Weierstrass Institute for
Applied Analysis and Stochastics

Leibniz Association

Non-asymptotic confidence bounds via multiplier bootstrap

Mayya Zhilova (joint with Vladimir Spokoiny)

CRC 649 "Economic Risk" Conference

Motzen, 2014

Outline

Generalized regression

- A random sample of a fixed size n from unknown distribution:

$$\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n,$$

$$Y_i \sim P_i\big(f(X_i)\big), \text{ independent.}$$

- Deterministic design $X_1, \ldots, X_n \in \mathbb{R}^d$.

- Unknown function $f(\cdot) : \mathbb{R}^d \to \mathbb{R}$.

- Mean regression

$$Y_i \sim f(X_i) + \varepsilon_i, \quad \varepsilon_i \text{ are i.i.d.}$$

- Logistic regression

$$Y_i \sim Bernoulli(f(X_i)).$$

- Quantile regression

$$P\left(Y_i - f(X_i) < 0\right) = \tau$$

for $\tau \in (0, 1)$.

- Let $\boldsymbol{Y} \sim I\!P$ (unknown).
- Fix some known parametric family:

$$\{I\!P(\boldsymbol{\theta}) \ll \mu_0, \boldsymbol{\theta} \in \Theta \subset I\!R^p\}\,.$$

- The target parameter $\boldsymbol{\theta}^* \in \Theta$ is defined by the projection:

$$\begin{aligned} \boldsymbol{\theta}^* &\stackrel{\text{def}}{=} \operatorname*{argmin}_{\boldsymbol{\theta}} KL(I\!P, I\!P(\boldsymbol{\theta})) \\ &= \operatorname*{argmin}_{\boldsymbol{\theta}} \int \log\left\{\frac{dI\!P}{dI\!P(\boldsymbol{\theta})}\right\} dI\!P. \end{aligned}$$

To find a sharp confidence set $\mathcal{E}$ s.t. for a given coverage probability $1-\alpha \in (0,1)$ and for a fixed sample size n :

$$I\!P\{\boldsymbol{\theta}^* \in \mathcal{E}\} \geq 1-\alpha.$$

Outline

Maximum likelihood method:

- Log-likelihood function

$$L(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \log \frac{d\mathbb{P}(\boldsymbol{\theta})}{d\mu_0}(\boldsymbol{Y})$$

$$= \sum_{i=1}^{n} \log \frac{dP_i(\boldsymbol{\theta})}{d\mu_0}(Y_i).$$

- The target parameter:

$$\boldsymbol{\theta}^* \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta}} \mathbb{E} L(\boldsymbol{\theta}) = \operatorname*{argmin}_{\boldsymbol{\theta}} KL(\mathbb{P}, \mathbb{P}(\boldsymbol{\theta})).$$

- (Quasi) Maximum likelihood estimate (qMLE):

$$\widetilde{\boldsymbol{\theta}} \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\theta}).$$

Wilks phenomenon for the likelihood ratio:

For $I\!P \in \{P(\boldsymbol{\theta})\}$

$$L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*) \approx \|\boldsymbol{\xi}\|^2/2, \qquad \boldsymbol{\xi} \stackrel{w}{\to} \mathcal{N}(0, \boldsymbol{I}_p)$$

with $n \to \infty$.

p is the dimension of parametric set Θ,

$\boldsymbol{\xi} \stackrel{\text{def}}{=} D_0^{-1} \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}^*)$ normalized score,

$D_0^2 \stackrel{\text{def}}{=} -\nabla^2_{\boldsymbol{\theta}} I\!E L(\boldsymbol{\theta}^*)$ full Fisher Information matrix.

Likelihood-based confidence set:

$$\mathcal{E}(\mathfrak{z}) \stackrel{\text{def}}{=} \{\boldsymbol{\theta} : L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}) \leq \mathfrak{z}\},$$

the level $\mathfrak{z} > 0$ can be found by the Wilks theorem:

$$I\!P\left\{\boldsymbol{\theta}^* \in \mathcal{E}(\mathfrak{z})\right\} \to P\left\{\chi^2_p \leq 2\mathfrak{z}\right\}.$$

- The likelihood ratio is asymptotically pivotal,
- $\mathcal{E}(\mathfrak{z})$ does not require to estimate the variance,
- pivotality fails under model misspecification $I\!P \notin \{I\!P(\boldsymbol{\theta})\}$,
- the convergence speed is slow.

Example: Mean Gaussian regression

$$Y_i = \theta^* + \varepsilon_i, \quad Y_i \sim \mathcal{N}(\theta^*, 1), \text{ i.i.d.}$$

- The log-likelihood function is quadratic:

$$L(\theta) = -\frac{1}{2}\sum_{i=1}^{n}(Y_i - \theta)^2 + C$$

- The MLE:

$$\widetilde{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i \sim \mathcal{N}(\theta^*, \frac{1}{\sqrt{n}})$$

- Wilks Theorem:

$$2\left\{L(\widetilde{\theta}) - L(\theta^*)\right\} = n(\widetilde{\theta} - \theta^*)^2 \sim \chi^2_1.$$

If $Y_i \nsim \mathcal{N}(\theta^*, 1)$, then the distribution of $L(\widetilde{\theta}) - L(\theta^*)$ differs from χ^2_1.

- It holds with probability $\geq 1 - 2\mathrm{e}^{-\mathrm{y}}$ (conditions come later)

$$\left| 2\left\{ L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*) \right\} - \|\boldsymbol{\xi}\|^2 \right| \leq \Delta_{\mathtt{sq}}(\mathrm{y}) \asymp \sqrt{(p+\mathrm{y})^3/n},$$

$$\left| \sqrt{2\left\{ L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*) \right\}} - \|\boldsymbol{\xi}\| \right| \leq \Delta(\mathrm{y}) \asymp (p+\mathrm{y})/\sqrt{n}.$$

- Deviation bound for the approximating term:

$$I\!P\left(\|\boldsymbol{\xi}\|^2 \geq p + 6\mathrm{y} \right) \leq 2\mathrm{e}^{-\mathrm{y}}.$$

- Allows the model to be misspecified,
- finite sample result,
- distribution of $\|\boldsymbol{\xi}\|$ in general depends on $I\!P$ and $\boldsymbol{\theta}^*$,
- the deviation bound is not sharp enough for a confidence set.

- Estimation of the distribution of the likelihood ratio statistic with a multiplier bootstrap procedure.

- Theoretical justification of the multiplier bootstrap using:

  - Uniform non-asymptotic local quadratic approximation of the likelihood ratio statistic (Wilks-type bound) for the Y and bootstrap worlds.

  - Non-asymptotic Gaussian approximation to connect the Y and the bootstrap worlds.

Outline

The idea:

To mimic the distribution of $L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*)$ using a multiplier bootstrap:

- Take an i.i.d. sample $u_1, \ldots, u_n$ independent of the data $\boldsymbol{Y}$:
  $I\!\!E(u_i) = \mathrm{Var}(u_i) = 1\,,\ I\!\!E \exp(u_i) < \infty$ (e.g. $\sim exp(1),\ \mathcal{N}(1,1)$).

- Bootstrap the likelihood function:

$$L^{\circledcirc}(\boldsymbol{\theta}) \stackrel{\mathrm{def}}{=} \sum_{i=1}^{n} \log \frac{dP_i^{\circledcirc}(\boldsymbol{\theta})}{d\mu_0}(Y_i) u_i$$

  $^{\circledcirc}$ denotes the conditional probability with the fixed sample $\boldsymbol{Y}$.

“$\boldsymbol{Y}$ world”	conditional “bootstrap world“
quasi MLE	
$\widetilde{\boldsymbol{\theta}} \stackrel{\mathrm{def}}{=} \operatorname{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\theta})$	$\widetilde{\boldsymbol{\theta}}^{\circ} \stackrel{\mathrm{def}}{=} \operatorname{argmax}_{\boldsymbol{\theta}} L^{\circ}(\boldsymbol{\theta})$
target	
$\boldsymbol{\theta}^* \stackrel{\mathrm{def}}{=} \operatorname{argmax}_{\boldsymbol{\theta}} I\!\!E L(\boldsymbol{\theta})$	$\widetilde{\boldsymbol{\theta}} \stackrel{\mathrm{def}}{=} \operatorname{argmax}_{\boldsymbol{\theta}} I\!\!E^{\circ} L^{\circ}(\boldsymbol{\theta})$
likelihood ratio	
$L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*)$	$L^{\circ}(\widetilde{\boldsymbol{\theta}}^{\circ}) - L^{\circ}(\widetilde{\boldsymbol{\theta}})$

- The true point in bootstrap world is exactly qMLE $\widetilde{\boldsymbol{\theta}}$.
- The bootstrap side is computable!
- The “bootstrap world” is inside of the parametric model, which, however, may be wrong.

