Locally adaptive confidence bands

Angelika Rohde
joint work with Tim Patschkowski

Albert-Ludwigs-Universität Freiburg

Luminy, December 2017

Introduction

- Model of density estimation $X_1, \ldots, X_n \overset{iid}{\sim} \mathbb{P}_p$
- $\mathbb{P}_p$ probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with Lebesgue density p (unknown to the statistician)
- $p \in \mathcal{P}$ (some nonparametric class, i.e. some massive parameter set)

Introduction

- Bickel and Rosenblatt (1973):

  Under certain regularity assumptions, the kernel density estimator

$$\widehat{p}_n(t, h_n) := \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - t}{h_n}\right)$$

with bandwidth $h_n = n^{-\delta}$ and $1/5 < \delta < 1/2$ satisfies

$$\lim_{n\to\infty} \mathbb{P}_p\left(\sup_{t\in[0,1]} \frac{|\widehat{p}_n(t, h_n) - p(t)|}{\widehat{\sigma}_n(t)} \le z_n(z)\right) = e^{-2e^{-z}}$$

with

$$\widehat{\sigma}_n(t) = \left(\frac{\widehat{p}_n(t, h_n)\|K\|_2^2}{nh_n}\right)^{1/2}, \quad z_n(z) = \left((2\delta \log n)^{1/2} + \frac{z + \log\left(\frac{\|K'\|_2}{2\pi\|K\|_2}\right)}{(2\delta \log n)^{1/2}}\right)$$

Introduction

- $z_n(z)$ does not depend on p
- If $z = z_\alpha$ is chosen such that the RHS equals $1 - \alpha$, then

$$\mathbb{P}_p^{\otimes n}\Big(p(t) \in [\widehat{p}_n(t, h_n) - \widehat{\sigma}_n(t) z_n(z_\alpha), \widehat{p}_n(t, h_n) + \widehat{\sigma}_n(t) z_n(z_\alpha)] \text{ for all } t \in [0,1]\Big) \approx 1 - \alpha$$

That is,

$$\Bigg([\widehat{p}_n(t, h_n) - \widehat{\sigma}_n(t) z_n(z_\alpha), \widehat{p}_n(t, h_n) + \widehat{\sigma}_n(t) z_n(z_\alpha)] \Bigg)_{t \in [0,1]}$$

is an asymptotic $(1 - \alpha)$-confidence band

Two questions are immediate ...

- Uniformity (in p) of the asymptotic coverage property?
- Choice of h_n? A suitable choice of the bandwidth h_n depends on the density's regularity ...

Introduction

- Typical nonparametric classes: "Smoothness classes",
  such as Hölder balls $\mathcal{H}(\beta, L)$, or some union of Hölder balls
- Hölder norm

$$\|p\|_\beta = \|p\|_{\sup} + \sup_{x \neq y} \frac{|p(x) - p(y)|}{|x - y|^\beta}, \quad \text{for } 0 < \beta \leq 1$$

- Hölder ball

$$\mathcal{H}(\beta, L) = \left\{ p : \mathbb{R} \to \mathbb{R} : \ \|p\|_\beta \leq L \right\}$$

- Probability densities within this Hölder ball

$$\mathcal{P}(\beta, L) = \left\{ p : \mathbb{R} \to \mathbb{R} : \ p \text{ Lebegue density}, \ \|p\|_\beta \leq L \right\}$$

Kernel density estimators with bandwidth $h_n \sim n^{-\frac{1}{2\beta+1}}$ are minimax optimal

Introduction

Adaptive **estimation**

- The smoothness parameter β plays the crucial role in the construction of the estimator, yet unknown to the statistician in many applications
- An estimator $\widehat{p}_n$ is called adaptive up to a logarithmic factor within a range of adaptation $[\beta_*,\ \beta^*]$, if for any $\beta \in [\beta_*,\ \beta^*]$ there exist $c_\beta > 0$ and $\gamma_\beta > 0$, such that for all $n \in \mathbb{N}$

$$\sup_{p \in \mathcal{P}(\beta, L)} \mathbb{E}_p \left(\widehat{p}_n(t) - p(t)\right)^2 \leq c_\beta \cdot n^{-\frac{2\beta}{2\beta+1}} \cdot (\log n)^{\gamma_\beta}$$

  $\longrightarrow$ (Almost) optimal rates of convergence over $\mathcal{P}(\beta, L)$ simultaneously for a large range of the parameter β
- Adaptive estimation is well understood in a large variety of nonparametric models

Is it possible to take advantage of adaptive density estimators for **inference**?

- Goal: Construction of confidence bands for a large union of models which
  - maintain given coverage probability over the full model
  - shrink at fastest possible rate over all submodels

Introduction

Honesty

For any interval $[a, b]$ and any significance level $\alpha \in (0, 1)$, a confidence band for p, described by a family of random intervals

$$C_n(t, \alpha), \;\; t \in [a, b],$$

is said to be *(asymptotically) honest* with respect to $\mathcal{P}$ if the coverage inequality

$$\liminf_{n} \inf_{p \in \mathcal{P}} \mathbb{P}_p^{\otimes n}\Big(p(t) \in C_n(t, \alpha) \text{ for all } t \in [a, b]\Big) \geq 1 - \alpha$$

is satisfied.

