Robust Change-Point Tests for Time Series

Herold Dehling

RUHR-UNIVERSITÄT BOCHUM

Joint work with Roland Fried (Dortmund), Aeneas Rooch (Bochum), Martin Wendler (Bochum), and Murad S. Taqqu (Boston)

Research supported by DFG-Sonderforschungsbereich 823
Statistik nichtlinearer dynamischer Prozesse

Recent Advances and Trends in Time Series Analysis
BIRS Workshop, Banff Centre, Alberta, April 27 – May 2, 2014

Outline

1. Change-Point Test for Time Series - Some Examples
    - CUSUM Test
    - Wilcoxon Change-Point Test
    - Hodges-Lehmann Change-Point Test
2. Three Relevant Processes
    - Two-Sample U-Statistics Process
    - Two-Sample Empirical U-Process
    - Two-Sample Empirical U-Quantiles
3. Dependent Data: SRD and LRD
4. Two-Sample U-Statistic Process
    - Short-Range Dependent Data
    - Long-Range Dependent Data
5. Two-Sample Empirical U-Process: SRD Data
6. Two-Sample Empirical U-Quantile Process: SRD Data
    - Application to Hodges-Lehmann Change-Point Test
7. Conclusions

Change-Point Tests for Time Series

Observations are generated by a stochastic process $(X_i)_{i \geq 1}$,

$$X_i = \mu_i + \epsilon_i,$$

- $(\mu_i)_{i \geq 1}$ are unknown signals,
- $(\epsilon_i)_{i \geq 1}$ is a stationary noise process, satisfying $E(\epsilon_i) = 0$.

Based on observations $X_1, \ldots, X_n$, we wish to test the hypothesis

$$H : \mu_1 = \ldots = \mu_n$$

against the alternative

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

i.e. that there is a level shift somewhere in the process.

Inspiration: Two-Sample Problem/Tests

When the change point k is known, we have a two-sample problem with the samples $X_1, \ldots, X_k$, and $X_{k+1}, \ldots, X_n$. The corresponding two-sample tests provide inspiration for change-point tests; e.g.

- Gauß test: $\frac{1}{k}\sum_{i=1}^{k} X_i - \frac{1}{n-k}\sum_{i=k+1}^{n} X_i$
- Wilcoxon test: $\sum_{i=1}^{k}\sum_{j=k+1}^{n} 1_{\{X_i \leq X_j\}}$
- Hodges-Lehmann test: $\mathrm{med}\{X_j - X_i : 1 \leq i \leq k, k+1 \leq j \leq n\}$

When the change-point is unknown [this is our assumption], we use the supremum of these statistics, taken over all $k \in \{1, \ldots, n-1\}$ and properly weighted, as change-point test statistic.

Example: CUSUM Test

The CUSUM test is the most popular change-point test; it arises from the two-sample Gauß test:

$$\frac{1}{k}\sum_{i=1}^{k} X_i - \frac{1}{n-k}\sum_{i=k+1}^{n} X_i = \frac{n}{k(n-k)}\sum_{i=1}^{k}(X_i - \bar{X})$$

Observe that, under the null hypothesis of no change,

$$\sum_{i=1}^{k}(X_i - \bar{X}) = \sum_{i=1}^{k}(X_i - \mu) - \frac{k}{n}\sum_{i=1}^{n}(X_i - \mu).$$

Thus, the asymptotic distribution of $\max_{1\le k\le n}\sum_{i=1}^{k}(X_i - \bar{X})$ can be derived from a functional limit theorem for the partial sum process

$$\sum_{i=1}^{[n\lambda]}(X_i - \mu),\ 0 \le \lambda \le 1.$$

Wilcoxon Change-Point Test – Two-Sample U-Process

Deriving the asymptotic distribution of the Wilcoxon change point test, requires the study of the process $\sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} 1_{\{X_i \le X_j\}}$, or more generally, of the two-sample U-processes

$$\sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} h(X_i, X_j), \; 0 \le \lambda \le 1,$$

with kernel $h : \mathbb{R}^2 \to \mathbb{R}$. The Wilcoxon, and the CUSUM test statistic are obtained when

$$\begin{aligned} h(x, y) &= 1_{\{x \le y\}} \\ h(x, y) &= y - x. \end{aligned}$$

