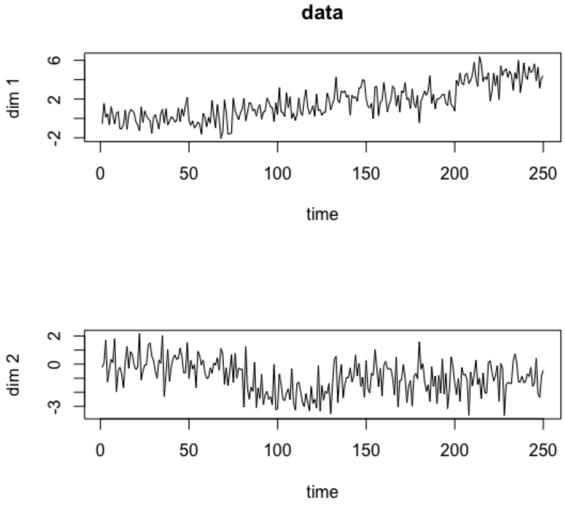
MOSUM procedure for multivariate mean-changes

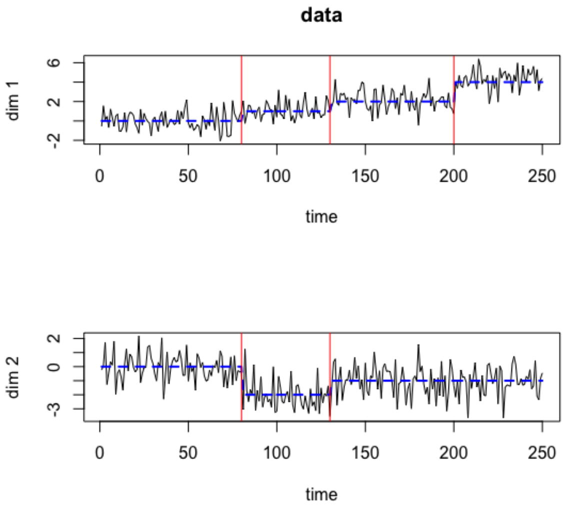
Niclas Stoffregen

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG
MATH

29. April 2023

Example


data
dim 1
6
2
-2
0
50
100
150
200
250
time
dim 2
2
0
-3
0
50
100
150
200
250
time



Two-dimensional data with changes in $k_1 = 80$, $k_2 = 130$, $k_3 = 200$.

A posteriori multiple change point detection problem in p-dimensional data

$$\mathbf{X}_t = \sum_{j=1}^{q+1} \boldsymbol{\mu}^{(j)} 1_{\{k_{j-1}<t\leq k_j\}} + \boldsymbol{\varepsilon}_t,$$

- $\boldsymbol{\mu}^{(j)} \in \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_P\}$ with $\boldsymbol{\mu}^{(j)} \neq \boldsymbol{\mu}^{(j+1)}$.
- number of structural breaks $q = q_n$, the change points $0 = k_0 < k_1 < \dots < k_q < k_{q+1} = n$ are unknown.
- centered stationary error sequence $(\boldsymbol{\varepsilon}_t)$ fulfills different assumptions, easiest case $\boldsymbol{\varepsilon}_t \overset{iid.}{\sim} (0, \Sigma)$.
- $\mathbf{d}_j := \boldsymbol{\mu}^{(j+1)} - \boldsymbol{\mu}^{(j)}$ denotes magnitude of the j-th mean change.

Agenda

MOSUM procedure

MOSUM (= MOving SUM) procedure aims to detect changes in the mean

Formally, the test statistic is given by

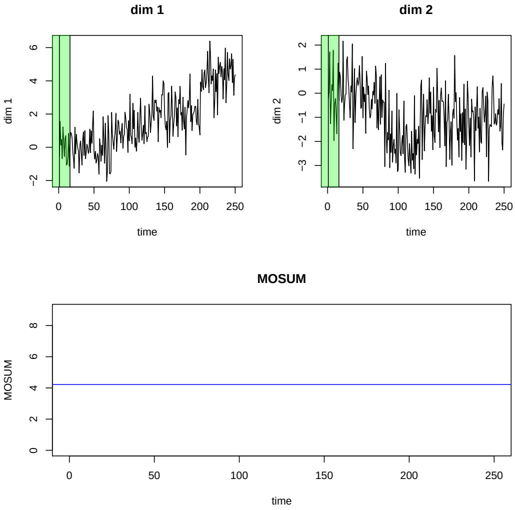
$$\|\hat{\Sigma}_k^{-1/2}\mathbf{T}_k(G)\|,$$

where $\hat{\Sigma}_k$ denotes an estimator for the long-run covariance which can depend on time and

$$\mathbf{T}_k(G) = \frac{1}{\sqrt{2G}}\left(\sum_{i=k+1}^{k+G}\mathbf{X}_i - \sum_{i=k-G+1}^{k}\mathbf{X}_i\right).$$

Idea: choose bandwidth G and for given point in time k, compare empirical means of $X_{k-G+1},\dots,X_k$ and $X_{k+1},\dots,X_{k+G}$, great difference indicates a change

Example



Open questions:

- **How to determine threshold indicated by the blue line?**
- How to determine actual estimator as threshold is exceeded in intervals not single time points and noise can cause false positives?
- How to choose covariance estimator and bandwidth to detect all changes?

