

Modeling of gradual epidemic changes

Alfredas Račkauskas, Aurelijus Tamulis

Vilnius University, Department of Mathematics and Informatics

Naugarduko 24, LT-03225 Vilnius

E-mail: alfredas.rackauskas@mif.vu.lt, aurelijus.tamulis@gmail.com

Abstract. The article is devoted to analysis of epidemic changes, when transition between regimes is gradual. The consistency of CUSUM, Uniform Increments (UI) and Dyadic Increments (DI) statistics is shown. The comparison of the size-adjusted power of the tests is presented graphically.

Keywords: change point, gradual change, epidemic alternative, power analysis.

1 Introduction

Many articles analyze change with epidemic alternative, when switching between regimes is instant. Our purpose is to expand the theory in cases, where regime switching is gradual or of another form. We consider the following model under H_0 :

$$X_i = \mu + \epsilon_i, \quad \forall i \in [1; n]$$

whereas under H_1 :

$$X_i = \begin{cases} \mu + \epsilon_i, & 1 \leq i \leq k^*, \\ a_i + \epsilon_i, & k^* + 1 \leq i \leq k^* + l^*, \\ \mu + \epsilon_i, & k^* + l^* + 1 \leq i \leq n \end{cases} \quad (1)$$

where ϵ_i , $i \geq 1$, are i.i.d. random variables with $E\epsilon_i = 0$, $E\epsilon_i^2 < \infty$. The paper is organized as follows. In Section 2, consistency of CUSUM, UI and DI statistics is shown. In Section 3, graphical representation of power analysis is provided.

2 Consistency

We start with consistency of classical CUSUM statistics

$$T_n = \frac{1}{n^{1/2}} \max_{1 \leq k \leq n} \left| \sum_{i=1}^k X_i - \frac{k}{n} \sum_{i=1}^n X_i \right|. \quad (2)$$

Convergence of T_n under H_0 is proved by W. Ploberger and W. Krämer [2]. We consider conditions for divergence of CUSUM statistics under H_1 .

By $\xrightarrow[n \rightarrow \infty]{P}$ we denote convergence in probability.

Theorem 1. Assume that for model (1) the following condition is satisfied:

$$\max \left\{ k^*/n, 1 - (k^* + \ell^*)/n \right\} \frac{\ell^*}{n^{1/2}} \left| \frac{1}{l^*} \sum_{i=k^*+1}^{k^*+l^*} a_i - \mu \right| \rightarrow \infty$$

as $n \rightarrow \infty$. Then

$$T_n \xrightarrow[n \rightarrow \infty]{P} \infty.$$

Proof. If $k^* > n - k^* - \ell^*$, we have

$$\begin{aligned} T_n &\geq \frac{1}{n^{1/2}} \left| \sum_{i=1}^{k^*} X_i - \frac{k^*}{n} \sum_{i=1}^n X_I \right| = \frac{1}{n^{1/2}} \left| \sum_{i=1}^{k^*} \varepsilon_i - \frac{k^*}{n} \sum_{i=1}^n \varepsilon_i + \frac{k^*}{n} l^* \mu - \frac{k^*}{n} \sum_{i=k^*+1}^{k^*+l^*} a_i \right| \\ &\geq \frac{k^* \ell^*}{n^{3/2}} \left| \mu - \frac{1}{\ell^*} \sum_{i=k^*+1}^{k^*+\ell^*} a_i \right| - V_n, \end{aligned}$$

where

$$V_n = \frac{1}{n^{1/2}} \max_{1 \leq k \leq n} \left| \sum_{i=1}^k \varepsilon_i - \frac{k}{n} \sum_{i=1}^n \varepsilon_i \right|.$$

By the classical Donsker invariance principle $V_n = O_P(1)$. If $k^* \leq n - k^* - \ell^*$, we have

$$T_n \geq \frac{1}{n^{1/2}} \left| \sum_{i=1}^{k^*+\ell^*} X_i - \frac{k^*+\ell^*}{n} \sum_{i=1}^n X_I \right| \geq \frac{(n - (k^* + \ell^*)) \ell^*}{n^{3/2}} \left| \frac{1}{l^*} \sum_{i=k^*+1}^{k^*+l^*} a_i - \mu \right| - V_n.$$

Finally, we complete the proof by using the assumption of Theorem.