- Wild bootstrap:
  [Wu, 1986], [Beran, 1986], [Hardle and Mammen, 1993], [Mammen, 1993];
  covers our approach for the case of a linear model and quadratic log-likelihood.
- Multiplier bootstrap for resampling of an objective function:
  - [Jin, Ying and Wei, 2001] use the multiplier bootstrap procedure for perturbing U-process of the k-th degree in order to estimate the covariance matrix of the U-process' minimiser.
  - [Lavergne and Patilea, 2013] use multiplier bootstrap for resampling of the smooth minimum distance statistic.
- Non-asymptotic results (very few):
  - [Arlot, Blanchard and Roquian, 2010] consider a sample of high dimensional i.i.d. Gaussian vectors, a non-asymptotical confidence set is constructed using exchangeable weighted bootstrap scheme.
  - [Chernozhukov, Chetverikov and Kato, 2013] approximate non-asymptotically the distribution of the maximum of a sum of independent high-dimensional vectors using Gaussian multiplier bootstrap.

We use the following approximating diagram:

$$\sqrt{2\Big\{L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*)\Big\}} \overset{(p+\mathrm{y})/\sqrt{n}}{\approx} \|\boldsymbol{\xi}\|$$

$$\wr\!\wr$$

$$\sqrt{2\Big\{L^\circledcirc(\widetilde{\boldsymbol{\theta}}^\circledcirc) - L^\circledcirc(\widetilde{\boldsymbol{\theta}})\Big\}} \overset{(p+\mathrm{y})/\sqrt{n}}{\approx} \|\boldsymbol{\xi}^\circledcirc\|$$

The approximation $\approx$ holds between the distributions on two different probability spaces:

$$\mathcal{L}\left(\|\boldsymbol{\xi}\|\right) \approx \mathcal{L}\left(\|\boldsymbol{\xi}^\circledcirc\| \,\middle|\, \boldsymbol{Y}\right).$$

Validity of the bootstrap [Spokoiny and Zhilova, 2014a]

It holds for all $\alpha \in (0,1)$ and the computable bootstrap $(1-\alpha)$-quantile

$$\mathfrak{z}_\alpha^\circ \stackrel{\mathrm{def}}{=} \min\left\{\mathfrak{z} \geq 0 : I\!P^\circ\left(\sqrt{2\left\{\sup_{\boldsymbol{\theta}} L^\circ(\boldsymbol{\theta}) - L^\circ(\widetilde{\boldsymbol{\theta}})\right\}} > \mathfrak{z}\right) \leq \alpha\right\}$$

$$I\!P\left(\sqrt{2\{L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*)\}} > \mathfrak{z}_\alpha^\circ(1+\delta) + \Delta_1\right) - \alpha \leq \Delta_2,$$

$$I\!P\left(\sqrt{2\{L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*)\}} > \mathfrak{z}_\alpha^\circ(1-\delta) - \Delta_1\right) - \alpha \geq -\Delta_2,$$

$\delta, \Delta_{1,2} \asymp \{\log(n)\}^{a_k}/n^{b_k}$, $a_k \geq 0, b_k > 0$

deterministic, explicit up to a generic constant.

- For a linear w.r.t. u_i function in the bootstrap world its deterministic part has an “unbiased” $\boldsymbol{Y}$ -random equivalent. For instance, for the likelihood process:

$$\mathbb{E}^{\circ} L^{\circ}(\boldsymbol{\theta}) = L(\boldsymbol{\theta}), \qquad \mathbb{E}^{\circ}\bigl\{L^{\circ}(\boldsymbol{\theta}) - L^{\circ}(\widetilde{\boldsymbol{\theta}})\bigr\} = L(\boldsymbol{\theta}) - L(\widetilde{\boldsymbol{\theta}}),$$

  and for the Hessian matrix:

$$\mathfrak{D}^2(\boldsymbol{\theta}) \stackrel{\mathrm{def}}{=} -\nabla^2_{\boldsymbol{\theta}} \mathbb{E}^{\circ} L^{\circ}(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \nabla^2_{\boldsymbol{\theta}} \ell_i(\boldsymbol{\theta}).$$

- The local geometry on Θ around target points is defined with matrices D_0^2 and $\mathfrak{D}_0^2 \stackrel{\mathrm{def}}{=} \mathfrak{D}^2(\widetilde{\boldsymbol{\theta}})$. The following bound holds with exponentially high probability:

$$\sup_{\boldsymbol{\theta} \in Loc(\boldsymbol{\theta}^*)} \bigl\| D_0^{-1} \mathfrak{D}^2(\boldsymbol{\theta}) D_0^{-1} - \boldsymbol{I}_p \bigr\|_{\infty} \leq \mathtt{C}(p + \mathtt{y})/\sqrt{n}.$$

  It allows to use the same (deterministic) Fisher information matrix in the quadratic expansion of the likelihood processes.

- Local quadratic approximations hold locally around the target points:

$\forall \boldsymbol{\theta} : \|D_0(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| \le \mathtt{r}_0$ and some $0 < \mathtt{r}_0 \le \mathtt{C}\sqrt{p + \mathtt{y}}$:

$$L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}^*) \overset{\Delta_{\mathrm{sq}}(\mathtt{y})}{\approx} \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}^*)^\top (\boldsymbol{\theta} - \boldsymbol{\theta}^*) - \|D_0(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|^2/2$$
$$\overset{\boldsymbol{\theta} := \widetilde{\boldsymbol{\theta}}}{\approx} \|\boldsymbol{\xi}\|^2/2,$$

$$L^{\circledcirc}(\boldsymbol{\theta}) - L^{\circledcirc}(\widetilde{\boldsymbol{\theta}}) \overset{\Delta_{\mathrm{sq}}(\mathtt{y})}{\approx} \nabla_{\boldsymbol{\theta}} L^{\circledcirc}(\widetilde{\boldsymbol{\theta}})^\top (\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}) - \|D_0(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}})\|^2/2$$
$$\overset{\boldsymbol{\theta} := \widetilde{\boldsymbol{\theta}}^{\circledcirc}}{\approx} \|\boldsymbol{\xi}^{\circledcirc}\|^2/2,$$

where $\boldsymbol{\xi} \overset{\mathrm{def}}{=} D_0^{-1} \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}^*)$, $\boldsymbol{\xi}^{\circledcirc} \overset{\mathrm{def}}{=} D_0^{-1} \nabla_{\boldsymbol{\theta}} L^{\circledcirc}(\widetilde{\boldsymbol{\theta}})$.

- The approximating term $\|\boldsymbol{\xi}^{\circledcirc}\|^2$ is of the 2-d order w.r.t. u_i :

  for $\ell_i(\boldsymbol{\theta}) \stackrel{\mathrm{def}}{=} \log \frac{dP_i(\boldsymbol{\theta})}{d\mu_0}(Y_i)$

$$\|\boldsymbol{\xi}^{\circledcirc}\|^2 = \sum_{i,j=1}^{n} \nabla_{\boldsymbol{\theta}} \ell_i(\widetilde{\boldsymbol{\theta}})^\top \nabla_{\boldsymbol{\theta}} \ell_j(\widetilde{\boldsymbol{\theta}})(u_i - 1)(u_j - 1),$$

$$\|\boldsymbol{\xi}\|^2 = \sum_{i,j=1}^{n} \bigl\{\nabla_{\boldsymbol{\theta}} \ell_i(\boldsymbol{\theta}^*) - \nabla_{\boldsymbol{\theta}} I\!\!E \ell_i(\boldsymbol{\theta}^*)\bigr\}^\top \bigl\{\nabla_{\boldsymbol{\theta}} \ell_j(\boldsymbol{\theta}^*) - \nabla_{\boldsymbol{\theta}} I\!\!E \ell_j(\boldsymbol{\theta}^*)\bigr\}.$$

Let $\breve{\boldsymbol{\xi}} \sim \mathcal{N}(0, \operatorname{Var} \boldsymbol{\xi})\,, \quad \breve{\boldsymbol{\xi}}^{\circ} \sim \mathcal{N}(0, \operatorname{Var} \boldsymbol{\xi}^{\circ})$

	Wilks approximation		Gaussian approximation	
$\sqrt{2\{L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*)\}}$	$\approx$	$\|\boldsymbol{\xi}\|$	$\overset{\mathcal{L}}{\approx}$	$\|\breve{\boldsymbol{\xi}}\|$
				$\wr$
$\sqrt{2\{L^{\circ}(\widetilde{\boldsymbol{\theta}}^{\circ}) - L^{\circ}(\widetilde{\boldsymbol{\theta}})\}}$	$\approx$	$\|\boldsymbol{\xi}^{\circ}\|$	$\overset{\mathcal{L}^{\circ}}{\approx}$	$\|\breve{\boldsymbol{\xi}}^{\circ}\|$

For $\breve{\xi} \sim \mathcal{N}(0, \operatorname{Var}\xi)$ $\mathcal{L}(\|\xi\|) \approx \mathcal{L}(\|\breve{\xi}\|)$.

