Filtered derivative with p-value method for multiple change-points detection

Pierre, R. BERTRAND[1,2] and Mehdi FHIMA[2]
[1] *INRIA Saclay, APIS Team*
[2] *Laboratoire de Mathématiques, UMR CNRS 6620*
& University Clermont-Ferrand II, France

Introduction

In different applications (health, finance,...), abrupt changes on the spectral density of long memory processes provide relevant information. In this work, we concern ourself with off-line detection. However, our method is close to the sliding window which is typically a sequential analysis method.

We model data by Gaussian processes with locally stationary and long memory increments. By using a wavelet analysis, one obtains a series with short memory. We compare numerically the efficiency of different methods for off-line detection of these changes, namely penalized least square estimators introduced by Bai and Perron (1998) versus a modification of the filtered derivative introduced by Basseville and Nikiforov (1993). The enhancement consists in computing the p-value of every change point and then apply an adaptive strategy.

Since estimation of abrupt changes on spectral density is a specific problem, we first study a more standard model. In Section 1, we concern ourself to off-line detection of abrupt changes in the mean of independent Gaussian variables with known variance and we numerically compare the efficiency of the different estimators in this case. In Section 2, we recall the definition of Gaussian processes with locally stationary increments and the properties of their wavelet coefficients. Then, we compare the different off-line detection methods on simulated locally fBm and present some results on real data.

1 A toy model: off-line detection of abrupt change in the mean or independent Gaussian variables

Let $(X_i)_{i=1,\dots,N}$ be a sequence of independent Gaussian r.v. with mean μ_i and a known variance σ^2. We assume that the map $i \mapsto \mu_i$ is piecewise constant, *i.e.* there exists a configuration $0 = \tau_0 < \tau_1 < \cdots < \tau_K < \tau_{K+1} = N$ such that $\mu_i = \mu_k$ for $\tau_k \leq i < \tau_{k+1}$. The integer K corresponds to the number of changes. However, in any real life situation, the number of abrupt changes K is unknown, leading to a problem of model selection.

There is a huge literature on this problem, see for instance the textbook of Basseville & Nikiforov (1993). Popular methods are those based on penalized least square criterion (PLSC). We refer to Birgé & Massart (2006) for a good summary of the problem. Other classical references are Lavielle & Moulines (2000) or Lebarbier (2005) or Lavielle & Teyssière (2006).

From a numerical point of view, the least square methods are based on dynamic programming algorithm. Thus we have to compute a matrix of size N. Therefore, the time and memory complexity of these algorithms is in $O(N^2)$, which becomes an important limitation with the computer progress. This has lead us to investigate the properties of a different algorithm.

Filtered derivative with p-value method (FDp-VM)

Filtered derivative method is based on the difference between the empirical mean computed on two sliding windows respectively at the right and at the left of the index k, both of size A, see [1, 4]. This difference corresponds to a sequence $(D(A,k))_{A\leq k\leq N-A}$ defined by $D(A,t) = \hat{\mu}(A,t) - \hat{\mu}(A,t-A)$ where $\hat{\mu}(A,k) = \frac{1}{A}\sum_{j=k+1}^{k+A} X_j$ is the empirical mean of X on the (sliding) box $[k+1,\, k+A]$. These

quantities can easily be calculated by recurrence with complexity $O(N)$. It suffices to remark that $D(A,k+1) = D(A,k) + X_{k+A+1} - 2X_k + X_{k-A+1}$.

From the other hand, note that $(D(A,k))_{A\leq k\leq N-A}$ is a sequence of centered r.v., except in the vicinity of a change point τ_k. In this case, there appears a "hat-function" of size $\delta_k := (\mu_{k+1} - \mu_k)$ approximatively located between $\tau_k - A$ and $\tau_k + A$, see [5]. Nevertheless, this sequence presents also hats at points different from $(\tau_1, \ldots, \tau_K)$, see Figure 2 below. There are false alarms.

