Optimal change–point estimation from indirect observations

A. Goldenshluger | A. Tsybakov | A. Zeevi

Haifa University | Université Paris VI | Columbia University

Problem formulation

- White noise convolution model

$$dY(x) = (K * f)(x)dx + \epsilon dW(x), \quad x \in \mathbb{R},$$

where $f \in \mathbb{L}_2(\mathbb{R})$ is an unknown function, $K \in \mathbb{L}_1(\mathbb{R})$, and $W(\cdot)$ is the standard two–sided Wiener process on $\mathbb{R}$.

- Assumption: f is smooth apart from a discontinuity jump of the first kind at $\theta \in [0, 1]$.

- Goal: estimate the change–point θ from observations of the trajectory $\{Y(t), t \in \mathbb{R}\}$.

Motivation

- Asymptotic equivalence between idealized white noise model and nonparametric regression, density estimation...

- Edge detection in imaging

$$Y = K * f + \epsilon,$$

  Y is the degraded image, K is the point spread function (blur), f is the true image, ϵ is the random noise (point degradation).

- Detection of abrupt changes in linear dynamic systems.

Minimax framework

Let $\hat{\theta}$ be an estimator of θ, and let $\mathcal{G}$ be a class of functions having a single change–point $\theta \in [0, 1]$.

- Maximal risk

$$R_\epsilon[\hat{\theta}; \mathcal{G}] = \sup_{f \in \mathcal{G}} \left\{ \mathbb{E}_f |\hat{\theta} - \theta|^2 \right\}^{1/2}.$$

- An estimator $\tilde{\theta}$ is rate optimal if

$$R_\epsilon[\tilde{\theta}; \mathcal{G}] \asymp R^*_\epsilon[\mathcal{G}] := \inf_{\hat{\theta}} R_\epsilon[\hat{\theta}; \mathcal{G}], \quad \epsilon \to 0.$$

- Goal: is to develop rate–optimal estimators of the change–point θ.

Related literature

- Change–point estimation in direct models

  Korostelev (1987), Yin (1988), Müller (1992),

  Wang (1995), Gijbels, Hall & Kneip (1999),

  Antoniadis & Gijbels (2002), ...

For a class $\mathcal{G}$ of functions with single change–point and sufficiently smooth (at least Lipschitz) otherwise, $R_\epsilon^*[\mathcal{G}] \asymp \epsilon^2$.

- Change–point estimation from indirect data

  – Neumann (1997), density deconvolution model.

  – Raimondo (1998), estimation of change–points in derivatives, regression model.

Neumann (1997)

- Density deconvolution model

$$Y_i = X_i + \epsilon_i, \;\; i = 1, \dots, n, \;\; X_i \overset{iid}{\sim} f, \;\; \xi_i \overset{iid}{\sim} K, \; X \perp \epsilon,$$

$$c(1+|\omega|)^{-\beta} \le |\widehat{K}(\omega)| \le C(1+|\omega|)^{-\beta}, \quad \beta > 0.$$

$\beta > 1/2 \;\Rightarrow\; K \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$ – most interesting case.

- Functional class $\mathcal{G}$

Let $[f](x) = f(x+) - f(x-)$, $\|f\|_\infty \le c_1$, and assume that

$\exists\theta : |[f](\theta)| \ge c_2$, and $|f(x) - f(y)| \le c_3|x - y|$, if $\theta \notin [x, y,]$.

- Result

$$R^*_\epsilon[\mathcal{G}] \asymp \begin{cases} n^{-1/(2\beta+1)}, & 0 < \beta \le 1/2, \\ n^{-2/(2\beta+3)}, & \beta > 1/2. \end{cases}$$

Raimondo (1998)

- Regression model

$$Y_i = f(i/n) + \epsilon_i, \quad i = 1, \ldots, n, \ \ \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2).$$

- Functional class $\mathcal{G}$

  For integer $\beta \geq 0$, $\exists \theta \in [a, b] \subset [0, 1]$ such that

$$|[f^{(\beta)}](\theta)| \geq c_1, \ \ |f^{(\beta)}(x) - f^{(\beta)}(y)| \leq c_2 |x - y|, \ \ \forall x < y, \theta \notin [x, y].$$

- Result

$$R^*_\epsilon[\mathcal{G}] \asymp n^{-1/(2\beta+1)}, \quad \text{for integer } \beta \geq 0.$$

Remarks

- Discrepancy in results

  When K is the Green function of a linear differential operator, deconvolution is equivalent to estimation of the change–point in a derivative of f from direct data.

