s.t.} \quad T_n(Y, \vartheta) \leq q \quad (\text{MJ})
\]

Goal: calibrate \( q \), s.t. \( P(\hat{K}(q) \neq K) \) is minimal

model selection error
Controlling the model selection error

Solve the (nonconvex) optimization problem

$$\inf_{\vartheta \in S} \# J(\vartheta) \quad \text{s.t.} \quad T_n(Y, \vartheta) \leq q \quad \text{(MJ)}$$

Goal: calibrate $q$, s.t. $P(\hat{K}(q) \neq K)$ is minimal

Decompose $P(\hat{K}(q) \neq K)$ into

$$P(\hat{K}(q) < K) \text{ oversmoothing (later)}$$

and $P(\hat{K}(q) > K) \text{ undersmoothing}$
Theory: Bounds for $K$, Overestimation/Undersmoothing

Solve the (nonconvex) optimization problem

$$\inf_{\vartheta \in S} \#J(\vartheta) \quad \text{s.t.} \quad T_n(Y, \vartheta) \leq q \quad \text{(MJ)}$$

Note (follows directly from the definition)

$$P(\hat{K}(q) > K) \leq P(T_n(Y, \vartheta) > q)$$
Theory: Bounds for $K$, Overestimation/Undersmoothing

Solve the (nonconvex) optimization problem

$$\inf_{\vartheta \in S} \#J(\vartheta) \quad \text{s.t.} \quad T_n(Y, \vartheta) \leq q \quad (MJ)$$

Note (follows directly from the definition)

$$P(\hat{K}(q) > K) \leq P(T_n(Y, \vartheta) > q)$$

$T_n$ can be (asymptotically) bounded in distribution by

$$M = \sup_{0 \leq s \leq t \leq 1} \left\{ \frac{|B(s) - B(t)|}{\sqrt{s-t}} - \sqrt{2 \log \frac{e}{t-s}} \right\}$$

Dümbgen/Spokoiny, 2001, AoS

Asymptotic distribution depends on

$$\log(\tau_i - \tau_{i-1})$$
cf. Zhang/Siegmund’07

Simple Strategy: Use (empirical) quantile of $M$ as a choice of $q$
Theory: Bounds for $K$, Overestimation/Undersmoothing

In the gaussian case it holds uniformly over $S$

\[
P(\hat{K}(q) > K) \leq P(T_n(Y, \vartheta) > q) \leq P(M > q) =: \alpha(q)
\]

overestimation error
IV. SMUCE in Action
SMUCE $\hat{\phi}$ with confidence set and true signal.

Undersmoothing control: $P(\hat{K}(q) > K) \leq \alpha$

alpha 0.02

blue arrows: confidence intervals for jump location
Theory: Bounds for $K$, Underestimation/Oversmoothing

Given $\vartheta(\cdot) \in S$.

- $\Delta$ smallest jump size
- $\lambda$ smallest interval length between two successive jumps

**Theorem (Underestimation/Oversmoothing Control)**

Let $\hat{K}(q)$ the SMUCE for $K$. Then

$$P \left( \hat{K}(q) < K \right) \leq 2K \left[ \exp \left( -\frac{1}{8} \left( \frac{\eta}{2\sqrt{2}} - q - \sqrt{2 \log \frac{2e}{\lambda}} \right)^2 \right) + \exp \left( -\frac{\eta^2}{16} \right) \right]$$

where $\eta = \sqrt{n} \frac{\lambda \Delta}{\sigma}$.

