Bootstrap for Dependent Hilbert Space-Valued Random Variables

Martin Wendler
joint work with H. Dehling and O. Sharipov

RUHR-UNIVERSITÄT BOCHUM

Worshop on Recent Advances and Trends in Time Series Analysis at BIRS
April 27 to May 2, 2014
Outline

Theory

CLT in Hilbert Space under Dependence
Bootstrap in Hilbert Space under Dependence

Application

Cramér-von Mises-Statistic
General Degenerate V-Statistics
Hilbert space:
- $H$ complete and separable vector space
- $\langle \cdot, \cdot \rangle$ inner product
- $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$ norm

Observations:
- $(X_n)_{n \in \mathbb{N}}$ $H$-valued sequence of random variables
- stationary
- $E \| X_n \|^2 < \infty$
- $S_n = \sum_{i=1}^{n} X_i$
Absolute Regularity

**Definition**

Absolute regularity coefficient \((\beta_m)_{m \in \mathbb{N}}\) of process \((\xi_n)_{n \in \mathbb{Z}}\) given by

\[
\beta_m = \left| E \sup_{\mathcal{A} \in \mathcal{F}_m^\infty} (P(A|\mathcal{F}_{-\infty}^0) - P(A)) \right|
\]

where \(\mathcal{F}_a^b\) is the \(\sigma\)-field generated by \(\xi_a, \ldots, \xi_b\), and \((\xi_n)_{n \in \mathbb{N}}\) called **absolutely regular**, if \(\beta_m \to 0\)

- satisfied for linear processes only under extra conditions
- not satisfied for dynamical systems
- difficult to check in practice
Aproximating Functionals

**Definition**

$\{X_n\}_{n \in \mathbb{N}}$ called *1-approximating functional* on a process $\{\xi_n\}_{n \in \mathbb{Z}}$, if there exists a sequence $\{a_m\}_{m \in \mathbb{N}}$ with $a_m \to 0$ as $m \to 0$ and for every $m$ a function $f_m : S^{2m+1} \to H$ such that

$$E \| X_0 - f_m(\xi_{-m}, \ldots, \xi_m) \| \leq a_m \quad \text{for all } m \in \mathbb{N}.$$ 

- linear processes
- GARCH-processes
- dynamical systems, $X_{n+1} = T(X_n)$ for piecewise smooth and expanding map $T : [0, 1] \to [0, 1]$
Central Limit Theorem

**Theorem**

1. $E \|X_1\|^{2+\delta} < \infty$ for a $\delta > 0$,
2. $(X_n)_{n \in \mathbb{Z}}$ be 1-approximating with $\sum_{m=1}^{\infty} (a_m)^{\delta/(1+\delta)} < \infty$,
3. $(\xi_n)_{n \in \mathbb{Z}}$ absolutely regular with $\sum_{m=1}^{\infty} (\beta_m)^{\delta/(2+\delta)} < \infty$.

Then $S_n = \sum_{i=1}^{n} X_i$ satisfies CLT, i.e.

$$\frac{1}{\sqrt{n}} (S_n - ES_n) \Rightarrow N,$$

$N$ Gaussian r.v. with mean 0 and covariance operator $V$ defined by

$$\langle V(x), y \rangle = \sum_{j=-\infty}^{\infty} E\langle X_0, x \rangle \langle X_j, y \rangle.$$
Nonoverlapping Block Bootstrap

Construction of new samples by drawing blocks of length \( p = p(n) \) with replacement \( k = \left\lfloor \frac{n}{p} \right\rfloor \) times, so for \( i, j = 1, 2, \ldots, k \):

\[
P \left[ \left( X_{(i-1)p+1}, \ldots, X_{ip}^* \right) = \left( X_{(j-1)p+1}, \ldots, X_{jp} \right) \right] = \frac{1}{k}
\]

Conditions on the block length:

- \( p(n) \to \infty \)
- \( p(n) \leq C n^{1-\epsilon} \) for some \( \epsilon > 0 \)
- \( p(n) = p(2^l) \) for \( 2^l < n \leq 2^{l+1}, \quad l = 1, 2, \ldots \)

Known result only for stationary bootstrap, see Politis and Romano (1994)
Bootstrap Notation

Bootstrapped probability and expectation:

\[ P^* \left( (X_1^*, \ldots, X_{kp}) \in M \right) := P \left( (X_1^*, \ldots, X_{kp}) \in M \mid X_1, \ldots, X_n \right) \]

\[ E^* \left[ g(X_1^*, \ldots, X_{kp}) \right] := E \left[ g(X_1^*, \ldots, X_{kp}) \mid X_1, \ldots, X_n \right] \]

Bootstrapped partial sum:

\[ S_n^* := \sum_{i=1}^{kp} X_i^*, \quad \bar{X}_n^* = \frac{1}{kp} S_n^* \]

Bootstrapped expectation of mean:

\[ E^* \left[ \bar{X}_n^* \right] = \frac{1}{kp} \sum_{i=1}^{kp} X_i =: \bar{X}_{n,kp} \]
Bootstrap Consistency

Theorem

► $E \| X_1 \|^{2+\delta} < \infty$ for a $\delta > 0$,
► $\sum_{m=1}^{\infty} (a_m)^{\delta'/(1+\delta')} < \infty$, $\sum_{m=1}^{\infty} m^{3/2} a_m < \infty$ for a $\delta' \in (0, \delta)$,
► $\sum_{m=1}^{\infty} (\beta_m)^{\delta'/(2+\delta')} < \infty$, $\sum_{m=1}^{\infty} m \beta_m < \infty$.