In order to eliminate these false alarms, we propose to calculate the p-values α_k associated to all detected change points $(\tilde{\tau}_k)_{1\leq k\leq Kmax}$. Then, we keep only the point corresponding to a p-value lesser than a critical level α_{critic}. For all $k \in [1, Kmax]$, the p-values are calculated by using the formula $\alpha_k = \phi\left(\frac{1}{\overline{\sigma}_k}\sqrt{\frac{A_k}{2}}|D(A_k,\tilde{\tau}_k)|\right)$ where A_k is an adaptive window chosen as the minimum of the distances between $\tilde{\tau}_k$ and these two neighbours $\tilde{\tau}_{k-1}$ and $\tilde{\tau}_{k+1}$, that is $A_k = |\tilde{\tau}_k - \tilde{\tau}_{k-1}| \wedge |\tilde{\tau}_{k+1} - \tilde{\tau}_k|$. The function ϕ denotes the complementary cumulative distribution function of the normal law and $\overline{\sigma}_k$ denotes the empirical variance of X on the box $(\tilde{\tau}_{k-1}, \tilde{\tau}_{k+1})$.

Remark that K_{max} is an integer fixed by the user. It represents the maximal number of change points. As soon as possible, K_{max} should be chosen bigger than the true number of change points K. By convention, one sets $\tau_0 = \tilde{\tau}_0 = 0$ and $\tau_{K+1} = \tilde{\tau}_{Kmax+1} = N$.

So, the novelty of this work consists in discriminating between true and false alarms by attributing a p-value to each detected change point. By keeping the change points with a p-value smaller than α_{critic}, one obtains the same precision as Lavielle & Teyssière (2006) or Lebarbier (2005). Also, the main advantage of this method lies in its memory and time complexity in $O(N)$.

A numerical simulation

At first, we give an example on one sample. In the next subsection, this example is plainly confirmed by Monte-Carlo simulations. To begin with, for $N = 5000$ we have simulated one replication of a sequence of Gaussian random variable $X_0, \ldots, X_N$ with variance $\sigma^2 = 1$ and mean $\mu(i) = g(i/N)$ where g is a piecewise-constant function with five change points such as $\delta_k \in [0.5, 1.25]$, see Figure 1 below.

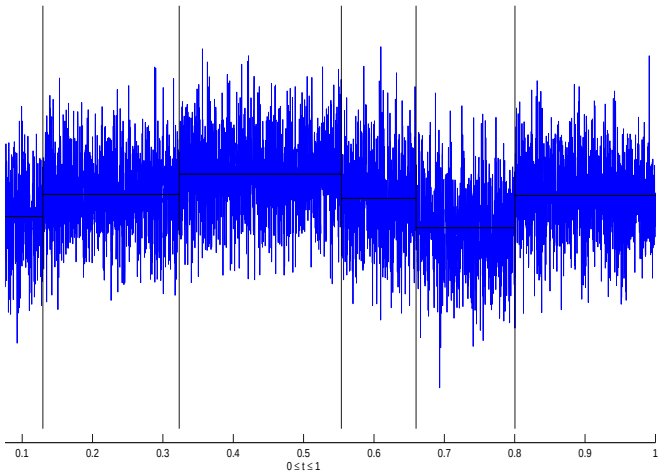

Figure 1: *The sequence $(X_i)_{0\leq i\leq N}$ with change points represented by the vertical line and means represented by the horizontal line .*

On this sample, we have computed the function $k \mapsto |D(A,k)|$ with $A = 300$, see Figure 2.

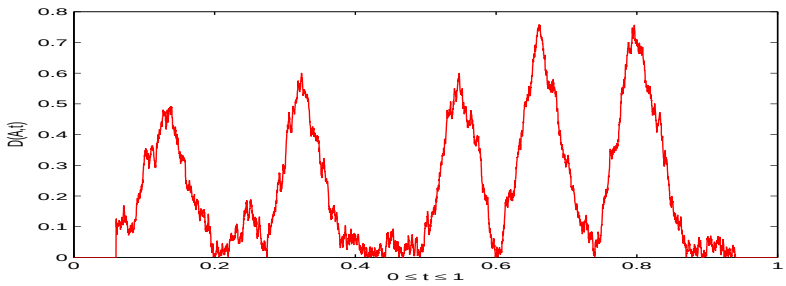

Figure 2: *The hat function.*

Both estimators penalized least square criterion (PLSC) and filtered derivative with p-value $\alpha_{critic} = 10^{-4}$ provide good results, see Figure 3 and the Monte-Carlo simulation below.

