

The Parameters Optimization of Filtered Derivative for Change Points Analysis

Mohamed Elmi¹

¹ University of Djibouti, Djibouti

Correspondence: Mohamed Elmi, University of Djibouti, Djibouti. E-mail: mahamed_elmi_abdillahi@univ.edu.dj

Received: February 7, 2014 Accepted: April 22, 2014 Online Published: June 11, 2014

doi:10.5539/ijsp.v3n3p29 URL: <http://dx.doi.org/10.5539/ijsp.v3n3p29>

Abstract

Let $\mathbf{X} = (X_1, X_2, \dots, X_N)$ be a time series. That is a sequence random variable indexed by the time $t = (1, 2, \dots, N)$, we suppose that the parameters of \mathbf{X} are piecewise constant. In other words, it exists a subdivision $\tau = (\tau_1 < \tau_2 < \dots < \tau_K)$ such that X_i is a family of independent and identically distributed (i.i.d) random variables for $i \in (\tau_k, \tau_{k+1}]$, and $k = 0, 1, \dots, K$ where by convention $\tau_0 = 0$ and $\tau_{K+1} = N$. The preceding works such that (Bertrand, 2000) control the probability of false alarms for minimizing the probability of type I error of change point analysis. The novelty in this work is to control the number of false alarms. We give an bound of number of false alarms and the necessary condition for number of no detection. In other hand, we know the filtered derivative (Basseville & Nikirov, 1993) depends the parameters such that the threshold and the window, we give in order to choose the optimal parameters. We compare the results of Filtered Derivative optimized parameters and the Penalized Square Error methods in particularity the adaptive method of (Lavielle & Teysnière, 2006).

Keywords: times series, filtered derivative, change points, false alarms, no-detections

1. Introduction

The problem of change detection is much studied in the literature, it exists two types of change points detection: The on-line detection or sequential points analysis and the off-line detection or change points analysis. For an updated overview, we can see the textbooks (Basseville & Nikirov, 1993; Csorgo & Horváth, 1997), or (Husková & Meintanis, 2006b; Gombay & Serban, 2009). Many applications use the change points analysis such as health, medicine and civil engineering and the sequential analysis such as fault detection, finance, surveillance and security system. Many methods exists but we often use: The penalized least square error (PLS) (Lavielle & Moulines, 2000) and the filtered derivative (FD) (Basseville & Nikirov, 1993). The calculus of PLS need a matrix of size $\mathcal{O}(n^2)$, and that FD is of order $\mathcal{O}(n)$. To improve the FD-method, two methods are developed: The filtered derivative with p-value (FDpV) (Bertrand, Fhima, & Guillin, 2011) and the filtered derivative and false discovery rate (FDqV) (Elmi, 2014). For the PLS-method, many authors are proposed (Lavielle & Moulines, 2000; Lavielle & Teysnière, 2006) the choices of penalized parameter for performing the PLS-method. For FD-method, there were no papers which mentioned this. Recall, the algorithm FD depends the window and the threshold and by consequent his performance depends the optimization of these parameters. In this work, we give the reasonable choices these parameters. We organised our paper in this way: section 1 is the introduction, section 2 describes the art of change points detection and the criteria of measure. Section 3 recall the methods of change point analysis. In the section 4, we discuss how to control the number of false alarms and numbers of no-detections. The section 5 contains numerical comparison of FD and PLS adaptive methods. Finally the appendix contains all proofs, propositions and lemmas used in this work.

2. The Art of Change Points Detection

The following subsection describes the problem of abrupt changes and different criteria used in the literature.

2.1 Problem of Change Points Detection: The Model

- $\mathbf{X} = (X_1, X_2, \dots, X_N)$ is a family of independent random variables indexed by the time.
- There exists a subdivision $\tau = (\tau_1, \dots, \tau_K)$ with $\tau_k \in \{1, \dots, N\}$ and $0 < \tau_1 < \tau_2 < \dots < \tau_K < N$.

- A configuration a K change points $\tau = (\tau_1, \dots, \tau_K)$ enlarged by convention by adding $\tau_0 = 0$ and $\tau_{K+1} = N$.
- Associated to the configuration of mean values (μ_0, \dots, μ_K) with $X_t \sim \mathcal{N}(\mu_k, 1)$, for $t \in (\tau_k, \tau_{k+1}]$ and for all $k = 0, \dots, K$.
- For notational convenience, we also define the configuration of shifts, for $k = 1, \dots, K$, $(\delta_1, \dots, \delta_K)$ where $\delta_k = \mu_{k+1} - \mu_k$.
- Let us define the minimal distance between to consecutive change points by $L_0 = \inf\{|\tau_{k+1} - \tau_k| \mid k = 0, \dots, K\}$, and
- The minimal absolute value of the shifts by $\delta_0 = \inf\{|\delta_k|, k = 1, \dots, K\}$.

Let us also recall the definition of the cumulative distribution function for standard Gaussian law

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp^{-\frac{u^2}{2}} du \quad \text{and} \quad \Psi(x) = 1 - \Phi(x). \quad (1)$$

All this paper, we use a following simulation:

2.2 Simulation

For $n=10000$, we have done one realization of a sequence of Gaussian random variable (X_1, \dots, X_n) with variance $\sigma^2 = 1$ constant and mean μ have different values. we consider seven change points at time $\tau = (2000, 2500, 3000, 4000, 7000, 8000, 9000)$ with means $\mu = (2.5, 2, 3, 4.5, 3, 3.5, 4, 5)$ and $\delta = (0.5, 1, 1.5, 1.5, 1.5, 0.5, 1)$. Below, we give a drawing for change points analysis.

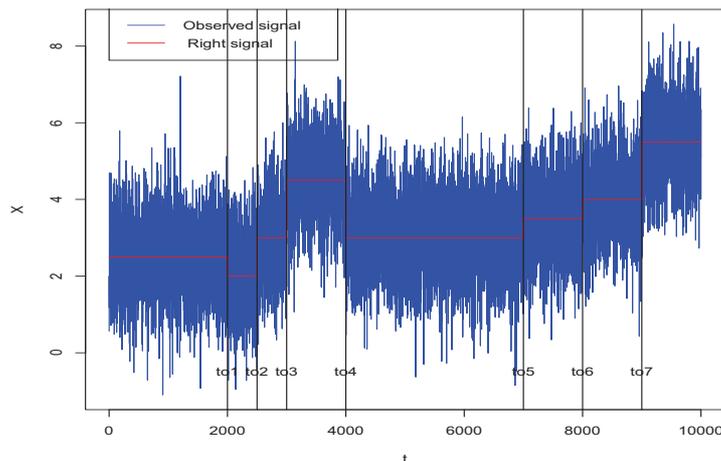


Figure 1. The observed signal and the right signal

We will use this model all this paper.