- $\|\xi\| = \sup_{\gamma \in I\!R^p, \|\gamma\|=1} \gamma^\top \xi \approx \sup_{\gamma \in \mathrm{G}_\gamma} \gamma^\top \xi$

  for a finite grid G_γ on the sphere $\{\gamma \in I\!R^p : \|\gamma\| = 1\}$.

- $I\!P(\|\xi\| > t) = I\!E\, \mathbb{1}(\|\xi\| > t)$.

- $\mathbb{1}\left(\sup_{\gamma \in \mathrm{G}_\gamma} \gamma^\top \xi > t\right) \approx f_\beta(\xi, t)$,

$$f_\beta(\xi, t) \stackrel{\mathrm{def}}{=} \frac{\sum_{\gamma \in \mathrm{G}_S(\varepsilon)} \exp\{\beta(\gamma^\top \xi - t)\}}{1 + \sum_{\gamma \in \mathrm{G}_S(\varepsilon)} \exp\{\beta(\gamma^\top \xi - t)\}}, \ \beta > 0.$$

$f_\beta(\xi, t)$ is a composition of logistic function and smooth max function:

$$\mathbb{1}(\|\xi\| > t) \approx \frac{\exp\{\beta(\|\xi\| - t)\}}{1 + \exp\{\beta(\|\xi\| - t)\}} \qquad \sup_{\gamma \in \mathrm{G}_\gamma} \gamma^\top \xi \approx \frac{1}{\beta} \log\Big\{ \sum_{\gamma \in \mathrm{G}_\gamma} \exp(\beta \gamma^\top \xi) \Big\}$$

For $\breve{\xi} \sim \mathcal{N}(0, \operatorname{Var} \xi)$ $\mathcal{L}(\|\xi\|) \approx \mathcal{L}(\|\breve{\xi}\|)$.

- $\|\xi\| = \sup_{\gamma \in I\!R^p, \|\gamma\|=1} \gamma^\top \xi \approx \sup_{\gamma \in \mathrm{G}_\gamma} \gamma^\top \xi$

  for a finite grid G_γ on the sphere $\{\gamma \in I\!R^p : \|\gamma\| = 1\}$.

- $I\!P(\|\xi\| > t) = I\!E\, \mathbb{1}(\|\xi\| > t)$.

- $\mathbb{1}\left(\sup_{\gamma \in \mathrm{G}_\gamma} \gamma^\top \xi > t\right) \approx f_\beta(\xi, t)$,

$$f_\beta(\xi, t) \stackrel{\mathrm{def}}{=} \frac{\sum_{\gamma \in \mathrm{G}_S(\varepsilon)} \exp\{\beta(\gamma^\top \xi - t)\}}{1 + \sum_{\gamma \in \mathrm{G}_S(\varepsilon)} \exp\{\beta(\gamma^\top \xi - t)\}}, \ \beta > 0.$$

For $\gamma \in I\!R^p$

$$\left| f_\beta(\gamma + \xi, t) - f_\beta(\gamma, t) - \xi^\top \nabla_\gamma f_\beta(\gamma, t) - \frac{1}{2} \xi^\top \nabla^2_\gamma f_\beta(\gamma, t) \xi \right| \leq 0.7 \beta^3 \|\xi\|^3$$

- By Lindeberg's telescopic sum device [Lindeberg, 1922]:

$$\left| I\!E f_\beta(\boldsymbol{\xi}, z) - I\!E f_\beta(\breve{\boldsymbol{\xi}}, z) \right| \leq \frac{\beta^3}{8} \sum_{i=1}^{n} I\!E \left(\|\boldsymbol{\xi}_i\|^3 + \|\breve{\boldsymbol{\xi}}_i\|^3 \right),$$

where

$$\begin{aligned} \boldsymbol{\xi}_i &\stackrel{\mathrm{def}}{=} D_0^{-1} \left\{ \nabla_{\boldsymbol{\theta}} \ell_i(\boldsymbol{\theta}^*) - \nabla_{\boldsymbol{\theta}} I\!E \ell_i(\boldsymbol{\theta}^*) \right\}, \\ \breve{\boldsymbol{\xi}}_i &\sim \mathcal{N}(0, \operatorname{Var} \boldsymbol{\xi}_i). \end{aligned}$$

$$\begin{aligned}
I\!P\left(\|\boldsymbol{\xi}\| \geq z + \sqrt{p}\Delta\right) &\leq I\!P\left(\sup_{\boldsymbol{\gamma}\in \mathrm{G}_S(\varepsilon)} \boldsymbol{\gamma}^\top(\boldsymbol{\xi} - \overline{\Delta}) \geq (1-\varepsilon)z\right) \\
&\leq I\!P\left(\|\breve{\boldsymbol{\xi}}\| \geq z\left(1 - \frac{2\varepsilon}{1+\varepsilon}\right) - \sqrt{p}\Delta\right) \\
&\quad + \left[1 + \mathrm{card}\{\mathrm{G}_S(\varepsilon)\}\right] \mathrm{e}^{-\beta\Delta} + \mathtt{C}_3 \frac{\left(\beta \mathtt{C}_2 \|\Sigma^{1/2}\|\right)^3}{4\sqrt{n}}.
\end{aligned}$$

Similarly for the lower bound:

$$\begin{aligned}
I\!P\left(\|\boldsymbol{\xi}\| \geq z - \sqrt{p}\Delta\right) &\geq I\!P\left(\|\breve{\boldsymbol{\xi}}\| \geq z\left(1 + \frac{2\varepsilon}{1-\varepsilon}\right) + \sqrt{p}\Delta\right) \\
&\quad - \left[1 + \mathrm{card}\{\mathrm{G}_S(\varepsilon)\}\right] \mathrm{e}^{-\beta\Delta} - \mathtt{C}_3 \frac{\left(\beta \mathtt{C}_2 \|\Sigma^{1/2}\|\right)^3}{4\sqrt{n}}.
\end{aligned}$$

- Finally we compare the norms of two Gaussian vectors: $\|\breve{\boldsymbol{\xi}}\|$ and $\|\breve{\boldsymbol{\xi}}^{\circledcirc}\|$:

$$\begin{array}{ccc} \|\boldsymbol{\xi}\| & \overset{\mathcal{L}}{\approx} & \|\breve{\boldsymbol{\xi}}\| \\ & & \wr\wr \\ \|\boldsymbol{\xi}^{\circledcirc}\| & \overset{\mathcal{L}^{\circledcirc}}{\approx} & \|\breve{\boldsymbol{\xi}}^{\circledcirc}\| \end{array}$$

Gaussian approximation

i.e. their covariance matrices:

$$\mathrm{Var}(\breve{\boldsymbol{\xi}}) = D_0^{-1} \mathrm{Var}\left(\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}^*)\right) D_0^{-1},$$
$$\mathrm{Var}^{\circledcirc}(\breve{\boldsymbol{\xi}}^{\circledcirc}) = D_0^{-1} \mathrm{Var}^{\circledcirc}\left(\nabla_{\boldsymbol{\theta}} L^{\circledcirc}(\widetilde{\boldsymbol{\theta}})\right) D_0^{-1}.$$

Let $\Sigma \stackrel{\text{def}}{=} \operatorname{Var}(\boldsymbol{\xi}),\ \Sigma^{\circledcirc} \stackrel{\text{def}}{=} \operatorname{Var}^{\circledcirc}(\boldsymbol{\xi}^{\circledcirc})$.

$$D_0 \Sigma D_0 = \sum_{i=1}^{n} I\!E \left\{ \nabla_{\boldsymbol{\theta}} \ell_i(\boldsymbol{\theta}^*) \nabla_{\boldsymbol{\theta}} \ell_i(\boldsymbol{\theta}^*)^\top \right\} - \sum_{i=1}^{n} I\!E \left\{ \nabla_{\boldsymbol{\theta}} \ell_i(\boldsymbol{\theta}^*) \right\} I\!E \left\{ \nabla_{\boldsymbol{\theta}} \ell_i(\boldsymbol{\theta}^*) \right\}^\top,$$

$$D_0 \Sigma^{\circledcirc} D_0 = \sum_{i=1}^{n} \nabla_{\boldsymbol{\theta}} \ell_i(\boldsymbol{\theta}^*) \nabla_{\boldsymbol{\theta}} \ell_i(\boldsymbol{\theta}^*)^\top.$$

It holds:

$$\left\| \Sigma^{-1/2} \Sigma^{\circledcirc} \Sigma^{-1/2} - \boldsymbol{I}_p \right\|_\infty \leq \mathtt{C}_{\mathcal{I}}^2 \delta_\xi^2 \asymp (p + \mathtt{y})/n^{1/2},$$

the terms $\mathtt{C}_{\mathcal{I}}^2, \delta_\xi^2$ are responsible for the modelling bias, they are proportional to the error terms δ and Δ_1 in the statement on the bootstrap validity above.