Introduction

Adaptivity

If $\mathcal{P}$ is some class of densities within a union of Hölder balls $\mathcal{H}(\beta, L)$ with fixed radius $L > 0$, the confidence band is called *globally adaptive* if for every $\beta > 0$ and for every $\varepsilon > 0$ there exists some constant $c > 0$, such that

$$\limsup_{n} \sup_{p \in \mathcal{H}(\beta,L) \cap \mathcal{P}} \mathbb{P}_p^{\otimes n} \left(\sup_{t \in [a,b]} |C_n(t, \alpha)| \geq c \cdot r_n(\beta) \right) < \varepsilon.$$

$r_n(\beta) =$ minimax-optimal rate of convergence for estimation under supremum norm loss over $\mathcal{H}(\beta, L) \cap \mathcal{P}$, possibly up to some logarithmic factor

Introduction

Negative Result

If $\mathcal{P}$ equals the set of *all* densities contained in

$$\bigcup_{0<\beta\leq\beta^*} \mathcal{H}(\beta, L),$$

honest and adaptive confidence bands provably do not exist!

- Low (1997): Honest random-length intervals for a probability density at a fixed point cannot have smaller expected width than fixed-length confidence intervals with the size corresponding to the lowest regularity under consideration

Not even possible to construct a family of random intervals $C_n(t, \alpha), t \in [a, b]$, whose expected length shrinks at the fastest possible rate simultaneously over two distinct nested Hölder balls with fixed radius, and which is at the same time asymptotically honest for the union $\mathcal{P}$ of these Hölder balls

Introduction

Numerous attempts to tackle this adaptation problem in alternative formulations

- Genovese and Wasserman (2008) relax the coverage property
  $\rightarrow$ do not require the confidence band to cover the function itself but a simpler surrogate function capturing the original function's significant features
- Under qualitative shape constraints, Hengartner and Stark (1995), Dümbgen (1998, 2003), and Davies, Kovacz and Meise (2009) achieve adaptive inference
- Picard and Tribouley (2000) investigate on pointwise adaptive confidence intervals under a *self-similarity condition* on the parameter space
- Under *such a condition*, Giné and Nickl (2010) even develop asymptotically honest confidence bands for probability densities whose width is adaptive to the global Hölder exponent
- ...

Preliminaries

- $X_1, \cdots, X_n \overset{iid}{\sim} \mathbb{P}_p$
- Kernel density estimator

$$\widehat{p}_n(\cdot, h) = \frac{1}{n} \sum_{i=1}^{n} K_h\left(X_i - \cdot\right)$$

with bandwidth $h > 0$ and rescaled kernel

$$K_h(\cdot) = h^{-1} K(\cdot/h)$$

- K is a symmetric kernel of bounded variation supported on $[-1, 1]$
- K is said to be of order $l \in \mathbb{N}$, if

$$\int x^j K(x)\,\mathrm{d}x = 0 \;\text{ for } 1 \leq j \leq l, \quad \int x^{l+1} K(x)\,\mathrm{d}x = c \;\text{ with } c \neq 0$$

Globally adaptive confidence bands

- Giné and Nickl (2010) consider **globally** adaptive confidence bands over

$$\widetilde{\mathcal{P}} = \bigcup_{\beta_* \le \beta \le \beta^*} \Big\{ p \in \mathcal{P}(\beta, L) : p \ge \delta \text{ on } [-\varepsilon, 1+\varepsilon], \quad c \cdot h^\beta \le \|K_h * p - p\|_{\sup} \le C \cdot h^\beta \text{ for all } h \le h_0 \Big\}$$

  for some constants $C > c > 0$ and $0 < \varepsilon < 1$
- $\beta^* = l + 1$ with l the order of the kernel
- Their honest and globally adaptive confidence bands are of the form

$$C_n(t, \alpha) = \left[\widehat{p}_n(t) - \sqrt{\widehat{p}_n(t)} \cdot \widehat{\Delta}_n(\alpha),\ \widehat{p}_n(t) + \sqrt{\widehat{p}_n(t)} \cdot \widehat{\Delta}_n(\alpha) \right],\ t \in [0,1],$$

- Stochastic order of the confidence band's width is determined by $\widehat{\Delta}_n(\alpha)$, which is independent of t
- For every $\beta > 0$ and for every $\varepsilon > 0$ there exist $c, \kappa > 0$, such that

$$\limsup_{n \to \infty} \sup_{p \in \mathcal{H}(\beta, L) \cap \widetilde{\mathcal{P}}} \mathbb{P}_p \left(\widehat{\Delta}_n(\alpha) \ge c \cdot n^{-\frac{\beta}{2\beta+1}} \cdot (\log n)^\kappa \right) < \varepsilon$$

Admissible functions

Remark

If $K(\cdot) = \frac{1}{2}1\{\cdot \in [-1,1]\}$ is the rectangular kernel, the set of all twice differentiable densities $p \in \mathcal{P}(2, L)$ that are supported in a fixed compact interval $[a, b]$ satisfies $c \cdot h^2 + o(h^2) \leq \|K_h * p - p\|_{\sup}$ with a constant $c > 0$.