- Note: $K \leq 1/\lambda$, prior information on $\lambda, \Delta$ sufficient.
Distributional overestimation bound

\[ P_{\theta}(\hat{K}(q) > K) \leq P(T_n > q) \leq P(M \geq q) \leq \alpha(q) \]

+ Exponential underestimation bound

\[ P_{\theta}(\hat{K}(q) < K) \text{ only depending on } n, \lambda, \Delta, q \]

Gives \[ P(\hat{K}(q) = K) \geq 1 - \alpha(q) - \exp(n, \lambda, \Delta, q) \]

Incorporate knowledge about
smallest scale \( \lambda \)
minimal signal strength \( \Delta \)
Distributional overestimation bound

\[ P_{\theta}(\hat{K}(q) > K) \leq P(T_n > q) \leq P(M \geq q) \leq \alpha(q) \]

+ Exponential underestimation bound

\[ P_{\theta}(\hat{K}(q) < K) \] only depending on \( n, \lambda, \Delta, q \)

Gives \( P(\hat{K}(q) = K) \geq 1 - \alpha(q) - \exp(n, \lambda, \Delta, q) \)

Incorporate knowledge about smallest scale \( \lambda \) minimal signal strength \( \Delta \)

Can be used to

- obtain uniform/honest confidence sets
- obtain uniform convergence for jump locations (not shown)
- determine \( q \) (later)
Sequentially Honest Confidence Sets

\[ P (\vartheta \in \mathcal{C}(q)) = P \left( T_n(Y, \vartheta) \leq q, \ K \leq \hat{K}(q) \right) \geq P (T_n(Y, \vartheta) \leq q) - P \left( \hat{K}(q) < K \right). \]

**Theorem**

Consider a sequence of nested models \( S_n \subset S \) s.t.

\[ \frac{n}{\log n} \Delta_n^2 \lambda_n \rightarrow \infty, \quad \text{as} \ n \rightarrow \infty, \]

then the confidence level is kept uniformly asymptotically over this sequence.

- Confidence bands: obtained from the graphs of

**Confidence set:**

\[ \mathcal{C}(q) = \{ \vartheta \in S[0, 1] : \vartheta \ has \ \hat{K} \ jumps \ and \ T_n(Y, \vartheta) \leq q \} \]
$P(\hat{K}(q) < K)$
$P(\hat{K}(q) > K)$
$P(\hat{K}(q) \neq K)$

Prior Information $\Delta \geq 0.2$, $K \leq 10$

$P(\hat{K}(q) \neq K) \sim 0.5$
**Example:** A novel acylated gramicidin A derivative (Diederichsen lab)

Time trace (grey) of conductance for the acylated gA derivative, Vm = 50 mV, 21s, SMUCE (blue solid line).

Smoothness guarantee: \( P(\hat{K}(q) > K) \leq 0.05 \)

Zooming in: 52-56s
SMUCE (blue solid line).
Example: A novel acylated gramicidin A derivative (Diederichsen lab)

Time trace (grey) of conductance for the acylated gA derivative, $V_m = 50 \text{ mV}, 21s$, SMUCE (blue solid line).

\[
\text{Smoothness guarantee: } P(\hat{K}(q) > K) \leq 0.05
\]

(B) Histogram of raw data. 
(C) Histogram for SMUCE with state boundaries (brown, dashed vertical lines).

Hotz et al., 2013, *(IEEE Trans. NanoBioscience)*
Current method: Semiautomatic

8,000 clicks per hand „ClampFit“

Sabine Bosk and Conrad Weichbrodt (Steinem Lab)
Comparison: $\log_{10}(\text{event length})$

- Manual analysis overestimates event lengths due to many missed events.
- Automatic analysis suggests 2 dynamics.

- Magenta: Overlap of manual and SMUCE average open time
  - Manual: 3.48 s
  - SMUCE: 0.71 s

$\Rightarrow$ manual analysis overestimates event lengths due to many missed events
$\Rightarrow$ automatic analysis suggest 2 dynamics
V. Remarks
Multiscale detection of vanishing signals I

normal observations, $\sigma = 1$

Detection boundary:
Control false alarm and sensitivity:
Any signal has to satisfy

$$\left( \frac{\Delta_n}{\sigma} \right)^2 |I_n| \geq 2 n^{-1} \log(1/|I_n|)$$

Detection boundary depends on

- noise level $\sigma$
- sample size $n$
- signal strength/height $\Delta_n$
- signal width (scale) $|I_n|$