Δ_1, Δ_2 are also determined by $\mathtt{C}_{\mathcal{I}}^2$ and the errors induced by $\mathbb{I}\big(\sup_{\mathrm{G}_\gamma} \boldsymbol{\gamma}^\top \boldsymbol{\xi} > t\big) \approx f_\beta(\boldsymbol{\xi}, t)$.

Conditions include

- Independency of the observations Y_i .
- Bounded exponential moments of gradient of $L(\boldsymbol{\theta}) - I\!\!E L(\boldsymbol{\theta})$ up to 3-d order.
- Small modelling bias:

$$\sum_{i=1}^{n} \| D_0^{-1} I\!\!E \nabla_{\boldsymbol{\theta}} \ell_i(\boldsymbol{\theta}^*) \|^2 \leq \delta_\xi^2 = \mathtt{C}(p + \mathtt{y})/n^{1/2}$$

  for $\ell_i(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \log \frac{dP_i(\boldsymbol{\theta})}{d\mu_0}(Y_i)$, $D_0^2 \stackrel{\text{def}}{=} -\nabla_{\boldsymbol{\theta}}^2 I\!\!E L(\boldsymbol{\theta}^*)$.

- $\| D_0 \operatorname{Var}\left(\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}^*)\right)^{-1} D_0 \|_\infty \leq \mathtt{C}_{\mathcal{I}}^2$.

  ($\delta_\xi^2 = 0, \mathtt{C}_{\mathcal{I}}^2 = 1$ if the parametric model is true)

I.i.d. errors, misspecified distribution

$$Y_i = 2 + \varepsilon_i, \;\; \varepsilon_i \sim Lap(0, 2^{-1/2}), \text{ i.i.d.}, \mathrm{Var}(\varepsilon_i) = 1,$$

$$L(\theta) = -\sum_{i=1}^{n} (Y_i - \theta)^2/2 \quad \text{corresponds to } \varepsilon_i \sim \mathcal{N}(0,1).$$

Quantiles of $\sqrt{2}\{L(\widetilde{\theta}) - L(\theta^*)\}^{1/2}$ are estimated from 5000 $\boldsymbol{Y}$ samples of length 50.

Quantiles of $\sqrt{2}\{L^{\circledcirc}(\widetilde{\theta}^{\circledcirc}) - L^{\circledcirc}(\widetilde{\theta})\}^{1/2}$ are estimated from 5000 $\{u_i\} \sim exp(1)$ samples and 50 $\boldsymbol{Y}$ samples of length 50.

Confidence level	0.99	0.95	0.90	0.85	0.80	0.75
$\sqrt{2}\{L(\widetilde{\theta}) - L(\theta^*)\}^{1/2}$	2.55	1.99	1.64	1.44	1.29	1.16
$\sqrt{2}\{L^{\circledcirc}(\widetilde{\theta}^{\circledcirc}) - L^{\circledcirc}(\widetilde{\theta})\}^{1/2}$, `Mean`	2.63	1.92	1.58	1.38	1.22	1.09
$\sqrt{2}\{L^{\circledcirc}(\widetilde{\theta}^{\circledcirc}) - L^{\circledcirc}(\widetilde{\theta})\}^{1/2}$, `St.dev.`	0.39	0.27	0.22	0.19	0.17	0.15

Comparison of the empirical quantiles for the sample size $n = 50$

Heteroscedastic errors, misspecified distribution

$$Y_i = 2 + \sigma_i \varepsilon_i, \quad \varepsilon_i \sim Lap(0, 2^{-1/2}), \quad \sigma_i = 2 - (i \mod 4)/2,$$

$$L(\theta) = -\sum_{i=1}^{n} (Y_i - \theta)^2/2 \quad \text{corresponds to} \quad \sigma_i \varepsilon_i \sim \mathcal{N}(0,1).$$

Quantiles of $\sqrt{2}\{L(\widetilde{\theta}) - L(\theta^*)\}^{1/2}$ are estimated from 5000 $\boldsymbol{Y}$ samples of length 50.

Quantiles of $\sqrt{2}\{L^{\circledcirc}(\widetilde{\theta}^{\circledcirc}) - L^{\circledcirc}(\widetilde{\theta})\}^{1/2}$ are estimated from 5000 $\{u_i\} \sim exp(1)$ samples and 50 $\boldsymbol{Y}$ samples of length 50.

Confidence level	0.99	0.95	0.90	0.85	0.80	0.75
$\sqrt{2}\{L(\widetilde{\theta}) - L(\theta^*)\}^{1/2}$	5.06	3.86	3.87	2.85	2.52	2.27
$\sqrt{2}\{L^{\circledcirc}(\widetilde{\theta}^{\circledcirc}) - L^{\circledcirc}(\widetilde{\theta})\}^{1/2}$, Mean	5.26	3.87	3.19	2.78	2.46	2.20
$\sqrt{2}\{L^{\circledcirc}(\widetilde{\theta}^{\circledcirc}) - L^{\circledcirc}(\widetilde{\theta})\}^{1/2}$, St.dev.	0.88	0.65	0.54	0.46	0.41	0.37

Comparison of the empirical quantiles for the sample size $n = 50$

Biased model

$$Y_i = \beta \sin(X_i) + \varepsilon_i, \quad \varepsilon_i \sim Lap(0, 2^{-1/2}), \text{ i.i.d.},$$

X_i are equidistant in $[0, 2\pi]$,

$$L(\theta) = -\sum_{i=1}^{n} (Y_i - \theta)^2/2 \;\Rightarrow\; \theta^* = 0,$$

$\beta > 0$ is an amplitude.

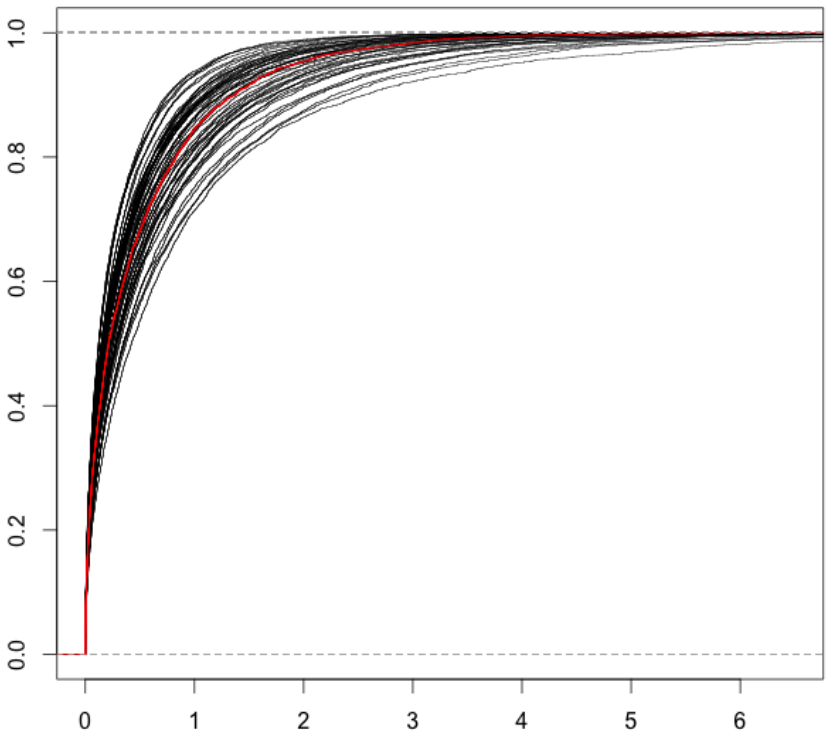
Let us check how the confidence set for θ^* changes with β.

