ERNST MORITZ ARNDT
UNIVERSITÄT GREIFSWALD

Wissen
lockt.
Seit 1456

Subsampling and Self-Normalization under Long Range Dependence

Martin Wendler
joint work with Annika Betken (Bochum)

Workshop: Recent developments in statistics for complex dependent data
Loccum, August 27 - 30, 2015

Institut für
Mathematik und Informatik

Outline

Long Range Dependence

long range depenent process: subordinated Gaussian sequences

- $(\xi_n)_{n\in\mathbb{N}}$ stationary, Gaussian
- $\text{Var}(\xi_n) = 1$
- $\gamma(k) := \text{Cov}(\xi_1, \xi_{k+1}) = k^{-D}L_\gamma(k)$
- $D \in (0, 1]$, L_γ slowly varying

observations:

- $X_n = G(\xi_n) + \mu_n$
- G measurable, $\mu_n \in \mathbb{R}$
- G, μ_n unknown

Change Point Problem

$$X_n = G(\xi_n) + \mu_n$$

hypothesis:

$$\mathbf{H} : \mu_1 = \ldots = \mu_n$$

against alternative:

$$\mathbf{A} : \mu_1 = \ldots = \mu_k \neq \mu_{k+1} = \ldots = \mu_n \text{ for some } k \in \{1, \ldots, n-1\}\,.$$

classical change point tests (e.g. CUSUM) affected by

- long range dependence
- heavy tailed distributions

Robust Test Statistic

based on Wilcoxon two-sample statistic: $R_i = \text{rank}(X_i) = \sum_{j=1}^{n} 1_{\{X_j \le X_i\}}$

$$W_n(k) = \sum_{i=1}^{k} R_i - \frac{k}{n} \sum_{i=1}^{n} R_i$$

Theorem (Dehling, Rooch, Taqqu, 2013)

under **H** *and if* $mD < 1$

$$n^{-2+(mD)/2} L_\gamma^{-m/2}(n) \max_{k=1,\ldots,n} |W_n(k)| \Rightarrow C_m C_G \sup_{0 \le \lambda \le 1} \left| Z_{m,H}(\lambda) - \lambda Z_{m,H}(1) \right|$$

- Hermite rank: $m \in \mathbb{N}$
- Hurst coefficient: $H = 1 - \frac{mD}{2}$
- Hermite process: $Z_{m,H}$

Robust Test Statistic

based on Wilcoxon two-sample statistic: $R_i = \text{rank}(X_i) = \sum_{j=1}^{n} 1_{\{X_j \leq X_i\}}$

$$W_n(k) = \sum_{i=1}^{k} R_i - \frac{k}{n} \sum_{i=1}^{n} R_i$$

Theorem (Dehling, Rooch, Taqqu, 2013)

under **H** *and if* $mD < 1$

$$n^{-2+(mD)/2} L_\gamma^{-m/2}(n) \max_{k=1,\ldots,n} |W_n(k)| \Rightarrow C_m C_G \sup_{0 \leq \lambda \leq 1} \left| Z_{m,H}(\lambda) - \lambda Z_{m,H}(1) \right|$$

- Hermite rank: $m \in \mathbb{N}$ unknown!
- Hurst coefficient: $H = 1 - \frac{mD}{2}$ unknown!
- Hermite process: $Z_{m,H}$ L_γ, G unknown!

Self-Normalized Test Statistic I

$$G_n(k) = \frac{\sum_{i=1}^{k} R_i - \frac{k}{n}\sum_{i=1}^{n} R_i}{\left\{\frac{1}{n}\sum_{t=1}^{k} S_t^2(1,k) + \frac{1}{n}\sum_{t=k+1}^{n} S_t^2(k+1,n)\right\}^{1/2}}$$

with $S_t(j,k) = \sum_{h=j}^{t}\left(R_h - \bar{R}_{j,k}\right)$ with $\bar{R}_{j,k} = \frac{1}{k-j+1}\sum_{t=j}^{k} R_t$

Theorem (Betken, 2014)

$$\max_{k\in\{\lfloor n\tau_1\rfloor,\ldots,\lfloor n\tau_2\rfloor\}} \left|G_n(k)\right|$$

$$\Rightarrow \sup_{\tau_1\le\lambda\le\tau_2} \frac{\left|Z_{m,H}(\lambda)-\lambda Z_{m,H}(1)\right|}{\left\{\int_0^{\lambda}(Z_{m,H}(r)-\frac{r}{\lambda}Z_{m,H}(\lambda))^2 dr+\int_0^{1-\lambda}(Z^{\star}_{m,H}(r)-\frac{r}{1-\lambda}Z^{\star}_{m,H}(1-\lambda))^2 dr\right\}^{1/2}}$$

with $Z^{\star}_{m,H}(r) = Z_{m,H}(1) - Z_{m,H}(1-r)$

Self-Normalized Test Statistic II

similar, but not robust statistic: Shao (2011)

no scaling needed, but shape dependent on unknown parameters

- L_γ has not to be known
- m, H still have to be estimated

use resampling methods?