Ingster’93, Dümbgen, Donoho/Jin’04 (AoS), Walther’08 (AoS), Frick et al.’14 (JRSS-B), Enikeeva et al.’15 (arXiv)
normal observations, \( \sigma = 1 \)

Detection boundary:
Control false alarm and sensitivity:
Any signal has to satisfy

\[
\left( \frac{\Delta_n}{\sigma} \right)^2 |I_n| \geq 2 n^{-1} \log(1/|I_n|)
\]

Detection boundary depends on

- noise level \( \sigma \)
- sample size \( n \)
- signal strength/height \( \Delta_n \)
- signal width (scale) \( |I_n| \)

intermediate scale scenario
normal observations, $\sigma = 1$

The needle in a haystack

Detection boundary:
Control false alarm and sensitivity:
Any signal has to satisfy

$$ \left( \frac{\Delta_n}{\sigma} \right)^2 \left| I_n \right| \geq 2 n^{-1} \log(1/|I_n|) $$

Detection boundary depends on
- noise level $\sigma$
- sample size $n$
- signal strength/height $\Delta_n$
- signal width (scale) $|I_n|$
Multiscale Detection of Vanishing Signals II

- SMUCE is capable of detecting multiple change-points simultaneously \textit{at the same optimal detection rate} (in terms of the smallest interval and jump size) as a single change-point.
- The constants differ that bound the size of the signals that can be detected. These increase with the complexity of the problem:
  - $\sqrt{2}$ for a single change point
  - 4 for a bounded (but unknown) number of change-points
  - 12 for an unbounded number of change-points.
- Jeng/Cai/Li’10 (JASA) achieve for \textit{sparse} step functions the optimal constant $\sqrt{2}$. Sparsity enters explicitly their estimator. We do not make any sparsity assumptions on the true signal. SMUCE adapts automatically to sparseness. A similar phenomenon occurs for density bump detection (Dümbgen/Walther’08).
Computation

- SMUCE can be computed by dynamic programing (Friedrich et al., 2008) in $O(n^2)$.
- The particular structure of the problem allows for pruning steps, similar to (Killick et al., 2011).
- Number of intervals in the dynamic program is of order

$$n^2 \sum_{k=1}^{\hat{K}+1} (\hat{\tau}_k - \hat{\tau}_{k-1})^2 \approx n^2 / \hat{K} \quad \text{(for equidistant change-points)}.$$
VI. Extensions
Heteroscedastic Data: Example

(Pein et al.‘15)

Figure: Signal (black line), variance (dotted black line), SMUCE (blue line) and H-SMUCE (red line), both with $\alpha = 0.1$. 

SMUCE
(const variance)

H-SMUCE requires additionally $\lambda \geq C \log n / n$
$\alpha = 0.3$

H-SMUCE (red)  
true signal (black)

variance $\sigma^2(t)$
Quantile Regression

- Quantile change-point regression: Let $\xi_\beta$ the $\beta$ quantile of $Z_i$.

$$Y_i = \begin{cases} 1 & \text{if } Z_i \leq \xi_\beta \\ 0 & \text{otherwise} \end{cases}$$

- Amounts to Bernoulli regression with $\mathbb{E}Y_i = \beta$
Quantile Regression (Example)

Figure: from top left to bottom right: Cauchy data; Cauchy data (magnification); true median function $\vartheta$; median estimate $\hat{\vartheta}$
Piecewise linear functions: Example
Inference on "Qualitative Features" of 1-D Signals

Change-point Inference:
- Simultaneous change-point detection on all scales

Mode Inference:
- Topological data analysis (TDA) on all scales


Persistence diagrams [Cohen-Steiner et al., 2005]
Persistence diagrams [Cohen-Steiner et al., 2005]
Stability theorem

- Recall: Bottleneck distance is sup-norm distance of persistence diagrams.

Theorem (Cohen-Steiner et al.’05,…, Ghrist’08)

\[ d_\infty(Dgm(f), Dgm(g)) \leq \|f - g\|_\infty \]
Stability theorem

- Recall: Bottleneck distance is sup-norm distance of persistence diagrams.