Numerical results (III). Biased model

$n = 50$

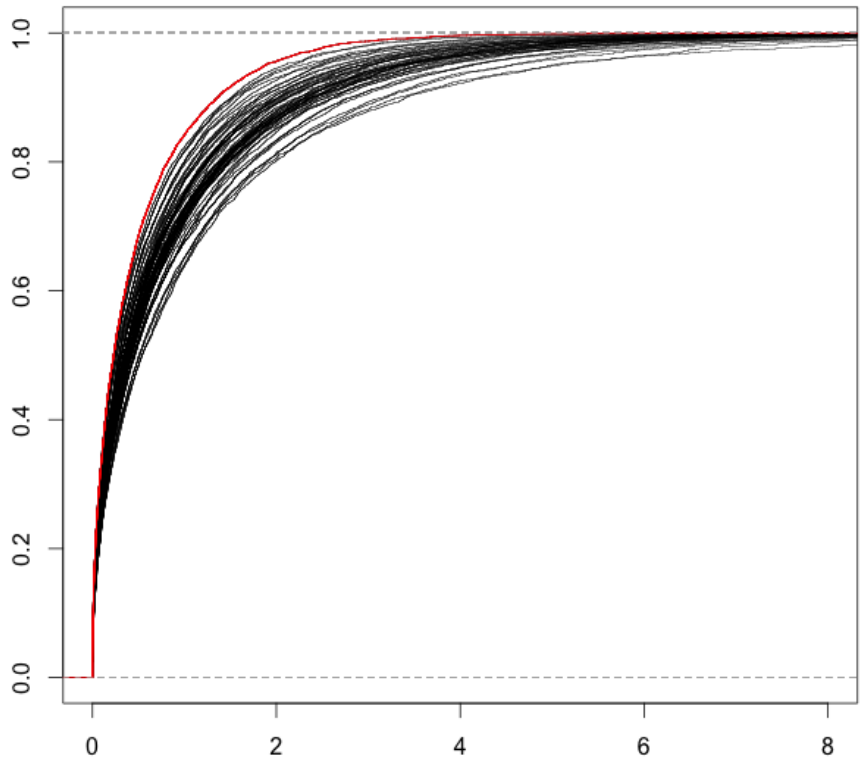
— empirical distribution function of $L(\widetilde{\theta}) - L(\theta^*)$ estimated with 5000 $\mathbf{Y}$ samples

— 50 empirical distribution functions of $L^{\circledcirc}(\widetilde{\theta}^{\circledcirc}) - L^{\circledcirc}(\widetilde{\theta})$ estimated with 5000 $u_1, \ldots, u_n \sim exp(1)$ samples



$Y_i = 0.25 \sin(X_i) + Lap(0, 2^{-1/2})$

the "small modelling bias" case

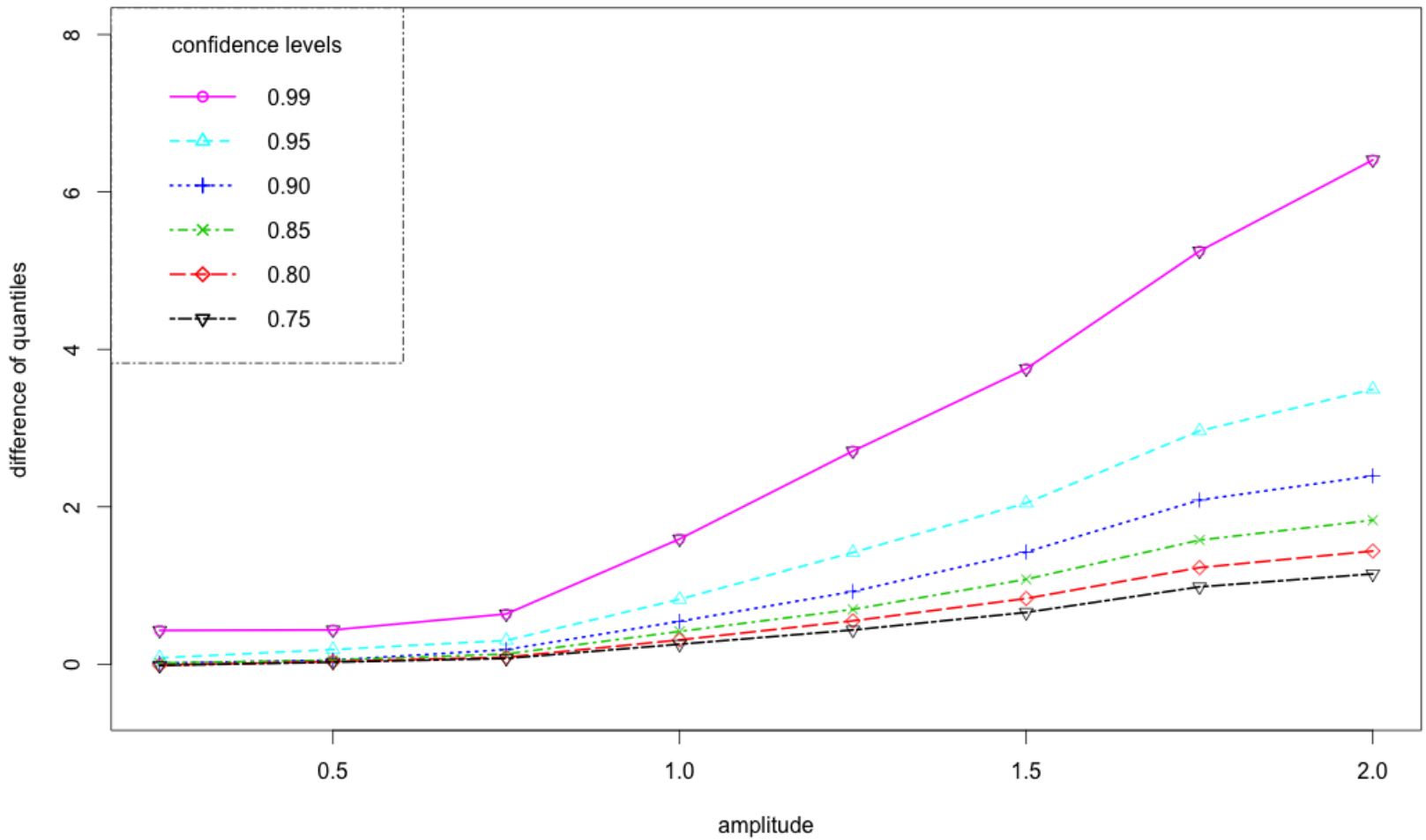


$Y_i = 1.25 \sin(X_i) + Lap(0, 2^{-1/2})\,,$

"small modelling bias" condition is not fulfilled

Numerical results (III). Biased model

The difference $\big($"Bootstrap quantile" $-$ "$\boldsymbol{Y}$-quantile"$\big)$ growing with modelling bias:



Logistic regression with bias

$$Y_i \sim Bernoulli(\beta X_i),$$

X_i are equidistant in $[0, 2]$,

$$\beta \in (0, 1/2],$$

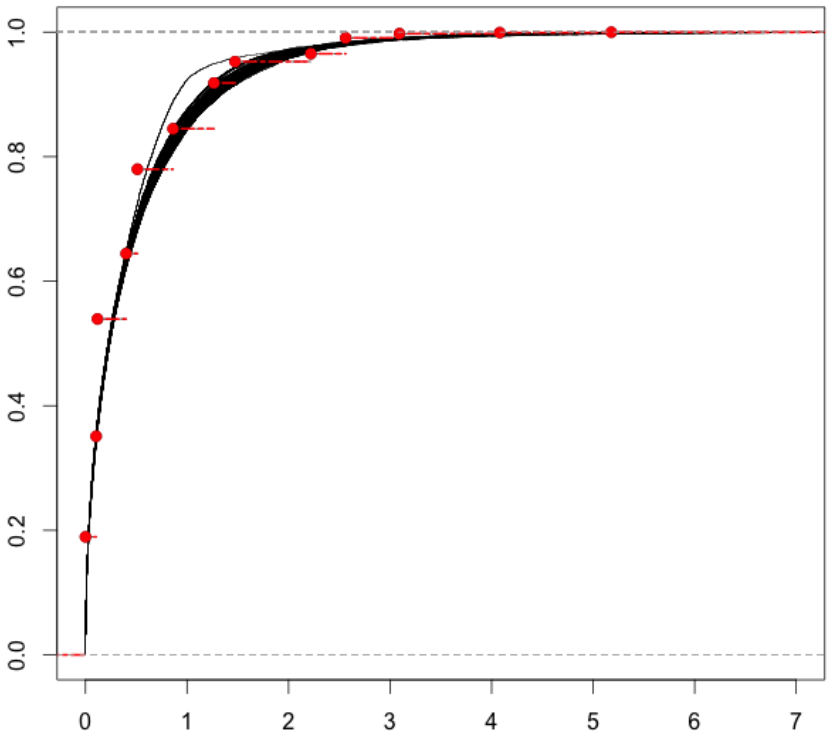
$$L(\theta) = \sum_{i=1}^{n} \left\{ Y_i \theta - \log(1 + \mathrm{e}^{\theta}) \right\},$$