Then almost surely

$$
\frac{1}{\sqrt{kp}} \left( S_n^* - kp \bar{X}_{n,kp} \right) \Rightarrow^* N,
$$

$N$ Gaussian r.v. with mean 0 and covariance operator $V$ defined by

$$
\langle V(x), y \rangle = \sum_{j=-\infty}^{\infty} E \langle X_0, x \rangle \langle X_j, y \rangle.
$$
Idea of Proof

1. prove convergence of blocks

\[
E^* \left[ f_m \left( \frac{1}{\sqrt{p}} \left( \sum_{i=1}^{p} X_i^* - p\bar{X}_{n,kp} \right) \right) \right] = \frac{1}{k} \sum_{j=1}^{k} f_m \left( \frac{1}{\sqrt{p}} \left( \sum_{i=(j-1)p+1}^{jp} X_i - p\bar{X}_{n,kp} \right) \right) \xrightarrow{n \to \infty} E \left[ f_m (N) \right]
\]

almost surely for countable collection \((f_m)_{m \in \mathbb{N}}\) of Lipschitz-continuous, bounded functions \(f_m : H \to \mathbb{R}\),

2. Varadarajan (1958):

\[
\frac{1}{\sqrt{p}} \left( \sum_{i=1}^{p} X_i^* - p\bar{X}_{n,kp} \right) \Rightarrow^* N
\]
3. $k$ fixed: as blocks are conditionally independent

$$
\frac{1}{\sqrt{kp}} (S_n^* - kp\bar{X}_{n,kp}) = \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \left( \frac{1}{\sqrt{p}} \sum_{i=(j-1)p+1}^{jp} X_i^* - p\bar{X}_{n,kp} \right)
$$

$$
\Rightarrow \frac{1}{\sqrt{k}} \sum_{j=1}^{k} N_j \approx N
$$

where $N_1, \ldots, N_k$ are independent copies of $N$

4. $k \to \infty$: similar to Bickel, Freedman (1981)
Test Statistic

Real-valued observations $Y_n$ with marginal distribution function $F$

Hypothesis: $H_0 : F = F_0$

Test: weighted $L^2$-distance of empirical distribution function $F_n$ with

$$F_n(t) := \frac{1}{n} \sum_{i=1}^{n} 1 \{ Y_i \leq t \}$$

Test statistic: for bounded, integrable weight function $w$

$$T_n := \int (F_n(t) - F_0(t))^2 w(t) dt$$
Hilbert Space

\( H \): space of measurable functions \( f : \mathbb{R} \to \mathbb{R} \) with \( \langle f, f \rangle < \infty \), where

\[
\langle f, g \rangle := \int f(t)g(t)w(t)dt
\]

Lipschitz-continuous mapping \( \mathbb{R} \to H \) with

\[
y \to 1\{y \leq \cdot\}
\]

For \( (Y_n)_{n \in \mathbb{N}} \) 1-approximating, \( \mathbb{R} \)-valued sequence:

1. \( X_n = 1\{Y_n \leq \cdot\} \) 1-approximating \( H \)-valued sequence
2. by CLT and continuous mapping theorem for Gaussian r.v. \( N \)

\[
nT_n = \left\| \sqrt{n}(F_n - F_0) \right\|^2 = \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - EX_i) \right\|^2 \Rightarrow \| N \|^2
\]
Bootstrap Version

Bootstrapped empirical distribution function $F_n^*$ (partial sum in $H$):

$$F_n^*(t) := \frac{1}{pk} \sum_{i=1}^{pk} 1\{Y_i^* \leq t\}$$

Bootstrapped Cramer-von Mises-statistic:

$$T_n^* := \int (F_n^*(t) - F_n(t))^2 w(t) dt$$

Almost surely as $n \to \infty$

$$pkT_n^* = \left\| \sqrt{pk} (F_n^* - F_n) \right\|^2 \Rightarrow^* \|N\|^2$$
Definition

**Definition**

\( h : \mathbb{R}^2 \rightarrow \mathbb{R} \) symmetric, measurable function, \((Y_n)_{n \in \mathbb{N}} \) \( \mathbb{R} \)-valued r.v.,

\[
V_n := \frac{1}{n^2} \sum_{i,j=1}^{n} h(Y_i, Y_j)
\]

called von Mises-statistic with kernel \( h \)