- Recent results

  It is common belief that the rate $n^{-1/(2\beta+1)}$ is optimal for estimation of the change–point in β-th derivative: Gijbels, Hall & Kneip (1999), Wang (1999), Huh & Carriere (2002), Park & Kim (2004)

- How to construct rate optimal estimates?

Functional class $\mathcal{F}_m(L,a)$

- Class $\mathcal{F}_1(L,a)$

  Let $a, L > 0$. We say $f \in \mathcal{F}_1 = \mathcal{F}_1(L,a)$ if $f \in \mathbb{L}_2(\mathbb{R})$, $\exists \theta \in [0,1]$ s.t. $|[f](\theta)| \geq a$, and f is Lipschitz with constant L on every interval that does not contain θ.

- Class $\mathcal{F}_m(L,a), \;\; m > 1$

  Let $a, L > 0$, $m > 1$. We say $f \in \mathcal{F}_m = \mathcal{F}_m(L,a)$ if $f \in \mathbb{L}_2(\mathbb{R})$,

  (i) $\exists \theta \in [0,1]$ s.t. $|[f](\theta)| \geq a$,

  (ii) $f'(x)$ exists for all $x \neq \theta$, $[f'](\theta) = 0$ so that function $g_f(x) := f'(x)$, $x \neq \theta$, and $g_f(\theta) := f'(\theta\pm)$ is continuous,

$$\int_{-\infty}^{\infty} |\widehat{g}_f(\omega)||\omega|^{m-1} d\omega \leq L.$$

Main result

Theorem Let the Fourier transform $\hat{K}(\cdot)$ of K satisfy

$$\underline{\kappa}(1+|\omega|^2)^{-\beta/2} \le |\widehat{K}(\omega)| \le \overline{\kappa}(1+|\omega|^2)^{-\beta/2}, \quad \forall \omega \in \mathbb{R}.$$

Then for sufficiently small ϵ

- if $\beta > 1/2$ then

$$R_\epsilon^*[\mathcal{F}_m(a,L)] \asymp a^{-1} L^{\frac{2\beta-1}{2\beta+2m+1}} \epsilon^{\frac{2m+2}{2m+2\beta+1}}.$$

- if $\beta < 1/2$ then

$$c_1 \epsilon^{2/(2\beta+1)} \le R_\epsilon[\mathcal{F}_m(a,L)] \le C_1 \epsilon^{2/(2\beta+1)} \left(\ln \epsilon^{-1}\right)^{\frac{1-2\beta}{2(2\beta+1)}}.$$

- if $\beta = 1/2$ then

$$c_2 \epsilon (\ln \epsilon^{-1})^{-1/2} \le R_\epsilon[\mathcal{F}_m(a,L)] \le C_2 \epsilon.$$

Remarks

- Optimal rate of convergence (ignoring $\ln$–factors)

$$R_\epsilon^*[\mathcal{F}_m(a,L)] \asymp \min\Big\{\epsilon^{(2m+2)/(2m+2\beta+1)}, \epsilon^{2/(2\beta+1)}\Big\}.$$

  The elbow corresponds to $\beta > 1/2$, $0 < \beta \leq 1/2$.

- The optimal rates are determined by
  - ill–posedness of convolution, β
  - smoothness of f away from the change–point (in contrast to direct observations)

- The optimal rate approaches the parametric one, ϵ, as $m \to \infty$, for $\beta > 1/2$.

