Nonparametric regression with irregular error distributions - an extreme value approach

(work in progress)

Holger Drees, Natalie Neumeyer, Leonie Selk

supported by DFG via Research Unit 1735

University of Hamburg, Department of Mathematics

Symposium on
Recent Advances in Extreme Value Theory
Lisbon, March 19, 2013

Outline

1. Nonparametric regression models
   - Nonparametric mean and endpoint regression
   - (Ir)regularity of models
2. Regression estimators based on local maxima
   - Naive estimators
   - Local linear estimator
   - The general case
3. Application to analysis of error distribution

Nonparametric mean regression models

$$Y_i = g(x_i) + \varepsilon_i, \qquad 1 \le i \le n,$$

ε_i iid errors with $E(\varepsilon_i) = 0$

g “smooth” mean regression function

standard nonparametric regression estimators: local averages

$$\hat{g}(x) = \sum_i w(x - x_i) Y_i$$

with $w(t) \begin{cases} \text{large} \\ \text{small} \end{cases}$ if $|t| \begin{cases} \text{small} \\ \text{large} \end{cases}$

Estimators “optimal” if distribution of ε_i “regular”.

Nonparametric mean regression models

$$Y_i = g(x_i) + \varepsilon_i, \qquad 1 \le i \le n,$$

ε_i iid errors with $E(\varepsilon_i) = 0$

g "smooth" mean regression function

standard nonparametric regression estimators: local averages

$$\hat{g}(x) = \sum_i w(x - x_i) Y_i$$

with $w(t) \begin{cases} \text{large} \\ \text{small} \end{cases}$ if $|t| \begin{cases} \text{small} \\ \text{large} \end{cases}$

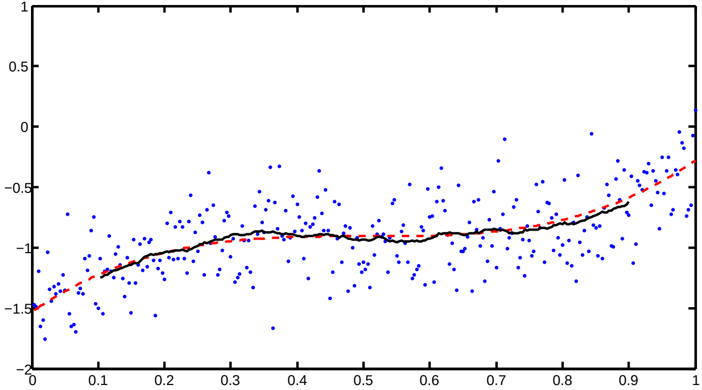
Estimators "optimal" if distribution of ε_i "regular".

Nonparametric mean regression models: Example

$\varepsilon_i = \tilde{\varepsilon}_i - E(\tilde{\varepsilon}_i)$ where $\tilde{\varepsilon}_i$ have cdf

$$F(y) = \mathrm{e}^{-|y|^{\alpha}}, \qquad y \le 0, \quad \alpha = 4$$

$g(x) = 5\left(x - \frac{1}{2}\right)^3$



Nonparametric boundary regression models

$$Y_i = g(x_i) + \varepsilon_i, \qquad 1 \le i \le n,$$

ε_i iid errors with right endpoint of support equal 0

g "smooth" regression function, describing right endpoint of support of Y_i

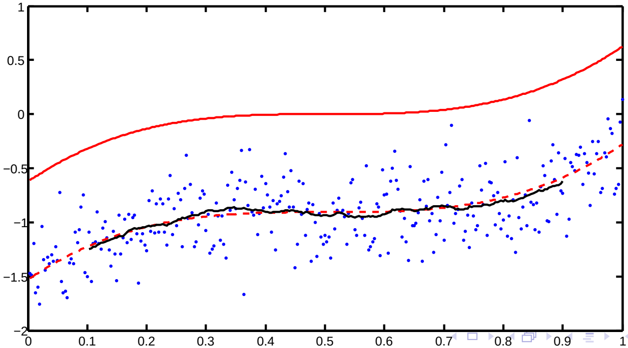
Nonparametric boundary regression models

$$Y_i = g(x_i) + \varepsilon_i, \qquad 1 \leq i \leq n,$$

ε_i iid errors with right endpoint of support equal 0

g "smooth" regression function, describing right endpoint of support of Y_i

Example: ε_i with cdf $F(y) = \mathrm{e}^{-|y|^\alpha}$, $\qquad y \leq 0, \quad \alpha = 4$



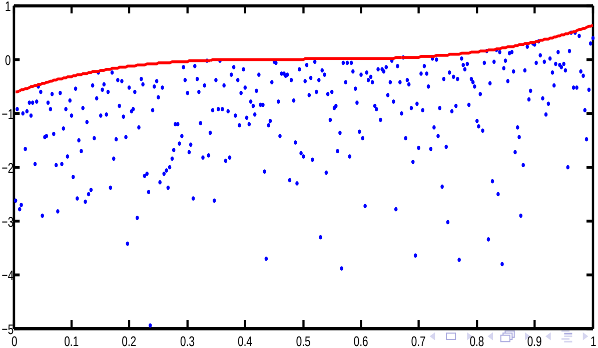
Nonparametric boundary regression models

$$Y_i = g(x_i) + \varepsilon_i, \qquad 1 \le i \le n,$$

ε_i iid errors with right endpoint of support equal 0

g "smooth" regression function, describing right endpoint of support of Y_i

Example: ε_i with cdf $F(y) = \mathrm{e}^{-|y|^\alpha}, \qquad y \le 0, \quad \alpha = 1$



Model with equidistant design

$$Y_i = g\left(\frac{i}{n}\right) + \varepsilon_i, \qquad 1 \le i \le n,$$

ε_i iid with cdf $\quad F(y) = 1 - c|y|^\alpha + o\big(|y|^\alpha\big)$

α controls regularity:

$\alpha > 2$: regular model,
optimal estimators can be constructed from local averages (if error distribution known);
classical nonparametric rates of convergence

$\alpha < 2$: irregular model,
estimators based on local extrema yield better rates of convergence

$\alpha = 2$ critical value separating realm of regular models from irregular models

Model with equidistant design

$$Y_i = g\left(\frac{i}{n}\right) + \varepsilon_i, \qquad 1 \le i \le n,$$

ε_i iid with cdf $\quad F(y) = 1 - c|y|^\alpha + o\big(|y|^\alpha\big)$

α controls regularity:

$\alpha > 2$: regular model,
optimal estimators can be constructed from local averages (if error distribution known);
classical nonparametric rates of convergence

$\alpha < 2$: irregular model,
estimators based on local extrema yield better rates of convergence

$\alpha = 2$ critical value separating realm of regular models from irregular models

(Ir)regularity from EVT point of view

$$Y_i = g\left(\frac{i}{n}\right) + \varepsilon_i, \qquad 1 \le i \le n,$$

ε_i iid with cdf $\quad F(y) = 1 - c|y|^\alpha + o\big(|y|^\alpha\big)$

$\alpha > 2$	$\alpha < 2$
ML estimator of $\gamma = 1/\alpha$ asymptotically normal	ML estimator of γ not as. normal ($1 \le \alpha < 2$) or not even defined ($\alpha < 1$)
extreme quantile estimators based on EVT approximation outperform empirical quantiles	extreme quantile est. based on EVT approx. does not work well; empirical quantiles converge at faster rate

(Ir)regularity in more general experiments

General parametric model with observations

$$Y_i \text{ iid with density } f_\vartheta, \quad \vartheta \in \Theta \subset \mathbb{R}^d$$

w.r.t. some σ-finite measure μ.

Consider

$$s_\vartheta := \sqrt{f_\vartheta}$$

as element of Banach space $L_2(\mu)$.

Model is called regular if $\vartheta \mapsto s_\vartheta$ differentiable in $L_2(\mu)$.

Then

- ML estimator behaves nicely,
- characterization of optimal estimators and tests (in minimax sense or in sense of convolution theorem) well known.

(Ir)regularity in more general experiments

General parametric model with observations

$$Y_i \text{ iid with density } f_\vartheta, \quad \vartheta \in \Theta \subset \mathbb{R}^d$$

w.r.t. some σ-finite measure μ.

Consider

$$s_\vartheta := \sqrt{f_\vartheta}$$

as element of Banach space $L_2(\mu)$.

Model is called regular if $\vartheta \mapsto s_\vartheta$ differentiable in $L_2(\mu)$.

Then

- ML estimator behaves nicely,
- characterization of optimal estimators and tests (in minimax sense or in sense of convolution theorem) well known.

(Ir)regularity in location models

Special case:

$$Y_i = \vartheta + \varepsilon_i, \qquad 1 \leq i \leq n,$$

ε_i iid with density $f \quad \Longrightarrow \quad Y_i$ iid with density $f_\vartheta = f(\cdot - \vartheta)$

Model regular if f is sufficiently "smooth"

- regular models:
  - normal distribution
  - Weibull $f(y) = \alpha \exp(-|y|^{\alpha+1}) 1_{(-\infty,0)}(y)$ for $\alpha > 2$
- irregular models:
  - uniform distribution on $[\vartheta, 1+\vartheta]$ (or $[0,\vartheta]$ or $[-\vartheta,\vartheta]$)
  - Weibull $f(y) = \alpha \exp(-|y|^{\alpha+1}) 1_{(-\infty,0)}(y)$ for $\alpha \leq 2$

(Ir)regularity in location models

Special case:

$$Y_i = \vartheta + \varepsilon_i, \qquad 1 \leq i \leq n,$$

ε_i iid with density $f \quad \Longrightarrow \quad Y_i$ iid with density $f_\vartheta = f(\cdot - \vartheta)$

Model regular if f is sufficiently "smooth"

- regular models:
  - normal distribution
  - Weibull $f(y) = \alpha \exp(-|y|^{\alpha+1}) 1_{(-\infty,0)}(y)$ for $\alpha > 2$
- irregular models:
  - uniform distribution on $[\vartheta, 1+\vartheta]$ (or $[0,\vartheta]$ or $[-\vartheta,\vartheta]$)
  - Weibull $f(y) = \alpha \exp(-|y|^{\alpha+1}) 1_{(-\infty,0)}(y)$ for $\alpha \leq 2$

Nonparametric endpoint regression models

$$Y_i = g(x_i) + \varepsilon_i, \qquad 1 \leq i \leq n,$$

If g smooth, then regression model locally around fixed x similar to location model.

Heuristically, concept of (ir)regularity carries over.

Hence for

$$\varepsilon_i \text{ iid with cdf} \qquad F(y) = 1 - c|y|^\alpha + o\big(|y|^\alpha\big)$$

leads to

- regular model for $\alpha > 2$
- irregular model for $\alpha < 2$

Here mainly interested in uniform estimation of g in irregular model

Related literature

- Hall, Van Keilegom (2009): minimax rates (under quadratic loss) and minimax rate optimal estimators of $g(x)$ for fixed x
- Müller, Wefelmeyer (2010): similar results if error distribution is symmetric
- estimation of g based on iid observations (X_i, Y_i) with support $\{(x, y) \in [0, 1] \times [0, \infty) \mid y \leq g(x)\}$:

  Härdle, Park, Tsybakov (1995): special case of $\alpha = 1$, L_1-type loss
  Hall, Nussbaum, Stern (1997): conditional density with power behavior at boundary; asymptotics only for x fixed
- similar papers in Poisson process model
- Huge literature for models with monotone boundary function g (frontier estimation); problem much easier then.

