Bootstrap pro závislá data a detekce změn

Zuzana Prášková, MFF UK Praha

ROBUST 2016, 11.-16.9. 2016, Rejhotice

Why bootstrap?

- exact distribution of a statistic under consideration is difficult to compute
- asymptotic distribution exists but not in an explicit form
- bootstrap can provide better approximations to the exact distribution than the asymptotic one
- bootstrap can reduce bias of estimators
- ...

- Efron (1979) for iii sample $X_1, \ldots X_n$: Monte Carlo from the empirical distribution (repeating random sampling with replacement)
  - many variants and modifications
- Dependent observations
  - model-based methods: model fitting, resampling of residuals
  - model-free methods: overlapping (moving) or non-overlapping blocks, resampling of blocks

    (+)dependency structure is saved within the blocks

    (-)dependency structure is violated between the blocks

    (-)regularly spaced data are assumed
  - Lahiri (2003), Härdle, Horowitz, and Kreiss (2003), Paparoditis and Politis (2009), Kreiss, Paparoditis (2011),...

The basic resampling algorithm (non-random block length):

1. Let $b \in \mathbb{N}, b \ll n, L = \left\lfloor \frac{n}{b} \right\rfloor$ and $k = n - L \cdot b$. Define discrete uniform, independent random variables $t_1, t_2, \dots, t_{L+1}$ taking values in the set $I_{n,l}$ where
   - $I_{n,l} = \{1, 2, \dots, n - b + 1\}$ for overlapping blocks
   - $I_{n,l} = \{1, b + 1, 2b + 1, \dots, (L - 1)b + 1\}$ for non-overlapping blocks
2. Lay the blocks $(X_{t_s}, X_{t_s+1}, \dots, X_{t_s+b-1}), s = 1, \dots, L + 1$ end to end in the order sampled together and discard the last $b - k$ observations to form a bootstrap pseudo-series $X_1^*, \dots, X_n^*$.

The block bootstrap approximation of the distribution $\mathcal{L}_n = \mathcal{L}\left(c_n(T_n - \nu)\right)$ is then given by $\mathcal{L}_n^* = \mathcal{L}\left(c_n(T_n^* - \nu^*)\right)$, where $T_n^* = T_n(X_1^*, \dots, X_n^*)$ and ν^* denotes bootstrap parameter.

Block bootstrap with random block length (stationary bootstrap):

1. The lengths b_i of the blocks are i.i.d. random variables having a geometric distribution with parameter $p \in (0, 1)$.
2. The first b_1 pseudo-observations of the bootstrap time series $X_1^*, \ldots, X_n^*$ consist of observations $X_{t_1}, \ldots, X_{t_1+b_1}$, the next b_2 bootstrap observations are the observations of the second sampled block of random length b_2 and so on. The bootstrap data generating process is stopped once n bootstrap observations have been generated.

Circular block bootstrap

"Wrapping" the data before blocking:

$$X_i = X_{(i \bmod n)}, \quad i > n,$$
$$X_0 = X_n,$$

then define block $(X_i, X_{i+1}, \dots, X_{i+b-1})$ for any $i = 1, \dots, n$ and any block length $b > 0$

Advantage of circular block bootstrap:

- resulting bootstrap series is automatically centered around the sample mean
- an automatic procedure was developed for estimation of optimal length of blocks, Politis and White (2004, 2009)
  - A. Patton in Matlab code
  - R-package *np*

Dependent wild bootstrap Shao (2010)

$X_1, \dots, X_n$ dependent, satisfy model $X_t = \mu + \varepsilon_t$
bootstrap model:

$$X_t^* = \bar{X}_n + (X_t - \bar{X}_n) \cdot Z_t, \quad t = 1, \dots, n,$$

where $\bar{X}_n$ is the sample mean and $\{Z_t\}$ are random variables satisfying the following assumption:

- $\{Z_t\}$ is independent of $\{X_t\}$, $\mathrm{E}\, Z_t = 0$, $\mathrm{Var} Z_t = 1$ for $t = 1, \dots, n$
- $\{Z_t\}$ is stationary with autocovariance function $cov(Z_t, Z_{t+k}) = a(\frac{k}{m})$, where $a(\cdot)$ is a kernel and $m = m_n$ is a bandwidth

