A Simple Way to Improve Some Online Change Point Detectors in Time Series

STAT556 Final Project

Man Fung (Heman) Leung

Spring, 2022

University of Illinois Urbana-Champaign

Outline

ILLINOIS

This project is based on a new example in a manuscript that I am revising. Our major reference in the review is Gösmann et al. (2021) because they mentioned the possibility of our example in their outlook; see also Dette and Gösmann (2020).

1. Basics of change point (CP) detection in time series
2. A simple way to improve some existing detectors
3. Simulations

Types of problem

ILLINOIS

Retrospective (offline, not our focus)

- Data are completely available before the analysis
- Single CP vs multiple CPs

Sequential (online)

- Data arrive consecutively and any new point can be a CP
- Closed-end vs open-end
- Major part of the 20th century: control charts
  - optimized for a minimal detection delay
  - but usually do not control the false alarm rate (type I error)

Comment

There are papers that minimized the detection delay subject to a false alarm rate control in the 20th century. However, we could not pinpoint the origin from our major reference.

Setup

ILLINOIS

The following setup was originally introduced in Chu et al. (1996), who use initial data sets and therefrom employ invariance principles to control the type I error.

- $\{X_t\}_{t\in\mathbb{Z}}$: a d-dimensional time series
- F_t: the distribution function of X_t
- $\theta_t = \theta(F_t)$: a p-dimensional parameter of F_t
- m: initial sample size (stable observations)
- Target: a decision rule for

$$H_0 : \theta_1 = \cdots = \theta_m = \theta_{m+1} = \theta_{m+2} = \cdots,$$

against

$$H_1 : \exists k^* \in \mathbb{Z}^+ : \theta_1 = \cdots = \theta_{m+k^*-1} \neq \theta_{m+k^*} = \theta_{m+k^*+1} = \cdots.$$

Common detectors are usually comparing estimators from different subsamples of the data. Let $\hat{\theta}_a^b$ be an estimator of θ based on the subsample $X_a, \ldots, X_b$.

- (Ordinary) CUSUM investigates

$$\hat{\theta}_1^m - \hat{\theta}_{m+1}^{m+k}.$$

- Page-CUSUM uses a function of

$$\{\hat{\theta}_1^m - \hat{\theta}_{m+j+1}^{m+k}\}_{j=0,\ldots,k-1}.$$

- Gösmann et al. (2021) propose using a function of

$$\{\hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k}\}_{j=0,\ldots,k-1}.$$

Comment on the detectors

Comment

When the change point is far away from X_{m+1}, $\hat{\theta}_{m+1}^{m+k}$ maybe 'corrupted' by pre-change observations. Gösmann et al. (2021) point out that Page-CUSUM and their proposal are able handle this problem. However, we can see that this comes at the cost of computational complexity.

A specific detector

ILLINOIS

Gösmann et al. (2021) propose

$$\hat{E}_m(k) = m^{-1/2} \max_{j=0}^{k-1} (k-j) \sqrt{\left(\hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k}\right)^\top \hat{\Sigma}_m^{-1} \left(\hat{\theta}_1^{m+j} - \hat{\theta}_{m+j+1}^{m+k}\right)}$$

Several more elements are needed:

- $\hat{\Sigma}_m$: a long-run variance (LRV) estimator;
- $w(\cdot)$: a threshold function; and
- $c(\alpha)$: a critical value such that the test is level α (as $m \to \infty$).

A CP is detected if $w(k/m)\hat{E}_m(k) > c(\alpha)$.

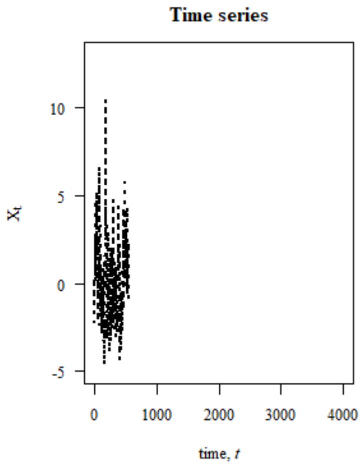
Comment on $\hat{\Sigma}_m$

Although the LRV has not been formally introduced in our course, it is not a completely new concept; see the asymptotic variance in the central limit theorem (CLT) for m-dependent sequence, or the spectral density.

Comment on $w(\cdot)$ (and $c(\alpha)$)

Restrictions on $w(\cdot)$ are needed so that $\sup_{k=1}^{\infty} w(k/m)\hat{E}_m(k)$ converges to some limiting distribution. Gösmann et al. (2021) consider a family of functions such that the limiting distribution can be simulated more easily. We guess some $w(\cdot)$ may lead to a smaller detection delay but searching one seems difficult.

The setting of this example ($\delta = 3$ is used) will be introduced in our simulations.



For presentation purpose, suppose $d = 1$ and $\theta = \mathbb{E}(X)$.