Related literature

- Hall, Van Keilegom (2009): minimax rates (under quadratic loss) and minimax rate optimal estimators of $g(x)$ for fixed x
- Müller, Wefelmeyer (2010): similar results if error distribution is symmetric
- estimation of g based on iid observations (X_i, Y_i) with support $\{(x, y) \in [0, 1] \times [0, \infty) \mid y \leq g(x)\}$:

  Härdle, Park, Tsybakov (1995): special case of $\alpha = 1$, L_1-type loss
  Hall, Nussbaum, Stern (1997): conditional density with power behavior at boundary; asymptotics only for x fixed
- similar papers in Poisson process model
- Huge literature for models with monotone boundary function g (frontier estimation); problem much easier then.

Related literature

- Hall, Van Keilegom (2009): minimax rates (under quadratic loss) and minimax rate optimal estimators of $g(x)$ for fixed x
- Müller, Wefelmeyer (2010): similar results if error distribution is symmetric
- estimation of g based on iid observations (X_i, Y_i) with support $\{(x,y) \in [0,1] \times [0,\infty) \mid y \leq g(x)\}$:

  Härdle, Park, Tsybakov (1995): special case of $\alpha = 1$, L_1-type loss
  Hall, Nussbaum, Stern (1997): conditional density with power behavior at boundary; asymptotics only for x fixed
- similar papers in Poisson process model
- Huge literature for models with monotone boundary function g (frontier estimation); problem much easier then.

Naive estimator

Assume that regression fct. g belongs to following Hölder class

$$\mathcal{H}_{\beta,L} := \Big\{h : [0,1] \to \mathbb{R} \mid h \text{ is } s := \lfloor \beta - \rfloor \text{ times differentiable}, \; |h^{(s)}(x) - h^{(s)}(y)| \le L|x-y|^{\beta-s} \; \forall x, y \in [0,1]\Big\}$$

Approximate g locally in neighborhood of x by $g(x)$.

Then

$$Y_i \approx g(x) + \varepsilon_i$$

for all $i \in I_n(x) := \big\{i \in \{1, \ldots, n\} \mid |i/n - x| \le h_n\big\}$ for suitable $h_n \downarrow 0$.

Because $\varepsilon_i < 0$ but $\max_{i \in I_n(x)} \varepsilon_i \to 0$ in probability if $|I_n(x)| \sim 2nh_n \to \infty$, define

$$\hat{g}_n(x) := \max_{i \in I_n(x)} Y_i$$

Naive estimator

Assume that regression fct. g belongs to following Hölder class

$$\begin{aligned} \mathcal{H}_{\beta,L} \quad := \quad & \big\{ h : [0,1] \to \mathbb{R} \mid h \text{ is } s := \lfloor \beta - \rfloor \text{ times differentiable,} \\ & |h^{(s)}(x) - h^{(s)}(y)| \le L|x-y|^{\beta - s} \; \forall x, y \in [0,1] \big\} \end{aligned}$$

Approximate g locally in neighborhood of x by $g(x)$.

Then

$$Y_i \approx g(x) + \varepsilon_i$$

for all $i \in I_n(x) := \big\{ i \in \{1, \ldots, n\} \mid |i/n - x| \le h_n \big\}$ for suitable $h_n \downarrow 0$.

Because $\varepsilon_i < 0$ but $\max_{i \in I_n(x)} \varepsilon_i \to 0$ in probability if $|I_n(x)| \sim 2nh_n \to \infty$, define

$$\hat{g}_n(x) := \max_{i \in I_n(x)} Y_i$$

Naive estimator: uniform asymptotics

$$\hat{g}_n(x) := \max_{i \in I_n(x)} Y_i \quad \text{with} \quad I_n(x) := \big\{ i \in \{1, \dots, n\} \mid |i/n - x| \le h_n \big\}$$

Theorem

If $h_n \to 0$, $nh_n \to \infty$, then

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}_n(x) - g(x)| = O(h_n^\beta) + O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{1/\alpha}\right)$$

uniformly for all $g \in \mathcal{H}_{\beta,L}$ for all fixed $\beta \in (0,1]$, $L > 0$.

Bandwidth $h_n = (n/\log n)^{-1/(\alpha\beta+1)}$ leads to optimal rate $(n/\log n)^{-\beta/(\alpha\beta+1)}$;
faster than usual nonparametric rate $n^{-\beta/(2\beta+1)}(\log n)^\tau$ if $\alpha < 2$

Naive estimator: uniform asymptotics

$$\hat{g}_n(x) := \max_{i \in I_n(x)} Y_i \quad \text{with} \quad I_n(x) := \big\{ i \in \{1, \ldots, n\} \mid |i/n - x| \le h_n \big\}$$

Theorem

If $h_n \to 0$, $nh_n \to \infty$, then

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}_n(x) - g(x)| = O(h_n^\beta) + O_P\Big(\Big(\frac{|\log h_n|}{nh_n}\Big)^{1/\alpha}\Big)$$

uniformly for all $g \in \mathcal{H}_{\beta,L}$ for all fixed $\beta \in (0,1]$, $L > 0$.

Bandwidth $h_n = (n/\log n)^{-1/(\alpha\beta+1)}$ leads to optimal rate $(n/\log n)^{-\beta/(\alpha\beta+1)}$;
faster than usual nonparametric rate $n^{-\beta/(2\beta+1)}(\log n)^\tau$ if $\alpha < 2$

Proof

$$
\begin{aligned}
|\hat{g}_n(x) - g(x)| &= \Big| \max_{i \in I_n(x)} g(i/n) + \varepsilon_i - g(x) \Big| \\
&\leq \max_{i \in I_n(x)} |g(i/n) - g(x)| + \min_{i \in I_n(x)} |\varepsilon_i| \\
&= O(h_n^\beta) + \min_{i \in I_n(x)} |\varepsilon_i|
\end{aligned}
$$

uniformly in x. Let $k_n = \lfloor 2nh_n \rfloor - 1, \quad l_n = \lfloor n/k_n \rfloor \sim 1/(2h_n)$.