2.3 The Criteria of Measure

We suppose that the number K is unknown, and the goal of the off-line detection is to estimate the instants $\tau = (\tau_1, \dots, \tau_K)$ and the values of the mean $(\mu_0, \mu_1, \dots, \mu_K)$. we set $\hat{\tau} = (\hat{\tau}_1, \dots, \hat{\tau}_{\hat{K}})$ and $(\hat{\mu}_0, \hat{\mu}_1, \dots, \hat{\mu}_{\hat{K}})$ the corresponding estimated.

2.4 Criterion

For measuring the quality of the estimation of parameters, we use the integrated square error (ISE). So we define the ISE as:

$$ISE = \sum_{i=1}^N \left\{ \left[\hat{l}(t) - l(t) \right]^2 \right\}$$

with the signal

$$l(t) = \sum_{k=0}^K \mu_k \times \mathbf{1}_{(\tau_k, \tau_{k+1}]}(t)$$

and the estimated signal

$$\widehat{l}(t) = \sum_{k=0}^{\widehat{K}} \widehat{\mu}_k \times \mathbf{1}_{(\widehat{\tau}_k, \widehat{\tau}_{k+1}]}(t)$$

As, a result for one replication is not significant, we make $M = 1,000$ replications and thus we use the mean integrated square error (MISE).

3. Methods of Off-Line Detection

The most popular methods used are: The Penalized Least Square Error (PLS) (Lavielle & Moulines, 2000), the Filtered Derivative with p-value (FDpV) (Bertrand, Fhima, & Guillin, 2011) and the Filtered Derivative with False Discovery Rate (FDqV) (Elmi, 2014). We make drawings with following methods using the simulation of the subsection (1.2).

3.1 PLS-Method

For PLS-method, we have to search the instants of change points which minimize the contrast function defined as

$$Q(\tau) = \sum_{k=0}^K \sum_{t=\tau_k+1}^{\tau_{k+1}} |X_t - \widehat{\mu}_k|^2 \quad (2)$$

Two cases are studied:

- The case where K is know (Bai & Perron, 1998) use the dynamical program method for estimating the instants of ruptures and the mean values corresponding.
- In the case where K is unknown, many authors (such that Lavielle & Moulines, 2000; Lavielle & Teyssière, 2006; Birgé & Massart, 2007) are proposed different values of penalized parameter for the performance of this method.

In Lavielle and Moulines (2000), the proposed choice of penalized parameter is:

$$\beta_1 = \frac{2\sigma^2(\log n)}{n}.$$

The inconvenient is to over-estimate the number of change points.

In Birgé and Massart (2007), we have

$$\beta_2 = \frac{\sigma^2}{n} \times \left[2 + 5 \times \log\left(\frac{n}{K}\right) \right]$$

We can only apply the times series with constant variance.

Lavielle and Teyssière (2006) give an adaptive-method for estimating the number in following manner

– For $1 < K < K_{max}$, we adjust the model $f(K) = a \times K + b \times \log(K) + e_K$ with the contrast function and e_K a sequence independent of random variable following the gaussian law standard.

– We evaluate the probability that $Q(K)$ follows this model.

– The estimated number of change points will be the highest value of K such the p-value corresponding is the smaller value of a given threshold.

For more details see Lavielle and Teyssière (2006).

For an illustration, we draw the following figure.

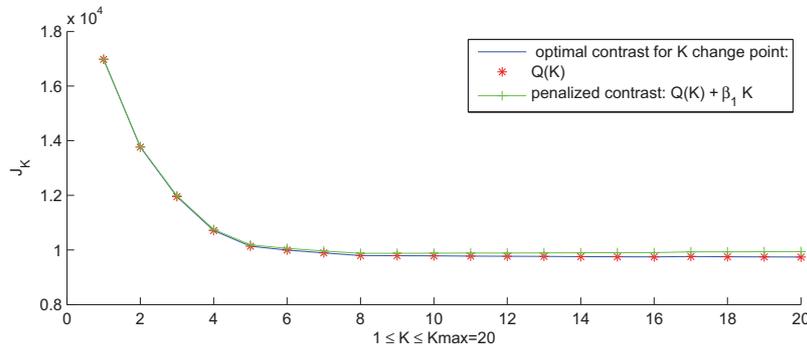


Figure 2. Blue with red crosses: The contrast function $Q(\hat{\tau}^K)$; green: The penalized contrasted $pen(K)$

3.2 FDpV-Method

The FDpV-method has two steps:

1) The step 1 is based for detection of potential changes points. For this we use the Filtered Derivative (Basseville & Nikirov, 1993) define as it follows:

$$FD(t, A) = \widehat{\mu}(t + 1, t + A) - \widehat{\mu}(t - A, t), \text{ for } A < t < N - A \tag{3}$$

where

$$\widehat{\mu}(t + 1, t + A) := A^{-1} \sum_{j=t+1}^{t+A} X_j$$

is the classical empirical mean.

In this case Without noise, the function $j \rightarrow FD(j, A)$ presents a hat centered at $\tau = j$ that is the top of the hat corresponding at the right change point. The height of the hat is exactly the size of the change on the mean, and the spread is $2A$. When the signal is random, the true μ at the right (resp left) on $(\tau - A, \tau + A)$ is replace by $\widehat{\mu}$ on $(\tau - A, \tau + A)$. The estimate mean $\widehat{\mu}$ is fluctuating around μ . In order to reduce the noise due to the sampling fluctuation, we filters the signal by replacing the true value mean at the right and the left at the point j by its estimated at the right and the left at point j on $(\tau - A, \tau + A)$, and we take the difference of these two quantities.