$$\Rightarrow \quad \theta^* = \log\{\beta/(1-\beta)\}.$$

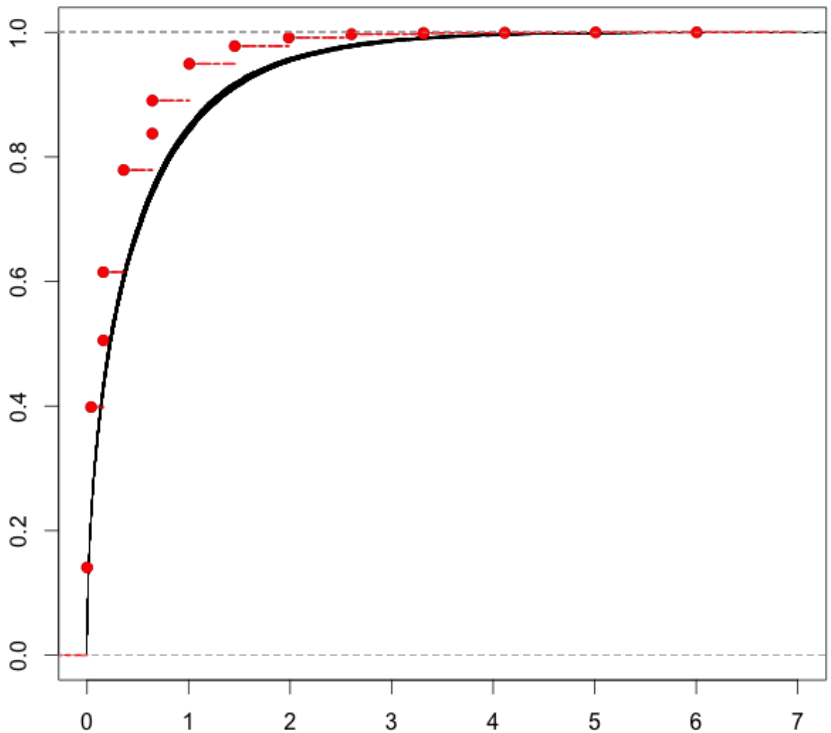
$n = 50$

— empirical distribution function of $L(\widetilde{\theta}) - L(\theta^*)$ estimated with 10000 $\boldsymbol{Y}$ samples

— 50 empirical distribution functions of $L^{\circledcirc}(\widetilde{\theta}^{\circledcirc}) - L^{\circledcirc}(\widetilde{\theta})$ estimated with 10000 $u_1, \ldots, u_n \sim exp(1)$ samples



$Y_i \sim Bernoulli(0.1X_i)$



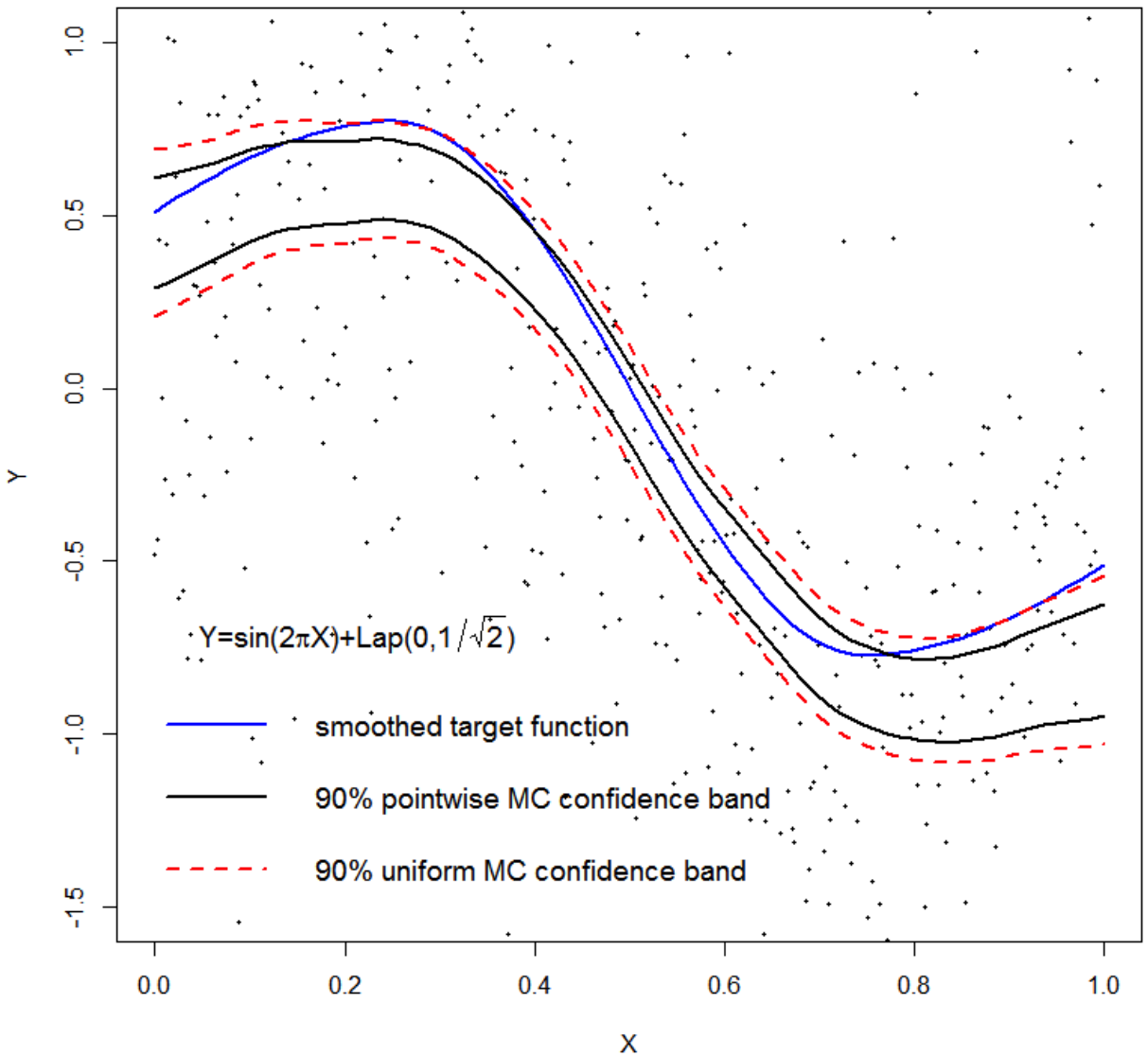
$Y_i \sim Bernoulli(0.5X_i)$

Outline

For a fixed sample size n construct a simultaneous confidence set for $f(x)$, when $x \in \mathcal{X} \subset \mathbb{R}^d$, with a given confidence level $1-\alpha$ and a minimal possible diameter.

$$\mathbb{P}\left\{\bigcap_{x\in\mathcal{X}} \{f(x) \in \mathcal{E}_x\}\right\} \geq 1-\alpha.$$

Uniform vs. pointwise confidence corridors



Simulated data:

$Y_i = \sin(2\pi X_i) + \varepsilon_i$,

$\varepsilon_i \sim Lap(0, 1/\sqrt{2})$ i.i.d.,

X_i are equidistant on $[0, 1]$,

sample size $= 500$,

confidence sets are based on the Nadaraya-Watson estimator with parabolic weights and bandwidth $= 0.25$.

Local likelihood function for $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$, $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$:

$$L(\boldsymbol{\theta}, \boldsymbol{x}) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \ell(Y_i, f(X_i, \boldsymbol{\theta})) w_i(\boldsymbol{x}),$$

$$\ell(Y_i, f(X_i, \boldsymbol{\theta})) \stackrel{\text{def}}{=} \frac{dP_i(f(X_i, \boldsymbol{\theta}))}{d\mu_0}(Y_i),$$

$f(\boldsymbol{x}, \boldsymbol{\theta})$ is a known smooth function $\mathbb{R}^d \times \mathbb{R}^p \mapsto \mathbb{R}$,

$w_i(\boldsymbol{x}) \stackrel{\text{def}}{=} K\left(\frac{\boldsymbol{x} - X_i}{h}\right)$ are localising weights,

$K(\boldsymbol{x}) \geq 0$ is a symmetric kernel function,

$h > 0$ is a fixed bandwidth (however, the results here are easily extended to the uniform in bandwidth bounds, when h lies on a fine finite grid).