$$
\begin{aligned}
& P\Big\{ \max_{x \in [h_n, 1-h_n]} \min_{i \in I_n(x)} |\varepsilon_i| > u \Big\} \\
\leq \;& P\Big\{ \max_{1 \leq j \leq n} \min_{j \leq i \leq j+k_n} |\varepsilon_i| > u \Big\} \\
\leq \;& P\Big\{ \max_{0 \leq l \leq l_n} M_{n,l} > u \Big\} \\
\leq \;& P\Big\{ \max_{0 \leq l \leq l_n, l \text{ even}} M_{n,l} > u \Big\} + P\Big\{ \max_{0 \leq l \leq l_n, l \text{ odd}} M_{n,l} > u \Big\}
\end{aligned}
$$

with $\quad M_{n,l} := \max_{lk_n < j \leq (l+1)k_n} \min_{j \leq i \leq j+k_n} |\varepsilon_i|$.

Proof

$$\begin{aligned} |\hat{g}_n(x) - g(x)| &= \left| \max_{i \in I_n(x)} g(i/n) + \varepsilon_i - g(x) \right| \\ &\le \max_{i \in I_n(x)} |g(i/n) - g(x)| + \min_{i \in I_n(x)} |\varepsilon_i| \\ &= O(h_n^\beta) + \min_{i \in I_n(x)} |\varepsilon_i| \end{aligned}$$

uniformly in x. Let $k_n = \lfloor 2nh_n \rfloor - 1, \quad l_n = \lfloor n/k_n \rfloor \sim 1/(2h_n)$.

$$\begin{aligned} & P\Big\{ \max_{x \in [h_n, 1-h_n]} \min_{i \in I_n(x)} |\varepsilon_i| > u \Big\} \\ \le \quad & P\Big\{ \max_{1 \le j \le n} \min_{j \le i \le j+k_n} |\varepsilon_i| > u \Big\} \\ \le \quad & P\Big\{ \max_{0 \le l \le l_n} M_{n,l} > u \Big\} \\ \le \quad & P\Big\{ \max_{0 \le l \le l_n, l \text{ even}} M_{n,l} > u \Big\} + P\Big\{ \max_{0 \le l \le l_n, l \text{ odd}} M_{n,l} > u \Big\} \end{aligned}$$

with $\quad M_{n,l} := \max_{lk_n < j \le (l+1)k_n} \min_{j \le i \le j+k_n} |\varepsilon_i|$.

Proof (cont.)

Because $M_{n,l} = \max_{lk_n < j \le (l+1)k_n} \min_{j \le i \le j+k_n} |\varepsilon_i|$ are iid for l even

$$P\Big\{ \max_{0 \le l \le l_n, l \text{ even}} M_{n,l} > u \Big\} = 1 - \big(P\{M_{n,0} \le u\}\big)^{\lfloor l_n/2 \rfloor + 1},$$

and likewise for maximum over odd l.

If $M_{n,0} > u$, then consider first j such that $\min_{j \le i \le j+k_n} |\varepsilon_i| > u$:

$$\begin{aligned} P\{M_{n,0} > u\} &= P\Big\{ \min_{1 \le i \le 1+k_n} |\varepsilon_i| > u \Big\} + \sum_{j=2}^{k_n} P\Big\{ |\varepsilon_{j-1}| \le u, \min_{j \le i \le j+k_n} |\varepsilon_i| > u \Big\} \\ &= (1 - F_{|\varepsilon|}(u))^{k_n+1} + (k_n - 1) F_{|\varepsilon|}(u) \big(1 - F_{|\varepsilon|}(u)\big)^{k_n+1}. \end{aligned}$$

Combine everything to obtain

$$P\Big\{ \sup_{x \in [h_n, 1-h_n]} \min_{i \in I_n(x)} |\varepsilon_i| > u \Big\} \le 2\Big(1 - \Big(1 - (1 + k_n F_{|\varepsilon|}(u))(1 - F_{|\varepsilon|}(u))^{k_n}\Big)^{\lfloor l_n/2 \rfloor + 1}\Big)$$

from which the assertion easily follows.

Proof (cont.)

Because $M_{n,l} = \max_{lk_n < j \le (l+1)k_n} \min_{j \le i \le j+k_n} |\varepsilon_i|$ are iid for l even

$$P\Big\{ \max_{0 \le l \le l_n, l \text{ even}} M_{n,l} > u \Big\} = 1 - \big(P\{M_{n,0} \le u\} \big)^{\lfloor l_n/2 \rfloor + 1},$$

and likewise for maximum over odd l.
If $M_{n,0} > u$, then consider first j such that $\min_{j \le i \le j+k_n} |\varepsilon_i| > u$:

$$\begin{aligned} P\{M_{n,0} > u\} &= P\Big\{ \min_{1 \le i \le 1+k_n} |\varepsilon_i| > u \Big\} + \sum_{j=2}^{k_n} P\Big\{ |\varepsilon_{j-1}| \le u, \min_{j \le i \le j+k_n} |\varepsilon_i| > u \Big\} \\ &= (1 - F_{|\varepsilon|}(u))^{k_n+1} + (k_n - 1) F_{|\varepsilon|}(u) \big(1 - F_{|\varepsilon|}(u)\big)^{k_n+1}. \end{aligned}$$

Combine everything to obtain

$$P\Big\{ \sup_{x \in [h_n, 1-h_n]} \min_{i \in I_n(x)} |\varepsilon_i| > u \Big\} \le 2 \bigg(1 - \Big(1 - (1 + k_n F_{|\varepsilon|}(u))(1 - F_{|\varepsilon|}(u))^{k_n} \Big)^{\lfloor l_n/2 \rfloor + 1} \bigg)$$

from which the assertion easily follows.

Proof (cont.)