The reason is that due to the constraint of being a probability density, $\|p''\|_{\sup}$ is bounded away from zero uniformly over this class, in particular p'' cannot vanish everywhere.

- x_0 point of maximum

$$\frac{1}{b-a} \leq p(x_0) - \underbrace{p(b)}_{=0} = (x_0 - b)\underbrace{p'(x_0)}_{=0} - \frac{1}{2}\underbrace{(x_0 - b)^2}_{\leq (b-a)^2} p''(x_i)$$

- Hence,

$$\|p''\|_{\sup} \geq \frac{2}{(b-a)^3}.$$

Locally adaptive confidence bands

- However, even one small wiggly part of the density inhibits stronger performance of the procedure in smooth segments
- Ideally, a confidence band is automatically tighter in regions where the unknown density is smooth and wider in less smooth parts

Our goal / challenges

(i) Find a proper notion of a locally adaptive confidence band
(ii) Design a suitably restricted class of densities tailored to *local* adaptation
(iii) Construct a locally adaptive confidence band
(iv) Prove honesty of the confidence band (calibration)
(v) Analyze the performance of the confidence band

Some notation

- Hölder class to the parameter $\beta > 0$ on the open interval $U \subset \mathbb{R}$

$$\mathcal{H}_U(\beta) = \left\{ p : U \to \mathbb{R} : \|p\|_{\beta,U} < \infty \right\}$$

with Hölder norm

$$\|p\|_{\beta,U} = \sum_{k=0}^{\lfloor\beta\rfloor} \|p^{(k)}\|_U + \sup_{\substack{x,y \in U \\ x \neq y}} \frac{|p^{(\lfloor\beta\rfloor)}(x) - p^{(\lfloor\beta\rfloor)}(y)|}{|x-y|^{\beta-\lfloor\beta\rfloor}} < \infty$$

- Corresponding Hölder ball with radius $L > 0$:

$$\mathcal{H}_U(\beta, L) = \left\{ p : U \to \mathbb{R} : \|p\|_{\beta,U} \leq L \right\}$$

With this definition of $\|\cdot\|_{\beta,U}$, the Hölder balls are nested, that is

$$\mathcal{H}_U(\beta_2, L) \subset \mathcal{H}_U(\beta_1, L)$$

for $0 < \beta_1 \leq \beta_2 < \infty$ and $|U| < 1$

- $\mathcal{H}_U(\infty, L) = \bigcap_{\beta>0} \mathcal{H}_U(\beta, L)$ and $\mathcal{H}_U(\infty) = \bigcap_{\beta>0} \mathcal{H}_U(\beta)$

Locally adaptive confidence bands

- What is locally adaptive confidence band?

Disregarding any measurability issues, we call a confidence band $(C_{n,\alpha}(t))_{t\in[0,1]}$ *locally adaptive* if for every $\varepsilon > 0$ there exists some $c > 0$, such that

$$\sup_{\substack{U\subset[a,b]:\\ U \text{ open interval}}} \limsup_{n\to\infty} \sup_{\substack{p\in\mathcal{P}:\\ p_{|U_\delta}\in\mathcal{H}_{U_\delta}(\beta,L^*)}} \mathbb{P}_p^{\otimes n}\Big(|C_{n,\alpha}(t)| \geq c\cdot r_n(\beta) \ \text{ for some } t\in U\Big) < \varepsilon \tag{1}$$

for any $\delta > 0$, ideally for any β in the range of adaptation.

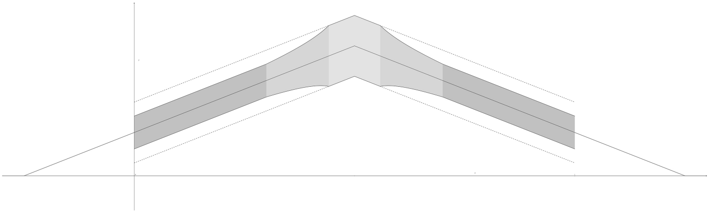
- Typically, $r_n(\beta')/r_n(\beta) \to 0$, whenever $\beta' > \beta$, implying that (1) guarantees significantly tighter confidence bands in case of inhomogeneous smoothness than the corresponding global condition

Locally adaptive confidence bands

Kind of adaptivity that the construction should reveal for the triangular density and for fixed sample size n

Shaded area: sketches the intended locally adaptive confidence band
Dashed line: globally adaptive band

- The triangular density is not smoother than Lipschitz at its maximal point but infinitely smooth at both sides
- The region where globally and locally adaptive confidence bands coincide up to logarithmic factors (light gray regime) should shrink as the sample size increases



Localizing the self-similarity condition

- Design a restricted class of densities $\mathcal{P}$
  - for which both honesty and local adaptivity are achievable
  - which is sufficiently massive for statistical purposes

- A localized self-similarity condition reads as follows:

  For any nondegenerate interval $(u, v) \subset [0, 1]$, there exists some $\beta \in [\beta_*, \beta^*]$ with $p_{|(u,v)} \in \mathcal{P}_{(u,v)}(\beta, L^*)$ and

$$c \cdot h^{\beta} \leq \sup_{s \in (u+h, v-h)} \left| (K_h * p)(s) - p(s) \right| \tag{2}$$

  for all $h \leq h_0 \vee (v - u)$

Remark

Inequality (2) can be satisfied only for

$$\widetilde{\beta} = \widetilde{\beta}_p(U) = \sup \left\{ \beta \in (0, \infty] : p_{|U} \in \mathcal{H}_U(\beta) \right\}$$

Localizing the self-similarity condition

(i) In contrast to the observation in Remark above, for any density p, $\|p''\|_U$ may vanish for subintervals U within the support of p.

$\Rightarrow$ The lower bound condition (2) is violated on such subintervals U for every $\beta \in (0, \beta^*]$. [Recall that the kernel K is symmetric and hence of order $l \geq 1$.]

Example

Assume that the kernel K is of order $l \geq 1$, and recall $\beta^* = l + 1$. Then (2) excludes for instance the triangular density

$$p(t) = \max\{1 - |t - 1/2|, 0\}, \quad t \in \mathbb{R}, \tag{3}$$

because the second derivative exists and vanishes when restricted to any open interval $U \subset [0, 1/2) \cup (1/2, 1]$.

- In general, if p restricted to U is a polynomial of order at most l, (2) is violated as the left-hand side is not equal to zero.

Localizing the self-similarity condition

(ii) For $p \in \mathcal{P}(\beta_*, L)$ and any fixed $h > 0$, the map

$$t \mapsto \|K_{2^{-j}} * p - p\|_{(t-h+2^{-j}, t+h-2^{-j})}$$

is continuous for any natural number j with $2^{-j} < h$. At the same time, the map

$$t \mapsto \sup\left\{ \beta \leq \beta^* : p_{|(t-h,t+h)} \in \mathcal{H}_{(t-h,t+h)}(\beta, L)\right\} \tag{4}$$

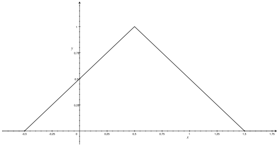
may be discontinuous, in which case the Local self-similarity condition 2 is violated.

Example (continued)

We consider again the triangular density in (3). Then,

$$\sup\left\{ \beta \leq \beta^* : p_{|(t-h,t+h)} \in \mathcal{H}_{(t-h,t+h)}(\beta, 1)\right\} = \begin{cases} 1 & \text{if } t \in \left(\frac{1}{2} - h, \frac{1}{2} + h\right) \\ \beta^* & \text{if } t \in [0,1] \setminus \left(\frac{1}{2} - h, \frac{1}{2} + h\right). \end{cases}$$

Localizing the self-similarity condition



- Localized version rules out examples which seem to be typical to statisticians

In view of these deficiencies, a condition like (2) is insufficient for statistical purposes

Localizing the self-similarity condition

β^*-capped Hölder norm

$$\|p\|_{\beta,\beta^*,U} = \sum_{k=0}^{\lfloor \beta\wedge\beta^*\rfloor} \left\|p^{(k)}\right\|_U + \sup_{\substack{x,y\in U\\ x\neq y}} \frac{\left|p^{(\lfloor \beta\wedge\beta^*\rfloor)}(x) - p^{(\lfloor \beta\wedge\beta^*\rfloor)}(y)\right|}{|x-y|^{\beta-\lfloor \beta\wedge\beta^*\rfloor}},$$

for $\beta > 0$, U bounded and open

- $0 < \beta_1 \leq \beta_2 < \infty$ and $|U| \leq 1$:

$$\|\cdot\|_{\beta_1,\beta^*,U} \leq \|\cdot\|_{\beta_2,\beta^*,U}$$

- $\mathcal{H}_{\beta^*,U}(\beta) = \{p : U \to \mathbb{R} : \|p\|_{\beta,\beta^*,U}$ is well-defined and $< \infty\}$
- $\mathcal{H}_{\beta^*,U}(\infty, L) = \bigcap_{\beta>0} \mathcal{H}_{\beta^*,U}(\beta, L)$
- $\mathcal{H}_{\beta^*,U}(\infty) = \bigcap_{\beta>0} \mathcal{H}_{\beta^*,U}(\beta)$

  $\hookrightarrow$ $\sup\{\beta \in (0,\infty] : p_{|U} \in \mathcal{H}_{\beta^*,U}(\beta)\} \in (0,\beta^*] \cup \{\infty\}$

Admissible densities

Assumption

For sample size $n \in \mathbb{N}$, some $0 < \varepsilon < 1$, $0 < \beta_* < 1$, and $L^* > 0$, a density p is said to be *admissible* if $p \in \mathcal{P}_{(-\varepsilon,1+\varepsilon)}(\beta_*, L^*)$ and the following holds true.