To derive threshold use asymptotic result under the no-change situation. The following assumptions must hold for the error distribution and the bandwidth:

Assumption 1

Let the following assumptions hold:

1. Let the error sequence (ε_t) fulfill strict stationarity.
2. There exists a p-dimensional standard Wiener process $(\mathbf{W}(k))_{1\leq k\leq n}$ and $\nu > 0$ such that (possibly after changing the probability space)
$$\left\|\Sigma^{-\frac{1}{2}}\sum_{i=1}^{k}\varepsilon_i - \mathbf{W}(k)\right\| = O\left(k^{\frac{1}{2+\nu}}\right) \text{ a.s.,}$$
where Σ denotes the positive definite long-run covariance matrix.

Remark:

- Property (2) can be derived for several error sequences, a prominent example are specific strongly mixing sequences (see [1], Regularity Conditions 4.1)

Assumption 2

Let the bandwidth G depend on n, $G = G(n)$. For $n \to \infty$ and $\nu > 0$ (same as in Assumption 1), assume that

$$\frac{n}{G} \to \infty,$$

$$\frac{n^{\frac{2}{2+\nu}} \log(n)}{G} \to 0.$$

Let $T_n(G) = \max_{G \leq k \leq n-G} \left\| \Sigma^{-1/2} \mathbf{T}_k(G) \right\|$, then the following result holds:

Theorem 1

Let Assumption 1 hold for the error sequence and Assumption 2 for the bandwidth G. Then in the no-change situation,

$$a(n/G)\, T_n(G) - b(n/G) \xrightarrow{\mathcal{D}} E$$

with E as Gumbel distributed random variable, i.e. $P(E \leq y) = \exp(-2\exp(-x))$ and with

$$a(x) = \sqrt{2\log(x)},$$

$$b(x) = 2\log(x) + \frac{p}{2}\log\log(x) - \log\left(\frac{2}{3}\Gamma\left(\frac{p}{2}\right)\right),$$

where Γ denotes the gamma function.

Theorem 1 is stated for the true long-run covariance, but that result does not change when using an estimator fulfilling Assumption 3.

Assumption 3

The estimator $\hat{\Sigma}_{k,n}$ of the long-run covariance matrix Σ can depend on k and satisfies

$$\max_{G \leq k \leq n-G} \left\| \hat{\Sigma}_{k,n}^{-1/2} - \Sigma^{-1/2} \right\| = o_P\left(\left(\log(n/G) \right)^{-1} \right)$$

in the situation that there do not exist any change points.

Open questions:

- How to determine threshold indicated by the blue line? ($\checkmark$)
- **How to determine actual estimator as threshold is exceeded in intervals not single time points and noise can cause false positives?**
- How to choose covariance estimator and bandwidth to detect all changes?

Observe that the MOSUM statistic decomposes into signal and noise term analogously to the underlying model:

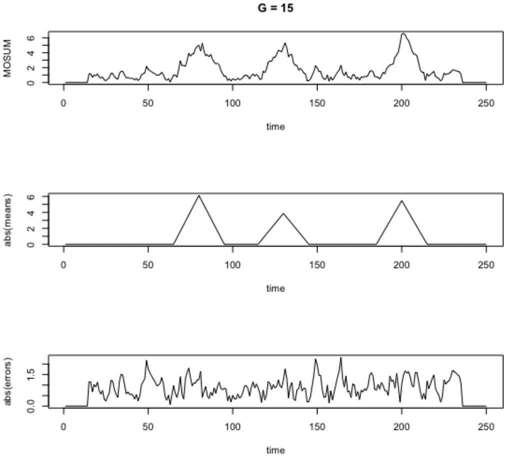
Lemma 1

The MOSUM statistic $\mathbf{T}_k(G) = \mathbf{m}_k + \boldsymbol{\Lambda}_k$ decomposes into a piecewise linear signal term $\mathbf{m}_k = \mathbf{m}(k, G)$ and a centered noise term $\boldsymbol{\Lambda}_k$ with

$$\sqrt{2G}\,\mathbf{m}_k = \begin{cases} (G - k + k_j)\mathbf{d}_j, & \text{for } k_j < k \leq k_j + G, \\ 0, & \text{for } k_j + G < k \leq k_{j+1} - G \\ (G + k - k_{j+1})\mathbf{d}_{j+1}, & \text{for } k_{j+1} - G < k \leq k_{j+1}. \end{cases}$$

and

$$\sqrt{2G}\,\boldsymbol{\Lambda}_k = \sum_{i=k+1}^{k+G} \varepsilon_i - \sum_{i=k-G+1}^{k} \varepsilon_i.$$



MOSUM statistic decomposition for $G = 15$ and true covariance.