CLT for the mean of stationary sequence

Consider a stationary sequence $\{X_t\}_{t\in\mathbb{Z}}$ with mean μ and autocovariance function γ_k. Under some suitable conditions (e.g., Hannan (1979)),

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma^2),$$

where $\sigma^2 = \sum_{k\in\mathbb{Z}} \gamma_k$ is called the LRV.

If the parameter of interest is not the mean, we can use the influence function (assuming its existence) to define the corresponding LRV.

Some LRV estimators

ILLINOIS

Consistent (small-b)

- Subsampling estimator, e.g., overlapping batch means (Meketon and Schmeiser, 1984):

$$\hat{\sigma}^2_{n,\mathrm{obm}} = \frac{\sum_{i=\ell_n}^{n} \left(\sum_{j=i-\ell_n+1}^{i} X_j - \ell_n \bar{X}_n\right)^2}{\sum_{i=\ell_n}^{n} \ell_n}.$$

- Kernel estimator, e.g., Bartlett kernel estimator (Newey and West, 1987):

$$\hat{\sigma}^2_{n,\mathrm{bart}} = \sum_{k=-\ell_n}^{\ell_n} \left(1 - \frac{|k|}{\ell_n}\right) \frac{1}{n} \sum_{i=|k|+1}^{n} (X_i - \bar{X}_n)(X_{i-|k|} - \bar{X}_n).$$

Inconsistent (fixed-b)

- Self-normalizer (Shao, 2010)

Difficulties with LRV estimation

ILLINOIS

In online CP detection, a LRV estimator often does not perform well because

- the estimator is not robust to CP; or
- the estimator cannot be updated quickly.

Some existing solutions

- Gösmann et al. (2021) (use $\hat{\sigma}_m^2$): inefficient
- Dehling et al. (2020) (split the time series into 3): less efficient
- Shao and Zhang (2010) (account for the CP in SN): slower
- Chan and Yau (2017) (update $\hat{\sigma}_n^2$ recursively): not robust

Comment

Gösmann et al. (2021) describe the use of $\hat{\sigma}_m^2$ as the standard approach. Indeed, several papers that we have read do the same.

A simple improvement

Difference statistics: add robustness to $\hat{\sigma}_n^2$ through

$$D_i = \sum_{j=0}^{g} d_j X_{i-jh}$$

- $\{d_j\}$: differencing sequence; g: differencing order; h: lag
- Chan (2021) proves the optimality in an offline setting

Window decomposition: allow recursive update of $\hat{\sigma}_n^2$ through

$$W_n(i,j) = T\left(d_n^T(i,j)\right) S\left(d_n^S(i,j)\right).$$

- $T(\cdot)$: tapering function; $S(\cdot)$: subsampling function
- $d_n^T(i,j)$ and $d_n^S(i,j)$: distances between times i and j when the sample size is n
- compatible with difference statistics

Consider a Bilinear model

$$X_i = (0.9 + 0.1\varepsilon_i)X_{i-1} + \varepsilon_i, \qquad \text{where} \qquad \varepsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0,1).$$

Let $\mu_i = \mathbb{E}(X_i)$. We are interested in testing

$$H_0 : \mu_1 = \cdots = \mu_m = \mu_{m+1} = \cdots,$$

against

$$H_1 : \mu_1 = \cdots = \mu_{m+k^*-1} \neq \mu_{m+k^*} = \mu_{m+k^*+1} = \cdots.$$

With 100 burn-in and $m = 400$ initial observations, we simulate 1000 replications of H_1 by

$$X_t^{(\delta)} = X_t + \delta \mathbb{1}_{t \geq m+k^*}.$$

The online CP detector in Gösmann et al. (2021) can be written as

$$\hat{E}_m(k) = m^{-1/2} \max_{j=0}^{k-1} (k-j) \left| \bar{X}_1^{m+j} - \bar{X}_{m+j+1}^{m+k} \right| / \hat{\sigma},$$

where $\bar{X}_a^b = (b-a+1)^{-1} \sum_{i=a}^{b} X_i$ and $\hat{\sigma}^2$ is a LRV estimator.

Following examples in Gösmann et al. (2021), we use

- threshold function: $w(t) = (1+t)^{-1}$;
- nominal size: $\alpha = 0.05$; and
- stopping point: $n^* = 4000$.

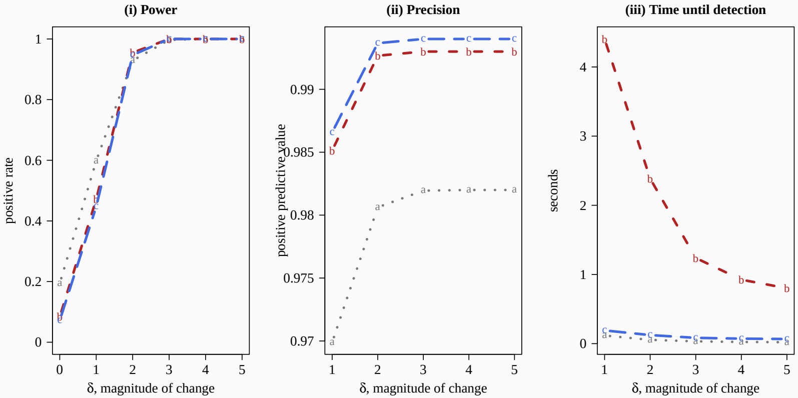


Figure 1: Online CP detection at 5% nominal size using different LRV estimation methods: (a) *fix* (dotted gray); (b) *offline* (dashed red); (c) *online* (longdash blue).