Next we analyze uniform increments and dyadic increments statistics introduced by A. Račkauskas and Ch. Suquet [3, 4] and [5]. For $0 \leq \alpha < 1/2$, the uniform increments statistic $UI(n, \alpha)$ is defined by

$$UI(n, \alpha) = \max_{1 \leq \ell < n} \ell^{-\alpha} (1 - \ell/n)^{-\alpha} \max_{0 \leq k \leq n-\ell} \left| \sum_{i=k+1}^{k+\ell} X_i - \frac{\ell}{n} \sum_{i=1}^n X_i \right|.$$

To define dyadic increments statistics let $D_j = \{(2\ell - 1)2^{-j}, 1 \leq \ell \leq 2^{j-1}\}$ be the dyadic numbers of level j , $j = 1, 2, \dots$. Put for $r \in D_j$, $j \geq 1$, $r^+ = r + 2^{-j}$, $r^- = r - 2^{-j}$. Then the dyadic increments statistic $DI(n, \alpha)$ is defined by

$$DI(n, \alpha) = \max_{1 \leq j \leq \log n} 2^{-\alpha j} \max_{r \in D_j} \left| \sum_{i=nr^-}^{nr} X_i - \sum_{i=nr}^{nr^+} X_i \right|.$$

Convergence of both uniform increments and dyadic increments statistics under H_0 is shown in [4]. Here we consider conditions for consistency of these statistics under H_1 .

Theorem 2. Let $0 \leq \alpha < 1/2$. For the model (1) assume that

$$\lim_{t \rightarrow \infty} t^{1/(1/2-\alpha)} P(|\varepsilon_1| > t) = 0. \tag{3}$$

If

$$\frac{\ell^{*(1-\alpha)}(1 - \ell^*/n)^{1-\alpha}}{n^{1/2-\alpha}} \left| \frac{1}{\ell^*} \sum_{i=k^*+1}^{k^*+\ell^*} a_i - \mu \right| \rightarrow \infty \tag{4}$$

as $n \rightarrow \infty$, then under H_1

$$n^{-1/2+\alpha} UI(n, \rho) \xrightarrow[n \rightarrow \infty]{P} \infty \tag{5}$$

as $n \rightarrow \infty$.

Proof. We have

$$UI(n, \alpha) \geq [\ell^*(1 - \ell^*/n)]^{1-\alpha} \left| \mu - \frac{1}{\ell^*} \sum_{i=k^*+1}^{k^*+\ell^*} a_i \right| - R_n,$$

where

$$R_n = \max_{1 \leq \ell < n} \ell^{-\alpha} (1 - \ell/n)^{-\alpha} \max_{0 \leq k \leq n-\ell} \left| \sum_{i=k+1}^{k+\ell} \varepsilon_i - \frac{\ell}{n} \sum_{i=1}^n \varepsilon_i \right|.$$

Under condition (4) we have $R_n = O_P(n^{1/2-\alpha})$ (see, e.g., [4]), hence we complete the proof of (5).

Theorem 3. Let $0 \leq \alpha < 1/2$. For the model (1) assume that

$$\lim_{t \rightarrow \infty} t^{1/(1/2-\alpha)} P(|\varepsilon_1| > t) = 0. \tag{6}$$

and

$$\lim_{n \rightarrow \infty} \ell^*/n = 0. \tag{7}$$

If

$$\lim_{n \rightarrow \infty} 2^{-\alpha} n^{-1/2+\alpha} \ell^{*1-\alpha} \left| \frac{1}{\ell^*} \sum_{i=k^*+1}^{k^*+\ell^*} a_i - \mu \right| = \infty \tag{8}$$

is satisfied then

$$n^{-1/2} DI(n, \rho) \xrightarrow[n \rightarrow \infty]{P} \infty \tag{9}$$

as $n \rightarrow \infty$.

Proof. Define

$$a_{nk} = \begin{cases} a_k, & \text{if } k = k^* + 1, \dots, k^* + \ell^*, \\ \mu, & \text{if } k = 1, \dots, k^*, k^* + \ell^* + 1, \dots, n. \end{cases}$$

We have

$$DI(n, \alpha) = \max_{1 \leq j \leq \log n} 2^{\alpha j} \max_{r \in D_j} \left| \sum_{i=nr^-}^{nr} (\varepsilon_i + a_{ni}) - \sum_{i=nr}^{nr^+} (\varepsilon_i + a_{ni}) \right|.$$

By triangle inequality it holds

$$DI(n, \alpha) \geq DI^{(1)}(n, \alpha) - DI^{(2)}(n, \alpha),$$

where

$$DI^{(1)}(n, \alpha) = \max_{1 \leq j \leq \log n} 2^{\alpha j} \max_{r \in D_j} \left| \sum_{i=nr^-}^{nr} a_{ni} - \sum_{i=nr}^{nr^+} a_{ni} \right|$$

and

$$DI^{(2)}(n, \alpha) = \max_{1 \leq j \leq \log n} 2^{\alpha j} \max_{r \in D_j} \left| \sum_{i=nr^-}^{nr} \varepsilon_i - \sum_{i=nr}^{nr^+} \varepsilon_i \right|.$$

Under condition (7) it holds that $DI^{(2)}(n, \alpha) = O_P(n^{1/2})$ (see, e.g., [4]). So it remains to check, that

$$\lim_{n \rightarrow \infty} n^{-1/2} DI^{(1)}(n, \alpha) = \infty$$

follows from (8).