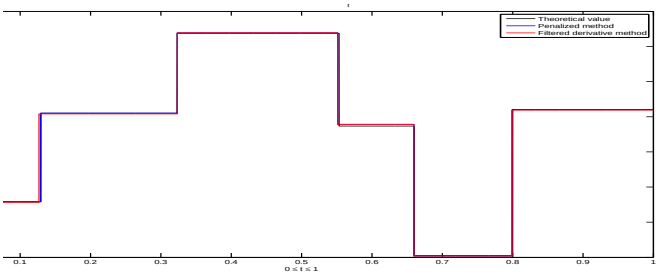

Figure 3: *Theoretical value of the piecewise-constant function g (black), and its estimators given by PLSC method (blue) and Filtered derivative method with p-value method (red).*

Monte-Carlo simulation

In this subsection, we have made $M = 1000$ simulations of independent copies of sequences of Gaussian r.v. $X_0^{(k)}, \ldots, X_N^{(k)}$ with variance $\sigma^2 = 1$ and mean $\mu(i) = g(i/N)$, for $k = 1, \ldots, M$. On each sample, we apply the FDp-V method and the PLSC method. We find the good number of changes in 98.1% of all cases for the first method and in 97.9% for the second one.

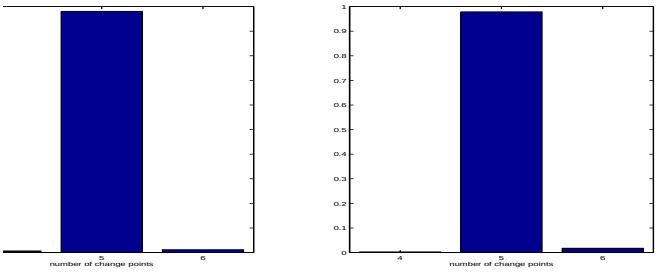

Figure 4: *Distribution of the estimated number of change points $\widehat{K}$ for $M = 1000$ realizations. Left: Using PLSC method. Right: Using Filtered derivative method*

Then, we compute the mean errors. There are two kinds of mean error :

- Mean Integrate Square Error (MISE) defined as $MISE = I\!E\left(\frac{1}{N+1}\sum_{i=0}^{N}|\widehat{g}(i/N) - g(i/N)|^2\right) =$

$I\!E\|\widehat{g} - g\|^2_{L^2([0,1])}$ which corresponds to the $L^2([0,1])$ norm of the difference between the true function g and the estimate function $\widehat{g}$. The estimate function is obtained in two steps : first we estimate the configuration of change points $(\hat{\tau}_k)_{k=1,\dots,\hat{K}}$, then we estimate the value of $\widehat{g}$ between two successive change points as the empirical mean.

- Square Error on Change Points (SECP) defined as $SECP = I\!E\left(\sum_{k=1}^{K} |\widehat{\tau}_k - \tau_k|^2\right)$, in the case where we have found the good number of abrupt changes.

We have the following results by Monte Carlo simulation

	Square Error on Change Points	Mean Integrated Squared Error
FDp-V method	1.1840×10^{-4}	0.0107
PLSC method	1.2947×10^{-4}	0.0114

Table 1: Errors (SECP & MISE) given by FDp-V method and PLSC method

Next, we compare the mean time complexity and the mean memory complexity. We have written the two programs in Matlab and have runned it with computer system which has the following characteristics: 1.8GHz processor and 512MB memory. The results concerning time and memory complexity are given in Table 2.

	Memory allocation (in Megabytes)	CPU time (in second)
FDp-V method	0.04 MB	1 s
PLSC method	200 MB	240 s

Table 2: Memory and time complexity of Filtered derivative method and PLSC method

A First conclusion

On the one hand, both methods have the same accuracy in terms of percentage of selection of the exact model, Square Error on the configuration of change points or MISE. On the other hand, the filtered derivative with p-value is less expensive in terms of time complexity and memory complexity. Indeed, algorithm based on Minimization of penalized least square criterion can use 39% of computer memory, while Filtered derivative method only needs 0.008%. This plainly confirms the difference of time and memory complexity, *i.e.* $O(N^2)$ *versus* $O(N)$.