- Local maximum likelihood estimate:

$$\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}) \stackrel{\mathrm{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta}, \boldsymbol{x}),$$

$$\widetilde{f}(\boldsymbol{x}) \stackrel{\mathrm{def}}{=} f(\boldsymbol{x}, \widetilde{\boldsymbol{\theta}}(\boldsymbol{x})).$$

- Target parameters (smoothed versions):

$$\boldsymbol{\theta}^*(\boldsymbol{x}) \stackrel{\mathrm{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} I\!\!E L(\boldsymbol{\theta}, \boldsymbol{x}),$$

$$f^*(\boldsymbol{x}) \stackrel{\mathrm{def}}{=} f(\boldsymbol{x}, \boldsymbol{\theta}^*(\boldsymbol{x})).$$

- Pointwise confidence sets

$$\mathcal{E}(\mathfrak{z}, \boldsymbol{x}) \stackrel{\text{def}}{=} \left\{ \boldsymbol{\theta} : L(\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{x}) - L(\boldsymbol{\theta}, \boldsymbol{x}) \le \mathfrak{z} \right\} \subset I\!R^p,$$

$$\mathcal{E}_f(\mathfrak{z}, \boldsymbol{x}) \stackrel{\text{def}}{=} \{ f(\boldsymbol{x}, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathcal{E}(\mathfrak{z}, \boldsymbol{x}) \} \subset I\!R.$$

- Uniform (simultaneous) confidence corridor:

$$\mathcal{E}_f(\mathfrak{z}) \stackrel{\text{def}}{=} \bigcup_{\boldsymbol{x} \in \mathcal{X}} \{ \boldsymbol{x} \times \mathcal{E}_f(\mathfrak{z}, \boldsymbol{x}) \} \subset I\!R^d \times I\!R.$$

- Aim: to build for each $\boldsymbol{x} \in \mathcal{X}$ a quantile function $\mathfrak{z}(\alpha, \boldsymbol{x})$ s.t.

$$I\!P \left(\bigcap_{\boldsymbol{x} \in \mathcal{X}} \left\{ f^*(\boldsymbol{x}) \in \mathcal{E}_f(\mathfrak{z}(\alpha, \boldsymbol{x}), \boldsymbol{x}) \right\} \right) \ge 1 - \alpha.$$

For a smooth $f(\cdot, \boldsymbol{\theta})$

$$f^*(\boldsymbol{x}) \in \mathcal{E}_f(\mathfrak{z}, \boldsymbol{x}) \;\Leftrightarrow\; \boldsymbol{\theta}^*(\boldsymbol{x}) \in \mathcal{E}(\mathfrak{z}, \boldsymbol{x})$$
$$\Leftrightarrow\; L(\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{x}) - L(\boldsymbol{\theta}^*(\boldsymbol{x}), \boldsymbol{x}) \le \mathfrak{z}.$$

Therefore, the initial problem translates into finding $\mathfrak{z}(\alpha, \boldsymbol{x})$ s.t.

$$\mathbb{P}\left(\sup_{\boldsymbol{x} \in \mathcal{X}} \left\{ L(\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{x}) - L(\boldsymbol{\theta}^*(\boldsymbol{x}), \boldsymbol{x}) - \mathfrak{z}(\alpha, \boldsymbol{x}) \right\} \le 0 \right) \ge 1 - \alpha.$$

Similarly to the parametric approach define the bootstrap objects for the local likelihood conditioned on $\boldsymbol{Y}$:

$$L^{\circledcirc}(\boldsymbol{\theta}, \boldsymbol{x}) \stackrel{\mathrm{def}}{=} \sum_{i=1}^{n} \ell(Y_i, f(X_i, \boldsymbol{\theta})) w_i(\boldsymbol{x}) u_i.$$

The target parameter

$$\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}) = \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} I\!\!E^{\circledcirc} L^{\circledcirc}(\boldsymbol{\theta}, \boldsymbol{x}).$$

Maximum likelihood ratio for the bootstrap case:

$$L^{\circledcirc}(\widetilde{\boldsymbol{\theta}}^{\circledcirc}, \boldsymbol{x}) - L^{\circledcirc}(\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{x}) \stackrel{\mathrm{def}}{=} \sup_{\boldsymbol{\theta} \in \Theta} L^{\circledcirc}(\boldsymbol{\theta}, \boldsymbol{x}) - L^{\circledcirc}(\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{x}).$$

$$\sqrt{2\{L(\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}),\boldsymbol{x}) - L(\boldsymbol{\theta}^*(\boldsymbol{x}),\boldsymbol{x})\}} \overset{\bigcap_{x\in\mathcal{X}}}{\approx} \|\boldsymbol{\xi}(\boldsymbol{x})\|$$

$$\wr\wr\, w_X$$

$$\sqrt{2\{L^{\circledcirc}(\widetilde{\boldsymbol{\theta}}^{\circledcirc}(\boldsymbol{x}),\boldsymbol{x}) - L^{\circledcirc}(\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}),\boldsymbol{x})\}} \overset{\bigcap_{x\in\mathcal{X}}}{\approx} \|\boldsymbol{\xi}^{\circledcirc}(\boldsymbol{x})\|$$

- $\overset{\bigcap_{x\in\mathcal{X}}}{\approx}$ denotes the uniform Wilks-type approximation,
- $\overset{w_X}{\approx}$ means approximation in distribution jointly in $\boldsymbol{x}\in\mathcal{X}$.
- The normalized scores and full Fisher Information matrices are:

$$\boldsymbol{\xi}(\boldsymbol{x}) \overset{\text{def}}{=} D_0^{-1}(\boldsymbol{x}_0)\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}^*(\boldsymbol{x}),\boldsymbol{x}) \qquad \boldsymbol{\xi}^{\circledcirc}(\boldsymbol{x}) \overset{\text{def}}{=} \mathfrak{D}_0^{-1}(\boldsymbol{x})\nabla_{\boldsymbol{\theta}} L^{\circledcirc}(\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}),\boldsymbol{x})$$

$$D_0^2(\boldsymbol{x}) \overset{\text{def}}{=} -\nabla^2_{\boldsymbol{\theta}} I\!\!E L(\boldsymbol{\theta}^*(\boldsymbol{x}),\boldsymbol{x}) \qquad \mathfrak{D}_0^2(\boldsymbol{x}) \overset{\text{def}}{=} -\nabla^2_{\boldsymbol{\theta}} I\!\!E^{\circledcirc} L^{\circledcirc}(\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}),\boldsymbol{x})$$

Let $\mathrm{G}_{\mathcal{X}}(\varepsilon)$ denote a finite ε-net on $\mathcal{X}$: $\forall \boldsymbol{x} \in \mathcal{X}\ \exists \boldsymbol{x}_0 \in \mathrm{G}_{\mathcal{X}}(\varepsilon) : \|\boldsymbol{x} - \boldsymbol{x}_0\| \leq \varepsilon$.

1. Pointwise step

For each $\boldsymbol{x} \in \mathrm{G}_{\mathcal{X}}(\varepsilon)$ compute an empirical distribution of the ratio $\sqrt{2\{L^{\circ}(\widetilde{\boldsymbol{\theta}}^{\circ}(\boldsymbol{x}), \boldsymbol{x}) - L^{\circ}(\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{x})\}}$, which yields the pointwise quantile function $\mathfrak{z}^{\circ}(\alpha, \boldsymbol{x})$ for each $\alpha \in (0,1)$ and $\boldsymbol{x} \in \mathrm{G}_{\mathcal{X}}(\varepsilon)$.

2. Correction for the uniformity

By an iterative procedure find a maximum value $\alpha(\varepsilon)$ s.t.

$$\mathbb{P}\left(\sup_{\boldsymbol{x} \in \mathrm{G}_{\mathcal{X}}(\varepsilon)} \left[\sqrt{2\{L^{\circ}(\widetilde{\boldsymbol{\theta}}^{\circ}(\boldsymbol{x}), \boldsymbol{x}) - L^{\circ}(\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{x})\}} - \mathfrak{z}^{\circ}(\alpha(\varepsilon), \boldsymbol{x}) \right] \leq 0 \right) \geq 1 - \alpha.$$

The value $\alpha(\varepsilon)$ is a corrected confidence level for $\sup_{\boldsymbol{x} \in \mathrm{G}_{\mathcal{X}}(\varepsilon)}$.

It holds $\alpha(\varepsilon)/\alpha \geq \mathtt{c}\,|\mathrm{G}_{\mathcal{X}}(\varepsilon)|^{-1}$.