Hodges-Lehmann Test – Two-Sample U-Quantiles

Deriving the asymptotic distribution of the Hodges-Lehmann test, requires the study of the process

$$\mathrm{med}\{X_j - X_i : 1 \le i \le [n\lambda] < j \le n\},\ 0 \le \lambda \le 1.$$

This leads to the two-sample empirical U-process

$$U_{[n\lambda],n-[n\lambda]}(x) = \frac{1}{[n\lambda](n-[n\lambda])} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} 1_{\{h(X_i,X_j)\le x\}},\ 0 \le \lambda \le 1,$$

and the two-sample empirical U-quantile process

$$Q_{[n\lambda],n-[n\lambda]}(p) = \inf\{x : U_{[n\lambda],n-[n\lambda]}(x) \ge p\},\ 0 \le \lambda \le 1.$$

Summary: Three Processes for Change-Point Tests

Given a kernel $h : \mathbb{R}^2 \to \mathbb{R}$, we investigate the processes

- Two-Sample U-Process

$$\sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} h(X_i, X_j),\ 0 \le \lambda \le 1$$

- Two-Sample Empirical U-Process

$$U_{[n\lambda],n-[n\lambda]}(x) = \frac{1}{[n\lambda](n-[n\lambda])} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} 1_{\{h(X_i,X_j)\le x\}},\ 0 \le \lambda \le 1$$

- Two-Sample Empirical U-Quantile Process

$$Q_{[n\lambda],n-[n\lambda]}(p) = \inf\{x : U_{[n\lambda],n-[n\lambda]}(x) \ge p\},\ 0 \le \lambda \le 1.$$

Asymptotic Theory/Simulations: Outlook

We develop asymptotic theory for the three processes introduced above, in the following cases:

1. Dependence
   - Short range dependent (SRD) noise
   - Long range dependent (LRD) noise
2. Hypothesis/Alternative
   - Under the hypothesis of stationarity
   - Under local alternatives
3. Assumptions on Kernels
   - General kernels
   - Specific examples, such as $h(x, y) = 1_{\{x \le y\}}$.

Simulations show that our tests have better power than the CUSUM test when the data are heavy-tailed, while they are still quite efficient when the data are Gaussian.

Dependence I: Short Range Dependent (SRD) Data

We assume that the noise process $(\epsilon_i)_{i\geq 1}$ is a functional of an absolutely regular (β-mixing) process:

$$\epsilon_i = f(Z_i, Z_{i-1}, \ldots),$$

where

1. $(Z_i)_{i\in\mathbb{Z}}$ is an absolutely regular process with mixing coefficients β_k
2. $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ is a 1-approximating functional, i.e.

$$E|\epsilon_i - f_m(Z_i, Z_{i-1}, \ldots, Z_{i-m})| \leq a_m,$$

   for some $f_m : \mathbb{R}^{m+1} \to \mathbb{R}$, and $a_m \to 0$, as $m \to \infty$.

Examples:

- ARMA processes, linear processes with summable coefficients
- Dynamical systems, e.g. expanding maps of the unit interval

Dependence II: Long Range Dependent (LRD) Data

We assume that $(\epsilon_i)_{i\geq 1}$ is a Gaussian subordinated process, i.e. that

$$\epsilon_i = H(\xi_i),$$

where

1. $(\xi_i)_{i\geq 1}$ is a stationary Gaussian process with $\xi_i \sim N(0,1)$, and with non-summable autocovariance function

   $$r(k) = \mathrm{Cov}(\xi_i, \xi_{i+k}) = k^{-D} L(k),$$

   where $0 < D < 1$ and L is a slowly varying function.
2. $H : \mathbb{R} \to \mathbb{R}$ is a measurable function.

