A Bayesian approach to change points detection in time series

Ali Mohammad-Djafari and Olivier Féron
Laboratoire des Signaux et Systèmes,
Unité mixte de recherche 8506 (CNRS-Supélec-UPS)
Supélec, Plateau de Moulon, 3 rue Joliot Curie, 91192 Gif-sur-Yvette, France.

Abstract. Change points detection in time series is an important area of research in statistics, has a long history and has many applications. However, very often change point analysis is only focused on the changes in the mean value of some quantity in a process. In this work we consider time series with discrete point changes which may contain a finite number of changes of probability density functions (pdf). We focus on the case where the data in all segments are modeled by Gaussian probability density functions with different means, variances and correlation lengths. We put a prior law on the change point occurances (Poisson process) as well as on these different parameters (conjugate priors) and give the expression of the posterior probability distributions of these change points. The computations are done by using an appropriate Markov Chain Monte Carlo (MCMC) technique.

The problem as we stated can also be considered as an unsupervised classification and/or segmentation of the time series. This analogy gives us the possibility to propose alternative modeling and computation of change points.

Keywords. Change point analysis, Bayesian classification and segmentation, Time series analysis.

1 Introduction

Figure 1 shows typical change point problems we consider in this work. Note that, very often people consider problems in which there is only one change point [1]. Here we propose to consider more general problems with any number of change points. However, very often the change point analysis problems need online or real time detection algorithms [2, 3, 4, 5], while here, we focus only on off line methods where we assume that we have gathered all the data and we want to analyze it to detect change points who have been occurred during the observation time. Also, even if we consider here change point estimation of 1-D time series, we can extend the proposed method to multivariate data, for example the images where the change point problems become equivalent to segmentation. One more point to position this work is that, very often the models used in change point problems assume to know perfectly the model of the signal in each segment, i.e., a linear or nonlinear regression model [5, 6, 7, 8, 9], while here, we use a probabilistic model for the signals in each segment which
gives probably more generality and applicability when we do not know perfectly those models.

Fig. 1: Change point problems description: In the first row, only mean values of the different segments are different. In the second row, only variances are changed. In the third row only the correlation strengths are changed. In the fifth row, the whole nature shape of their probability distribution have been changed. The last row show the change points $t_n$.

More specifically, we model the time series by a hierarchical Gauss-Markov modeling with hidden variables which are themselves modeled by a Markov model. Though, in each segment which corresponds to a particular value of the hidden variable, the time series is assumed to be modeled by a stationary Gauss-Markov model. However, we chose a simple parametric model defined only with three parameters of mean $\mu$, variance $\sigma^2 = 1/\tau$ and a parameter $\rho$ measuring the local correlation strength of the neighboring samples.

The choice of the hidden variable is also important. We have studied three different modeling: i) change point time instants $t_n$, ii) classification labels $z_n$ or iii) a Bernoulli variable $q_n$ which is always equal to zero except when a change point occurs.

The rest of the paper is organized as follows: In the next section we introduce the notations and fix the objectives of the paper. In section 3 we consider the model with explicit change point times as the hidden variables and propose particular modeling for them and an MCMC algorithm to compute their \textit{a posteriori} probabilities. In sections 4 and 5 we consider the two other aforementioned models. Finally, we show some simulation results and present our conclusions and perspectives.
2 Notations, modeling and classical methods

We note by $\mathbf{x} = [x(t_0), \ldots, x(t_0 + T)]'$ the vector containing the data observed from time $t_0$ to $t_0 + T$. We note by $\mathbf{t} = [t_1, \ldots, t_N]'$ the unknown change points and note $\mathbf{x} = [\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N]'$ where $\mathbf{x}_n = [x(t_n), x(t(n+1)), \ldots, x(t_{n+1})]'$, $n = 0, \ldots, N$ represent the data samples in each segment. In the following we will have $t_{N+1} = T$.

We model the data $\mathbf{x}_n = [x(t_n), x(t(n+1)), \ldots, x(t_{n+1})]'$, $n = 0, \ldots, N$ in each segment by a Gauss-Markov chain:

\[ p(x(t_n)) = \mathcal{N}(\mu_n, \sigma_n^2) \]
\[ p(x(t_n + l)|x(t_n + l - 1)) = \mathcal{N}(\rho_n x(t_n + l - 1) + (1 - \rho_n)\mu_n, \sigma_n^2(1 - \rho_n^2)), \]
with $l = 1, \ldots, t_{n+1} - t_n + 1 = \text{dim} [\mathbf{x}_n]$.