Observe that algorithms based on penalized least square are considered by Davis et al. (2008) as maximizing *a posteriori* (MAP) criterion, whereas filtered derivative is based on sliding window and could be adapted to sequential detection, see for instance Bertrand (2000) and Bertrand & Fleury (2008).

2 Segmentation on the spectral density estimation of some long memory processes

Our model

Let X be a Gaussian centered process with stationary increments, it is known, see Cramér & Leadbetter (1967), that this process has the harmonizable representation $X(t) = \int_{I\!R} \left(e^{it\xi} - 1\right) f^{1/2}(\xi) dW(\xi)$ for all $t \in I\!R$ where $W(dx)$ is a complex Brownian measure such that $X(t)$ is a real number for all $t \in I\!R$. This process has long memory, but its wavelet coefficient $d_\psi(a,b)$ is a short memory Gaussian process where, for a scale and a shift $(a,b) \in I\!R_+^* \times I\!R$ and a wavelet ψ with a compact time support $[L_1, L_2]$, one has defined

$$d_\psi(a,b) \quad := \quad a^{-1/2} \int_{I\!R} \psi\left(\frac{t-b}{a}\right) X(t)\, dt \tag{1}$$

Moreover, for any fixed scale a, $b \mapsto d_\psi(a,b)$ is a stationary, centered, Gaussian process with variance $\mathcal{I}_\psi(a) := \int_{I\!R} |\widehat{\psi}(x)|^2 f(x/a)\, dx$, see Bardet & Bertrand (2007). Next, we assume that the signal is a Gaussian process, centered, with locally stationary increments given by the representation formula

$$X(t) = \int_{I\!R} \left(e^{it\xi} - 1\right) f^{1/2}(t,\xi)\, dW(\xi), \qquad \text{for all} \quad t \in I\!R, \tag{2}$$

where $\xi \mapsto f(t,\xi)$ is an even and positive function, called spectral density piecewise constant, i.e., there exists a partition $\tau_1 < \tau_2 < \cdots < \tau_K$ and Hurst parameters $H = (H_0, H_1, \ldots, H_K)$ such that $f(t,\xi) = f_k(\xi) = C(H_k)|\xi|^{-2H_k-1}$ for $t \in [\tau_k, \tau_{k+1}[$ and $C(H_k) = \pi^{-1} H_k \Gamma(2H_k)\sin(\pi H_k)$. Thus the series $\log|d_\psi(a,b_i)|^2$ where $d_\psi(a,b)$ is defined by (1) and $b_i = i \in I\!N$ has a piecewise constant mean μ_i and a finite known variance, more precisely, one has $\log|d_\psi(a,b_i)|^2 = \mu_i + \zeta_i$ where ζ_i are weakly dependent r.v. of law $\ln|U|^2$ with $U \sim \mathcal{N}(0,1)$ and $\mu_i = \ln \int_{I\!R} |\widehat{\psi}(x)|^2 f_k(x/a)\, dx$ if $b_i \in [\tau_k - aL_1, \tau_{k+1} - aL_2]$.

Numerical simulation

First, for $T = 10^5$, we have simulated one realization of process $(X(t))_{t\in[0,T]}$ with five change points $\tau = \{12500, 25496, 43045, 70083, 82040\}$ and Hurst parameters $H = (0.55, 0.67, 0.53, 0.61, 0.7, 0.57)$. Let us stress that, after having changed the scale in order to obtain Hurst index belonging to $(0,1)$, the configuration of change points and means is the same as in Section 1.

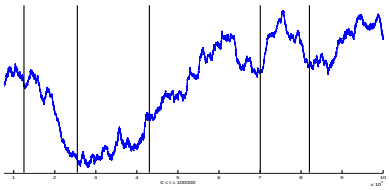

Figure 5: *One replication of the process* $(X(t))_{t\in[0,T]}$

Next, for a frequency $1/a = 0.2$Hz and by using the Daubechies wavelet of order 6, we have calculated the wavelet coefficients $(d_\psi(a,b_0), \ldots, d_\psi(a,b_N))$ with $b_k = k$. Figure 5 below displays the sequence $(Y_0, \ldots, Y_N)$ where $Y_k = \log|d_\psi(a,b_k)|^2$. Then, we have calibrated the Filtered derivative algorithm with $A = 500$ and $\alpha_{critic} = 10^{-11}$. We observe that the detected change points perfectly fit the theoretical configuration of changes.