- Consistent with results of Neumann (1997):

$$m = 1,\ \beta > 1/2 \ \Rightarrow \ R_\epsilon^*[\mathcal{F}_1(a,L)] \asymp (\epsilon^2)^{\frac{2}{2\beta+3}}.$$

Probe functional

- Kernel: $\varphi \in \mathbb{L}_2(\mathbb{R})$, $\widehat{\varphi} \in C^\infty$ is even, non-negative, and for small $\eta > 0$

$$\widehat{\varphi}(\omega) = \begin{cases} 1, & 1/3 + \eta \le |\omega| \le 2/3 - \eta, \\ 0, & |\omega| \notin [1/3, 2/3]. \end{cases}$$

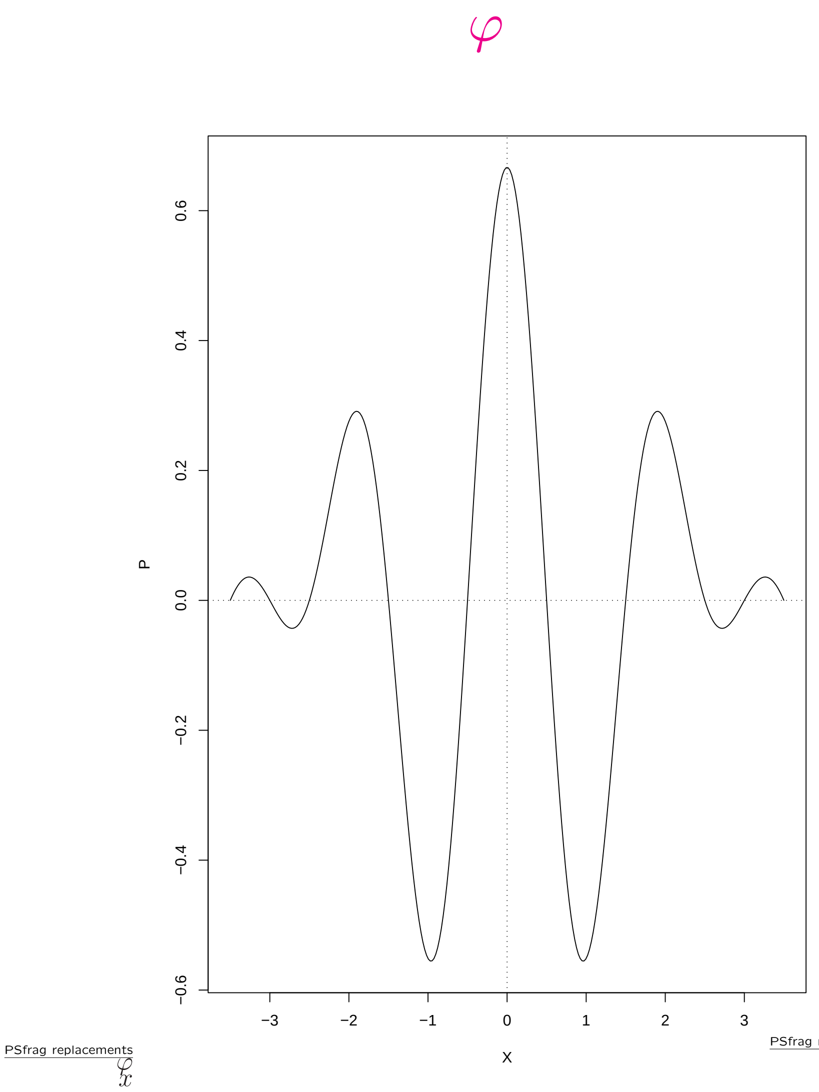
- Probe functional: for $h > 0$ let

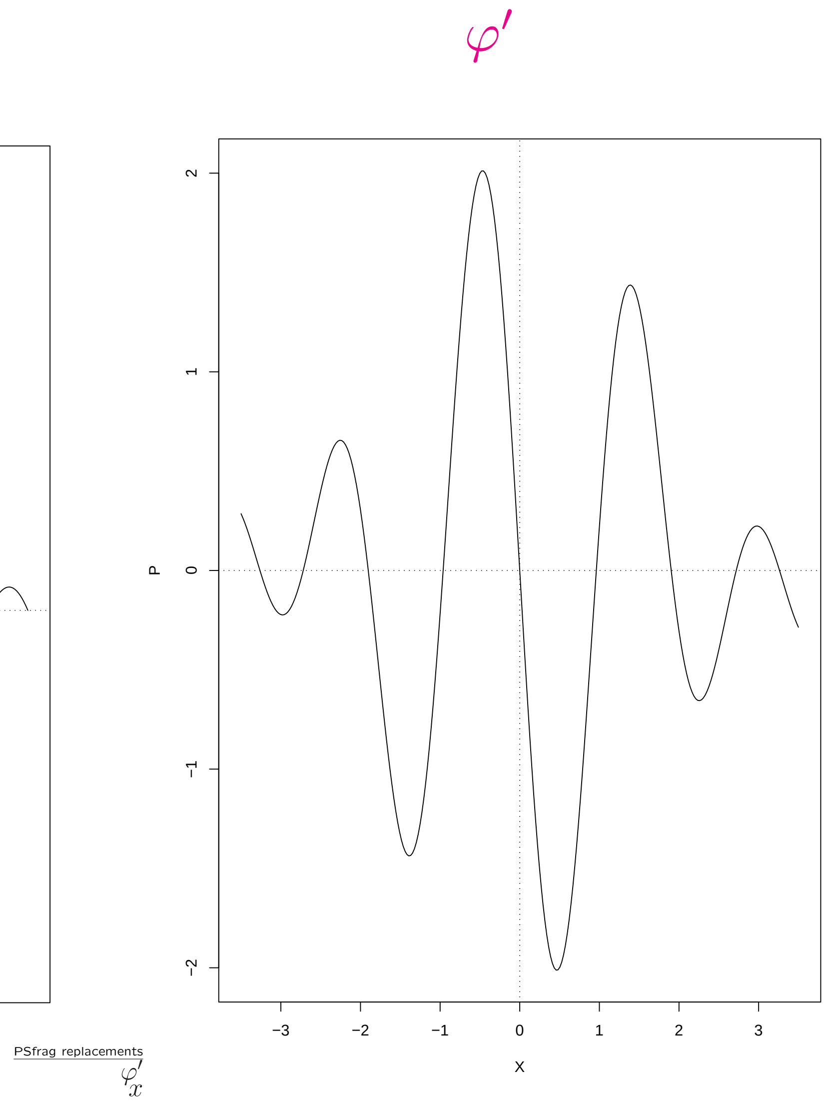
$$\ell_h(t) = \langle f, \psi_t \rangle := \frac{1}{h^3} \int_{-\infty}^{\infty} f(x) \varphi'' \Big(\frac{x-t}{h} \Big) dx, \quad t \in \mathbb{R}.$$

- $\ell_h(t)$ is a smoothed second derivative of f at t:

  $\ell_h(t) \to f''(t)$ as $h \to 0$, provided that f is twice continuously differentiable at t.

Kernel φ





Probe functional properties

- Good detection ability

  The point of zero–crossing of $\ell_h(t)$ (minimum of $|\ell_h(t)|$) is indicative of the change–point location: for small $\delta > 0$

$$\inf_{t:\delta<|t-\theta|<ch} \Big\{ |\ell_h(t)| - |\ell_h(\theta)| \Big\} \geq C_1 \delta h^{-3}.$$

- Can be estimated with high accuracy

  There exists an unbiased estimator $\tilde{\ell}_h(t)$ of $\ell_h(t)$ such that

$$\mathbb{P}\Big\{ \sup_{t\in B} |\tilde{\ell}_h(t) - \ell_h(t)| \geq \lambda \Big\} \leq C_2 \frac{h^{\beta+3/2}}{\epsilon} |B| \lambda \exp\Big\{ -C_3 \frac{\lambda^2 h^{2\beta+5}}{\epsilon^2} \Big\}.$$

  for any $\lambda \geq C_4 \epsilon^2 h^{-2\beta-5}$.