Because $M_{n,l} = \max_{lk_n < j \le (l+1)k_n} \min_{j \le i \le j+k_n} |\varepsilon_i|$ are iid for l even

$$P\Big\{ \max_{0 \le l \le l_n, l \text{ even}} M_{n,l} > u \Big\} = 1 - \big(P\{ M_{n,0} \le u \} \big)^{\lfloor l_n/2 \rfloor + 1},$$

and likewise for maximum over odd l.
If $M_{n,0} > u$, then consider first j such that $\min_{j \le i \le j+k_n} |\varepsilon_i| > u$:

$$\begin{aligned} P\{M_{n,0} > u\} &= P\Big\{ \min_{1 \le i \le 1+k_n} |\varepsilon_i| > u \Big\} + \sum_{j=2}^{k_n} P\Big\{ |\varepsilon_{j-1}| \le u, \min_{j \le i \le j+k_n} |\varepsilon_i| > u \Big\} \\ &= (1 - F_{|\varepsilon|}(u))^{k_n+1} + (k_n - 1) F_{|\varepsilon|}(u) \big(1 - F_{|\varepsilon|}(u)\big)^{k_n+1}. \end{aligned}$$

Combine everything to obtain

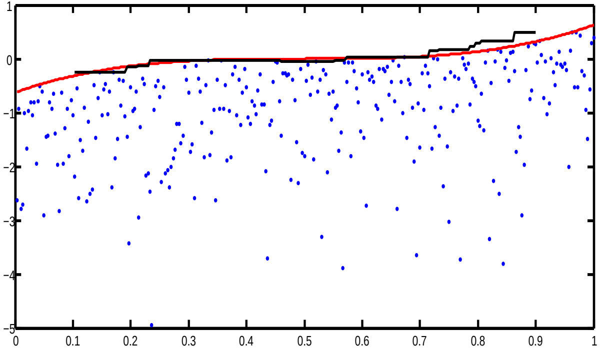
$$P\Big\{ \sup_{x \in [h_n, 1-h_n]} \min_{i \in I_n(x)} |\varepsilon_i| > u \Big\} \le 2 \bigg(1 - \Big(1 - (1 + k_n F_{|\varepsilon|}(u))(1 - F_{|\varepsilon|}(u))^{k_n} \Big)^{\lfloor l_n/2 \rfloor + 1} \bigg)$$

from which the assertion easily follows.

Naive estimator

Remark: Proof yields more precise asymptotics (not needed in what follows)

Example: ε_i with cdf $F(y) = e^{-|y|^\alpha}$, $y \leq 0$, $\alpha = 1$



Performance not so good where function changes more quickly

Local linear approximation

If regression function g belongs to $\mathcal{H}_{\beta,L}$ for some $\beta > 1$ (i.e. it is smoother), then it can be locally approximated by linear function $g(t) \approx a_0 + a_1(t - x)$ for $|x - t| \leq h_n$;

approximation more accurate on same neighborhood of x, resp. larger neighborhood can be used to obtain same accuracy:

$$Y_i \approx a_0 + a_1\left(\frac{i}{n} - x\right) + \varepsilon_i \quad \text{if} \quad \left|\frac{i}{n} - x\right| \leq h_n,$$

where h_n can now be chosen larger.

Similar as before this leads to optimization problem to minimize a_0 under the constraints $a_0 + a_1\left(\frac{i}{n} - x\right) > Y_i$, i.e.

$$\hat{g}_n(x) := \min\left\{a_0 \in \mathbb{R} \mid \exists a_1 \in \mathbb{R} \text{ s.t. } a_0 + a_1\left(\frac{i}{n} - x\right) > Y_i \ \forall i \in I_n(x)\right\}.$$

Local linear approximation

If regression function g belongs to $\mathcal{H}_{\beta,L}$ for some $\beta > 1$ (i.e. it is smoother), then it can be locally approximated by linear function $g(t) \approx a_0 + a_1(t - x)$ for $|x - t| \leq h_n$;

approximation more accurate on same neighborhood of x, resp. larger neighborhood can be used to obtain same accuracy:

$$Y_i \approx a_0 + a_1\left(\frac{i}{n} - x\right) + \varepsilon_i \quad \text{if} \quad \left|\frac{i}{n} - x\right| \leq h_n,$$

where h_n can now be chosen larger.

Similar as before this leads to optimization problem to minimize a_0 under the constraints $a_0 + a_1\left(\frac{i}{n} - x\right) > Y_i$, i.e.

$$\hat{g}_n(x) := \min\left\{a_0 \in \mathbb{R} \mid \exists a_1 \in \mathbb{R} \text{ s.t. } a_0 + a_1\left(\frac{i}{n} - x\right) > Y_i \; \forall i \in I_n(x)\right\}.$$

Uniform asymptotics of improved estimator

$$\hat{g}_n(x) := \min\left\{a_0 \in \mathbb{R} \mid \exists a_1 \in \mathbb{R} \text{ s.t. } a_0 + a_1\left(\frac{i}{n} - x\right) > Y_i \; \forall i \in I_n(x)\right\}.$$

Theorem

If $h_n \to 0$, $nh_n \to \infty$, then

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}_n(x) - g(x)| = O(h_n^\beta) + O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{1/\alpha}\right)$$

uniformly for all $g \in \mathcal{H}_{\beta,L}$ for all fixed $\beta \in (1,2]$, $L > 0$.

Bandwidth $h_n = (n/\log n)^{-1/(\alpha\beta+1)}$ leads to optimal rate $(n/\log n)^{-\beta/(\alpha\beta+1)}$;
naive estimator yields slower rate $(n/\log n)^{-1/(\alpha+1)}$ (for optimal choice of h_n)

Uniform asymptotics of improved estimator

$$\hat{g}_n(x) := \min\left\{a_0 \in \mathbb{R} \mid \exists a_1 \in \mathbb{R} \text{ s.t. } a_0 + a_1\left(\frac{i}{n} - x\right) > Y_i \; \forall i \in I_n(x)\right\}.$$

Theorem

If $h_n \to 0$, $nh_n \to \infty$, then

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}_n(x) - g(x)| = O(h_n^\beta) + O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{1/\alpha}\right)$$

uniformly for all $g \in \mathcal{H}_{\beta,L}$ for all fixed $\beta \in (1,2]$, $L > 0$.