Two-Sample U-Statistics Process: SRD Data

We now consider the two-sample U-statistics process,

$$U_{[n\lambda],n-[n\lambda]} = \frac{1}{[n\lambda](n-[n\lambda])} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} h(X_i, X_j),\ 0 \leq \lambda \leq 1.$$

The analysis of this process uses the Hoeffding decomposition

$$h(x,y) = \theta + h_1(x) + h_2(y) + \psi(x,y),$$

where

$$\begin{aligned} \theta &= Eh(X,Y) \\ h_1(x) &= Eh(x,Y) - \theta \\ h_2(y) &= Eh(X,y) - \theta \\ \psi(x,y) &= h(x,y) - h_1(x) - h_2(y) - \theta \end{aligned}$$

and where X, Y are independent with the same distribution as X_1

Two-Sample U-Statistic Functional CLT: SRD Data

Theorem (D., Fried, Garcia, Wendler 2013)

Let $(X_i)_{i\geq 1}$ be a functional of an absolutely regular process. Then, under some technical conditions, the two-sample U-statistic process

$$\sqrt{n}\left(\lambda(1-\lambda)(U_{[n\lambda],n-[n\lambda]}-\theta)\right)_{0\leq\lambda\leq 1}$$

converges in distribution to $((1-\lambda)W_1(\lambda)+\lambda(W_2(1)-W_2(\lambda))_{0\leq\lambda\leq 1}$*, where* $(W_1(\lambda), W_2(\lambda))$ *denotes two-dimensional Brownian motion with covariance function*

$$E(W_i(\lambda)\,W_j(\mu)) = (\lambda\wedge\mu)\sum_{k\in\mathbb{Z}}\mathrm{Cov}(h_i(X_1), h_j(X_k)).$$

- Csörgő and Horváth (1988) proved this for IID data.

Idea of Proof

Using Hoeffding decomposition $h(x, y) = \theta + h_1(x) + h_2(y) + \psi(x, y)$, we obtain

$$\sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} (h(X_i, X_j) - \theta)$$

$$= (n - [n\lambda]) \sum_{i=1}^{[n\lambda]} h_1(X_i) + [n\lambda] \sum_{j=[n\lambda]+1}^{n} h_2(X_j) + \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} \psi(X_i, X_j)$$

- The first two terms can be treated using the functional CLT for partial sums of vectors $\sum_{i=1}^{[n\lambda]} (h_1(X_i), h_2(X_i))$.
- Need to show that $\sup_{0 \le \lambda \le 1} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} \psi(X_i, X_j) = o_P(n^{3/2})$. Note that the kernel $\psi(x, y)$ is degenerate, i.e. that $E\psi(x, Y) = E\psi(X, y) = 0$.

Asymptotic Distribution of Wilcoxon Test: LRD Data

Theorem (D., Rooch, Taqqu (2013))

Let $(X_i)_{i\geq 1}$ *be an LRD Gaussian subordinated process,* $X_k = G(\xi_k)$*, and assume that* X_k *has a continuous distribution function* F*. Define the Hermite coefficients*

$$J_q(x) = E\left(1_{\{G(\xi)\leq x\}} H_q(\xi)\right),$$

the Hermite rank of the class of functions $1_{\{G(\cdot)\leq x\}}$

$$m = \min\{q \geq 1 : J_q(x) \neq 0 \text{ for some } x\},$$

and assume that $0 < D < \frac{1}{m}$*. Then, in the space* $D[0,1]$*,*

$$\frac{1}{n\, d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} (1_{\{X_i \leq X_j\}} - \frac{1}{2}) \xrightarrow{\mathcal{D}} \frac{\int J_m(x) dF(x)}{m!} \left(Z_m(\lambda) - \lambda Z_m(1)\right).$$

Idea of Proof

- Observe that

$$\sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} 1_{\{X_i \le X_j\}} = [n\lambda] \sum_{j=[n\lambda]+1}^{n} F_{[n\lambda]}(X_j)$$
$$= [n\lambda](n - [n\lambda]) \int F_{[n\lambda]}(x) dF_{[n\lambda]+1,n}(x),$$

  where F_k denotes the empirical distribution function of the first k observations, and where $F_{k+1,n}$ denotes the empirical distribution function of the observations $X_{k+1}, \ldots, X_n$.
- Now apply the empirical process non-CLT for LRD Gaussian subordinated data of D., Taqqu (1989), and use integration by parts.

