Detection of Multiple Change Points by the Filtered Derivative and False Discovery Rate

Mohamed Elmi[1]

[1] University of Djibouti, Djibouti

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

Abstract

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a time series, that is a sequence of random variable indexed by the time $t = 1, 2, \ldots, n$. We assume the existence of a segmentation $\tau = (\tau_1, \tau_2, \ldots, \tau_n)$ such that X_i is a family of independent identically distributed (i.i.d) random variable for i $\in (\tau_k, \tau_k + 1]$, *and* $k = 0, \ldots, K$ where by convention τ_o and $\tau_{K+1} = N$. In the literature, it exist two main kinds of change points detections: The change points on-line and the change points off-line. In this work, we consider only the change point analysis (off-line), when number of change points is unknown. The result obtained is based on Filtered Derivative method where we use a second step based on False Discovery Rate. We compare numerically this new method with the Filtered Derivative with p-Value.

Keywords: off-line detection of multiple change points, filtered derivative method, false discovery rate, linear time and memory complexity

1. Introduction

Change-point detection is an important problem in many applications, and it has been well-studied for a long time, see e.g. the textbooks (Basseville & Nikirov, 1993; Brodsky & Darkhovsky, 1993; Csorgo & Horváth, 1997), or (Huskovà & Meintanis, 2006b; Gombay & Serban, 2009) for an updated overview. Depending on the method of data acquisition, there exist two different kinds of change detection: *A posteriori* or off-line change-point detection arises when the series of observations is complete at the time we process the data, whereas in sequential analysis, the detection is performed on line. In this work, we only consider the *a posteriori* problem. In this century, the state to the art method was the Penalized Least Square Criterion (PLS): When the number of change point is known, PLS minimizes a contrast function (Bai & Perron, 1998; Lavielle & Moulines, 2000). When the number of change point is unknown, many authors use the penalized version of the contrast function (Lavielle & Teyssière, 2006; Lebarbier, 2005). From a computational point of view, PLS method use the dynamic programming algorithms and it needs to compute a matrix. Therefore, the time and memory complexity of PLS algorithm is of order $O(n^2)$, where n denote the size of the dataset. Due to the *data deluge*, the size of datasets are larger and larger, then the computational complexity of statistical method has became a challenge. Cumulative sum can be iteratively computed and therefore leads to algorithms with both time and memory complexity of order $O(n)$. Among these methods, the Filtered Derivative has been introduced by Benveniste and Basseville (1984) and Basseville and Nikirov (1993). The advantage of Filtered Derivative method is the time and memory complexity, both of order $O(n)$. On the other hand, Filtered Derivative method leads to many false discoveries of change points. Recently, Bertrand, Fhima, and Guillin (2011) have introduced a method called Filtered Derivative with p-value (FDpV) (see below for more details). Change detection by FDpV method has been successfully applied to real life large datasets ($n = 120,000$ or $n = 40,000$) of heartbeat series (Bertrand, Fhima, & Guillin, 2011; Ayache & Bertrand, 2011; Khalfa, Bertrand, Boudet, Chamoux, & Billat, 2012). However, Step 2 of FDpV algorithm use single hypothesis tests, and therefore it does not allow to control the rate of false discoveries. In this work, we propose to replace the family of single hypothesis tests of Step 2 in FDpV method by the use of the False Discovery Rate. The False Discovery Rate (FDR) has been introduced for multiple tests (Benjamini & Hochberg, 1995). Moreover, we investigate the effect of adding a Step 3, for taking advantage of the enlargement of windows when the number of potential change point decreases.

The rest of this paper is structured as follows: Section 1 describes the problem and the comparison criterions.

Section 2 recall the methods (FDpV and PLS) used for off-line change detection. Section 3 described the new method proposed in this work (FDqV), then Section 4 contains the numerical comparison. Eventually, the choice of the extra-parameters for FDpV or FDqV method is discussed in Section 5.

2. Description of the Problem

In this section we describe the problem of change point analysis and we give some comparison's criterions. For sake of simplicity, we restrict ourselves to a toy model, since we still have checked on real life datasets the efficiency of FDpV method, see Bertrand, Fhima, and Guillin (2011), Ayache and Bertrand (2011) and Khalfa, Bertrand, Boudet, Chamoux, and Billat (2012).

2.1 Change Point Analysis: A Toy Model

Let $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ be a series indexed by the time $\mathbf{t} = 1, 2, \ldots, N$. We assume that there exists a segmentation $\tau = (\tau_1, \ldots, \tau_K)$ such that X_t is a family of independent identically distributed (iid) random variables for $t \in (\tau_k, \tau_{k+1}]$, and $k = 0, \ldots, K$, where by convention $\tau_0 = 0$ and $\tau_{K+1} = N$. The most simple model is X_t a sequence of independent Gaussian variable with $X_t \in \mathcal{N}(\mu(t), \sigma)$, where $\mathcal{N}(\mu, \sigma)$ denote the Gaussian law with mean μ and standard deviation $\sigma = 1$, and $t \mapsto \mu(t)$ is a piecewise constant map, that is $\mu(t) = \mu_k$ for all time $t \in (\tau_k, \tau_{k+1}]$. We will use this model in all the sequel of this work.