For detection change points, (Basseville & Nikirov, 1993; Benveniste & Basseville, 1984) choose a threshold C_1 and keep only the instants $\widehat{\tau}_k$ for $k = 1, \dots, \widehat{K}$, where \widehat{K} is the length of potential change points and such that $\max |FD(\widehat{\tau}_k)|$ exceed the threshold C_1 . At the end of the calculus, we have the instant of potential change points $\widehat{\tau}_k$ for $k = 1, \dots, \widehat{K}$. We have $K < \widehat{K}$. We draw below a figure for illustration.

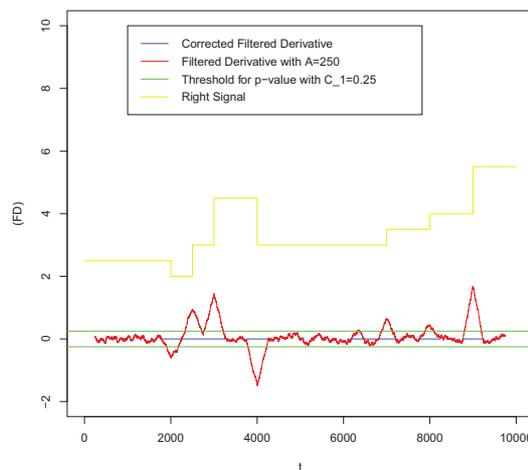


Figure 3. The right signal (red), the noisy signal (blue), and Filtered Derivative function (green)

Remark 1 A natural question is coming on : Does it exist in order to choose the optimal parameters of filtered derivative A and C_1 ? The goal of this work is to give a response of this question.

2) Recently, Bertrand, Fhima, and Guillin (2011) remark it exist false alarms at the end of the step 1, so for keeping only as possible the right change points, they have had an idea to add a second step. For this, they have compared pairwise the means estimated $\widehat{\mu}_{k-1} := \text{mean}(X, \tau_{k-1}^*, \tau_k^*)$ versus $\widehat{\mu}_k := \text{mean}(X, \tau_k^*, \tau_{k+1}^*)$. In other words they have done a test hypothesis where:

$$(H_{0,k}) : \widehat{\mu}_k = \widehat{\mu}_{k+1} \quad \text{versus} \quad (H_{1,k}) : \widehat{\mu}_k \neq \widehat{\mu}_{k+1}$$

In the sequel, they have calculated the p-value corresponding to each potential change points and they have chosen an critical p-value p^* for keeping only the p-value lesser than the critical p-value.

Remark 2 In (Bertrand, Fhima, & Guillin, 2011), the critical p-value chosen is arbitrary ($p^* = 10^{-6}$), so we can say that the problem of optimal p-value is open and we will try to do in future work.

3.3 FDqV-Method

In (Elmi, 2014), we have proposed a method for change points detection, this method use also the filtered derivative as step 1, but we have added a step 2, which allow us to detect as possible all change points right. The difference between the FDPV and FDqV is: The first use a single hypothesis for keeping all change points right and the second use a multiple test. The power of FDqV is established in (Elmi, 2014).

The algorithm of the step 2 of FDqV is:

- We put the p- values in this way $p_1^* \leq \dots \leq p_{K^*}^*$.
- We choose a critic value denote q^* .
- We eliminate all p-value such that $p_i^* > \frac{i}{K^*} q^*$.

At the end of the step 2, we obtain $(p_1^*, p_2^*, \dots, p_{K^*}^*)$ and the estimated instants $(\tau_1^*, \tau_2^*, \dots, \tau_{K^*}^*)$, with $K < K^*$.

Signal estimation after Step2 by FDqV method with parameters $A=250$ and $C_1=0.1$.

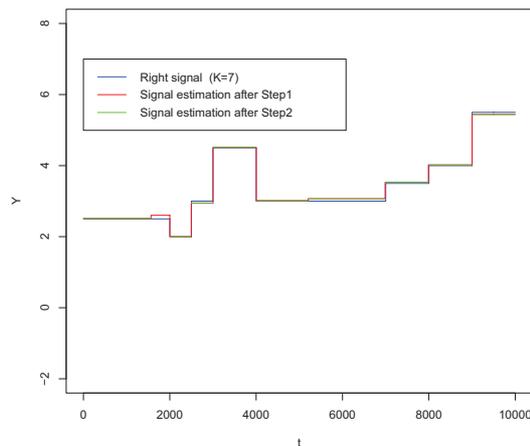


Figure 4. Signal reconstruction after step 2 by FDqV method

Now, we start the main part of this article.

4. The Choice of Parameters for Filtered Derivative Method

All change point method depends on extra-parameters, which have to be well chosen. The PLS method depends only on the penalization parameter β , different choices are possible see Section 3 above. The filtered derivative method depends on the parameters, namely the window size A and the threshold C_1 . Both FDPV and FDqV method use filtered derivative as Step 1, so they depends on the same extra-parameters A and C_1 . Moreover, FDPV and FDqV method add a Step 2, which depends on another extra-parameter, that is the critical p-value p^* or the q-value q^* . In Subsection, we discuss the different criterium. In Subsection, we give the bound of the type II error.

4.1 Choice the Extra-Parameters of FD

As pointed out in Subsection, the quality of a change point method can be evaluated by two criteria: i) the absolute value of the number of estimated change point minus the right number of change points $|\widehat{K} - K|$; ii) ISE or MISE. Both criterions lead to prefer detecting more potential change point than missing at least one. Indeed, the no detection of one change point could greatly impact the mean values $\widehat{\mu}_k$'s and after the ISE, but also the p-value p_k^* 's. Stress that this phenomenon does not more exist when we restrict ourselves to FD method with the number of change as criterion.

So, the type II error or probability of no detection (PND) should be controlled at a level close to zero. However, the previous remark address to detect the right change point at the right times. As pointed out in (Bertrand, 2000), when there is more no detection, we have: For each right change point τ_k , we define the local PND as

$$\beta_{loc}(\tau_k) = \mathbf{P}(B_k) \quad \text{where} \quad B_k = \{\forall k \in [\tau_k - A, \tau_k + A], |D(A, k)| < C_1\}.$$

Then with these notations, we can write the global PND in this manner

$$PND_{global} = \mathbf{P}\left(\bigcup_{k=1}^K B_k\right). \tag{4}$$

On the other hand, we define the probability of false alarm or probability of type I error as following:

$$\alpha(A, C_1) = \mathbb{P}(\tau(C_1, A) \leq N - A)$$

Where $\tau(A, C_1)$ is the first hitting time of C_1 and

$$\tau(A, C_1) := \inf\{k \geq A \text{ such that } FD(A, k) \geq C_1\} \tag{5}$$

However, the type I error is the probability of at least one false alarm and thus appears as a rough criterion see (Bertrand, 2000).