Theorem (Cohen-Steiner et al.’05,..., Ghrist’08)

\[ d_\infty(Dgm(f), Dgm(g)) \leq ||f - g||_\infty \]

Theorem (Bauer et al.’14)

Let \( f_n \) a sequence of regression functions with a rectangular bump of size \( \delta_n \), s.t. \( \delta_n^2 = o(\log n) \).

\[ Y_i = f(i/n) + \epsilon_i, \]

and \( \epsilon_i \sim N(0, \sigma^2) \), \( i = 1, ..., n \). Then there is no thresholding rule for the sup norm persistence diagram, which consistently detects this bump.

Recall: In this case a signal of size \( \delta_n \sim n^{-1/2} \) is detectable.
Stability theorem

- Recall: Bottleneck distance is sup-norm distance of persistence diagrams.

Theorem (Cohen-Steiner et al.’05,..., Ghrist’08)

\[ d_\infty(D_{gm}(f), D_{gm}(g)) \leq ||f - g||_\infty \]

Theorem (Bauer et al.’14)

Let \( f_n \) a sequence of regression functions with a rectangular bump of size \( \delta_n \), s.t. \( \delta_n^2 = o(\log n) \).

\[ Y_i = f(i/n) + \epsilon_i, \]

and \( \epsilon_i \sim N(0, \sigma^2) \), \( i = 1, ..., n \). Then there is no thresholding rule for the sup norm persistence diagram, which consistently detects this bump.

Presmoothing is an option (e.g. Bubenik et al.’10, Fasy et al.’14). We want to avoid this as it requires full reconstruction of the signal.
Modes and signatures for $d = 1$

- $X := \{ f : [0, 1] \rightarrow \mathbb{R} \text{ with a finite (but unknown) number of modes} \}$
- $X_k := \{ f \in X : J(f) \leq k \} \subset X$ be the class of functions with at most $k$ modes.
- For a metric $d$ define the $k$-th metric signature of $f \in X$ as

$$s_k(f) := \inf_{g \in X_k} d(f, g) \quad \text{for} \quad k \in \mathbb{N}_0,$$

i.e., the distance of $f$ to the best approximating function with $k$ modes (w.r.t. $d$).

- We will choose $d$ to be a (simplified) multiscale statistic.
Kolmorogov signatures

- Let $f, g \in X$, and let $F, G$ denote the respective antiderivatives. The Kolmogorov distance is defined as

$$d_K(f, g) := d_{\infty}(F, G).$$

We will use $d_K$ for inferring the number of modes.

- We consider the Kolmorogov signatures

$$s_k(f) := \inf_{g \in X_k} d_K(f, g) \quad \text{for} \quad k \in \mathbb{N}_0.$$

**Figure:** A function with exactly two modes (left) and its closest function with exactly one mode w.r.t. the Kolmogorov norm (right, in purple). The attendant Kolmogorov signature, $s_1$, for removing the smallest mode of $f$, can be read off from the light-blue areas.
Inference

- We do not aim to estimate the regression function $f$ itself but rather to infer directly the sequence of signatures $s_k(f)$ together with the number of modes $k$.
- An estimate for the sequence of signatures can be obtained by the empirical signatures

$$\hat{s}_k = \inf_{g \in X_k} d_K(Y, g)$$

**Theorem**

Assume that the noise $(\epsilon_i)$ is independently distributed with mean zero such that for some $\kappa > 0$, $v > 0$ and all $m \geq 2$

$$\mathbb{E} |\epsilon_i|^m \leq vm! \kappa^{m-2}/2 \text{ for all } i = 1, \ldots, n. \tag{1}$$

Then, for any $\delta > 0$ and any $f \in X$

$$\mathbb{P} \left( \max_{j \in \mathbb{N}_0} |s_j - \hat{s}_j| \geq \delta \right) \leq 2 \exp \left( -\frac{\delta^2 n}{2v + 2\kappa \delta} \right).$$