2.2 Comparison Criterions

Assume that we do not know in advance the number K of change points. We have to estimate the configuration of change $\tau = (\tau_1, \ldots, \tau_K)$ and the values of the mean $(\mu_0, \mu_1, \ldots, \mu_K)$. We denote the estimates by $\hat{\tau} = (\hat{\tau}_1, \ldots, \hat{\tau}_{\hat{K}})$ and $(\hat{\mu}_0, \hat{\mu}_1, \ldots, \hat{\mu}_K)$. Remark that the number of change points is unknown and estimated by $\widehat{K}$.

Criterion

1) The quality of estimation for one sample can be measured by two criterions:

i) $\hat{K} - K$

ii) The integrated square error (ISE). Actually, we can reformulate the problem as an estimation of a noisy signal. The signal is

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

where we have set by convention $\tau_0 = 0$ and $\tau_{K+1} = N$. The estimated signal is then

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

and the integral square error (ISE) by

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

2) However, a result on just one simulation is hazardous. So, we have to do M simulations, with e.g. $M = 1,000$ and calculate the mean integrated square error (MISE).

3) The second family of criterions is the time complexity and the memory complexity that is the mean CPU time for estimating $\widehat{s}$ and which quantity of memory is used.

Remark 1 Other criterium are possible, like the BIC (Schwarz) criterium based on the likelihood, see e.g. Ciuperca (2011a).

3. Some Methods for Change Point Analysis

In this section, we recall some methods for change point analysis: The Penalized Least Square Error (PLS) and the Filtered Derivative with p-value (FDpV).

3.1 Penalized Least Square Method

Set $\mathcal{S}_K = \{\tau \ such \ that \ length(\tau) = K, \ that \ is \ \tau = (\tau_1, \ldots, \tau_K)\}$ the set of all possible configuration of change of length K . Firstly, when the number of change points K is known, for each configuration of change $\tau \in \mathcal{S}_K$, we can

define

$$\widehat{\mu}_k = mean(X, (\tau_k, \tau_k + 1]) := \frac{1}{(\tau_{k+1} - \tau_k)} \sum_{i=\tau_k+1}^{\tau_{k+1}} X_i, \; for \; k = 0, \ldots, K \tag{1}$$

where $mean(X, Box)$ denotes the mean of the family X_t for the indices $t \in Box$. Next, we search the configuration of change $\widehat{\tau}_K \in \mathcal{S}_K$ which minimizes the square error $Q(\tau)$ defined by

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

and we denote it by $\widehat{\tau}^K$. Secondly, we consider that the number of change points K is unknown. We remark that the map $K \longmapsto Q(\widehat{\tau}^K)$ is decreasing. So minimizing the function $Q(\tau)$ with an unknown number of changes will lead to consider as optimal the trivial configuration of changes $\tau^\star = (1, 2, \ldots, N)$. To avoid this drawback, we add a penalty term proportional to the length of the change point configuration. Eventually we want to minimize

$$pen(K) = Q(\widehat{\tau}^K) + \beta \times K \; \text{ for } K = 0 \ldots, N.$$

Different choices of the penalty coefficient β are possible. In Lavielle and Moulines (2000), the following choice is proposed:

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

In Lebarbier (2005) and Birgé and Massart (2007), the proposed choice is

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

where σ^2 is the variance assumed to be constant and known and n the size of the series. In Figure 1 below, we have plotted the contrast function and the penalized contrast function (Lavielle & Moulines, 2000). We clearly see that the penalized contrast is almost horizontal, thus the minimum value is very fluctuating with respect to the choice of the parameter β.

Q(K) where Q is the contrast function

- Q(K) calculated for Dynamical Program method
- Penalized contrast function
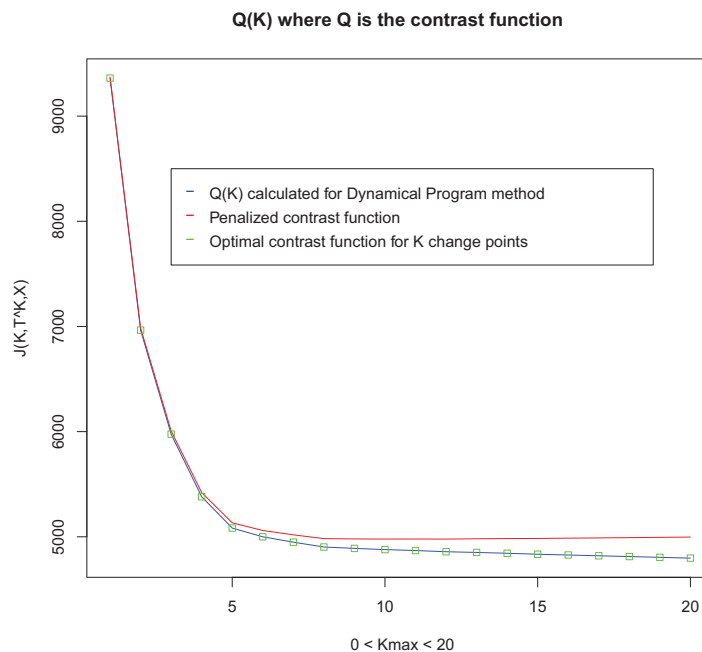- Optimal contrast function for K change points

Figure 1. Blue: Q(K) calculated with dynamical program method; red: the penalized contrast function; Green: the optimal contrast function for K change points

Let us stress that both time and memory complexity of PLS method is $O(n^2)$, see e.g. Lavielle and Moulines (2000), Lavielle and Teyssière, (2006), and Lebarbier (2005).