For dependent wild bootstrap
$\mathrm{E}^* X_t^* = \bar{X}_n, \mathrm{Var}^* X_t^* = (X_t - \bar{X}_n)^2$, and

$$\mathrm{Var}^* \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t^* \right) = \sum_{|k| \leq n-1} a\left(\frac{k}{M}\right) \widehat{R}(k) := \hat{\sigma}_n^2(m)$$

where

$$\widehat{R}(k) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n), k = 0, 1, \ldots, n-1 \\ \hat{R}(-k), k < 0 \end{cases}$$

$\hat{\sigma}_n^2(M)$ is a kernel estimator of the long-run variance
$\sigma^2 = \mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \right)$

Kernel estimators of long-run variance

Bartlett, Parzen, quadratic, flat-top,...

Optimal bandwidth automatic selection

- Andrews (1991) for Bartlett, Parzen, quadratic kernel
  R-package *sandwich*, Zeileis (2004)
- Politis (2003) for flat-top kernels
  R-package *iosmooth*

Application to change-point
Location model

$$X_t = \mu + e_t, \qquad t = 1, \dots, k_n,$$
$$= \mu + \delta + e_t, \qquad t = k_n + 1, \dots, n, \tag{1}$$

where $\mu \in \mathbb{R}$, $\delta = \delta_n \neq 0$ and $1 \leq k_n \leq n$ are parameters and $\{e_t\}_{t=1}^{\infty}$ are error terms
Test:

$$H_0 : k_n = n \text{ (no change) against } H_1 : k_n < n$$

Test statistic:

$$T_n = \max_{1 \leq k < n} \left\{ \sqrt{\frac{n}{k(n-k)}} |S_k| \right\},$$

S_k - k-th partial sum of OLS residuals from the model with $k_n = n$, i.e. $S_k = \sum_{t=1}^{k} (X_t - \bar{X}_n)$.

If H_0 holds then under various assumptions on errors (iid, linear process, strong mixing,...)

$$\lim_{n\to\infty} P\left(a(\log n) T_n \leq \sigma(x + b_1(\log n))\right) = \mathrm{e}^{-2\mathrm{e}^{-x}},$$

where

$$\begin{aligned} a(y) =& \sqrt{2 \log y} \\ b_p(y) =& 2 \log y + \frac{p}{2} \log \log y - \log\left(\Gamma\left(\frac{p}{2}\right)\right) \\ \sigma^2 =& \lim_{n\to\infty} \mathrm{Var}\left(\sqrt{n}\bar{X}_n\right) \end{aligned}$$

slow convergence, conservative critical values - use some variant of bootstrap!

- Kirch(2007) for block bootstrap, linear process, consistency Hušková, Kirch (2010) for circular block bootstrap (CB) and strong mixing sequences
- CB gives correct critical values also under alternative hypothesis
- Wild dependent bootstrap ?

asympt = 2.94				
ρ	MC	CB	CSB	DWB
0.7	1.85	1.53	1.45	1.79
0.5	2.17	1.86	1.79	2.27
0.3	2.38	2.15	2.10	2.72
0	2.33	2.32	2.32	2.91
-0.3	3.61	3.80	3.90	5.43
-0.5	3.82	3.70	3.81	5.80

Table: 90% quantiles of T_n statistic by Monte Carlo, Circular, Circular stationary and Dependent wild bootstrap, $n = 500$, location model, errors $AR(1)$ (M. Čellár, 2016)

asympt = 3.66				
ρ	MC	CB	CSB	DWB
0.7	2.21	1.97	1.88	2.42
0.5	2.63	2.31	2.24	2.94
0.3	2.80	3.63	2.16	3.42
0	2.78	2.78	2.78	3.61
-0.3	4.31	4.40	4.52	6.57
-0.5	4.82	4.42	4.55	7.03