Bandwidth $h_n = (n/\log n)^{-1/(\alpha\beta+1)}$ leads to optimal rate $(n/\log n)^{-\beta/(\alpha\beta+1)}$;
naive estimator yields slower rate $(n/\log n)^{-1/(\alpha+1)}$ (for optimal choice of h_n)

Further improvements for smoother functions

Suppose $g \in \mathcal{H}_{\beta,L}$ for some $\beta > 2$. Let $s = \lfloor \beta - \rfloor$. Then

$$g(t) \approx \sum_{j=0}^{s} a_j (t-x)^j$$

on neighborhood of x. However, for $x \notin \{1/n, 2/n, \dots, 1\}$

$$\tilde{g}_n(x) := \min\left\{ a_0 \in \mathbb{R} \mid \exists a_j \in \mathbb{R} \text{ s.t. } \sum_{j=0}^{s} a_j \left(\frac{i}{n} - x\right)^j > Y_i \ \forall i \in I_n(x) \right\} = -\infty$$

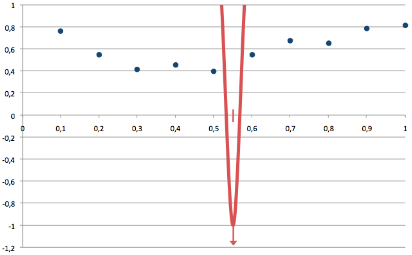
Further improvements for smoother functions

Suppose $g \in \mathcal{H}_{\beta,L}$ for some $\beta > 2$. Let $s = \lfloor \beta - \rfloor$. Then

$$g(t) \approx \sum_{j=0}^{s} a_j (t-x)^j$$

on neighborhood of x. However, for $x \notin \{1/n, 2/n, \ldots, 1\}$

$$\tilde{g}_n(x) := \min \left\{ a_0 \in \mathbb{R} \mid \exists a_j \in \mathbb{R} \text{ s.t. } \sum_{j=0}^{s} a_j \left(\frac{i}{n} - x \right)^j > Y_i \; \forall i \in I_n(x) \right\} = -\infty$$



New interpretation of naive and local linear estimator

With

$$\mathcal{P}_s := \{p \text{ polynomial of order } s \text{ on } [x - h_n, x + h_n] \mid p(i/n) \geq Y_i \; \forall\, i \in I_n(x)\}$$

one has

$$\hat{g}_n(x) = p_s(x) \qquad \text{where} \qquad p_s \in \mathcal{P}_s \text{ minimizes } \int_{x-h_n}^{x+h_n} p(t)\, dt$$

for $s = 0$ (naive est.) resp. $s = 1$ (local linear est.)

Reason: $\int_{x-h_n}^{x+h_n} p(t)\, dt = 2p(x)h_n$ for $p \in \mathcal{P}_s$ with $s \in \{0, 1\}$

This interpretation can be easily generalized to higher order polynomials:

$$\hat{g}_{n,s}(x) = p_s(x) \qquad \text{with} \qquad p_s = \arg\min_{p_s \in \mathcal{P}_s} \int_{x-h_n}^{x+h_n} p(t)\, dt$$

Other variant:

$$\hat{g}^*_{n,s}(x) = p^*_s(x) \qquad \text{with} \qquad p^*_s = \arg\min_{p \in \mathcal{P}_s} \sum_{i \in I_n(x)} p(i/n)$$

New interpretation of naive and local linear estimator

With

$$\mathcal{P}_s := \{p \text{ polynomial of order } s \text{ on } [x-h_n, x+h_n] \mid p(i/n) \geq Y_i \ \forall\, i \in I_n(x)\}$$

one has

$$\hat{g}_n(x) = p_s(x) \qquad \text{where} \qquad p_s \in \mathcal{P}_s \text{ minimizes } \int_{x-h_n}^{x+h_n} p(t)\, dt$$

for $s = 0$ (naive est.) resp. $s = 1$ (local linear est.)

Reason: $\int_{x-h_n}^{x+h_n} p(t)\, dt = 2p(x)h_n$ for $p \in \mathcal{P}_s$ with $s \in \{0, 1\}$

This interpretation can be easily generalized to higher order polynomials:

$$\hat{g}_{n,s}(x) = p_s(x) \qquad \text{with} \qquad p_s = \arg\min_{p_s \in \mathcal{P}_s} \int_{x-h_n}^{x+h_n} p(t)\, dt$$

Other variant:

$$\hat{g}^*_{n,s}(x) = p^*_s(x) \qquad \text{with} \qquad p^*_s = \arg\min_{p \in \mathcal{P}_s} \sum_{i \in I_n(x)} p(i/n)$$

New interpretation of naive and local linear estimator

With

$$\mathcal{P}_s := \{p \text{ polynomial of order } s \text{ on } [x-h_n, x+h_n] \mid p(i/n) \geq Y_i \; \forall\, i \in I_n(x)\}$$

one has

$$\hat{g}_n(x) = p_s(x) \qquad \text{where} \qquad p_s \in \mathcal{P}_s \text{ minimizes } \int_{x-h_n}^{x+h_n} p(t)\,dt$$

for $s = 0$ (naive est.) resp. $s = 1$ (local linear est.)