Subsampling

general assumptions:

- $(X_n)_{n\in\mathbb{N}}$ stationary process
- $T_n = T_n(X_1, \ldots, X_n)$ series of statistics
- $T_n \Rightarrow T$ for some random variable T

subsampling estimator:

$$\hat{F}_{l,n}(t) = \frac{1}{n-l+1} \sum_{j=1}^{n-l+1} 1_{\{T_l(X_i,\ldots,X_{i+l-1})\leq t\}}$$

with block length $l \to \infty$, $l/n \to 0$

estimator for $F_{T_n}(t) = P\left(T_n \leq t\right)$

Consistency under Short Range Dependence

Theorem (Politis, Romano, 1994)

If $(X_n)_{n\in\mathbb{N}}$ is strongly mixing, t is a point of continuity of F_T, then

$$\left(\hat{F}_{l,n}(t) - F_{T_n}(t)\right) \xrightarrow{\mathcal{P}} 0$$

sketch of proof:

- $\operatorname{Var}(\hat{F}_{l,n}(t)) \to 0$ because of strong mixing property
- $F_{T_n}(t) \to F_T(t)$ because $T_n \Rightarrow T$
- $\operatorname{E} \hat{F}_{l,n}(t) = F_{T_l}(t) \to F_T(t)$ because $l \Rightarrow \infty$

Consistency under Long Range Dependence

- Hall, Jing, Lahiri (1998): restrictive class of subordinated Gaussian processes, general statistics T_n
- Nordman, Lahiri (2005): linear LRD processes, sample mean
- Jach, McElroy, Politis (2012): subordinated Gaussian processes, G and T_n Lipschitz-continuous
- Beran, Feng, Ghosh, Kulik (2013): Gaussian processes, sample mean
- Zhang, Ho, Wendler, Wu (2013): subordinated linear LRD processes, sample mean

Assumptions I

assumption on statistic:
$T_n = T_n(X_1, \ldots, X_n) \Rightarrow T$

- standard assumption for subsampling

assumption on block length:
$l = l_n \to \infty$ and $l_n = \mathcal{O}\left(n^{(1+D)/2-\epsilon}\right)$ for an $\epsilon > 0$

- range for l depends on unknown parameter D
- but $D > 0$: $l \approx C\sqrt{n}$ always possible

Assumptions II

assumptions on process:

$X_n = G(\xi_n)$ for stationary, Gaussian process $(\xi_n)_{n\in\mathbb{N}}$ with

$$\gamma(k) := \mathrm{Cov}(\xi_1, \xi_{1+k}) = k^{-D} L_\gamma(k)$$

such that following conditions hold:

1. $D \in (0, 1]$ and L_γ slowly varying with

$$\max_{\tilde{k}\in\{k,\ldots,k+2l+1\}} \left| L_\gamma(k) - L_\gamma(\tilde{k}) \right| \leq K \frac{l}{k} \quad \text{for some } K < \infty$$

2. spectral density $f(x) = |x|^{D-1} L_f(x)$ of $(\xi_n)_{n\in\mathbb{N}}$, slowly varying function L_f bounded away from 0

- assumptions hold for fractional Gaussian noise
- assumptions hold for Gaussian FARIMA

Subsampling Consistency

subsampling estimator:

$$\hat{F}_{l,n}(t) = \frac{1}{n-l+1} \sum_{j=1}^{n-l+1} 1_{\{T_l(X_i,\ldots,X_{i+l-1}) \leq t\}}$$

Theorem (Betken, Wendler, 2015)

Under assumptions above for all points of continuity t of F_T, we have

$$\hat{F}_{l,n}(t) - F_{T_n}(t) \xrightarrow{\mathcal{P}} 0.$$

If F_T is continuous, then

$$\sup_{t \in \mathbb{R}} \left| \hat{F}_{l,n}(t) - F_{T_n}(t) \right| \xrightarrow{\mathcal{P}} 0.$$