Results when $m + k^* = 1001$

ILLINOIS

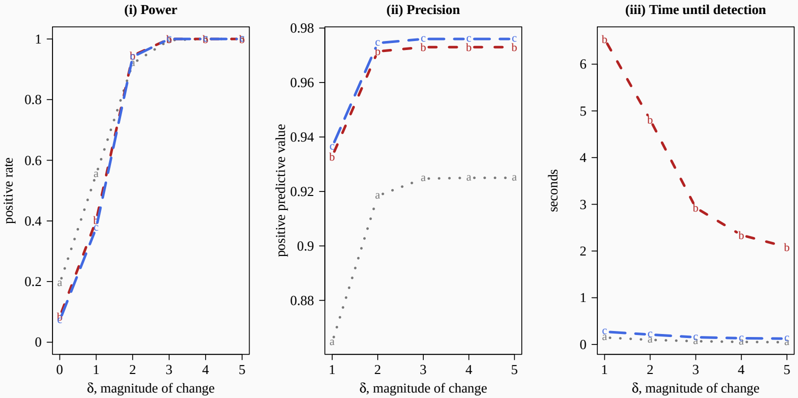


Figure 2: Online CP detection at 5% nominal size using different LRV estimation methods: (a) *fix* (dotted gray); (b) *offline* (dashed red); (c) *online* (longdash blue).

Results when $m + k^* = 1401$

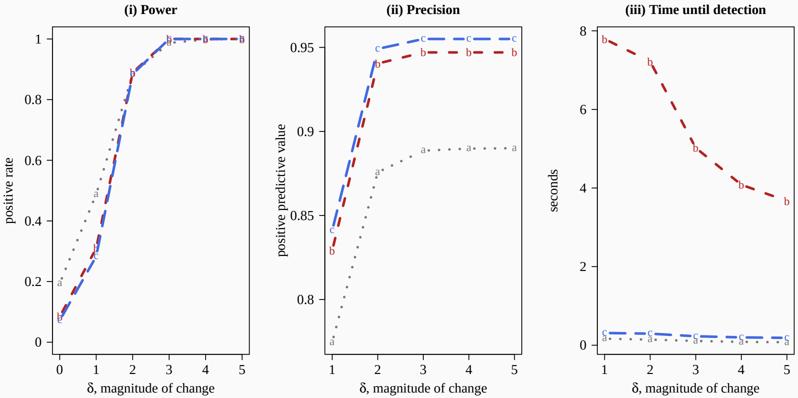


Figure 3: Online CP detection at 5% nominal size using different LRV estimation methods: (a) *fix* (dotted gray); (b) *offline* (dashed red); (c) *online* (longdash blue).

Chan, K. W. (2021), 'Optimal difference-based variance estimators in time series: A general framework', *To appear in The Annals of Statistics* (To appear).

Chan, K. W. and Yau, C. Y. (2017), 'Automatic optimal batch size selection for recursive estimators of time-average covariance matrix', *Journal of the American Statistical Association* **112**(519), 1076–1089.

Chu, C.-S. J., Stinchcombe, M. and White, H. (1996), 'Monitoring structural change', *Econometrica* **64**(5), 1045–1065.

Dehling, H., Fried, R. and Wendler, M. (2020), 'A robust method for shift detection in time series', *Biometrika* **107**(3), 647–660.

Dette, H. and Gösmann, J. (2020), 'A likelihood ratio approach to sequential change point detection for a general class of parameters', *Journal of the American Statistical Association* **115**(531), 1361–1377.

Gösmann, J., Kley, T. and Dette, H. (2021), 'A new approach for open-end sequential change point monitoring', *Journal of Time Series Analysis* **42**(1), 63–84.

Hannan, E. J. (1979), 'The central limit theorem for time series regression', *Stochastic Processes and their Applications* **9**(3), 281–289.

Meketon, M. S. and Schmeiser, B. (1984), Overlapping batch means: something for nothing?, *in* 'Proceedings of the Winter Simulation Conference', IEEE, pp. 226–230.

Newey, W. K. and West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**, 703–708.

Shao, X. (2010), 'A self-normalized approach to confidence interval construction in time series', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**(3), 343–366.

Shao, X. and Zhang, X. (2010), 'Testing for change points in time series', *Journal of the American Statistical Association* **105**(491), 1228–1240.