For any $t \in [0,1]$ and for any $h \in \mathcal{G}_\infty$ with

$$\mathcal{G}_\infty = \{2^{-j} : j \in \mathbb{N},\, j \geq j_{\min} = \lceil 2 \vee \log_2(2/\varepsilon) \rceil\},$$

there exists some $\beta \in [\beta_*, \beta^*] \cup \{\infty\}$ such that the following conditions are satisfied for $u = h$ or $u = 2h$:

$$p_{|(t-u,t+u)} \in \mathcal{H}_{\beta^*,(t-u,t+u)}(\beta, L^*) \tag{5}$$

and

$$\sup_{s \in (t-(u-g),t+(u-g))} |(K_g * p)(s) - p(s)| \geq \frac{g^\beta}{\log n} \tag{6}$$

for all $g \in \mathcal{G}_\infty$ with $g \leq u/8$.

Admissible densities

- Passing from the Hölder norm to the β^*-capped Hölder norm enlarges the set of densities under consideration
  - Densities which restricted to $[0,1]$ are described by a polynomial of order at most l are now included
  - The order l is a natural limit because a kernel of order l provides bias-free estimators for polynomials up to the order l, that is, for any $0 < h < 1/2$,

$$\mathbb{E}_p \ \widehat{p}_n(t,h) = p(t), \quad t \in [h, 1-h].$$

- We do not require (5) and (6) to hold for *every* $u = h$ but only for $u = h$ *or* $u = 2h$
  - Essential to incorporate densities with abrupt changes in the smoothness behavior
- $\mathscr{P}_n^{adm} \subset \mathscr{P}_{n+1}^{adm}$, $n \in \mathbb{N}$, permitting smaller and smaller Lipschitz constants

$$\mathscr{P}_n = \left\{ p \in \mathscr{P}_n^{adm} : \inf_{x \in [-\varepsilon, 1+\varepsilon]} p(x) \geq M \right\}$$

Admissible functions

Example (Triangular density)

If K is the rectangular kernel and L^* is sufficiently large, the triangular density is (eventually – for sufficiently large n) admissible.

- It is globally not smoother than Lipschitz, and the bias lower bound condition (6) is (eventually) satisfied for $\beta = 1$ and pairs (t, h) with $|t - 1/2| < (7/8)h$
- Although the bias lower bound condition to the exponent $\beta^* = 2$ is not satisfied for any (t, h) with $t \in [0, 1] \setminus (1/2 - h, 1/2 + h)$, these tuples (t, h) fulfill (5) and (6) for $\beta = \infty$, which is not excluded anymore
- Finally, if the conditions (5) and (6) are not simultaneously satisfied for some pair (t, h) with

$$\frac{1}{2} + \frac{7}{8}h < |t| < \frac{1}{2} + h,$$

then they are fulfilled for the pair $(t, 2h)$ and $\beta = 1$, because

$$|t - 1/2| < (7/8)2h$$

How massive is our set of admissible functions?

- Rectangular kernel $K_R(\cdot) = \frac{1}{2}1\{\cdot \in [-1, 1]\}$

Proposition (Lower pointwise risk bound)

The pointwise minimax rate of convergence remains unchanged when passing from the class $\mathcal{H}(\beta, L^*)$ to $\mathscr{P}_n^{adm} \cap \mathcal{H}(\beta, L^*)$, $\beta < 1$.

Proposition (Set of permanently excluded densities is pathological)

Let

$$\mathcal{R} = \bigcup_{n \in \mathbb{N}} \mathscr{P}_n^{adm}(K_R, \beta_*, L^*, \varepsilon).$$

Then, for any $t \in [0, 1]$, for any $h \in \mathcal{G}_\infty$ and for any $\beta \in [\beta_*, 1)$, the set

$$\mathcal{P}_{(t-h,t+h)}(\beta, L^*) \setminus \mathcal{R}_{|(t-h,t+h)}$$

is nowhere dense in $\mathcal{P}_{(t-h,t+h)}(\beta, L^*)$ with respect to $\| \cdot \|_{\beta,(t-h,t+h)}$.

Specific functions

Lemma

For all $\beta \in (0,1)$, the Weierstraß function W_β satisfies $W_{\beta|U} \in \mathcal{H}_U(\beta, L_W)$ with some Lipschitz constant $L_W = L_W(\beta)$ for every open interval U. For the rectangular kernel K_R and $\beta \in (0,1]$, the Weierstraß function fulfills the bias lower bound condition

$$\sup_{s(t-(h-g),t+h-g)} |(K_{R,g} * W_\beta)(s) - W_\beta(s)| > \left(\frac{4}{\pi} - 1\right) g^\beta$$

for any $t \in \mathbb{R}$ and for any $g, h \in \mathcal{G}_\infty$ with $g \leq h/2$.