Table: 95% quantiles of T_n statistic by Monte Carlo, Circular, Circular stationary and Dependent wild bootstrap, $n = 500$ (M. Čellár, 2016)

		n=100		n = 200		n = 500	
ρ	method	0.1	0.05	0.1	0.05	0.1	0.05
0.5	AS	0.052	0.008	0.085	0.023	0.277	0.120
	CB	0.302	0.165	0.360	0.219	0.586	0.446
	SB	0.318	0.179	0.377	0.226	0.590	0.444
	DWB	0.168	0.061	0.263	0.109	0.474	0.292
0	AS	0.189	0.054	0.519	0.331	0.963	0.903
	CB	0.447	0.312	0.733	0.621	0.989	0.980
	SB	0.449	0.311	0.732	0.609	0.990	0.979
	DWB	0.261	0.140	0.531	0.359	0.948	0.880
-0.5	AS	0.668	0.542	0.840	0.657	1.000	0.996
	CB	0.348	0.198	0.796	0.665	1.000	0.998
	SB	0.274	0.149	0.743	0.590	1.000	0.998
	DWB	0.115	0.031	0.452	0.231	0.929	0.836

Table: Achieved levels of power of bootstrap test procedures for $\delta = 0.5$ and $q_n = \frac{n}{2}$ (M. Čellár, 2016)

- Changes in regression model

$$\begin{aligned} Y_i &= \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, & i &= 1, \dots, k^* \\ &= \mathbf{x}_i^T (\boldsymbol{\beta} + \boldsymbol{\delta}_n) + \varepsilon_i, & i &= k^* + 1, \dots, n \end{aligned}$$

  where

  $1 < k^* \leq n$ is an unknown change point,

  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ are regressors,

  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, $\boldsymbol{\delta}_n = (\delta_{1n}, \dots, \delta_{pn})^T$ parameters,

  ε_i random errors

- Test hypothesis: $H_0 : k^* = n$ against $H_1 : k^* < n$

CUSUM test statistics - based on functionals of cumulative sums of LSE residuals

$$T_n(h) = \max_{0<k<n} \left\{ \frac{1}{nh^2(k/n)} \mathbf{S}_k^T \widehat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{S}_k \right\}$$

- $\mathbf{S}_k = \sum_{i=1}^k \mathbf{x}_i \widehat{\varepsilon}_i = \sum_{i=1}^k \mathbf{x}_i (y_i - \mathbf{x}_i^T \widehat{\beta})$, $\widehat{\beta}$ - LSE
- $\widehat{\boldsymbol{\Sigma}}_n$ - estimator of

$$\boldsymbol{\Sigma} = \lim_{n\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \varepsilon_i \right)$$

- h is a positive weight function defined on $(0,1)$

Theorem (asymptotic distribution of test statistic under H_0)

Let $\widehat{\mathbf{\Sigma}}_n - \mathbf{\Sigma} = o_P(1)$ and $h(t) = [t(1-t)]^\gamma,\ 0 \le \gamma < \frac{1}{2}, t \in (0,1)$. Then, as $n \to \infty$

$$T_n(h) \xrightarrow{\mathcal{D}} \sup_{0<t<1} \left\{ \sum_{j=1}^{p} B_j^2(t)/h^2(t) \right\}$$

where B_j are independent Brownian bridges.

- Large values of test statistics detect that the null hypothesis is violated
- Holds under various assumptions on regressors and errors
- Holds also for M-estimators and M-residuals
- Critical values have to be simulated

Dependent wild bootstrap

Under H_0

$$\mathbf{S}_k = \sum_{i=1}^{k} \mathbf{x}_i \widehat{\varepsilon}_i = \sum_{i=1}^{k} \mathbf{x}_i \varepsilon_i - \mathbf{C}_k \mathbf{C}_n^{-1} \sum_{i=1}^{n} \mathbf{x}_i \varepsilon_i, \quad \mathbf{C}_k = \sum_{i=1}^{k} \mathbf{x}_i \mathbf{x}_i^T$$

Bootstrap version $\mathbf{S}_k^*$: replace ε_i by ε_i^* in $\mathbf{S}_k$

- $\varepsilon_i^* = \widehat{\varepsilon}_i Z_i$
- Z_i independent of $\mathcal{F}(\mathbf{x}, \boldsymbol{\varepsilon}) = \mathcal{F}\{\mathbf{x}_1, \ldots, \mathbf{x}_n, \varepsilon_1 \ldots, \varepsilon_n\}$
- $\mathrm{E}\,|Z_i|^{2+\Delta} < \infty, \;\; \Delta > 0$
- $\mathrm{E}\, Z_i = 0, \mathrm{Var}\, Z_i = 1, \mathrm{Cov}(Z_i, Z_j) = a((i-j)/m), \; a(.)$ - kernel, $m = m(n)$ bandwidth
- $Z_i = Z_{i,n}$ is $m(n)$-dependent triangular array