Observation:

- signal has peaks at true change points
- local extremum of MOSUM statistic inside a specific environment is a good choice for estimator

A point $\hat{k}$ is a local extremum if it is the leftmost point (in the case of a tie) to maximise the absolute MOSUM statistic within its own ηG-environment, i.e.

$$\hat{k} = \arg \max_{\hat{k}-\eta G \leq k \leq \hat{k}+\eta G} \|\mathbf{T}_k(G)\|.$$

Open questions:

- How to determine threshold indicated by the blue line?($\checkmark$)
- How to determine actual estimator as threshold is exceeded in intervals not single time points and noise can cause false positives?($\checkmark$)
- **How to choose covariance estimator and bandwidth to detect all changes?**

Last question concerns consistency, has to be combined with assumptions on significance levels.

Observation

- threshold must be large enough to prevent false positives caused by noise
- threshold must not grow too fast to be able to detect small changes

Assumption 4

Let the sequence of significance levels fulfill

$$\alpha_n \to 0,$$

$$\frac{\log\log \frac{1}{\sqrt{1-\alpha_n}}}{a(n/G)} = O(1).$$

Assumption 5

Let the bandwidth fulfill

$$G \leq \min(k_1, n - k_q),$$
$$2G < \min_{0 \leq j \leq q} |k_{j+1} - k_j|,$$

and

$$\frac{\|\mathbf{d}_j\|^2 G}{\log(n/G)} \to \infty$$

for all $j = 1, \ldots, q$.

Observation

- for asymptotic consistency a well-behaving covariance estimator is needed in the case of using threshold from before

Assumption 6

The symmetric and positive definite estimator $\hat{\Sigma}_{k,n}$ of the long-run covariance matrix Σ can depend on k and satisfies

$$\max_{G \leq k \leq n-G} \left\| \hat{\Sigma}_{k,n}^{-1/2} - \Sigma^{-1/2} \right\| = o_P\left(\left(\log(n/G) \right)^{-1} \right)$$

in the no-change situation and satisfies

$$\max_{G \leq k \leq n-G} \left\| \hat{\Sigma}_{k,n} \right\| = O_P(1)$$

under the alternative.

Theorem 2

Let the above assumptions hold and let $\hat{k}_1 < \ldots < \hat{k}_{\hat{q}}$ be the change point estimators. Then for any $\tau > 0$, it holds

$$\lim_{n\to\infty} P\left(\max_{1\leq j\leq \min(\hat{q},q)} |\hat{k}_j - k_j| < \tau G, \hat{q} = q \right) = 1$$

Remark:

- often too restrictive to assume a covariance estimator to fulfill Assumption 6
- instead use 'weaker' assumption on covariance and adapt threshold to obtain results of Theorem 2

Let the long-run covariance estimator fulfill

$$\max_{G \leq k \leq n-G} \left\| \hat{\Sigma}_k^{-1} \right\| = O_P(1), \qquad \max_{j=1,\ldots,q} \max_{|k-k_j|<G} \left\| \hat{\Sigma}_k \right\| = O_P(1)$$

and let the threshold $\tilde{D}_n$ fulfill

$$\frac{\tilde{D}_n}{\sqrt{\log(n/G)}} \to \infty,$$
$$\frac{\tilde{D}_n}{\sqrt{G}\min_{i=1,\ldots,q} \|\mathbf{d}_i\|} \to 0$$

to reach consistency as in Theorem 2.

For better covariance estimator and assumptions on significance levels, the assumptions on signal strength are weaker since

$$D_n(G,\alpha_n) = O\left(\sqrt{\log(n/G)}\right)$$

and

$$\frac{\min_{j=1,\ldots,q} \|\mathbf{d}_j\|^2 G}{\log(n/G)} \to \infty.$$

In contrast, for weak assumptions on covariance estimator, the signal strength must grow faster than $\log(n/G)$ as

$$\frac{\tilde{D}_n}{\sqrt{\log(n/G)}} \to \infty,$$

$$\frac{\tilde{D}_n}{\sqrt{G}\min_{i=1,\ldots,q} \|\mathbf{d}_i\|} \to 0$$

- Many assumptions on bandwidth and signal strength
- In the case that all of them hold, consistency can be reached
- Question: What can we do if assumptions are violated for a fixed bandwidth G?
- Answer: Multiscale MOSUM-procedure
- Idea: Use multiple bandwidth to detect changes and prune down the set of possible candidates by using an information criterion

C. Kirch and K. Reckruehm.
Data segmentation for time series based on a general moving sum approach.
arXiv preprint arXiv:2207.07396, 2022.