Reason: $\int_{x-h_n}^{x+h_n} p(t)\,dt = 2p(x)h_n$ for $p \in \mathcal{P}_s$ with $s \in \{0,1\}$

This interpretation can be easily generalized to higher order polynomials:

$$\hat{g}_{n,s}(x) = p_s(x) \qquad \text{with} \qquad p_s = \arg\min_{p_s \in \mathcal{P}_s} \int_{x-h_n}^{x+h_n} p(t)\,dt$$

Other variant:

$$\hat{g}^*_{n,s}(x) = p^*_s(x) \qquad \text{with} \qquad p^*_s = \arg\min_{p \in \mathcal{P}_s} \sum_{i \in I_n(x)} p(i/n)$$

Uniform asymptotics of general estimators

$$\hat{g}_{n,s}(x) = p_s(x) \qquad \text{with} \qquad p_s = \arg\min_{p_s \in \mathcal{P}_s} \int_{x-h_n}^{x+h_n} p(t)\,dt$$

$$\hat{g}^*_{n,s}(x) = p^*_s(x) \qquad \text{with} \qquad p^*_s = \arg\min_{p \in \mathcal{P}_s} \sum_{i \in I_n(x)} p(i/n)$$

Theorem

Suppose $g \in \mathcal{H}_{\beta,L}$ *for some* $\beta \in (s, s+1]$. *If* $h_n \to 0$, $nh_n \to \infty$, *then*

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}_{n,s}(x) - g(x)| = O(h_n^\beta) + O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{1/\alpha}\right)$$

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}^*_{n,s}(x) - g(x)| = O(h_n^\beta) + O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{1/\alpha}\right)$$

uniformly for all $g \in \mathcal{H}_{\beta,L}$ *for all fixed* $L > 0$.

Bandwidth $h_n = (n/\log n)^{-1/(\alpha\beta+1)}$ leads to optimal rate $(n/\log n)^{-\beta/(\alpha\beta+1)}$

Uniform asymptotics of general estimators

$$\hat{g}_{n,s}(x) = p_s(x) \qquad \text{with} \qquad p_s = \arg\min_{p_s \in \mathcal{P}_s} \int_{x-h_n}^{x+h_n} p(t)\,dt$$

$$\hat{g}^*_{n,s}(x) = p^*_s(x) \qquad \text{with} \qquad p^*_s = \arg\min_{p \in \mathcal{P}_s} \sum_{i \in I_n(x)} p(i/n)$$

Theorem

Suppose $g \in \mathcal{H}_{\beta,L}$ for some $\beta \in (s, s+1]$. If $h_n \to 0$, $nh_n \to \infty$, then

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}_{n,s}(x) - g(x)| = O(h_n^\beta) + O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{1/\alpha}\right)$$

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}^*_{n,s}(x) - g(x)| = O(h_n^\beta) + O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{1/\alpha}\right)$$

uniformly for all $g \in \mathcal{H}_{\beta,L}$ for all fixed $L > 0$.

Bandwidth $h_n = (n/\log n)^{-1/(\alpha\beta+1)}$ leads to optimal rate $(n/\log n)^{-\beta/(\alpha\beta+1)}$

Empirical process of residuals

Goal: Analyse distribution of errors (e.g. test for parametric hypotheses)

If errors ε_i were observable, statistical analysis based on empirical distribution function

$$F_n(y) := \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,y]}(\varepsilon_i).$$

Use instead residuals

$$\hat{\varepsilon}_i := Y_i - \hat{g}_n(i/n), \qquad nh_n \le i \le n - nh_n,$$

with empirical distribution function

$$\hat{F}_n(y) := \frac{1}{m_n}\sum_{i \in I_n} 1_{(-\infty,y]}(\hat{\varepsilon}_i)$$

where $I_n := \{i \in \{1,\dots,n\} \mid nh_n \le i \le n - nh_n\}$ and $m_n := |I_n| \sim n$.

Empirical process of residuals

Goal: Analyse distribution of errors (e.g. test for parametric hypotheses)

If errors ε_i were observable, statistical analysis based on empirical distribution function

$$F_n(y) := \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,y]}(\varepsilon_i).$$

Use instead residuals

$$\hat{\varepsilon}_i := Y_i - \hat{g}_n(i/n), \qquad nh_n \leq i \leq n - nh_n,$$

with empirical distribution function

$$\hat{F}_n(y) := \frac{1}{m_n} \sum_{i \in I_n} 1_{(-\infty,y]}(\hat{\varepsilon}_i)$$

where $I_n := \big\{ i \in \{1, \ldots, n\} \mid nh_n \leq i \leq n - nh_n \big\}$ and $m_n := |I_n| \sim n$.

Empirical process of residuals: asymptotics

$$\hat{F}_n(y) := \frac{1}{m_n} \sum_{i \in I_n} 1_{(-\infty, y]}(\hat{\varepsilon}_i), \qquad \tilde{F}_n(y) := \frac{1}{m_n} \sum_{i \in I_n} 1_{(-\infty, y]}(\varepsilon_i).$$

Theorem

If $\max_{i \in I_n} |\hat{g}_n(i/n) - g(i/n)| = O_P\big(n^{-\beta/(\alpha\beta+1)}(\log n)^\tau\big)$ *for some* $\tau < 0$ *and* F *is Hölder continuous of order* $\alpha \wedge 1$*, then*

$$\sup_{y \in \mathbb{R}} \big|\hat{F}_n(y) - \tilde{F}_n(y)\big| = o_P\big(n^{-1/2}\big).$$

if $1/\beta < \alpha < 2 - 1/\beta$*. In particular,*

$$\sqrt{n}(\hat{F}_n - F) \longrightarrow B \circ F$$

weakly in $D(\mathbb{R})$ *for a Brownian bridge* B*.*

Proof

Conditions on α ensure that $a_n := n^{-\beta/(\alpha\beta+1)+\varepsilon} = o\big(n^{-1/(2(\alpha\wedge 1))}\big)$ for sufficiently small $\varepsilon < 0$. By the assumption on $\hat{g}_n$