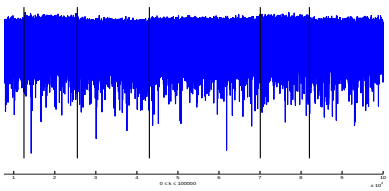

Figure 6: *Segmentation of the sequence* $(Y_0, \ldots, Y_N)$

Note that we can not use penalized least square criterion due to the size of data, indeed PLSC would have need 80GB which is almost 150 times our computer memory capacity.

3 Application to real data

Recent measurement methods allow us to access to electrocardiograms (ECG) for healthy people over a long period of time: marathon runners, daily (24 hours) records, etc. These large datasets allow us to characterize the variation of the heartbeat rate in the parasympathetic frequency band $(0.15\,Hz, 0.5\,Hz)$. According to the recommendations of the Task Force of Cardiologists [14], this frequency band corresponds to the parasympathetic system of control of the heartbeat. Moreover, the spectral density of the heart beat time series follows a power law, thus after having substract its mean, this series can be modelized by (2). Figure 7 provides an example of interbeat time series record on an healthy subject during 24 hours. We have calculated its wavelet coefficients and used the Filtered derivative algorithm with $A = 500$ and $\alpha_{critic} = 10^{-11}$. We obtain the following segmentation: $\tau = \{14435, 21903, 28003, 31984, 33377, 37274, 40470, 42306, 73153\}$

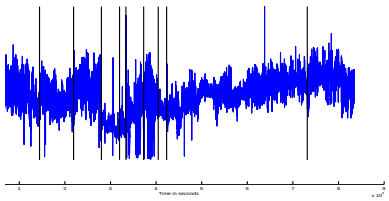

Figure 7: *Segmentation of the heart interbeat for healthy subjects during a period of 24 hours*

In future works, we will investigate sequential detection of change points of the Hurst index in connection with the cardiac behavior of sick subject.

References

[1] Antoch, J. & Hušková, M. (1994), *Procedures for the detection of multiple changes in series of independent observations*, in Proc. of the 5th Prague Symposium on Asymptotic Statistics, Physica Verlag, 3–20.

[2] Bai J. and Perron P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* 66, p. 47-78.

[3] Bardet, J.M. & Bertrand, P.R. (2007). Identification of the multiscale fractional Brownian motion with biomechanical applications. *J. of Time Series. Anal.* **28**, 1–52.

[4] M. Basseville & I. Nikiforov, (1993). *Detection of Abrupt Changes: Theory and Application.* Prentice Hall, Englewood Cliffs, NJ.

[5] Bertrand, P. R. (2000). A local method for estimating change points: the hat-function. *Statistics* **34**, 215–235.

[6] Bertrand, P. R. and Fleury, G. (2008). Detecting Small Shift on the Mean by Finite Moving Average, *International Journal of Statistics and Management System*, vol. **3**, No 1-2, pp.56-73

[7] Birgé L. & Massart, P. (2007). Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields* **138**, 33–73.

[8] Cramér, H. & Leadbetter, M. R. (1967). *Stationary and related stochastic processes. Sample function properties and their applications.* Wiley and Sons.

[9] Csörgo, M.; Horváth, L.; Limit Theorem in Change-Point Analysis, Wiley, (1997).

[10] Davis, R. A.; Lee, T. C. M.; Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. J. Amer. Statist. Assoc. 101, no. 473, 223–239.

[11] Lavielle, M. & Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *J. of Time Series Anal.* **21**, 33–59.

[12] Lavielle, M. & Teyssière, G. (2006). Detection of multiple change points in multivariate time series. *Lithuanian Math. J.* **46**, 287–306.

[13] Lebarbier, E. (2005). Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing* **85**, 717–736.

[14] Task force of the European Soc. Cardiology and the North American Society of Pacing and Electrophysiology (1996). Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Circulation 93, 1043–1065.