Remark 2 In sake of completness, let us remark that penalized least square as in Lavielle and Moulines (2000), Lavielle and Teyssière, (2006), and Lebarbier (2005) can be replaced by penalized least absolute deviations, see e.g. Ciuperca (2011b). However both time and memory complexity will be of order at least $O(n^2)$.

3.2 Description of the Filtered Derivative With p-Value Method (FDpV) (Bertrand, Fhima, & Guillin, 2011)

This method is a two steps procedure for change detection: Step 1 is based on Filtered Derivative and select a set of potential change points, whereas Step 2 calculate the p-value associated to each potential change point, for disentangling right change points and false alarms. More precisely, the method is defined as follows:

1) Step 1: Computation of the filtered derivative function. The filtered derivative function (Bertrand, 2000; Benveniste & Basseville, 1984; Basseville & Nikirov, 1993) is defined :

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

where

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

denote the empirical mean of the variables X_j on $(t + 1, t + A)$.

Next, remark that quantities $A \times FD(t, A)$ can be iteratively calculated by using

$$A \times FD(t + 1, A) = A \times FD(t, A) + X(t + 1 + A) - 2X(t + 1) + X(t - A). \tag{4}$$

Thus, the computation of the whole function $t \longmapsto FD(t)$ for $t \in [A, n-A]$ requires $O(n)$ operations and the storage of n real numbers.

2) The determination of the potential change points. Let us point that the absolute value of filtered derivative $|FD|$ presents hats at the vicinity of the change points see Figure 2 below.
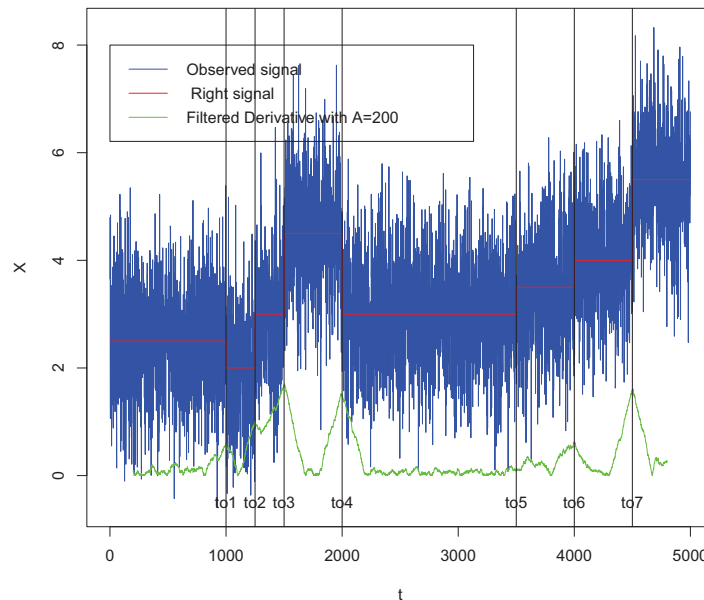

Figure 2. The right signal (red), the noisy signal (blue), and Filtered Derivative function (green)

Potential change points τ_k^*, for $k = 1, \ldots, K^*$, are selected as local maxima of the absolute value of the filtered derivative $|FD(t, A)|$ where moreover $|FD(\tau_k^*, A)|$ exceed a given threshold C_1. In Bertrand (2000) and Bertrand,

Fhima, and Guillin (2011), we have given the asymptotic distribution of the maximum $|FD|$ under the null hypothesis. Therefore, we can fix the error type at level $p_1^{\star}$, and then we can deduce the threshold C_1 corresponding to $\Pr(\max|FD(\tau_k, A)| > C_1) = p_1^*$. We can remark the existence of many local maxima in the vicinity of each right change point (see Figure 4 and Bertrand, 2000; Bertrand, Fhima, & Guillin, 2011 for theoretical explanation). On the other hand, if there is no noise that is when $\sigma = 0$, we get hats of width $2A$ and hight $\mu_{k+1} - \mu_k$ at each change point τ_k, see Figure 3.
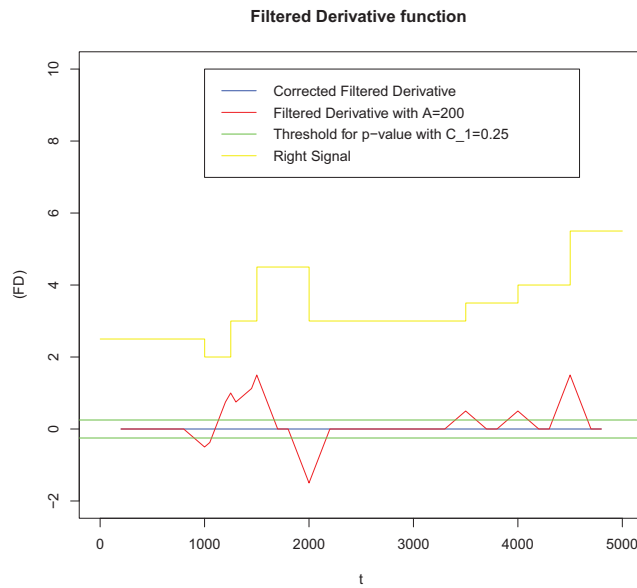