Estimator of the probe functional

- **Linear functional strategy**

If $\psi_t \in Range(K^*)$ then there exists $\gamma_t \in \mathbb{L}_2(\mathbb{R})$ s.t.

$$\ell_h(t) = \langle f, \psi_t \rangle = \langle f, K^* \gamma_t \rangle = \langle Kf, \gamma_t \rangle, \quad K^* \gamma_t = \psi_t.$$

For the convolution transform,

$$\widehat{\gamma}_t(\omega) = -(2\pi\omega)^2 e^{2\pi i \omega t} \frac{\widehat{\varphi}(\omega h)}{\widehat{K}(-\omega)} \ .$$

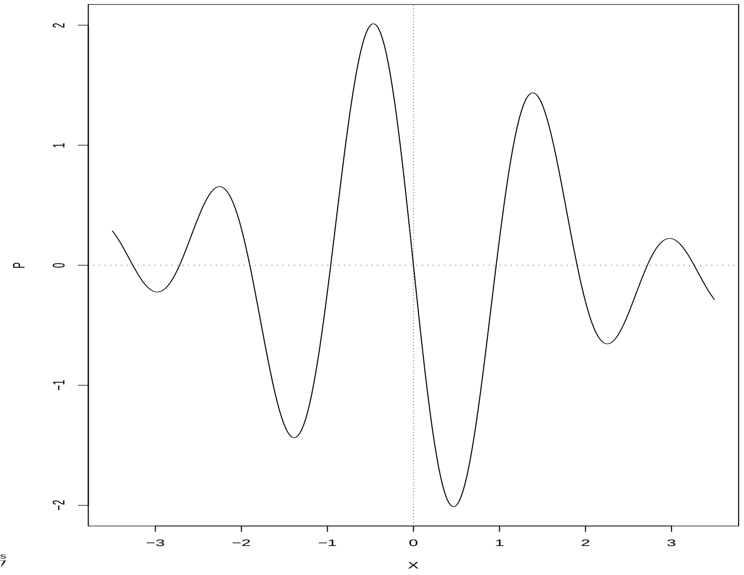
- **Estimator**

$$\tilde{\ell}_h(t) = \int_{-\infty}^{\infty} \gamma_t(x) dY(x), \quad t \in \mathbb{R}.$$

Localization step

- It is shown that

$$\ell_h(t) = -\frac{1}{h^2}[f](\theta)\varphi'\left(\frac{\theta - t}{h}\right) + \text{“something small”}, \quad \forall t \in \mathbb{R}.$$



PSfrag replacements
φ'_x

- Rough estimates

$$\hat{t}_* := \arg\min_{t\in[0,1]} \tilde{\ell}_h(t), \quad \hat{t}^* := \arg\max_{t\in[0,1]} \tilde{\ell}_h(t).$$

Change–point estimator

- **Rough estimates:** for $\hat{t} = \hat{t}_*$ or $\hat{t}^*$

$$\mathbb{P}\Big\{|\hat{t} - \theta| > c_1 h\Big\} \le c_2 \frac{h^{\beta-1/2}}{\epsilon} \exp\Big\{-c_3 \frac{h^{2\beta+1}}{\epsilon^2}\Big\}.$$

- **The estimator**

  Let $\hat{A}_h$ be the closed interval with the endpoints $\hat{t}_*$ and $\hat{t}^*$. Define

$$\tilde{\theta}_h = \arg\min_{t \in \hat{A}_h} |\tilde{\ell}_h(t)| \ .$$

- **Bandwidth choice**

$$h_* = \begin{cases} C_1^*(\epsilon/L)^{2/(2m+2\beta+1)}, & \beta > 1/2 \\ C_2^*(\epsilon\sqrt{\ln \epsilon^{-1}})^{2/(2\beta+1)}, & \beta \le 1/2 \end{cases}$$

Concluding remarks

- Optimal procedure balances

$$\underbrace{\delta h^{-3}}_{\delta-\text{separation rate}} \asymp \underbrace{h^{m-2}}_{\text{bias}} \asymp \underbrace{\epsilon h^{-\beta-5/2}}_{\text{stochastic error}}$$

- Another probe functional

  The search for maximum of a smoothed first derivative results in slower rate of convergence, $\epsilon^{(m+2)/(2m+2\beta+1)}$.

- Extension

  Estimation of derivatives in the regression setup with equidistant design.

  The change–point procedures in Raimondo (1998), Wang (1999), Huh & Carriere (2002), Park & Kim (2004) are *not rate optimal*, contrary to what is claimed in some of these papers.