**Assumptions:**

1. \( Eh(Y_1, Y_1) < \infty \)
2. \( h \) Lipschitz-continuous
3. \( h \) degenerate, i.e. \( E(h(x, Y_i)) = 0 \) for all \( x \in \mathbb{R} \)
4. \( h \) positive definite, i.e. for all \( c_i, y_i: \sum_{i,j=1}^{m} c_i c_j h(y_i, y_j) \geq 0 \)
Representation of Kernel

By extension of Mercer’s theorem (Sun, 2005):

\[ h(x, y) = \sum_{l=1}^{\infty} \lambda_l \Phi_l(x) \Phi_l(y) \]

with the following properties

- \( E (h(x, Y_1) \Phi_l(Y_1)) = \lambda_l \Phi_l(x) \)
- \( E \Phi_l(Y_1) = 0 \) for all \( l \in \mathbb{N} \)
- \( E \Phi_l^2(Y_1) = 1 \) for all \( l \in \mathbb{N} \)
- \( E \Phi_{l_1}(Y_1) \Phi_{l_2}(Y_1) = 0 \) for all \( l_1 \neq l_2 \)
- \( \lambda_l \geq 0 \) for all \( l \in \mathbb{N} \)
- \( \sum_{l=1}^{\infty} \lambda_l < \infty \)
Hilbert Space

$H$: Hilbert space of sequences $x = (x_i)_{i \in \mathbb{N}}$ with $\langle x, x \rangle < \infty$ and inner product

$$\langle x, z \rangle = \sum_{l=1}^{\infty} \lambda_l x_l z_l.$$ 

$\frac{1}{2}$-Hölder-continuous mapping $\mathbb{R} \to H$ with $y \to (\Phi_l(y))_{l \in \mathbb{N}}$

$$nV_n = \frac{1}{n} \sum_{i,j=1}^{n} h(Y_i, Y_j) = \frac{1}{n} \sum_{i,j=1}^{n} \sum_{l=1}^{\infty} \lambda_l \Phi_l(Y_i) \Phi_l(Y_j)$$

$$= \sum_{l=1}^{\infty} \lambda_l \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Phi_l(Y_i) \right) \left( \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Phi_l(Y_j) \right) = \left\| \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Phi_l(Y_i) \right)_{l \in \mathbb{N}} \right\|^2$$

so by by CLT for a Gaussian $H$-valued r.v. $N$: $nV_n \Rightarrow \|N\|^2$
Bootstrap Version

Bootstrapping and centering of $H$-valued sequence $((\Phi_l(Y_n))_{l \in \mathbb{N}})_{n \in \mathbb{N}}$:

$$V_n^* = \sum_{l=1}^{\infty} \lambda_l \left( \frac{1}{pk} \sum_{i=1}^{pk} \Phi_l(Y_i^*) - \frac{1}{pk} \sum_{i=1}^{pk} \Phi_l(Y_i) \right)^2$$

$$= \frac{1}{(pk)^2} \sum_{i,j=1}^{pk} h(Y_i^*, Y_j^*) - \frac{2}{(pk)^2} \sum_{i,j=1}^{pk} h(Y_i^*, Y_j) + \frac{1}{(pk)^2} \sum_{i,j=1}^{pk} h(Y_i, Y_j),$$

known before as artificial degeneration

Bootstrapping nondegenerate $V$-statistics (Sharipov, Wendler, 2012):

$$\tilde{V}_n^* = \frac{1}{(pk)^2} \sum_{i,j=1}^{pk} h(Y_i^*, Y_j^*) - V_n$$
Bootstrap Consistency

Theorem

1. \( h \) degenerate, Lipschitz-continuous, positive definite, \( \mathbb{E}h(Y_1, Y_1) < \infty \)

2. 1-approximation, \( \sum_{m=1}^{\infty} (a_m)^{\delta'/(1+2\delta')} < \infty \), \( \sum_{m=1}^{\infty} m^{3/2} \sqrt{a_m} < \infty \)

3. underlying sequence absolutely regular with \( \sum_{m=1}^{\infty} (\beta_m)^{\delta'/(1+\delta')} < \infty \), \( \sum_{m=1}^{\infty} m\beta_m < \infty \)

4. block length with \( p = O(n^{1-\epsilon}) \) for some \( \epsilon > 0 \) and \( p_n = p_{2^l} \) for \( n = 2^{l-1} + 1, \ldots, 2^l \)

Then almost surely \( nV_n \) and \( pkV_n^* \) converge to the same limit in distribution.

Similar result by Leucht, Neumann (2013) for dependent wild bootstrap
Summary

- nonoverlapping block bootstrap consistent in Hilbert space
- broad class of dependent processes, mild assumptions on block length
- new method to show bootstrap consistency of von Mises-statistics

Thank you for listening!

- Questions?
- Comments?
References