4.2 The Type I and II Errors at Step 1 (Filtered Derivative)

In the following proposition, we give an upper bound for PND_{global} .

Proposition 1 Assume there are K change points and a configuration of change points $\tau = (\tau_1, \tau_2, \dots, \tau_K)$, with means (μ_0, \dots, μ_K) and shifts $(\delta_1, \dots, \delta_K)$ as described in Subsection 2.1. Then

$$PND_{global} \leq K \times \beta^*(C_1, A).$$

where PND_{global} is defined by (4) and

$$\beta^*(C_1, A) := \Psi\left(\frac{\delta - C_1}{\sigma} \sqrt{\frac{A}{2}}\right) \times \Phi\left(\frac{C_1 - \delta/3}{\sigma} \sqrt{\frac{A}{2}}\right)^2. \tag{6}$$

and Φ and Ψ are given by (1).

Proof. Following (Bertrand, 2000, p. 222, Prop. 3.2), we have for each change point τ_k ,

$$\mathbf{P}(B_k) \leq \Psi\left(\frac{\delta_k - C_1}{\sigma} \sqrt{\frac{A}{2}}\right) \times \Phi\left(\frac{C_1 - \delta_k/3}{\sigma} \sqrt{\frac{A}{2}}\right)^2. \tag{7}$$

Next, by remarking that the right side of (7) is a decreasing function of δ_k and setting $\delta = \inf_{k=1, \dots, K} \delta_k$, we can deduce that

$$\mathbf{P}(B_k) \leq \beta^*(C_1, A) := \Psi\left(\frac{\delta - C_1}{\sigma} \sqrt{\frac{A}{2}}\right) \times \Phi\left(\frac{C_1 - \delta/3}{\sigma} \sqrt{\frac{A}{2}}\right)^2. \tag{8}$$

By consequent, we obviously obtain

$$PND_{global} \leq \sum_{k=1}^K \mathbf{P}(B_k)$$

which combined with 8) gives us the bound (6). This finishes the proof of Proposition 1. □

K is unknown, but is not variable. Thus, we will monitor the quantities $\beta^*(C_1, A)$, for instance we choose to set $\beta^*(C_1, A) = 10^{-4}$ and we can write $\ln \beta^*(C_1, A) = f\left(\frac{C_1}{\delta}, \frac{\delta}{\sigma} \sqrt{\frac{A}{2}}\right)$. Thus after the variables change we have $f(x, y) = \ln(10^{-4})$ with $x = \frac{C_1}{\delta}$ and $y = \frac{\delta}{\sigma} \sqrt{\frac{A}{2}}$ and $f(x, y) = \ln \Psi((1 - x) \times y) + 2 \ln \Phi((x - 1/3) \times y)$. This equation can be numerically solved, and we find couples solution of this equation. Since the map $C_1 \mapsto \beta^*(C_1, A)$ is decreasing, and we find an implicit function $A \mapsto C_1(A)$. After having controlled the PND, we can control the PFA. We know (Bertrand, 2000, p. 221, Prop. 3.1) that for all $\varepsilon > 0$ there exists a constant M_ε such that

$$\alpha \leq M_\varepsilon \times \alpha^*(C_1, A) := M_\varepsilon \times \left(\frac{n - A}{A}\right) \times \Psi\left(\frac{C_1}{\sigma} \sqrt{\frac{A}{2 + \varepsilon}}\right). \tag{9}$$

For instance, we can set $\varepsilon = 0.1$, next we plug the implicit relationship between A and C_1 inside (9) and we obtain a function $A \mapsto \alpha^*(C_1(A), A)$. The first idea is to make varying the parameter A in order to find the optimal value corresponding to a minimum of the map $A \mapsto \alpha^*(C_1(A), A)$. Unfortunately, the map $A \mapsto \alpha^*(C_1(A), A)$ is decreasing and reaches no minimum value.

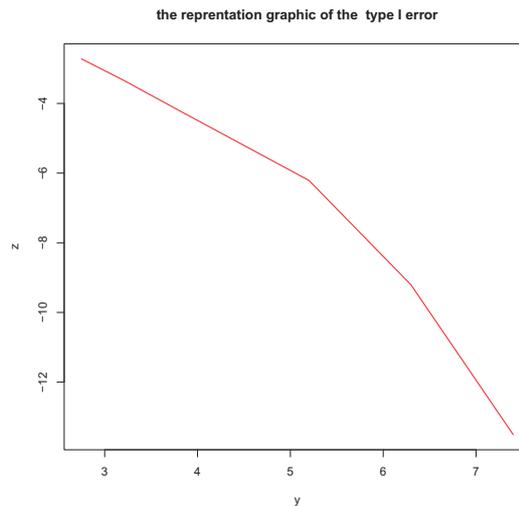


Figure 5. The graphic corresponding at the type I error, $y = \sqrt{\frac{A}{2}}$ and $z = \alpha^*(C_1(A), A)$

4.3 Necessary Condition of No-Detection

In this subsection, we draw a three figures for choosing a “good” window A . According to the figure below, we can choose A with the following condition:

$$2 \times A < L_0 := \inf\{|\tau_{k+1} - \tau_k|, k = 1, \dots, K\}. \tag{10}$$

With this drawings, we can say that A must verify $A < L_0/2$, because in the first, we detect all change points and others, we only detect two change points.