Construction of the confidence band

- The new confidence band is based on a kernel density estimator with variable bandwidth incorporating a localized but not the fully pointwise Lepski bandwidth selection procedure
- A suitable discretization of the interval $[a, b]$ and a locally constant approximation of both the density estimator and the (random) bandwidth allow to piece the segmentwise confidence statements together to obtain a continuum of confidence statements over $[a, b]$
- Due to the discretization, the band is computable and feasible from a practical point of view without losing optimality between the mesh points

Construction of the confidence band

- Divide the sample into two parts of equal size $\widetilde{n} = \lfloor n/2 \rfloor \to \widehat{p}_n^{(1)}, \widehat{p}_n^{(2)}$
- The interval $[0,1]$ is discretized into equally spaced grid points

$$T_n := \{k\delta_n : k \in \mathbb{Z}\} \cap [0,1]$$

- Let

$$\mathcal{J}_n := \left\{ j \in \mathbb{N}_0 \;:\; j_{\min} \leq j \leq j_{\max} := \log_2\left(\frac{\widetilde{n}}{(\log \widetilde{n})^{\kappa_2}}\right)\right\}$$

and

$$\mathcal{G}_n := \left\{ 2^{-j} \;:\; j \in \mathcal{J}_n \right\}$$

be the corresponding dyadic grid of bandwidths
- Define the set of admissible bandwidths for each $t \in [0,1]$ as

$$\mathcal{A}_n(t) := \Big\{ j \in \mathcal{J}_n : \max_{s \in (t - c\cdot 2^{-j}, t + c\cdot 2^{-j}) \cap T_n} \left| \widehat{p}_n^{(2)}(s, m') - \widehat{p}_n^{(2)}(s, m) \right| \leq c \sqrt{\frac{\log \widetilde{n}}{\widetilde{n} 2^{-m}}}$$

$$\text{for all } m, m' \in \mathcal{J}_n \text{ with } m > m' > j + 2 \Big\},$$

and choose

$$\widehat{j}_n(t) = \min \mathcal{A}_n(t), \quad t \in [0,1]$$

Construction of the confidence band

- For $t \in I_k := [(k-1)\delta_n, k\delta_n)$, set furthermore

$$\widehat{h}_n^{loc}(t) = \min\left\{2^{-\widehat{j}_n((k-1)\delta_n)},\ 2^{-\widehat{j}_n(k\delta_n)}\right\} 2^{-u_n}$$

  and

$$\widehat{p}_n^{loc}(t,h) := \widehat{p}_n^{(1)}(k\delta_n, h),$$

  where $u_n \asymp \log\log \widetilde{n}$

- Define the class of densities

$$\mathscr{P}_n := \left\{ p \in \mathcal{H}(\beta_*, L^*) : \inf_{[0-\varepsilon, 1+\varepsilon]} p(x) \geq M \right\} \cap \mathscr{P}_n^{adm}$$

Calibration to the level $1 - \alpha$

Theorem (Least favorable case)

For these estimators and normalizing sequences

$$A_n = c_3(-2\log\delta_n)^{1/2}, \quad B_n = \frac{3}{c_3}\left\{(-2\log\delta_n)^{1/2} - \frac{\log(-\log\delta_n) + \log 4\pi}{2(-2\log\delta_n)^{1/2}}\right\},$$

with $c_3 = \sqrt{2}/TV(K)$, it holds

$$\liminf_{n\to\infty} \inf_{p\in\mathscr{P}_n} \mathbb{P}_p^{\otimes n}\left(A_n\left(\sup_{t\in[0,1]} \frac{\left|\widehat{p}_n^{loc}(t, \widehat{h}_n^{loc}(t)) - p(t)\right|}{[\widetilde{n}\widehat{h}_n^{loc}(t)]^{-\frac{1}{2}}} - B_n\right) \le x\right) \geq 2\,\mathbb{P}\left(\sqrt{L^*}G \le x\right) - 1$$

for some standard Gumbel distributed random variable G.

Calibration to the level $1 - \alpha$

- We use non-asymptotic approximation techniques of Chernozhukov, Chetverikov and Kato (2014) to pass over to the supremum over a Gaussian process
- Whereas the related globally adaptive procedure of Giné and Nickl (2010) reduces to the limiting distribution of the supremum of a stationary Gaussian process, our locally adaptive approach leads to a highly non-stationary situation
- This highly non-stationary Gaussian process even depends on the unknown density p
- Identify a stationary process as a least favorable case by means of Slepian's comparison inequality
  $\rightarrow$ not dependent on p any longer

Locally adaptive confidence bands

- Confidence band:

$$C_n^{\alpha}(t) = \Bigg[\widehat{p}_n^{loc}(t, \widehat{h}_n^{loc}(t)) - \left(\frac{q_{1-\frac{\alpha}{2}}}{A_n} + B_n\right)\left(\widetilde{n}\widehat{h}_n^{loc}(t)\right)^{-\frac{1}{2}}, \\ \widehat{p}_n^{loc}(t, \widehat{h}_n^{loc}(t)) + \left(\frac{q_{1-\frac{\alpha}{2}}}{A_n} + B_n\right)\left(\widetilde{n}\widehat{h}_n^{loc}(t)\right)^{-\frac{1}{2}}\Bigg]$$

Corollary (Honesty)

The confidence band satisfies

$$\liminf_{n\to\infty} \inf_{p\in\mathscr{P}_n} \mathbb{P}_p^{\otimes n}\Big(p(t) \in C_n^{\alpha}(t) \text{ for every } \ t \in [0,1]\Big) \geq 1-\alpha.$$

Locally adaptive confidence bands

Performance

Theorem (Local adaptivity)

For every open interval $U \subset [0,1]$, and for any $\delta > 0$,

$$\limsup_{n\to\infty} \sup_{\substack{p\in\mathscr{P}_n:\\ p_{|U_\delta}\in\mathcal{H}_{U_\delta}(\beta,L^*)}} \mathbb{P}_p\left(\sup_{t\in U}|C_{n,\alpha}(t)| \geq \left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+1}}(\log n)^\gamma\right) = 0$$

for every $\beta \in [\beta_*, \beta^*]$ and $\gamma = \gamma(c_1)$, where U_δ is the open δ-enlargement of U.