Figure 3. Filtered Derivative function without noise ($\sigma = 0$)
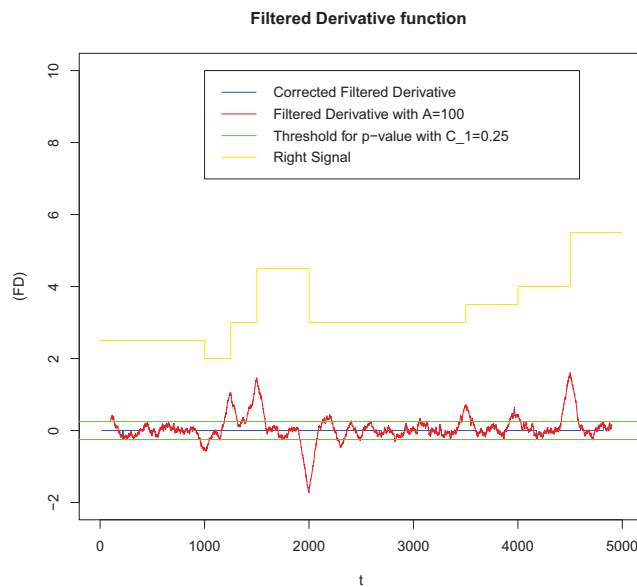

Figure 4. Filtered Derivative function with noise ($\sigma = 1$)

For this reason, we select as first potential change point τ_k^* the global maximum of the function $|FD_k(t, A)|$, then we define the function FD_{k+1} by putting to 0 a vicinity of width $2A$ of the point τ_k^* and we iterate this algorithm while $|FD_k(\tau_k^*, A)| > C_1$, see Bertrand, Fhima, and Guillin (2011). When there is noise (e.g. $\sigma = 1$), we get the following landscape, see Figure 4.

3) Step 2: Elimination of false alarm by p-value. A potential change point $\tau_k^\star$ can be an estimator of a right change point or a false alarm. In the first case, there exists an error of estimation on the location of the change. So we have to cancel a small vicinity of size ε_k around each point $\tau_k^\star$, Bertrand (2000) and Bertrand, Fhima, and Guillin (2011). Then, for each segment, we calculate an estimation of the mean

$$\widehat{\mu}_k := mean(X, \tau_k + \varepsilon_k, \tau_{k+1} - \varepsilon_{k+1}). \tag{5}$$

Next, we have to eliminate false detection in order to keep (as possible) only the right change points. In Bertrand, Fhima, and Guillin (2011), we use as Step 2 single hypothesis tests: For each potential change point $\tau_k^\star$, we test wether the parameter is the same for $t \in (\tau_{k-1}^\star + \varepsilon_{k-1}, \tau_k^\star - \varepsilon_k)$ and $t \in (\tau_k^\star + \varepsilon_k, \tau_{k+1}^\star - \varepsilon_{k+1})$ or not. More formally, for all $1 \le k \le K$, we apply the following hypothesis testing $(H_{0,k}) : \widehat{\mu}_k = \widehat{\mu}_{k+1}$　versus　$(H_{1,k}) : \widehat{\mu}_k \ne \widehat{\mu}_{k+1}$ where $\widehat{\mu}_k$'s are defined by (5). By using this second single hypothesis test, we calculate the p-values $p_1^\star, \dots, p_{K^\star}^\star$ associated to each potential change point $\tau_1^\star, \dots, \tau_{K^\star}^\star$.

4) Calculation of p-Value. We choose the statistic Student T. Indeed, under the null hypothesis, $t_k^\star$ has a Student distribution of degrees of freedom $d = N_k + N_{k-1} - 2$, where

$$t_k^\star \quad = \quad \frac{\widehat{\mu}_k - \widehat{\mu}_{k-1}}{\sqrt{\frac{S_{k-1}^2}{N_{k-1}} + \frac{S_k^2}{N_k}}}, \tag{6}$$

$\widehat{\mu}_k$'s are given by (5), $N_k = \left\{ (\tau_{k+1}^* - \varepsilon_{k+1}) - (\tau_k^* + \varepsilon_k) \right\}$, and $S_k^2 = \left\{ \left(\frac{1}{N_k} \sum_{t=\tau_k+\varepsilon_k}^{\tau_{k+1}-\varepsilon_{k+1}} X_t^2 \right) - \overline{X}_k^2 \right\}$. By construction, $d > 2A - (\varepsilon_{k-1} + 2\varepsilon_k + \varepsilon_{k+1})$, thus for $A > 30$ the distribution of $t_k^\star$ is approximatively Gaussian an we can set

$$p_k^\star \quad \approx \quad 2 \times \left\{ 1 - \Phi(|t_k|) \right\} \tag{7}$$

where Φ is the cumulative distribution function of the zero mean standard Gaussian law. Let us point a slight difficulty: Since τ_k^* maximizes the criterium $|FD_k(t, A)|$, τ_k^* is also a random variable. We avoid this drawback by canceling a small vicinity of size ε_k for each selected change point, see Formula (5) and Bertrand, Fhima, and Guillin (2011, Rem. 2.1, pp. 178-179) for details.