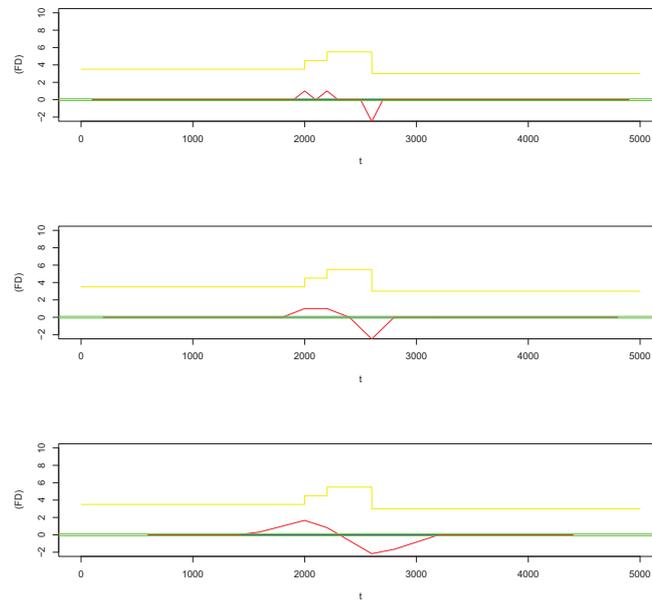


Figure 6. Red: the filtered derivative with $A=100$, $A=200$, $A=600$, Yellow: the right signal, Green: the threshold $C_1 = 0.1$

4.4 Control of Number of False Alarms

In this subsection, we want to control the number of false alarms (NFA) and not only the PFA (probability of false alarms). First, we can remark that the number of false alarm is always greater than the corresponding one when there is no change. Indeed, let us denote by \tilde{K} the number of change points select in step 1 (FD), then the number of false alarms is $(\tilde{K} - K)$. By using (Bertrand, 2000), we have

$$FD(A, t) = \Gamma(A, t) + \sum_{k=1}^K \delta_k \times g\left(\frac{t - \tau_k}{A}\right) \text{ for all } t$$

where

$$\Gamma(A, t) = A^{-1}[\widehat{S}_{t+A} + \widehat{S}_{t-A} - 2 \times \widehat{S}_t], \text{ and } \widehat{S}_t = \sum_{k=1}^t X_k$$

and

$$g(x) = \begin{cases} 1 - |x|: & \text{when } |x| \leq 1 \\ 0: & \text{when } |x| \geq 1 \end{cases}$$

Let us point that when there is no change, then $FD(A, t) = \Gamma(A, t)$ for all t , this implies that $(\tilde{K} - K) \leq \tilde{K}_0$, where \tilde{K}_0 denotes the number of change points detected by FD when there is no change. For example, using the simulation the subsection (1.2), we can see that $\tilde{K}_0 = 3$ (see drawings below and count the number τ^*).

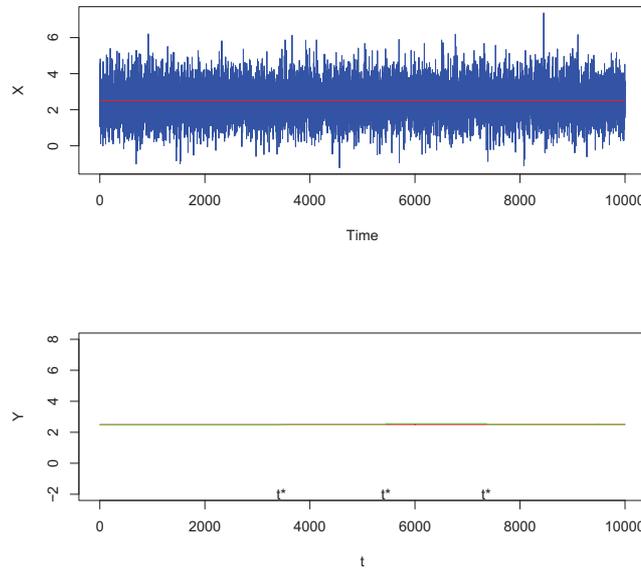


Figure 7. First drawing: The signal observed: blue, The right signal: red. Second drawing: The signal reconstruction: green

Thus, we can restrict ourselves for estimation the number of false alarms to the case without change. In this case there are only false detection, and we denote by RL_k the real variable corresponding to the run length before the k th false detection. For $k \geq 1$, we have $\tilde{\tau}_k = \sum_{j=1}^k RL_j(\omega)$. Next, we denote by $M(\omega)$ the number of potential change points. With these notation, we can state the following theorem that allow us to give a bound of number of false alarms:

Theorem 1 Assume there is no change, then

i) For all integer $L \leq N$

$$\mathbb{P}(M(\omega) \leq L - 1) \leq \varphi(A, C_1, N, L) \tag{11}$$

where

$$\varphi(A, C_1, N, L) := \sum_{(L_1, \dots, L_L), \sum L_j > N} \prod_{i=1}^L |L_i - 2A| \times \Psi\left(\frac{AC_1}{\sigma \sqrt{L_i}}\right)$$

ii) Moreover

$$\mathbb{E}M(\omega) \geq -\varphi(A, C_1, N, 0) - \sum_{L \geq 2} \varphi(A, C_1, N, L) \tag{12}$$

Proof. See Appendix. □

5. Discussion

5.1 For FD Method

5.1.1 The Choice of Parameter A

As stress above, the question of parameters which depends the FD method is important for its algorithm. In this work, we give the criteria for the choice of reasonable parameters A and C_1 . In the preceding section, we have established that for detecting of all right change points we must have $2 \times A < L_0$ with $L_0 := \inf\{|\tau_{k+1} - \tau_k|, k = 1, \dots, K\}$, see also Figure 5.

5.1.2 The Choice of C_1

In (Bertrand, 2000), we have $C_1 < \delta_o$ with $\delta_o = \inf\{|\delta_k|, k = 1, \dots, K\}$ where δ_k are the size of the average. An other hand in the Theorem 1, we have obtained a bound of number of false detection and its average using the function φ . For N, L fixed and A supposed verified the condition above, we can choose C_1 optimal. We remark that if C_1 is increasing, the function Ψ is decreasing and consequently the average of number of false alarms is decreasing, so it should to choose C_1 the greatest possible and C_1 must verify the condition was given by Bertrand (2000).

5.2 Simulation

We use the simulation of the subsection (1.2) and we make various drawings with different values C_1 and A .