Under the following configuration $r^-n \leq k^* < k^* + \ell^* \leq rn$, where $r \in D_j$ and j is such that $2^{-j-1} < \ell^*/n \leq 2^{-j}$ we obtain

$$\begin{aligned} n^{-1/2} DI^{(1)}(n, \alpha) &\geq n^{-1/2} 2^{\alpha j} \left| \sum_{i=nr^-}^{nr} a_{ni} - \sum_{i=nr}^{nr^+} a_{ni} \right| \\ &= n^{-1/2} 2^{\alpha j} \left| \sum_{i=k^*+1}^{k^*+\ell^*} a_i - \ell^* \mu + (2nr - nr^- - nr^+) \mu \right| \\ &= n^{-1/2} 2^{\alpha j} \ell^* \left| \frac{1}{\ell^*} \sum_{i=k^*+1}^{k^*+\ell^*} a_i - \mu \right| \\ &\geq 2^{-\alpha} n^{-1/2+\alpha} \ell^{*1-\alpha} \left| \frac{1}{\ell^*} \sum_{i=k^*+1}^{k^*+\ell^*} a_i - \mu \right|, \end{aligned}$$

since for configuration under investigation it holds $2^j \geq n/(2\ell^*)$. Under condition (8), we get the result (9).

The similar result holds for the configuration $rn \leq k^* < k^* + \ell^* \leq r^+n$:

$$n^{-1/2} DI^{(1)}(n, \alpha) \geq 2^{-\alpha} n^{-1/2+\alpha} \ell^{*1-\alpha} \left| \frac{1}{\ell^*} \sum_{i=k^*+1}^{k^*+\ell^*} a_i - \mu \right|,$$

Other possible configurations of $k^*, k^* + \ell^*$ with respect to dyadic number, say, r .

1. $r^-n \leq k^* \leq rn \leq k^* + \ell^* \leq r^+n$;
2. $k^* \leq r^-n \leq k^* + \ell^* \leq rn$;
3. $rn \leq k^* \leq r^+n \leq k^* + \ell^*$;

Due to dyadic structure the interval $[r_j^-, r_j^+]$ is a half of the interval $[r_{j-1}^-, r_{j-1}^+]$ and the configuration No. 1 becomes one of previously analysed configuration of the dyadic level $j-1$. The level of dyadic number is reduced as follows:

$$r = (2\ell - 1)2^{-j}, \quad r^- = (2\ell - 1)2^{-j} - 2^{-j} = (2\ell - 2)2^{-j} = (\ell - 1)2^{-j+1}$$

similarly

$$r^+ = (2\ell - 1)2^{-j} + 2^{-j} = (2\ell)2^{-j} = (\ell)2^{-j+1}$$

ℓ or $\ell - 1$ is odd number, we take this odd number and rewrite in such form: $2k - 1$.

$$(2k - 1)2^{-j+1} = r, \quad r \in D_{j-1}.$$

The same approach applies for the configuration No. 2. We may need apply several times the idea of the decreasing dyadic level for configurations No. 2 until we reach such j that for $r \in D_j$ is valid $r^-n \leq k^* < k^* + l^* \leq rn$. This level j is always reachable as for $j = 0$ ($r \in D_0$), $[r^-n, rn]$ or $[rn, r^+n]$ covers the whole interval. The configuration No. 3 are reduced to configuration No. 2 by changing dyadic interval of the same level.

$$r_0 = r + 2(2^{-j}) = (2\ell - 1)2^{-j} + 2(2^{-j}) = (2\ell + 1)2^{-j}.$$

The configuration No. 3 becomes the configuration No. 2 in respect of dyadic number r_0 . This completes the proof.

3 Graphical representation of power

We use graphical representation of the size-adjusted power of the tests for comparison. We explore the idea of visualization of power analysis developed by R. Davidson and J.G. MacKinnon [1]. It consists of plotting two empirical distribution functions: one empirical distribution function under H_0 another one under H_1 . These distribution functions are plotted on a $[0, 1] \times [0, 1]$ square. When the curve of one test is higher than the curve of another, it shows that the size-adjusted power of first test is higher than the power of another test.