In Bertrand, Fhima, and Guillin (2011), we only keep the change points corresponding to a p-value lesser than a fixed threshold $p_2^\star$. Consequently, Step 2 is much more selective and it allows us to deduce an estimator of the piecewise constant map $t \longmapsto \mu(t)$, see Figure 5 below.
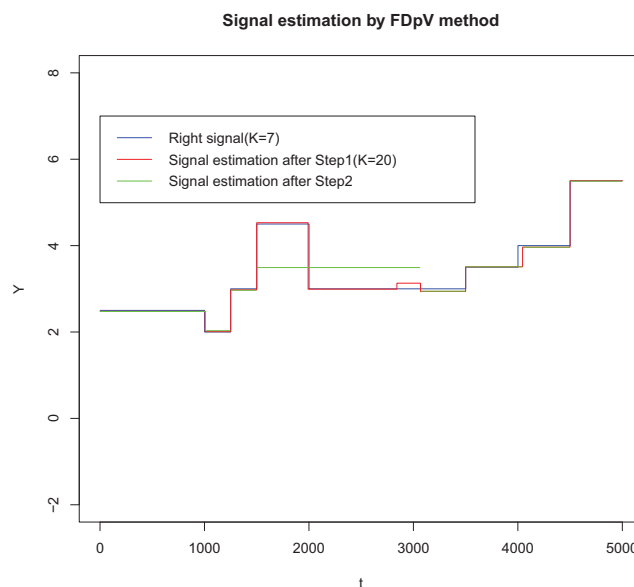

Figure 5. Signal reconstruction after Step2 by FDpV method

4. A New Method for Change Point Analysis: Filtered Derivative With a q False Discovery Rate (FDqV)

We propose a new method derived from the FDpV one: We replace Step 2 of FDpV by False Discovery Rate method (FDR) and we call this method FDqV.

Step 1. The first step is the same as in FDpV: We compute the filtered derivative function $t \mapsto FD(t, A)$ and then select the potential change points as the local maxima of the function $t \mapsto |FD(t, A)|$ reaching a threshold C_1.

Step 2. The novelty of this work is the use of False Discovery Rate thresholding procedure (Benjamini & Hochberg, 1995). The computation of p-value $p_k^\star$ is the same as for Step 2 of FDpV method. However, we then use a Bonferroni type multiple testing procedure:

(i) We tidy up p- value in the increasing order $p_{(1)}^* \leq \ldots \leq p_{(K^*)}^*$.

(ii) We choose a threshold q corresponding to the rate of false alarms or FDR.

(iii) We keep only the potential change points τ_i^* corresponding to a p-value $p_{(i)}^*$ such that $p_{(i)}^* \leq \dfrac{i}{K^*} q$.
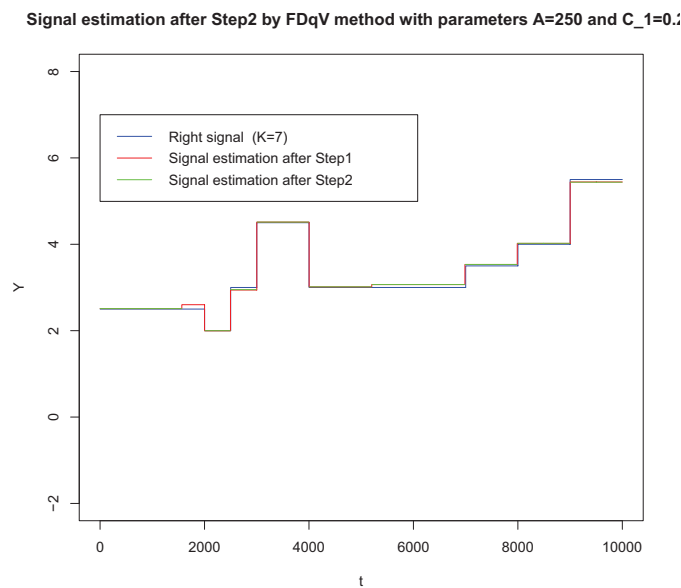

Figure 6. Signal reconstruction after Step 2 by FDqV method

Step 3. Let us point that Step 1 of FDpV or FDqV select potential change point as local maxima of the absolute value of the filtered derivative function, which is the difference of the mean estimated on sliding window of size A on a box at left and at right of the point t. Then Step 2 select some change points corresponding to a p-value smaller to a fixed threshold (FDpV) or a linear threshold (qFDR). In both case, the p-value is computed following the potential change point selected in Step 1. We recall here that these p-values are therefore calculated with windows larger than A. In other words, we have more information on the mean at Step 2 than at Step 1. This remark has suggested us to add a third step, which is the same as Step2 but with larger windows. More formally, Step 2 of FQqV can be seen as a map $FDR : (X, \tau^*, q) \longmapsto \widetilde{\tau}$, where $\widetilde{\tau} = (\widetilde{\tau}_1, \ldots, \widetilde{\tau}_{\widetilde{K}})$ with $\widetilde{K} \leq K^*$. Thus, at Step 3, we plug $\widetilde{\tau}$ instead τ^* as input of Step 2, that is $FDR: (X, \widetilde{\tau}, q) \longmapsto \widehat{\tau}$, where $\widehat{\tau} = (\widehat{\tau}_1, \ldots, \widehat{\tau}_{\widehat{K}})$ with $\widehat{K} \leq \widetilde{K} \leq K^*$.