General Two Sample U-Statistics Process: LRD Data

D., Rooch, Wendler (2014) study the U-statistics process

$$\sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} h(X_i, X_j),\ 0 \leq \lambda \leq 1,$$

for general kernels $h(x, y)$, using two different techniques:

- Representation as a functional of the empirical process of $X_1, \ldots, X_n$; then use integration by parts and empirical process non-CLT of D., Taqqu (1989)
- Hermite expansion of the kernel $h(x, y)$ into a series $\sum_{k,l} a_{kl} H_k(x) H_l(y)$; then prove a reduction principle, and finally use the multivariate non-CLT of Bai, Taqqu (2014).

Both approaches require different conditions on the kernels $h(x, y)$; together they cover most kernels arising in applications.

Wilcoxon Process Under Local Alternatives: LRD Data

We consider the process $W_n(\lambda) = \frac{1}{n\, d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} (1_{\{X_i \leq X_j\}} - \frac{1}{2})$ under local alternatives, where $X_i = \mu_i + G(\xi_i)$, and

$$A_\tau(n): \;\; \mu_i = \begin{cases} \mu & \text{for } i = 1, \ldots, [n\tau] \\ \mu + \frac{d_n}{n} c & \text{for } i = [n\tau]+1, \ldots, n. \end{cases}$$

Theorem (D., Rooch, Taqqu (2013+))

Let $(\xi_i)_{i \geq 1}$ be stationary Gaussian LRD, $G : \mathbb{R} \to \mathbb{R}$ measurable, $F(x) = P(G(\xi_1) \leq x)$ with bounded density $f(x)$. Let m be the Hermite rank of the class $1_{\{G(\xi_i) \leq x\}} - F(x)$, $x \in \mathbb{R}$, and assume $0 < D < \frac{1}{m}$. Then, under $A_\tau(n)$, the process $(W_n(\lambda))_{0 \leq \lambda \leq 1}$ converges weakly to

$$\left(\frac{\int_{\mathbb{R}} J_m(x) dF(x)}{m!} (Z_m(\lambda) - \lambda Z_m(1)) + c\phi_\tau(\lambda) \int_{\mathbb{R}} f^2(x) dx \right)_{0 \leq \lambda \leq 1},$$

where $\phi_\tau(\lambda) = \begin{cases} \lambda(1-\tau) & \text{for } \lambda \leq \tau \\ (1-\lambda)\tau & \text{for } \lambda \geq \tau. \end{cases}$

ARE of CUSUM and Wilcoxon Test: LRD Data

A similar limit theorem holds for the CUSUM process under local alternatives. Thus, we can calculate the asymptotic relative efficiency (ARE) of the CUSUM and the Wilcoxon change-point tests,

$$ARE(C, W) = \left(\frac{a_m \int_{\mathbb{R}} f^2(x)dx}{\int_{\mathbb{R}} J_m(x)dF(x)} \right)^{2/(mD)}$$

For Gaussian data, we obtain $ARE(C, W) = 1$; note that for i.i.d. Gaussian noise, we have $ARE(C, W) = \frac{\pi}{3}$.

Two Sample Empirical U-Process: SRD Data

$$U_{[n\lambda],n-[n\lambda]}(t) = \frac{1}{[n\lambda](n-[n\lambda])} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} 1_{\{g(X_i,X_j)\leq t\}}$$

Theorem (D., Fried, Wendler 2014+)

Let $(X_i)_{i\geq 1}$ be a functional of an absolutely regular process. Then, under some technical conditions,

$$\left(\sqrt{n}\lambda(1-\lambda)(U_{[n\lambda],n-[n\lambda]}(t) - U(t))\right)_{0\leq\lambda\leq 1} \xrightarrow{\mathcal{D}} (W(\lambda))_{0\leq\lambda\leq 1},$$

where $W(\lambda) = (1-\lambda)W_1(\lambda) + \lambda(W_2(1) - W_2(\lambda)$*, and where* $(W_1(\lambda), W_2(\lambda))$ *is 2-dimensional Brownian motion with covariance*

$$E(W_i(\lambda)\, W_j(\mu)) = (\lambda \wedge \mu) \sum_{k\in\mathbb{Z}} \mathrm{Cov}(h_i(X_1; t), h_j(X_k; t)).$$

Here, $U(t) = P(g(X,Y) \leq t)$*, and* $h_1(x;t)$*,* $h_2(y;t)$ *denote the terms in the Hoeffding decomposition of* $h(x,y,t) = 1_{\{g(x,y)\leq t\}}$.