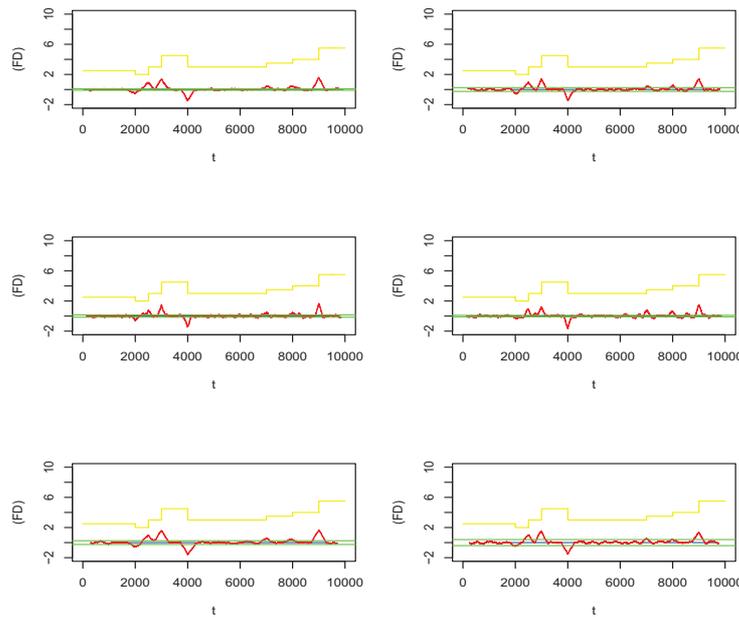


Figure 8. The Filtered Derivative with different parameters A and C_1

5.2.1 Comment

We notice through these drawings above that the number of false alarms and the number of no detections vary according to parameters A and C_1 . Thus, a choice of A and C minimizing these two points (NND and NFA) is imperative. It is what we are going to do after.

5.2.2 Numerical Estimation of NND and NFA

In this part, we want to have an estimation of NND and NFA. For this we make the calculation for Filtered Derivative method with different $A = 30$ to 500 and $C_1 = 0.1$ to 1 and we choose $K_{max} = 20$ (we suppose that the maximum number change points is equal 20). An each couple (A, C_1) , we make 1000 simulations for to have an number exact of no detection of change points and number of false alarms. Then, we can deduce the NND and NFA for each couple and we sum up the result in the followings arrays:

Table 1. Table of no detections of change points

A/C_1	0.1	0.2	0.25	0.3	0.4	0.5	0.7
30	0.176	0.307	0.50	0.65	1.05	1.53	2.46
40	0.098	0.243	0.411	0.5	1.02	1.49	2.50
60	0.045	0.132	0.231	0.395	0.874	1.47	2.616
100	0.008	0.053	0.0117	0.219	0.710	1.457	2.766
180	0.005	0.008	0.025	0.071	0.470	1.489	2.92
220	0.002	0.003	0.011	0.055	0.469	1.489	2.934
250	0	0	0.01	0.028	0.369	0.161	2.954
350	0	0	0	1.029	1.143	1.447	2.99
450	0	0	0	0.003	0.716	1.508	2.994
500	0	0	0	0	0.11	1.48	3

Table 2. Table of number of false alarms

A/C_1	0.1	0.2	0.25	0.3	0.4	0.5	0.7
30	13.51	13.281	13.471	13.622	14.026	14.506	14.796
40	13.166	13.217	13.385	13.469	14.002	14.461	5.810
60	13.019	13.105	13.205	13.369	13.848	10.077	1.714
100	12.834	13.028	13.143	13.052	5.168	2.163	1.283
180	12.976	12.217	6.762	2.948	1.433	1.518	1.0833
220	12.974	8.3033	3.614	1.686	1.269	1.407	1.035
250	12.974	5.752	2.429	1.281	1.227	1.393	1.041
350	10.822	2.331	1.189	0.010	0.291	1.335	1.005
450	4.333	1.296	1.0339	1.001	1.079	1.289	1
500	4.663	1.165	1.0219	0.994	1.030	1.167	0.986

Table 3. Table of integer square error

A/C_1	0.1	0.2	0.25	0.3	0.4	0.5	0.7
30	7947.31	7947.31	7945.21	7924.43	8078.99	7947.31	7807.86
40	7779.01	7779.01	7871.04	7779.01	7779.01	7779.01	4897.58
60	2475.03	7592.16	7651.39	7737.42	7737.42	7737.42	2419.67
100	7641.06	7565.51	7641.06	7758.73	4531.81	1860.44	1775.77
180	7748.16	7748.18	5540.20	3187.57	1567.89	888.73	2406.88
220	7880.80	6395.04	3818.09	2235.87	1404.52	788.74	2099.12
250	7992.98	5261.26	3094.07	1875.92	1579.60	916.71	2134.77
350	7712.28	3208.19	1983.91	1874.15	1600.91	515.51	2140.14
450	5908.22	2102.07	1837.95	2144.79	1584.63	810.66	2201.55
500	5079.89	2332.44	2223.88	1961.04	1654.26	861.96	2080.03

6. Comparison the Filtered Derivative With Parameters Optimized and Penalized Least Square Error (the Adaptive Method (Lavielle & Teysnière, 2006))

6.1 Monte Carlo Simulation

For comparing the both-methods (The filtered derivative with parameters optimized and adaptive method of (Lavielle & Teysnière, 2006), we choose the simulation of the subsection (1.2). For FD-method, the optimal parameters chosen are $A_{opt} = 250$ and $C_{1,opt} = 0.25$.

The criteria of comparison are the number of false alarms, the number of no-detection, and the mean square error. Firstly, for one replication, we obtain:

- For adaptive method, NFA = 1, NND = 4, ISE = 28670 (see figure below).

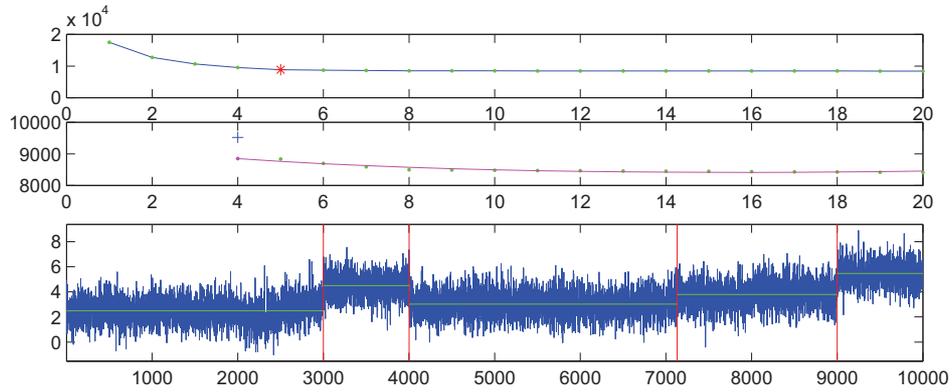


Figure 9. The adaptive method

- For filtered derivative with optimized parameters, NFA = 3, NND = 1, ISE = 3998.8 (see figure below).