$$\begin{aligned}\hat{F}_n(y) &= \frac{1}{m_n}\sum_{i\in I_n} 1_{(-\infty,y]}(\hat{\varepsilon}_i) = \frac{1}{m_n}\sum_{i\in I_n} 1_{(-\infty,y]}\big(\underbrace{\varepsilon_i + g(i/n)}_{=Y_i} - \hat{g}_n(i/n)\big)\\ &\leq \frac{1}{m_n}\sum_{i\in I_n} 1_{(-\infty,y]}\big(\varepsilon_i - a_n\big) = \tilde{F}(y+a_n)\end{aligned}$$

with probability tending to 1. Now

$$\sup_{y\in\mathbb{R}}(F(y+a_n) - F(y)) = O\big(a_n^{\alpha\wedge 1}\big) = o(n^{-1/2}).$$

and

$$E_n := \sqrt{n}(\tilde{F}_n - F) \to B \circ F$$

where the limit has continuous sample paths. Hence, uniformly for all $y \in \mathbb{R}$,

$$\begin{aligned}\sqrt{n}(\hat{F}_n(y) - \tilde{F}(y)) &\leq \sqrt{n}(\tilde{F}_n(y+a_n) - \tilde{F}(y))\\ &= E_n(y+a_n) - E_n(y) + \sqrt{n}(F(y+a_n) - F(y)) = o_p(1).\end{aligned}$$

Proof

Conditions on α ensure that $a_n := n^{-\beta/(\alpha\beta+1)+\varepsilon} = o\big(n^{-1/(2(\alpha\wedge 1))}\big)$ for sufficiently small $\varepsilon < 0$. By the assumption on $\hat{g}_n$

$$\begin{aligned} \hat{F}_n(y) &= \frac{1}{m_n}\sum_{i\in I_n} 1_{(-\infty,y]}(\hat{\varepsilon}_i) = \frac{1}{m_n}\sum_{i\in I_n} 1_{(-\infty,y]}\big(\underbrace{\varepsilon_i + g(i/n)}_{=Y_i} - \hat{g}_n(i/n)\big) \\ &\le \frac{1}{m_n}\sum_{i\in I_n} 1_{(-\infty,y]}\big(\varepsilon_i - a_n\big) = \tilde{F}(y+a_n) \end{aligned}$$

with probability tending to 1. Now

$$\sup_{y\in\mathbb{R}}(F(y+a_n) - F(y)) = O\big(a_n^{\alpha\wedge 1}\big) = o(n^{-1/2}).$$

and

$$E_n := \sqrt{n}(\tilde{F}_n - F) \to B\circ F$$

where the limit has continuous sample paths. Hence, uniformly for all $y\in\mathbb{R}$,

$$\begin{aligned} \sqrt{n}(\hat{F}_n(y) - \tilde{F}(y)) &\le \sqrt{n}(\tilde{F}_n(y+a_n) - \tilde{F}(y)) \\ &= E_n(y+a_n) - E_n(y) + \sqrt{n}(F(y+a_n) - F(y)) = o_p(1). \end{aligned}$$

Proof

Conditions on α ensure that $a_n := n^{-\beta/(\alpha\beta+1)+\varepsilon} = o\big(n^{-1/(2(\alpha\wedge 1))}\big)$ for sufficiently small $\varepsilon < 0$. By the assumption on $\hat{g}_n$

$$\begin{aligned}
\hat{F}_n(y) &= \frac{1}{m_n}\sum_{i\in I_n} 1_{(-\infty,y]}(\hat{\varepsilon}_i) = \frac{1}{m_n}\sum_{i\in I_n} 1_{(-\infty,y]}\big(\underbrace{\varepsilon_i + g(i/n)}_{=Y_i} - \hat{g}_n(i/n)\big) \\
&\leq \frac{1}{m_n}\sum_{i\in I_n} 1_{(-\infty,y]}\big(\varepsilon_i - a_n\big) = \tilde{F}(y+a_n)
\end{aligned}$$

with probability tending to 1. Now

$$\sup_{y\in\mathbb{R}}(F(y+a_n) - F(y)) = O\big(a_n^{\alpha\wedge 1}\big) = o(n^{-1/2}).$$

and

$$E_n := \sqrt{n}(\tilde{F}_n - F) \to B \circ F$$

where the limit has continuous sample paths. Hence, uniformly for all $y \in \mathbb{R}$,

$$\begin{aligned}
\sqrt{n}(\hat{F}_n(y) - \tilde{F}(y)) &\leq \sqrt{n}(\tilde{F}_n(y+a_n) - \tilde{F}(y)) \\
&= E_n(y+a_n) - E_n(y) + \sqrt{n}(F(y+a_n) - F(y)) = o_p(1).
\end{aligned}$$

Proof (cont.)

Likewise,

$$\sqrt{n}(\hat{F}_n(y) - \tilde{F}(y)) \geq E_n(y - a_n) - E_n(y) + \sqrt{n}(F(y - a_n) - F(y)) = o_p(1),$$

which concludes the proof.

Further results and open problems

Working group at HU Berlin: established estimators which adapt to unknown smoothness and attain pointwise minimax rates.

Open problems to be considered next in working group at Hamburg University:

- More precise analysis of estimation error to obtain similar results for empirical process of residuals under weaker conditions on α
- Analysis of tail empirical process of residuals
- Testing model assumptions on behavior of F at 0

Further results and open problems

Working group at HU Berlin: established estimators which adapt to unknown smoothness and attain pointwise minimax rates.

Open problems to be considered next in working group at Hamburg University:

- More precise analysis of estimation error to obtain similar results for empirical process of residuals under weaker conditions on α
- Analysis of tail empirical process of residuals
- Testing model assumptions on behavior of F at 0

Further results and open problems

Working group at HU Berlin: established estimators which adapt to unknown smoothness and attain pointwise minimax rates.

Open problems to be considered next in working group at Hamburg University:

- More precise analysis of estimation error to obtain similar results for empirical process of residuals under weaker conditions on α
- Analysis of tail empirical process of residuals
- Testing model assumptions on behavior of F at 0

Thank you for your attention!