Local constant regression:

$$Y_i = \sin(2\pi X_i) + \varepsilon_i, \quad \varepsilon_i \sim Lap(0, 2^{-1/2}), \text{ i.i.d.},$$

X_i are equidistant in $[0,1]$,

$$L(\theta) = -\frac{1}{2}\sum_{i=1}^{n}(Y_i - \theta)^2 w_i(x),$$

$$w_i(x) = \begin{cases} 1 - \{(x - X_i)/h\}^2, & \text{if } |(x - X_i)/h| \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Nadaraya-Watson estimate:

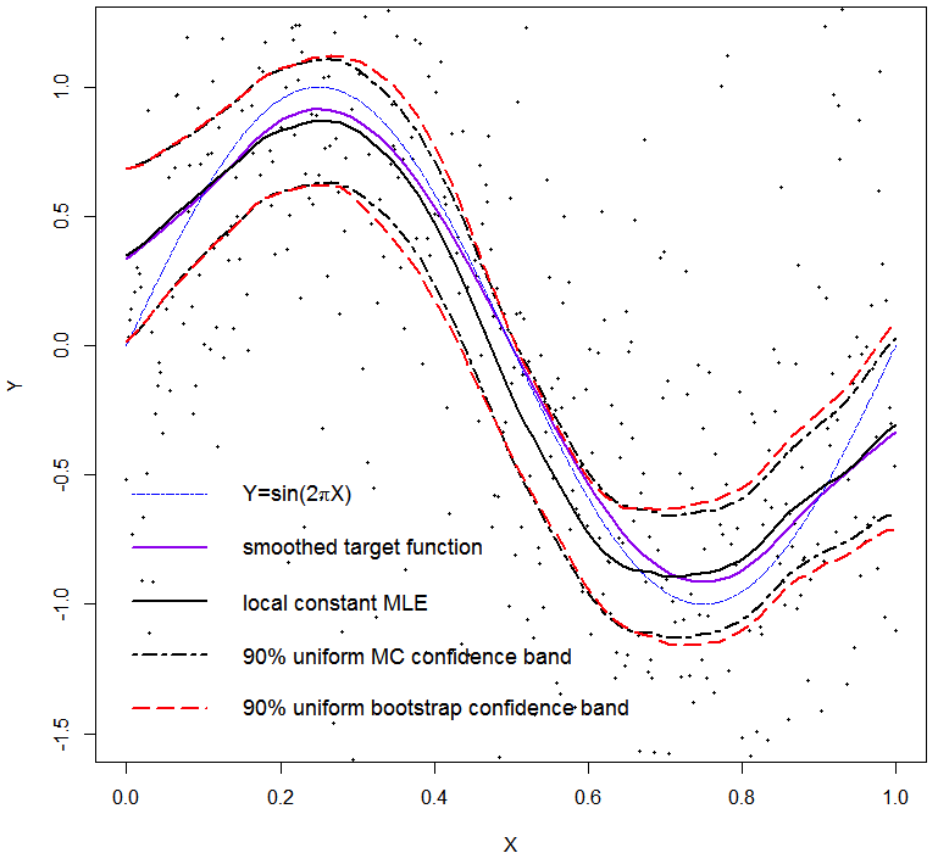
$$\widetilde{\theta}(x) = \sum_{i=1}^{n} Y_i \frac{w_i(x)}{\sum_{i=1}^{n} w_i(x)},$$

Smoothed target function:

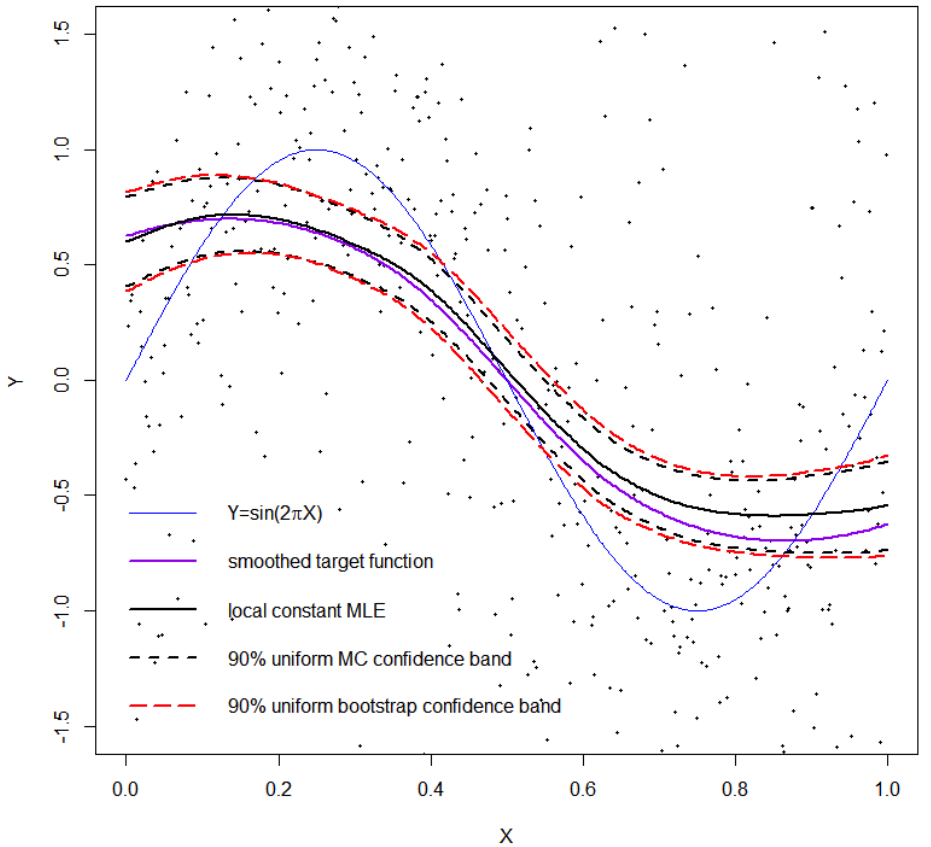
$$\theta^*(x) = \sum_{i=1}^{n} I\!\!E Y_i \frac{w_i(x)}{\sum_{i=1}^{n} w_i(x)}.$$

Numerical results. Uniform confidence corridors

$$Y_i = \sin(2\pi X_i) + \varepsilon_i, \quad \varepsilon_i \sim Lap(0, 2^{-1/2}), \text{ i.i.d.}, \quad n = 500\,.$$



$h = 0.15$



$h = 0.35$

Let $\boldsymbol{x}_G$ denote the projection of a point $\boldsymbol{x} \in \mathcal{X}$ on the grid $\mathrm{G}_{\mathcal{X}}(\varepsilon)$:

$$\boldsymbol{x}_G \stackrel{\mathrm{def}}{=} \operatorname{argmin}_{\boldsymbol{x}_1 \in \mathrm{G}_{\mathcal{X}}(\varepsilon)} \|\boldsymbol{x} - \boldsymbol{x}_1\| .$$

It holds with probability $\geq 1 - \mathtt{C} \left|\mathrm{G}_{\mathcal{X}}(\varepsilon)\right| \mathrm{e}^{-\mathtt{y}} = 1 - \mathtt{C} \exp\left\{-\mathtt{y} + \log \left|\mathrm{G}_{\mathcal{X}}(\varepsilon)\right|\right\}$

$$\sup_{\boldsymbol{x}:\|\boldsymbol{x}-\boldsymbol{x}_G\| \leq \varepsilon} \left| \sqrt{2\bigl\{L(\widetilde{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{x}) - L(\boldsymbol{\theta}^*(\boldsymbol{x}), \boldsymbol{x})\bigr\}} - \|\boldsymbol{\xi}(\boldsymbol{x}_G)\| \right| \leq \Delta(\varepsilon, \mathtt{y}),$$

where $\Delta(\varepsilon, \mathtt{y}) = \mathtt{C}(p + \mathtt{y})/\sqrt{(nh^d)}$.

Some further problems

- Adaptive choice of optimal bandwidth using a multiscale correction.
- Non-asymptotic bounds for Monte Carlo simulations (joint with Christian Bayer).
- Estimation of dynamic quantile/expectile curves (joint with project B1).
- Application to the multiple hypothesis testing: choice of a correction for multiplicity.

References

Arlot, S., Blanchard, G., and Roquain, E. (2010).

Some nonasymptotic results on resampling in high dimension. I. Confidence regions.

Ann. Statist., 38(1):51–82.

Beran, R. (1986).

Discussion: Jackknife, bootstrap and other resampling methods in regression analysis.

The Annals of Statistics, pages 1295–1298.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013).

Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors.

The Annals of Statistics, 41(6):2786–2819.

Hardle, W. and Mammen, E. (1993).

Comparing nonparametric versus parametric regression fits.

The Annals of Statistics, pages 1926–1947.

References


Jin, Z., Ying, Z., and Wei, L. J. (2001).
A simple resampling method by perturbing the minimand.
Biometrika, 88(2):381–390.


Lavergne, P. and Patilea, V. (2013).
Smooth minimum distance estimation and testing with conditional estimating equations: uniform in bandwidth theory.
Journal of Econometrics, 177(1):47–59.


Lindeberg, J. W. (1922).
Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung.
Mathematische Zeitschrift, 15(1):211–225.


Mammen, E. (1993).
Bootstrap and wild bootstrap for high dimensional linear models.
The Annals of Statistics, pages 255–285.

References

Spokoiny, V. (2012).

Parametric estimation. finite sample theory.

The Annals of Statistics, 40(6):2877–2909.

Spokoiny, V. (2014).

Wilks theorem for penalized maximum likelihood estimators.

Manuscript. arXiv:1205.0498.

Spokoiny, V. and Zhilova, M. (2014a).

Parametric confidence sets via multiplier bootstrap.

to appear.

Spokoiny, V. and Zhilova, M. (2014b).

Uniform confidence sets via multiplier bootstrap.

to appear.

References

Wu, C. F. J. (1986).

Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis.

The Annals of Statistics, 14(4):1261–1295+.

Thank you for your attention!