- Recall that the minimax-rate of convergence over $\mathscr{P}_{n|U_\delta} \cap \mathcal{H}_{\beta^*,U_\delta}(\beta, L^*)$ remains $n^{-\beta/(2\beta+1)}$

Locally adaptive confidence bands

Performance

Let $p \in \mathcal{H}(\beta, L)$ and $p_{|U_\delta} \in \mathcal{H}_{U_\delta}(\beta', L)$ for some $\beta' > \beta$ and some open interval $U \subset [0,1]$

- **Locally adaptive confidence bands:** the maximal width over U is of the stochastic order (up to logarithmic factors)

$$\mathcal{O}_{\mathbb{P}_p}\left(n^{-\frac{\beta'}{2\beta'+1}}\right)$$

- **Globally adaptive confidence bands:** only guarantee a width of stochastic order (up to logarithmic factors)

$$\mathcal{O}_{\mathbb{P}_p}\left(n^{-\frac{\beta}{2\beta+1}}\right)$$

Locally adaptive confidence bands

Performance

- Can we formulate the asymptotic statement not only for arbitrary but fixed intervals?
- The more observations are available the more localized and smaller are regions the statistician would like to learn about
- Is it even possible to adapt to some "local Hölder exponent"?

Locally adaptive confidence bands

Performance

Theorem

Attaining the minimax rates of convergence corresponding to the pointwise or local Hölder exponent (possibly inflated by some logarithmic factor) uniformly over $\mathscr{P}_n$ is an unachievable goal.

Is our confidence band adaptive to some notion of local Hölder regularity nevertheless?

- Set $h_{\beta,n} = 2^{-j_{\min}} \cdot \left(\frac{\log n}{n}\right)^{\frac{1}{2\beta+1}}$ (optimal adaptive bandwidth within $\mathcal{H}(\beta, L^*)$)

n-dependent local Hölder exponent

Define

$$\beta_{n,p}(t) = \sup\left\{\beta > 0 : \|p\|_{\beta,\beta^*,(t-h_{\beta,n},t+h_{\beta,n})} \le L^*\right\}.$$

(If the supremum is running over the empty set, we set $\beta_{n,p}(t) = 0$.)

- Roughly speaking, we intend the local Hölder exponent to be the maximal β such that the density attains this Hölder exponent within $(t - h_{\beta,n}, t + h_{\beta,n})$

Locally adaptive confidence bands

Performance

Theorem (Strong local adaptivity)

There exists some $\gamma = \gamma(c_1)$, such that

$$\limsup_{n\to\infty} \sup_{p\in\mathscr{P}_n} \mathbb{P}_p\left(\sup_{t\in[0,1]} |C_{n,\alpha}(t)| \cdot \left(\frac{\log n}{n}\right)^{-\frac{\beta_{n,p}(t)}{2\beta_{n,p}(t)+1}} \geq (\log n)^\gamma \right) = 0.$$

- Note that $\beta_{n,p}(t) = \infty$ is not excluded
- The confidence band attains even adaptively the parametric width $n^{-1/2}$ (up to logarithmic factors) if p can be locally represented as a polynomial of degree strictly less than β^*

Locally adaptive confidence bands

Performance

- The stochastic order of the width of our confidence band is (up to logarithmic factors) given by

$$n^{-\frac{\beta_{n,p}(t)}{2\beta_{n,p}(t)+1}}, \quad t \in [0,1],$$

Example: triangular density $p(t) = \max\{1 - |t - 1/2|, 0\}$, $L^* \geq 2$

- If $|t - 1/2\,| \geq 2^{-j_{\min}} \longrightarrow \beta_{n,p}(t) = \infty$ and $n^{-\frac{\beta_{n,p}(t)}{2\beta_{n,p}(t)+1}} = n^{-1/2}$
- If $|t - 1/2\,| \leq h_{1,n} \;\; \longrightarrow \beta_{n,p}(t) = 1$ and $n^{-\frac{\beta_{n,p}(t)}{2\beta_{n,p}(t)+1}} = n^{-1/3}$
- If $h_{1,n} < |t - 1/2\,| < 2^{-j_{\min}} \longrightarrow \beta_{n,p}(t)$ is the solution of $h_{\beta,n} = |t - 1/2|$