Bootstrap statistic

$$T_n^*(h) = \max_{0<k<n} \left\{ \frac{1}{nh^2(k/n)} \mathbf{S}_k^{*\,T} \widehat{\boldsymbol{\Sigma}}_n^{*-1} \mathbf{S}_k^* \right\}$$

$$\widehat{\boldsymbol{\Sigma}}_n^* = \mathrm{Var}^* \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \varepsilon_i^* \right)$$

Asymptotic results and consistency

- assumption of L_p-$m-$ approximability:
  For any $i \in \mathbb{Z}$, $\mathbf{x}_i = \mathbf{h}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_{i-1}, \ldots)$, where $\mathbf{h}$ is measurable, $\{\boldsymbol{\xi}_i\}_i$ is a sequence of i.i.d. random vectors and $E\|\mathbf{x}_i\|^{2+\Delta} < \infty$ for some $\Delta > 0$.
  For all $i \in \mathbb{Z}$,
  $$\sum_{m=1}^{\infty} \|\mathbf{x}_i - \mathbf{x}_i^{(m)}\|_p < \infty$$
  where
  $$\mathbf{x}_i^{(m)} = \mathbf{h}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_{i-1}, \ldots \boldsymbol{\xi}_{i-m+1}, \boldsymbol{\xi}_{i-m}^{(m)}, \boldsymbol{\xi}_{i-m-1}^{(m)}, \ldots),$$
  $\boldsymbol{\xi}_{i-m}^{(m)}, \boldsymbol{\xi}_{i-m-1}^{(m)}, \ldots$ are i.i.d. with the same distribution as $\boldsymbol{\xi}_i$ independent of $\{\boldsymbol{\xi}_i\}_i$
  $\mathbf{x}_i^{(m)}$-$m-$dependent

- $\{\mathbf{x}_i\} - L_p$-$m-$ approximable, $p = 2 + \Delta$
- $\{\varepsilon_i\} - L_p$-$m-$ approximable, $p = 2$
- finite moments up to $4 + \Delta$
- $\{\mathbf{x}_i\}$ and $\{\varepsilon_i\}$ mutually independent
- $\{Z_{i,n}\}$ m_n dependent, $m_n = o(n^{\Delta/(2+\Delta)})$

then

$$P^*(T_n^*(h) \leq x) \xrightarrow{P} P\Big(\sup_{0<t<1} \sum_{j=1}^{p} B_j^2(t)/h^2(t) \leq x\Big)$$

uniformly in x under H_0 and local alternatives $||\boldsymbol{\delta}|| = O(n^{-1/2})$
P^* is the conditional probability given $\mathbf{x}_i, y_i, i = 1, \ldots, n$ (resp $\mathbf{x}_i, \varepsilon_i, i = 1, \ldots, n$)

Crucial step: a conditional functional central limit theorem

Consider process

$$\mathbf{Y}_n(t) = \frac{1}{\sqrt{n}} \mathbf{\Sigma}_n^{*-\frac{1}{2}} \sum_{i=1}^{\lfloor nt \rfloor} \mathbf{x}_i \varepsilon_i Z_{i,n}, \quad t \in [0,1].$$

$$\mathbf{\Sigma}_n^* = \mathrm{Var}^* \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \varepsilon_i Z_{i,n} \right)$$

Then, as $n \to \infty$,

$$\{\mathbf{Y}_n(t), t \in [0,1]\} \overset{*}{\Rightarrow} \{\mathbf{W}_d(t), t \in [0,1]\} \text{ } \textit{almost surely } [P]$$

where $\{\mathbf{W}_d(t), t \in [0,1]\}$ is a standard d-dimensional Wiener process on $[0,1]$ and $\overset{*}{\Rightarrow}$ means the weak convergence with respect to P^*.