A Useful Lemma

ρ-mixing coefficient:

$$\rho(\mathcal{A}, \mathcal{B}) := \sup \operatorname{corr}(X, Y)$$

supremum taken over all $\mathcal{A}$-measurable X and all $\mathcal{B}$-measurable Y

Lemma

under our assumptions

$$\begin{aligned}\rho(k,l) &:= \rho\left(\sigma(\xi_i, 1 \le i \le l), \sigma(\xi_j, k+l+1 \le j \le k+2l)\right) \\ &\le C_1\,(k/l)^{-D} L_\gamma(k) + C_2 l^2 k^{-D-1} \max\{L_\gamma(k), 1\}.\end{aligned}$$

Proof of the Lemma

Kolmogorov, Rozanov (1960): $\exists\ a_1, a_2, \ldots, a_l, b_1, b_2, \ldots, b_l$:

$$\rho\big(\sigma(\xi_i, 1 \leq i \leq l), \sigma(\xi_j, k+l+1 \leq j \leq k+2l)\big)$$
$$= \mathrm{Cov}\left(\sum_{i=1}^{l} a_i\xi_i, \sum_{j=1}^{l} b_j\xi_{k+l+j}\right) = \sum_{i=1}^{l} a_i \sum_{j=1}^{l} b_j\gamma(k+l-i+j)$$

and $\mathrm{Var}(\sum_{i=1}^{l} a_i\xi_i) = \mathrm{Var}(\sum_{j=1}^{l} b_j\xi_{k+l+j}) = 1$
restrictions on the coefficients:

$$\sum_{i=1}^{l} a_i^2 \leq K_d, \qquad \sum_{j=1}^{l} b_j^2 \leq K_d,$$

$$\Big|\sum_{i=1}^{l} a_i\Big| \leq K_d' l^{D/2}, \qquad \Big|\sum_{j=1}^{l} b_j\Big| \leq K_d' l^{D/2}$$

Proof of Theorem

$$\mathrm{Var}(\hat{F}_{l,n}(t)) = \mathrm{Var}\left(\frac{1}{n-l+1}\sum_{j=1}^{n-l+1} 1_{\left\{T_l(X_i,\ldots,X_{i+l-1})\leq t\right\}}\right)$$
$$\leq \frac{4l}{n-l+1} + \frac{2}{n-l+1}\sum_{k=l}^{n-2l-1}\rho(k,l) \xrightarrow{n\to\infty} 0$$

Furthermore:

$$\mathrm{E}\left(\hat{F}_{l,n}(t)\right) - F_{T_n}(t) = F_{T_l}(t) - F_{T_n}(t)$$
$$= \left(F_{T_l}(t) - F_T(t)\right) - \left(F_{T_n}(t) - F_T(t)\right) \xrightarrow{n\to\infty} 0$$

because $T_n \Rightarrow T$, $T_l \Rightarrow T$

Simulations I

self-normalized, robust change point test statistic:

$$\sup_{k \in \{\lfloor n\tau_1 \rfloor, \ldots, \lfloor n\tau_2 \rfloor\}} |G_n(k)|$$

$$\text{with } G_n(k) = \frac{\sum_{i=1}^{k} R_i - \frac{k}{n} \sum_{i=1}^{n} R_i}{\left\{\frac{1}{n} \sum_{t=1}^{k} S_t^2(1,k) + \frac{1}{n} \sum_{t=k+1}^{n} S_t^2(k+1,n)\right\}^{1/2}}$$

Random variables

- $(\xi_n)_{n \in \mathbb{N}}$ fractional Gaussian noise
- $H \in \{0.6, 0.7, 0.8, 0.9\}$
- $G(x) = x$ (normal margins) or $G(x) = \frac{2}{\sqrt{3}}(\phi^{-1/3}(x) - \frac{3}{2})$ (Pareto(3, 1) margins)

critical values dependent on: Hermite rank m, Hermite coefficient H

Simulations II

first method: estimate H by local Whittle estimator, assume $m = 1$

second method: use subsampling

scenarios:

- sample size $n = 300$
- time of change $[n\tau]$ with $\tau = 0.25, 0.5$
- height of change $h = 0, 0.5, 1$

nominal level: $\alpha = 0.05$
5000 simulation runs per scenario
block length $l = [\sqrt{n}] = 17$
$\tau_1 = 1 - \tau_2 = 0.15$