We will analyse two computer-based modelling examples. The first example represents the gradual change that consists of three parts: increase period, stable period and decrease period. The second example represents gradual change that consists only of two parts: increase period and decrease period. The graph of the size-adjusted power of the examples is presented below in Fig. 1. Now we will describe these examples in details: We have generated random values from $N(1, 1)$. The length of total sample is 1024 (2^{10}). The length of epidemic change is 18.75% of the total sample in the first example and 12.5% in the second one. The epidemic change consists of three equal parts in the first example:

- a. gradual (45-degree line) increase of a mean from 1 to 1.35;
- b. stable period with a mean equal to 1.35;
- c. gradual (45-degree line) decrease of a mean from 1.35 to 1.

The epidemic change consists of two equal parts in the second example:

- a. gradual (45-degree line) increase of a mean from 1 to 1.5;
- b. gradual (45-degree line) decrease of a mean from 1.5 to 1.

The computer based modeling shows that the size-adjusted power of the CUSUM test is smaller than the power of the UI and the DI tests in both cases. The size-adjusted power of the UI and the DI tests is almost equal, but the DI test requires less operations and due to this performs much faster.

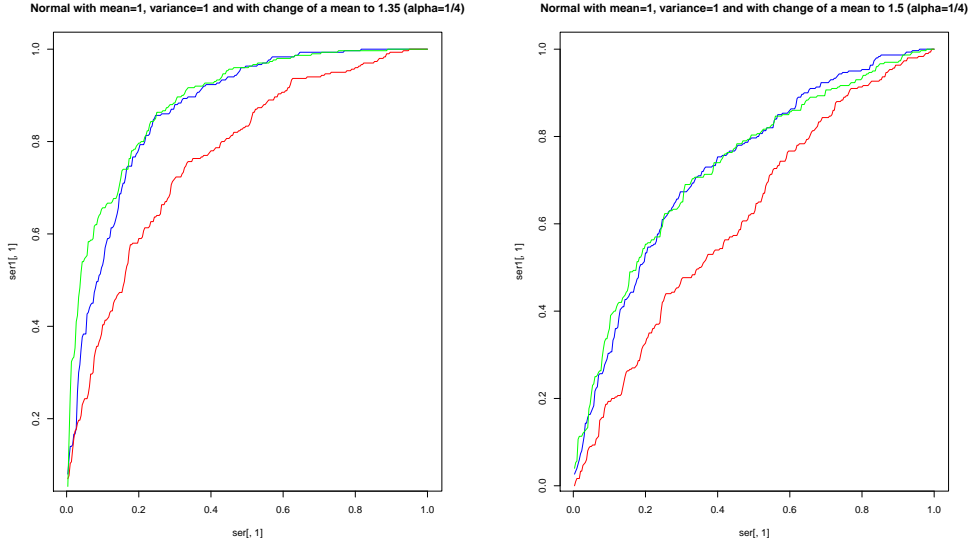


Fig. 1. Power plot with epidemic change in mean. CUSUM – red, DI – blue, UI – green line.

References

- [1] R. Davidson and J.G. MacKinnon. Graphical methods for investigating the size and power of hypothesis tests. *Manch. Sch.*, **66**(1):1–26, 1998.
- [2] W. Ploberger and W. Krämer. The cusum test with ols residuals. *Econometrica*, **60**(2):271–285, 1992.
- [3] A. Račkauskas and Ch. Suquet. Invariance principles for self-normalized partial sums processes. *Stoch. Proc. Appl.*, **95**:63–81, 2001.
- [4] A. Račkauskas and Ch. Suquet. Hölder norm test statistics for epidemic change. *J. Stat. Plan. Infer.*, **126**:495–520, 2004.
- [5] A. Račkauskas and Ch. Suquet. Testing epidemic changes of infinite dimensional parameters. *Stat. Inf. Stoch. Proc.*, **9**(1):111–134, 2006.

REZIUMĖ

Laipsniškų epideminių pasikeitimų tyrimas

A. Račkauskas, A. Tamulis

Straipsnis yra skirtas epideminių pasikeitimų analizei, kai perėjimas iš vienos būsenos į kitą yra laipsniškas. Įrodytas CUSUM, Tolygių priaugių, Diadinių priaugių testų suderintumas. Minėtų testų galia pavaizduota grafiškai.

Raktiniai žodžiai: pasikeitimo taškas, laipsniškas pasikeitimas, epideminė alternatyva, galios analizė.