$\mathbf{X} = (1, X_i), X_i \sim N(0,1), \varepsilon \sim AR(1)$ with the parameter ρ

quantiles	δ	90%	95%	99%
asymptotic		2.1080	2.5036	3.3621
$\rho = 0.3$				
simulated	$[0, 0]$	2.0604	2.4139	3.2456
bootstrap	$[0, 0]$	1.9363	2.2329	2.8990
bootstrap	$[0.25, 0.25]$	2.0088	2.3140	2.9731
bootstrap	$[0.5, 0.5]$	2.2187	2.5736	3.2827
$\rho = 0.5$				
simulated	$[0, 0]$	2.2670	2.6534	3.4631
bootstrap	$[0, 0]$	1.9858	2.3059	2.9615
bootstrap	$[0.25, 0.25]$	2.0862	2.4022	3.0988
bootstrap	$[0.5, 0.5]$	2.2436	2.5788	3.2634

Table: DWB: Asymptotic, simulated and bootstrap quantiles (based on 500 bootstrap samples and for 500 repetitions), $n = 250$

$\mathbf{X} =\sim N_2(0, V), \varepsilon \sim AR(1)$ with the parameter ρ

quantiles	δ	90%	95%	99%
asymptotic		2.1080	2.5036	3.3621
$\rho = 0.3$				
simulated	$[0, 0]$	1.8955	2.1969	2.8832
bootstrap	$[0, 0]$	1.9189	2.2235	2.8514
bootstrap	$[0.25, 0.25]$	2.0152	2.3279	2.9737
bootstrap	$[0.5, 0.5]$	2.3812	2.7713	3.5455
$\rho = 0.5$				
simulated	$[0, 0]$	1.8979	2.2323	2.9130
bootstrap	$[0, 0]$	1.8041	2.0885	2.6841
bootstrap	$[0.25, 0.25]$	1.9858	2.2880	2.9200
bootstrap	$[0.5, 0.5]$	2.3111	2.6815	3.4183

Table: DWB: Asymptotic, simulated and bootstrap quantiles (based on 500 bootstrap samples and for 500 repetitions), $n = 250$

$\mathbf{X} = (1, X_i), X_i \sim N(0,1), \varepsilon \sim AR(1)$ with the parameter ρ

asymptotic			bootstrap		
$\rho = 0.3$			$\rho = 0.3$		
$\boldsymbol{\delta}$	[0,0]	0.0458	$\boldsymbol{\delta}$	[0,0]	0.0732
	[0.25,0.25]	0.3892		[0.25,0.25]	0.4644
	[0.5,0.5]	0.9638		[0.5,0.5]	0.9550
$\rho = 0.5$			$\rho = 0.5$		
$\boldsymbol{\delta}$	[0,0]	0.0716	$\boldsymbol{\delta}$	[0,0]	0.0834
	[0.25,0.25]	0.3316		[0.25,0.25]	0.3654
	[0.5,0.5]	0.8840		[0.5,0.5]	0.8688

Table: DWB: Empirical level of rejection based on asymptotic and bootstrap critical values, nominal level $\alpha = 0.05$, 500 bootstrap samples and 5.000 simulations, $n = 250$

$\mathbf{X} = \sim N_2(0, V), \varepsilon \sim AR(1)$ with the parameter ρ

asymptotic			bootstrap		
$\rho = 0.3$			$\rho = 0.3$		
$\boldsymbol{\delta}$	[0,0]	0.0306	$\boldsymbol{\delta}$	[0,0]	0.0484
	[0.25,0.25]	0.8440		[0.25,0.25]	0.8742
	[0.5,0.5]	1.0000		[0.5,0.5]	1.0000
$\rho = 0.5$			$\rho = 0.5$		
$\boldsymbol{\delta}$	[0,0]	0.0238	$\boldsymbol{\delta}$	[0,0]	0.0690
	[0.25,0.25]	0.7538		[0.25,0.25]	0.8238
	[0.5,0.5]	1.0000		[0.5,0.5]	0.9996

Table: DWB: Empirical level of rejection based on asymptotic and bootstrap critical values, nominal level $\alpha = 0.05$, 500 bootstrap samples and 5.000 simulations, $n = 250$

Thank You for Your Attention!