**Stability theorem for general metrics:**

$$|s_k(g) - s_k(f)| \leq d(f, g)$$
Examples

Figure: true signal and data

Figure: empirical signatures $\hat{s}_k$ (blue) and true signatures (grey)
Estimating the number of modes

We estimate the number of modes by thresholding the empirical signatures \( \hat{s}_k \):

\[
\hat{k}(q) = \min \{ l \in \mathbb{N} : \hat{s}_l(Y) \leq q \}
\]

**Theorem (Overestimation of modes)**

Let \( k \) denote the true number of modes of \( f \) and set

\[
q(\alpha) := \frac{1}{n} \left( \sqrt{\log(\alpha/2) (\log(\alpha/2) \kappa^2 + 2nv)} + \kappa \log(\alpha/2) \right).
\]

Then,

\[
\max_{k \in \mathbb{N}_0} \sup_{f \in X_k} \mathbb{P} \left( \hat{k}(q(\alpha)) > k \right) \leq \alpha.
\]

**Theorem (Underestimation Bound and Consistency)**

Assume that \( f \in X_k \) is such that \( s_{k-1}(f) \geq \epsilon \), i.e. the smallest mode is larger than \( \epsilon \). Then,

\[
\mathbb{P} \left( \hat{k}(\epsilon/2) = k \right) \geq 1 - 2 \exp \left( -\frac{\epsilon^2 n}{8v + 4\kappa \epsilon} \right).
\]
Summary

- SMUCE: Multiscale Change Point Estimator in EFs:
  - $\ell_0$-minimisation under multiscale local likelihood constraint
  - model selection step + constraint estimation for „multiscale regressogram“

![Graphs showing mean, normal, variance, normal, Poisson, and binary distributions.](image-url)
Summary

- SMUCE: Multiscale Change Point Estimator in EFs:
  - \( \ell_0 \)-minimisation under multiscale local likelihood constraint
  - model selection step + constraint estimation for "multiscale regressogram"

- Computationally feasible: linear to quadratic time
Summary

- SMUCE: Multiscale Change Point Estimator in EFs:
  - $\ell_0$-minimisation under multiscale local likelihood constraint
  - model selection step + constraint estimation for „multiscale regressogram“
- Computationally feasible: linear to quadratic time
- Bounds for under/overestimation of $K$
  - controls model selection error $P(\hat{K}(q) \neq K)$
  - guide for thresholding
  - allows to incorporate prior information
  - sequentially honest confidence sets
Summary

- SMUCE: Multiscale Change Point Estimator in EFs:
  - $\ell_0$-minimisation under multiscale local likelihood constraint
  - model selection step + constraint estimation for „multiscale regressogram“
- Computationally feasible: linear to quadratic time
- Bounds for under/overestimation of $K$
  - controls model selection error $P(\hat{K}(q) \neq K)$
  - guide for thresholding
  - allows to incorporate prior information
  - sequentially honest confidence sets
- Obeys good performance confirmed by simulations (not shown)
  - optimal detection on (essentially) all scales
  - adapts automatically to sparseness ($p=n$ not $p >> n$)
  - (up to log) optimal estimation rates (not shown)
Summary

Extensions to

- Heterogeneous data (H-SMUCE), Pein et al.'15
- Higher selection power (FDR-based), Li et al.'14
- Inference for TDA: we have some answers for $d=1$
  TDA then relates to mode hunting
  direct estimation of KS signatures possible
  confidence statements for KS signatures/persistent barcodes
  computationally fast

Open issues:

- Much is unexplored: How does ITDA transfer to $d>1$?
  Conceptually, computationally?


---

**R-package StepR**

[www.stochastik.math.uni-goettingen.de/smuce](http://www.stochastik.math.uni-goettingen.de/smuce)

[www.stochastik.math.uni-goettingen.de/fdrs](http://www.stochastik.math.uni-goettingen.de/fdrs)

[www.stochastik.math.uni-goettingen.de/munk](http://www.stochastik.math.uni-goettingen.de/munk)