Two Sample Empirical U-Quantile Process: SRD Data

$$Q_{[n\lambda],n-[n\lambda]}(p) = \inf\{t : U_{[n\lambda],n-[n\lambda]}(t) \geq p\},$$

Theorem (D., Fried, Wendler 2014+)

Let $(X_i)_{i\geq 1}$ be a functional of an absolutely regular process. Then, under some technical conditions,

$$\sqrt{n}\left(\lambda(1-\lambda)\left(Q_{[n\lambda],n-[n\lambda]}(p) - Q(p)\right)\right)_{0\leq\lambda\leq 1} \xrightarrow{\mathcal{D}} (W(\lambda))_{0\leq\lambda\leq 1},$$

where $W(\lambda) = (1-\lambda)W_1(\lambda) + \lambda(W_2(1) - W_2(\lambda)$, and where $(W_1(\lambda), W_2(\lambda))$ is 2-dimensional Brownian motion with covariance

$$\mathrm{Cov}(W_i(\mu), W_j(\lambda)) = (\mu\wedge\lambda)\frac{1}{u^2(Q(p))}\sum_{k\in\mathbb{Z}} E(h_i(X_0; Q(p)), h_j(X_k; Q(p))).$$

Here $u(t) = \frac{d}{dt}U(t)$, and $Q(p) = U^{-1}(p) = \inf\{t : P(g(X,Y) \geq p)\}$.

Hodges-Lehmann Change-Point Test: SRD Data

Corollary (D., Fried, Wendler 2014+)

Let $(X_i)_{i\geq 1}$ be a functional of an absolutely regular process. Then

$$\frac{\hat{u}(0)}{\hat{\sigma}}\sqrt{n} \max_{1\leq k\leq n-1} \frac{k}{n}(1-\frac{k}{n})|\,\mathrm{med}\{(X_j - X_i) : 1 \leq i \leq k, k+1 \leq j \leq n\}|$$

converges in distribution to the Kolmogorov-Smirnov distribution.

- The density $u(t)$ of $Y - X$ can be estimated by a kernel density estimator, using the data $X_i - X_j$, $1 \leq i < j \leq n$.
- The asymptotic variance $\sigma^2 = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(F(X_0), F(X_k))$ can be estimated by a variety of techniques, e.g. by subsampling of $F_n(X_i), 1 \leq i \leq n$, using the edf F_n as approximation to F.

Conclusion

1. The analysis of some robust change-point tests for time series requires asymptotic theory of three processes
   - Two-sample U-statistics processes
   - Two-sample empirical U-process
   - Two-sample empirical quantile process
2. We develop asymptotic theory for these processes for
   - Short range dependent data
   - Long range dependent data,

   both under the null hypothesis as well as under local alternatives.
3. The asymptotic results are confirmed by simulations.
4. The Wilcoxon test has better power than the CUSUM test when the data are heavy tailed.
5. For Gaussian data, the efficiency loss is small. In the LRD case, ARE(CUSUM, Wilcoxon)=1.

References

- H.Dehling, A.Rooch, M.S.Taqqu (2013): Nonparametric Change-Point Tests for Long-Range Dependent Data. *Scandinavian Journal of Statistics* **40**, 153–173.
- H.Dehling, A.Rooch, M.S.Taqqu (2013): Power of Change-Point Tests for Long-Range Dependent Data. *arXiv*
- H.Dehling, A.Rooch, M.Wendler (2014): Two-Sample U-Statistic Processes for Long-Range Dependent Data. *arXiv*
- A.Betken (2014): Testing for Change-Points in Long-Range Dependent Time Series by Means of a Self-Normalized Wilcoxon Test, *arXiv*.
- H.Dehling, R.Fried (2012): Asymptotic Distribution of Two-Sample Empirical U-Quantiles with Application to Robust Tests for Shifts in Location. *Journal of Multivariate Analysis* **105**, 124–140.
- H.Dehling, R.Fried, I.Garcia, M.Wendler (2013): Change-Point Detection under Dependence Based on Two-Sample U-Statistics. *arXiv*
- H.Dehling, R.Fried, M.Wendler (2014): A Robust Method for Shift Detection in Time Series. *Work in Progress.*