To sum up, in Bertrand, Fhima, and Guillin (2011), we have made simple statistics tests and we compare the means pairwisely. So we choose a threshold of critical probability to eliminate the false alarms. The novelty in this work is the use of a multiple test (FDR) with FDR fixed at level q. Both time and memory complexity of FDqV remain of order $O(n)$.
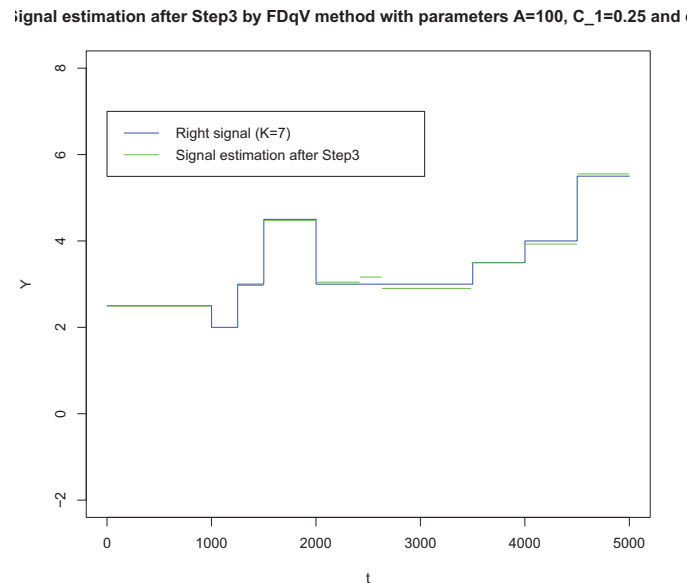
Figure 7. Signal reconstruction after Step3 by FDqV method

5. Numerical Comparisons

In this section, we compare numerically the FDpV method and the new proposed method FDqV. We use Monte Carlo simulation and via MISE.

5.1 Simulations Based on one Realization

Firstly, we select the simulation for one realization, which corresponds to Figure 2, and Figures 4–8 above. For $n = 5,000$, we have simulated one replication of a sequence of Gaussian random variables $(X_1, \ldots, X_n)$ with variance $\sigma^2 = 1$ and mean $\mu_t = f(t)$ where f is a piecewise constant function with four change points at times $\tau = (1000, 2000, 3500, 4500)$ with means $\mu = (2.5, 3, 4.5, 3, 3.5)$. Both FDpV and FDqV method depend on extra-parameters, namely the window size A, the threshold C_1 corresponding to Step 1, the maximum number changes K_{max} for Step 1, the threshold p_2^* corresponding to step 2 of FDpV, the uncertainties on the location of changes ε_k, and the threshold q of False Discovery Rate for Step 2 and Step 3 of FDqV. A brief discussion on the choice of the extra-parameters is postponed in Section 5. We have made the following choices: $A = 100$, $K_{max} = 15$, $C_1 = 0.1$, $p_2^* = 2 \times \left\{ 1 - \Phi(1.5) \right\} = 0.134$, $q = 0.1$.

5.2 Monte-Carlo Simulation

In this subsection, we made $M = 1,000$ simulations of independent iterations of sequences of Gaussian random variables $(X_0^j, \ldots, X_n^j)$ with variance $\sigma^2 = 1$ and mean $\mu_t = f(t)$, for $j = 1 \ldots, M$ and $t = 1, \ldots, N$. On each sample, we apply the FDpV method and the FDqV method with the extra-parameters given above. The mean value of $(\widehat{K} - K)$ is 3.38 with standard deviation (std) 1.64 for FDpV against a mean 2.84 with $std = 1.59$ for FDqV at Step 2, and mean $\hat{K} - K = 0.65$ with $std = 1.98$ for FDqV at Step 3. Thus we can see that than the number of false discovery is smaller by FDqV. Note that at Step1, we have $mean(\hat{K} - K) = 12$.

5.3 Mean Integrate Square Error (MISE)

For $M = 1,000$ Monte Carlo simulations, we obtain the following values of MISE:

  (i) for Filtered Derivative MISE=1419.12

  (ii) for FDpV MISE=189.59

  (iii) for FDqV (Step 2) MISE=148.75

  (iv) for FDqV (Step3) MISE=126.97

5.4 Numerical Conclusion

We clearly see that the overestimation of the number of change points is smaller for the method FDqV than the for

FPpV one. On the other hand, we can note that for the MISE criterion FDqV is better than FDpV, which is still better than Filtered Derivative.

6. How to Choose the Extra-Parameters?

In this section, we address the question of the choice of extra-parameters for FDpV method. Natural criterions are the error of type I and error of type II, so-called probability of false alarm (PFA), denoted α, and probability of non-detection (PND), denoted β.

6.1 Errors of Type I and Type II at Step 1

We stress that error of type II (PND) is more important than error of type I (PFA), at least for the ISE criterion: Indeed, just one change point missing increases strongly the error. On the other hand, as pointed out in Bertrand (2000), when there is more than one change, the notion of probability of non detection should be make more precise: For each right change point τ_k, we define the local PND as $\beta_{loc}(\tau_k) = \mathbf{P}(B_k)$ where $B_k = \{\forall k \in [\tau_k - A, \tau_k + A], |D(A, k)| < C_1\}$. Next, we can define the global PND as $PND_{global} = \mathbf{P}\left(\bigcup_{k=1}^{K} B_k\right)$. Next, we can obtain an upper bound for PND_{global}. On the one hand, let us denote by δ_k the size of the change on the mean at change point τ_k, more precisely $\delta_k = \mu_k - \mu_{k-1}$ for $k = 1, \ldots, K$. We have Bertrand (2000, Prop. 3.2, p. 222),

$$\mathbf{P}(B_k) \leq \Psi\left(\frac{\delta_k - C_1}{\sigma} \sqrt{\frac{A}{2}}\right) \times \Phi\left(\frac{C_1 - \delta_k/3}{\sigma} \sqrt{\frac{A}{2}}\right)^2 \tag{8}$$