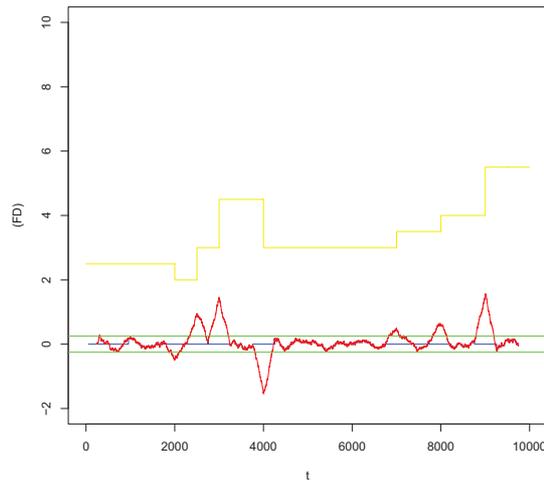


Figure 10. The filtered derivative with parameters optimized

For $M = 1,000$ replications, we obtain:

- For FD-method, we obtain $MISE = 3094.07$, the number of false alarms $NFA = 2.429$, the number of no detection $NND = 0.01$.
- For adaptive-method, $MISE = 29009$, $NFA = 0.600$, $NND = 2.250$.

6.2 Numerical Conclusion

It is clearly that the FD-method with parameters optimized is better the PLS-method adaptive (Lavielle & Teysière, 2006) for the criteria mean integrate square error. In other hand, the FD-method with parameters optimized has less no detection of points that the PLS-method adaptive but the firstly has many false alarms that the secondly. Stress that, for in forthcoming work, we will add in step 2 for FD-method with parameter optimized for having the number of false alarms at a level close to zero. In other words, we will optimize the FDqV (Elmi, 2014) for the q-value corresponds the false discovery rate.

7. Conclusion

In this work, we gave the reasonable parameters of filtered derivative method. We obtained these parameters by doing the simulations but if we consider the Theorem 1 and fix L, N and choose A in order (10) we can calculate C_1 theoretically. To do directly a theoretical calculus of A and C_1 is very difficult and not solution at this moment. In other hand, we can say that is better then to monitor the number of false alarms and number of no-detections

that to control the probability of false alarms and the probability of no-detection as done in the preceding works. A natural sequel will have to make the same for FDqV-method for keeping as possible the right number of change points. It will be interesting to search in manner to adapt these results for the times series weakly and strongly dependent.

Acknowledgements

The author thanks the anonymous reviewers for all remarks and suggestions for performing my paper. The author also thanks the president of University of Djibouti Dr Djama Mohamed Hassan for his encouragement along this work and finally the professor Mohamed Abdi Hachi for his helpful of learning the English courses.

References

- Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1), 47-78. <http://dx.doi.org/10.2307/2998540>
- Basseville, M., & Nikirov, I. (1993). *The detection of Abrupt Changes-Theory and Applications*. Information and System Sciences Series. Englewood Cliffs, NJ: Prentice-Hall.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B. Methodological*, 57(1), 289-300. <http://dx.doi.org/10.2307/2346101>
- Benveniste, A., & Basseville, M. (1984). Detection of abrupt changes in signals and dynamical systems: Some statistical aspects. *Analysis and optimization of systems*, 62, 145-155. Lecture Notes in Control and Inform. Sci. Berlin: Springer.
- Bertrand, P. (2000). A local method for estimating change points: The “hat-function”. *Statistics*, 34(3), 215-235. <http://dx.doi.org/10.1080/02331880008802714>
- Bertrand, P. R., Fhima, M., & Guillin, A. (2011). Off-line detection of multiple change points by the filtered derivative with p -value method. *Sequential Analysis*, 30(2), 172-207. <http://dx.doi.org/10.1080/07474946.2011.563710>
- Birgé, L., & Massart, P. (2007). Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields*, 138(1-2), 33-73. <http://dx.doi.org/10.1007/s00440-006-0011-8>
- Brodsky, B., & Darkhovsky, B. (1993). *Nonparametric Methods in Change-Points Problems*. The Netherlands: Kluwer Academic Publishers. <http://dx.doi.org/10.1007/978-94-015-8163-9>
- Csorgo, M., & Horváth, L. (1997). *Limit Theorem in Change-Point Analysis*. New York: Wiley.
- Elmi, M. (2014) Multiple change points detection by Filtered Derivative and False Discovery Rate. *International Journal of Statistics and Probability*, 3(1), 12-23. <http://dx.doi.org/10.5539/ijsp.v3n1p12>
- Gombay, E., & Serban, D. (2009). Monitoring parameter change in AR(p) time series models. *Journal of Multivariate Analysis*, 100, 715-725. <http://dx.doi.org/10.1016/j.jmva.2008.08.005>
- Husková, M., & Meintanis, S. (2006b). Change point analysis based on the empirical characteristic functions of ranks. *Sequential Analysis*, 25, 421-436.
- Lavielle, M., & Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *J. Time Ser. Anal.*, 21(1), 33-59.
- Lavielle, M., & Teyssière, G. (2006). Detection of multiple change points in multivariate time series. *Lithuanian Math. J.*, 46, 287-306. <http://dx.doi.org/10.1007/s10986-006-0028-9>
- Lebarbier, E. (2005). Detecting multiple change-points in the mean of gaussian process by model selection. *Signal Processing*, 85, 717-736. <http://dx.doi.org/10.1016/j.sigpro.2004.11.012>

Appendix

In this subsection, we give some technical lemmas and proposition useful for the proof of the main theorem.