Simulations III

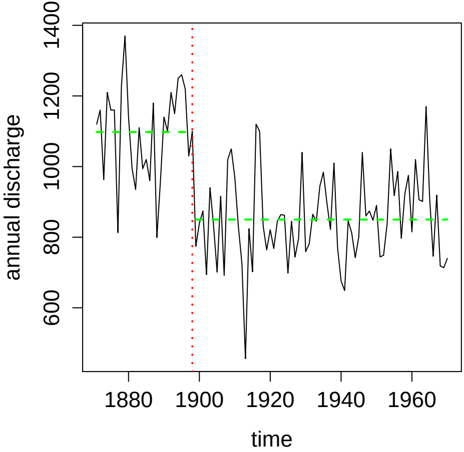
normal margins: whittle estimator/ subsampling

		$\tau = 0.25$		$\tau = 0.5$	
H	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$
0.6	0.044/0.064	0.21/0.31	0.52/0.74	0.42/0.57	0.86/0.96
0.7	0.053/0.070	0.11/0.17	0.27/0.42	0.23/0.31	0.61/0.76
0.8	0.048/0.067	0.08/0.12	0.14/0.23	0.14/0.21	0.36/0.49
0.9	0.057/0.074	0.07/0.10	0.10/0.16	0.12/0.17	0.31/0.40

Pareto margins:

		$\tau = 0.25$		$\tau = 0.5$	
H	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$
0.6	0.056/0.067	0.82/0.87	0.91/0.95	0.98/0.99	1.00/1.00
0.7	0.070/0.064	0.53/0.53	0.70/0.74	0.86/0.88	0.98/0.99
0.8	0.072/0.068	0.30/0.28	0.43/0.45	0.64/0.67	0.88/0.91
0.9	0.073/0.071	0.17/0.17	0.24/0.25	0.50/0.53	0.74/0.77

Data Example: Nile River Discharges



- 1871-1970
- for whole data set: $\hat{H} = 0.962$
- test with subsampling: change in 1899
- before, after change: $\hat{H} = 0.517$, $\hat{H} = 0.5$
- same findings: e.g. Cobb (1978), Balke (1993), Shao (2011)

Conclusion

Results:

- subsampling consistent under mild regularity assumptions on process
- for self-normalized robust change point test: similar results as with parameter estimation

Open Questions:

- choice of block length?
- long range dependent linear processes?

Thank you
for your attention!

References

- A. Kolmogorov, Y. A. Rozanov (1960). On strong mixing conditions for stationary Gaussian processes. *PTRF* 5, 204-208.
- G.W. Cobb (1978). The problem of the Nile: conditional solution to a changepoint problem. *Biometrika* 65, 243-251.
- N.S. Balke (1993). Detecting Level Shifts in Time Series. J. Bus. Econom. Statist. 11, 81-92.
- D.N. Politis, J.P. Romano (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* 4, 2031-2050.
- P. Hall, B.-Y. Jing, S.N. Lahiri (1998). On the sampling window method for long-range dependent data. *Statist. Sinica* 8, 1189-1204.
- D.J. Nordman, S.N. Lahiri (2005). Validity of the sampling window method for long-range dependent linear processes. *Econometric Theory* 21, 1087-1111.

- X. Shao (2011). A simple test of changes in mean in the possible presence of long-range dependence. *J. Time Ser. Anal.* 32, 598-606.
- A. Jach, T. McElroy, D.N. Politis (2012). Subsampling inference for the mean of heavy-tailed long-memory time series. *J. Time Ser. Anal.* 33, 96-111.
- J. Beran, Y. Feng, S. Ghosh, R. Kulik (2013). *Long-Memory Processes*. Springer.
- H. Dehling, A. Rooch, M.S. Taqqu (2013). Non-Parametric Change-Point Tests for Long-Range Dependent Data. *Scand. J. Statist.* 40, 153-173.
- T. Zhang, H.-C. Ho, M. Wendler, W.B. Wu (2013). Block sampling under strong dependence. *SPA* 123, 2323-2339.
- A. Betken (2014). Testing for change-points in long-range dependent time series by means of a self-normalized Wilcoxon test. *preprint* arXiv:1403.0265.
- A. Betken, M. Wendler (2015+). Subsampling for general statistics under long range dependence. *work in progress*.