$\Psi(x) = 1 - \Phi(x)$. Next, by remarking that the right side of (8) is a decreasing function of δ_k and setting $\delta = \inf_{k=1,\ldots,K} \delta_k$, we can deduce that

$$\mathbf{P}(B_k) \leq \beta^*(C_1, A) := \Psi\left(\frac{\delta - C_1}{\sigma} \sqrt{\frac{A}{2}}\right) \times \Phi\left(\frac{C_1 - \delta/3}{\sigma} \sqrt{\frac{A}{2}}\right)^2 \tag{9}$$

On the other hand, we obviously have

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

which combined with (9) gives us

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

The right number K is unknown, but fixed. Thus, we will control the quantities $\beta^*(C_1, A)$, for instance we choose to set $\beta^*(C_1, A) = 10^{-4}$. This equation can be numerically solved, since the map $C_1 \longmapsto \beta^*(C_1, A)$ is decreasing, and we find an implicit function $A \mapsto C_1(A)$. After having controlled the error of type II (PND), we can control the error of type I (PFA). We know Bertrand (2000, Prop. 3.1, p. 221) that for all $\varepsilon > 0$ there exists a constant M_ε such that

$$\alpha \leq M_\varepsilon \times \alpha^*(C_1, A) := M_\varepsilon \times \left(\frac{n - A}{A}\right) \times \Psi\left(\frac{C_1}{\sigma} \sqrt{\frac{A}{2 + \varepsilon}}\right). \tag{10}$$

For instance, we can set $\varepsilon = 0.1$, next we plug the implicit relationship between A and C_1 inside (10) and we obtain a function $A \longmapsto \alpha^*(C_1(A), A)$. The first idea is to make varying the parameter A in order to find the optimal value corresponding to a minimum of the map $A \longmapsto \alpha^*(C_1(A), A)$. Unfortunately enough, the map $A \longmapsto \alpha^*(C_1(A), A)$ is decreasing and reaches no minimum value.

6.1.1 Choice of the Window A

From the preceding subsection, we can get the feeling that the larger the window size A is, the smaller type I and type II errors will be. This reasoning holds true as long as

$$2 \times A < L_0 := \inf\{|\tau_{k+1} - \tau_k|, k = 1, \ldots, K\}. \tag{11}$$

Thus, we have to choose a parameter $A < L_0/2$, even if we do not exactly know the quantity L_0. Figures 8-10 below illustrate the necessary condition (11). We consider the case without noise, that is $\sigma = 0$, with three change

point at $\tau = (2000, 2200, 2600)$ thus $L_0 = 200$, and we make varying the window size A at $A = 100$, $A = 200$, and $A = 600$.

In Figure 8, we detect the three right change points. In Figure 9 and Figure 10, we only detect two change points. This plainly confirm the necessity of condition (11).
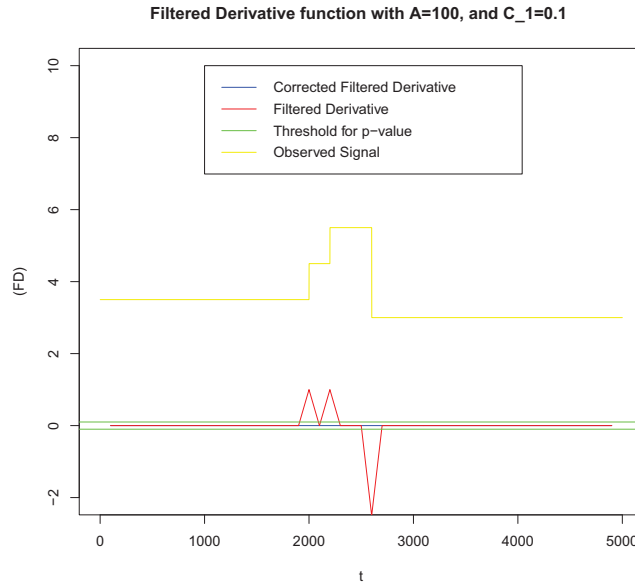

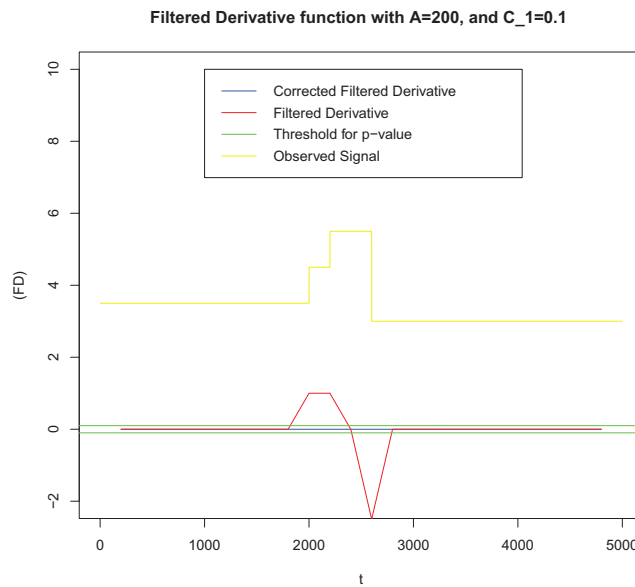
Figure 8. The Filtered Derivative without noise and A=100