Lemma 1 Let $L \in \mathbb{N}$, then

$$\mathbb{P}(M(\omega) \leq L - 1) = \mathbb{P}\left(\sum_{j=1}^L RL_j(\omega) > N\right)$$

Proof. We have

$$\left\{\sum_{j=1}^{L+1} RL_j(\omega) > N\right\} \Rightarrow \{M(\omega) \leq L\}$$

and

$$\left\{\sum_{j=1}^L RL_j(\omega) \leq N\right\} \Rightarrow \{M(\omega) \geq L\}$$

$$\mathbb{P}(M(\omega) < L) = 1 - \mathbb{P}(M(\omega) \geq L) \leq \mathbb{P}\left(\sum_{j=1}^L RL_j(\omega) > N\right)$$

Finally, we have the result above. □

Lemma 2 We suppose that σ is constant and know, then $\Gamma(A, k)$ is a family Gaussian such that $\Gamma(A, k) \sim \mathcal{N}(0, \sigma \sqrt{\frac{2}{A}})$.

Proof. See Bertrand (2000). □

Proposition 2 Let $L_j \in \mathbb{N}$,

$$\mathbb{P}(RL_j \leq L_j) = 0 \text{ when } L_j < 0$$

and

$$\mathbb{P}(RL_j \leq L_j) \leq |L_j - 2 \times A| \Psi\left(\frac{AC_1}{\sigma^2 \sqrt{L_j}}\right) \text{ when } L_j \geq 0$$

with

$$\Psi(x) = 1 - \Phi(x) \text{ and } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du.$$

Proof. From

$$(RL(A, C_1) > L_j) = \{\max(\Gamma(A, k), \Gamma(A, k + 1), \dots, \Gamma(A, L_j)) \leq C_1\}$$

We can write

$$\mathbb{P}(RL_j \leq L_j) = 1 - \mathbb{P}(RL_j > L)$$

and

$$\mathbb{P}(RL_j > L_j) = \mathbb{P}\left(\max_{t \in [k, L_j]} \Gamma(A, t) \leq C_1\right)$$

By scaling, we obtain

$$\max_{t \in [k, L_j]} \Gamma(A, t) \stackrel{\mathcal{L}}{=} \sigma A^{-1} \sqrt{L_j} \rho\left(\frac{A}{L_j}\right)$$

where

$$\rho(u) = \max\left[W\left(u + \frac{1}{A}\right) + W\left(u - \frac{1}{A}\right) - 2W(u)\right] \text{ for } u = \frac{A}{n}, \dots, 1 - \frac{A}{n}$$

ρ is the maximum of discrete Wiener Process, according to Lemma (3), we know

$$\left[W\left(u + \frac{1}{A}\right) + W\left(u - \frac{1}{A}\right) - 2W(u)\right] \sim \mathcal{N}\left(0, \sigma \sqrt{\frac{2}{A}}\right)$$

Then

$$\mathbb{P}\left(\max_{t \in [k, L_j]} \Gamma(A, t) \leq C_1\right) = \mathbb{P}\left(\sigma A^{-1} \sqrt{L_j} \rho\left(\frac{A}{L_j}\right) \leq C_1\right)$$

According to the following remark from (Csorgo & Horváth, 1997).

If $|I|$ is finite and $\forall i \in I, X_i \in \mathcal{N}(0, \sigma_i)$, then

$$\mathbb{P}(\sup_{i \in I} X_i \geq a) \leq |I| \Psi\left(\frac{a}{\sup_{i \in I} \sigma_i}\right)$$

Finally, we get

$$\mathbb{P}(RL_j \leq L_j) \leq |L_j - 2 \times A| \Psi\left(\frac{AC_1}{\sigma^2 \sqrt{L_j}}\right)$$

□

Proof. [Proof of Theorem 1]

i) First, we prove the upper bound (11).

$$\begin{aligned} \mathbb{P}\left(\sum_{j=1}^L RL_j(\omega) > N\right) &= \bigcup_{\{(L_1, \dots, L_L), \sum L_j > N\}} \{\forall j = 1, \dots, L, \text{ such that } RL_j = L_j\} \\ \mathbb{P}\left(\sum_{j=1}^L RL_j(\omega) > N\right) &= \mathbb{P}\left(\bigcup_{\{(L_1, \dots, L_L), \sum L_j > N\}} \{\forall j = 1, \dots, L, \text{ such that } RL_j = L_j\}\right) \end{aligned}$$

By independence of RL_j , we can write

$$\begin{aligned} \mathbb{P}\left(\sum_{j=1}^L RL_j(\omega) > N\right) &= \sum_{\{(L_1, \dots, L_L), \sum L_j > N\}} \mathbb{P}(\forall j = 1, \dots, L, \text{ such that } RL_j = L_j) \\ \mathbb{P}\left(\sum_{j=1}^L RL_j(\omega) > N\right) &= \sum_{\{(L_1, \dots, L_L), \sum L_j > N\}} \mathbb{P}\left(\bigcap_{j=1}^L \{RL_j = L_j\}\right) \end{aligned}$$

Using again independence of RL_j

$$\mathbb{P}\left(\sum_{j=1}^L RL_j(\omega) > N\right) = \sum_{\{(L_1, \dots, L_L), \sum L_j > N\}} \prod_{j=1}^L \mathbb{P}(\{RL_j = L_j\})$$

According to the proposition 2, we have

$$\mathbb{P}\left(\sum_{j=1}^L RL_j(\omega) > N\right) \leq \sum_{\{(L_1, \dots, L_L), \sum L_j > N\}} \prod_{i=1}^L |L_j - 2A| \Psi\left(\frac{AC_1}{\sigma \sqrt{L_j}}\right) := \varphi(A, C_1, N, L)$$

Finally, using the lemma 1 we have the result (11).

ii) We turn now to the upper bound (12). By definition

$$\begin{aligned} \mathbb{E}[M(\omega)] &= \sum_{L \geq 1} L \times \mathbb{P}(M(\omega) = L) \\ &= \sum_{L \geq 1} L \times [\mathbb{P}(M(\omega) \leq L) - \mathbb{P}(M(\omega) \leq L - 1)] \\ &= \sum_{L \geq 1} [L \times \mathbb{P}(M(\omega) \leq L)] - \sum_{L' \geq 0} [(L' + 1) \times \mathbb{P}(M(\omega) \leq L')] \\ &= -\mathbb{P}(M(\omega) \leq 0) - \sum_{L \geq 1} \mathbb{P}(M(\omega) \leq L) \end{aligned}$$

By using the bound (11), we straightforward deduce the bound (12). This finishes the proof of Theorem 1. □

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).