Locally adaptive confidence bands

Performance

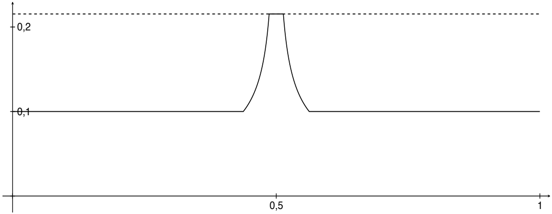


Abbildung: $n^{-\frac{\beta_{n,p}(t)}{2\beta_{n,p}(t)+1}}$ and $n^{-1/3}$, $t \in [0,1]$, $n = 200$, $j_{\min} = 4$

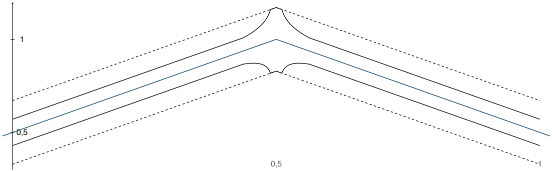


Abbildung: $p(t) \pm n^{-\frac{\beta_{n,p}(t)}{2\beta_{n,p}(t)+1}}$ and $p(t) \pm n^{-1/3}$, $t \in [0,1]$, $n = 200$, $j_{\min} = 4$

Thank you for your attention!

Localizing the self-similarity condition

Remark

(i) There exist functions $p : U \to \mathbb{R}$, $U \subset \mathbb{R}$ some interval, which are *not* Hölder continuous to their exponent $\widetilde{\beta}$.

Example:

$$W_1(\cdot) = \sum_{n=0}^{\infty} 2^{-n} \cos\left(2^n \pi \cdot\right) \quad \text{(Weierstraß function)}$$

Hardy (1916):

$$W_1(x+h) - W_1(x) = O\left(|h| \log\left(1/|h|\right)\right),$$

which implies the Hölder continuity to any parameter $\beta < 1$, hence $\widetilde{\beta} \geq 1$.

Hardy shows in the same reference that W_1 is nowhere differentiable, meaning that it cannot be Lipschitz continuous, that is $\widetilde{\beta} = 1$ but $W_1 \notin \mathcal{H}_U(\widetilde{\beta})$.

Localizing the self-similarity condition

Remark

(ii) It can also happen that $p_{|U} \in \mathcal{H}_U(\widetilde{\beta})$ but

$$\limsup_{\delta \to 0} \sup_{\substack{|x-y| \leq \delta \\ x,y \in U}} \frac{|p^{(\lfloor \widetilde{\beta} \rfloor)}(x) - p^{(\lfloor \widetilde{\beta} \rfloor)}(y)|}{|x-y|^{\widetilde{\beta} - \lfloor \widetilde{\beta} \rfloor}} = 0,$$

meaning that (2) is violated. In the analysis literature, the subset of functions in $\mathcal{H}_U(\widetilde{\beta})$ with this property is called little Lipschitz (or little Hölder) space.

As a complement of an open and dense set, it forms a nowhere dense subset of $\mathcal{H}_U(\widetilde{\beta})$.

Poiwise Hölder exponent

Pointwise Hölder exponent, Seuret and Lévy-Véhel

Let $p : \mathbb{R} \to \mathbb{R}$ be a function, $\beta > 0$, $\beta \notin \mathbb{N}$, and $t \in \mathbb{R}$. Then $p \in \mathcal{H}_t(\beta)$ if and only if there exists a real $R > 0$, a polynomial P with degree less than $\lfloor \beta \rfloor$, and a constant c such that

$$|p(x) - P(x-t)| \leq c|x-t|^{\beta}$$

for all $x \in (t-R, t+R)$. The pointwise Hölder exponent is denoted by

$$\beta_p(t) = \sup\{\beta : p \in \mathcal{H}_t(\beta)\}.$$

Local Hölder exponent

Local Hölder exponent, Seuret and Lévy-Véhel (2002)

Let $p : \Omega \to \mathbb{R}$ be a function and $\Omega \subset \mathbb{R}$ an open set. One classically says that $p \in \mathcal{H}_{loc}(\beta, \Omega)$, where $0 < \beta < 1$, if there exists a constant c such that

$$|p(x) - p(y)| \leq c|x - y|^{\beta}$$

for all $x, y \in \Omega$. If $m < \beta < m + 1$ for some $m \in \mathbb{N}$, then $p \in \mathcal{H}_{loc}(\beta, \Omega)$ means that there exists a constant c such that

$$|\partial^m p(x) - \partial^m p(y)| \leq c|x - y|^{\beta - m}$$

for all $x, y \in \Omega$. Set now

$$\beta_p(\Omega) = \sup\{\beta : p \in \mathcal{H}_{loc}(\beta, \Omega)\}.$$

Finally, the local Hölder exponent in t is defined as

$$\beta_p^{loc}(t) = \sup\{\beta_p(O_i) : i \in I\},$$

where $(O_i)_{i \in I}$ is a decreasing family of open sets with $\cap_{i \in I} O_i = \{t\}$. [By Lemma 2.1 in Seuret and Lévy-Véhel (2002), this notion is well defined, that is, it does not depend on the particular choice of the decreasing sequence of open sets.]