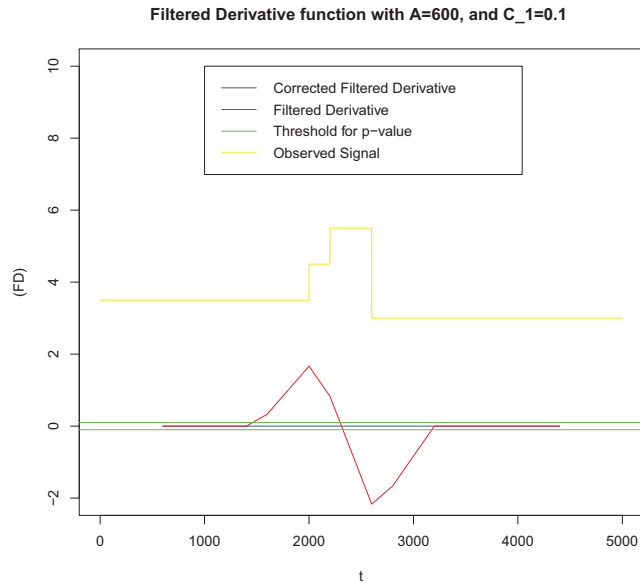

Figure 9. The Filtered Derivative without noise and A=200

Figure 10. The Filtered Derivative without noise and A=600

6.2 Error of Type I and Type II at Step 2

We can calculate t_k^* under both null and alternative assumption.

1) Under null assumption (H_0) : $\mu_k = \mu_{k+1}$, t_k^* approximatively follows a Gaussian law $\mathcal{N}(0, 1)$.

2) Under alternative assumption (H_1) with $\mu_{k+1} - \mu_k = \delta$, then $t_k^* \sim \mathcal{N}(0, 1) + \dfrac{\delta}{\sqrt{1/N_k + 1/N_{k+1}}}$.

Actually, we want to select all the right change points with as few as possible false alarms. So, we want to control the probability of non detection PND. For a single change point, let us fix a critical level t_c, then

$$
\begin{aligned}
\beta_k = PND_k &= \Pr(|t_k^*| < t_c) \\
&= \Pr\left(-t_c < \mathcal{N}(0, 1) + \frac{\delta}{\sqrt{1/N_k + 1/N_{k+1}}} < t_c\right) \\
&\simeq 1 - \Phi\left(\frac{\delta}{\sqrt{1/N_k + 1/N_{k+1}}} - t_c\right)
\end{aligned}
$$

On the other hand, the probability of one false alarm is

$$
\alpha = PFA = \Pr(|\mathcal{N}(0, 1)| > t_c) = 2 \times (1 - \Phi(t_c)).
$$

For example, when $\delta = 0.5$, $t_c = 1.5$, $N_k = N_{k+1} = 100$ then $\beta_k = 0.0207$ and $\alpha = 0.1336$.

7. Summary and Conclusions

In this work, it clearly appears the power FDqV method than FDpV method. However, the FDqV method is established by the simulations, in the future we will valid by the real data. The questions not developed in the literature are:

1) The FDpV and FDqV methods with the random variable weakly or strongly dependent.

2) The Choice of parameters such the window and the threshold, which depend the both methods.

All these questions are very difficult so more detail is needed. We will try to do in forthcoming work.

Acknowledgements

be done. The author also thanks the referee and assistant editor chief for the suggestions and remarks for improve my paper.

References

Ayache, A., & Bertrand, P. R. (2011). Discretization error of wavelet coefficient for fractal like process. *Advances in Pure and Applied Mathematics, 2*(2), 297-321. http://dx.doi.org/10.1515/apam.2011.005

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, Ser. B, Methodological, 57*(1), 289-300.

Benveniste, A., & Basseville, M. (1984). Detection of abrupt changes in signals and dynamical systems: some statistical aspects. In *Analysis and optimization of systems, Lecture Notes in Control and Inform. Sci., 62*, 145-155. http://dx.doi.org/10.1007/BFb0004951

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

Ciuperca, G. (2011a). A general criterion to determine the number of change-points. *Statistics and Probability Letters, 81*(8), 1267-1275. http://dx.doi.org/10.1016/j.spl.2011.03.027

Ciuperca, G. (2011b). Penalized least absolute deviations estimations for nonlinear model with change points. *Statistical Papers, 52*(2), 371-390.

Csorgo, M., & Horváth, L. (1997). *Limit Theorem in Change-Point Analysis.* New York: J. Wiley.

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

Huskovà, M., & Meintanis, S. (2006b). Change point analysis based on the empirical characteristic functions of ranks. *Sequential Analysis, 25*, 421-436.

Khalfa, N., Bertrand, P. R., Boudet, G., Chamoux, A., & Billat, V. (2012). Heart rate regulation processed through wavelet analysis and change detection. Some case studies. *Acta Biotheoretica, 60*(1-2), 109-129. http://dx.doi.org/10.1007/s10441-012-9154-4

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*(3), 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*(4), 717-736. http://dx.doi.org/10.1016/j.sigpro.2004.11.012