Then, we have

\[ p(\mathbf{x}_n) = p(x(t_n)) \prod_{l=1}^{t_n} p(x(t_n + l)|x(t_n + l - 1)) \]
\[ \exp \left\{ -\frac{1}{2\sigma_n^2} (x(t_n) - \mu_n)^2 \right\} \exp \left\{ -\frac{1}{2(\sigma_n^2(1 - \rho_n^2))} \sum_{l=1}^{t_n} [x(t_n + l) - \rho_n x(t_n + l - 1) - (1 - \rho_n)\mu_n]^2 \right\} \]
\[ = \mathcal{N}(\mu_n\mathbf{1}, \Sigma_n) \text{ with } \Sigma_n = \sigma_n^2 \text{Toeplitz}([1, \rho_n, \rho_n^2, \ldots, \rho_n^{t_n}]) \] (1)

Noting by $\mathbf{t} = [t_1, \ldots, t_N]$ the vector of the change points and assuming that the samples from any two segments are independent, we can write:

\[ p(\mathbf{x}|\mathbf{t}, \boldsymbol{\theta}, N) = \prod_{n=0}^{N} \mathcal{N}(\mu_n\mathbf{1}, \Sigma_n) \]
\[ = \left( \prod_{n=0}^{N} \Sigma_n^{-1/2} \right) \exp \left\{ -\frac{1}{2} \sum_{n=0}^{N} (\mathbf{x}_n - \mu_n\mathbf{1})\Sigma_n^{-1}(\mathbf{x}_n - \mu_n\mathbf{1}) \right\} \] (3)

where we noted $\boldsymbol{\theta} = \{\mu_n, \sigma_n, \rho_n, n = 0, \ldots, N\}$.

Note that

\[ -\ln p(\mathbf{x}|\mathbf{t}, \boldsymbol{\theta}, N) = \sum_{n=0}^{N} (l_n/2) \ln(2\pi) + \frac{1}{2} \sum_{n=0}^{N} \ln |\Sigma_n| - \frac{1}{2} \sum_{n=0}^{N} (\mathbf{x}_n - \mu_n\mathbf{1})\Sigma_n^{-1}(\mathbf{x}_n - \mu_n\mathbf{1}) \]

and when the data are i.i.d., ($\Sigma_n = \sigma_n I$) this becomes

\[ -\ln p(\mathbf{x}|\mathbf{t}, \boldsymbol{\theta}, N) = (T/2) \ln(2\pi) + \sum_{n=0}^{N} (l_n/2) \ln \sigma_n^2 - \sum_{n=0}^{N} \frac{||\mathbf{x}_n - \mu_n\mathbf{1}||^2}{2\sigma_n^2} \] (4)

Then, the inference problems we will be faced are the following:

1. **Learning**: Infer on $\boldsymbol{\theta}$ given a training set $\mathbf{x}$ and $\mathbf{t}$;

2. **Supervised estimation**: Infer on $\mathbf{t}$ given $\mathbf{x}$ and $\boldsymbol{\theta}$;

3. **Unsupervised estimation**: Infer on $\mathbf{t}$, $\boldsymbol{\theta}$ or jointly on $\mathbf{t}$ and $\boldsymbol{\theta}$ given $\mathbf{x}$.
The classical maximum likelihood estimation (MLE) approach for these problems becomes:
- Estimate $\theta$ given $x$ and $t$ by $\hat{\theta} = \arg \max_\theta \{ p(x|t, \theta) \}$
- Estimate $t$ given $x$ and $\theta$ by $\hat{t} = \arg \max_t \{ p(x|t, \theta) \}$
- Estimate $t$ and $\theta$ given $x$ by $(\hat{t}, \hat{\theta}) = \arg \max_{(t, \theta)} \{ p(x|t, \theta) \}$
- Estimate $t$ given $x$ by $\hat{t} = \arg \max_t \{ p(x|t) \}$ with $p(x|t) = \int p(x|t, \theta) d\theta$
- Estimate $\theta$ given $x$ by $\hat{\theta} = \arg \max_\theta \{ p(x|\theta) \}$ with $p(x|\theta) = \int p(x|t, \theta) dt$

However, we must be careful to check the boundedness of the likelihood function before using any optimization algorithm. The optimization with respect to $\theta$ when $t$ is known can be done easily, but the optimization with respect to $t$ is very hard and computationally costly.

3 Bayesian estimation of the change point time instants

In Bayesian approach, one assigns prior probability laws on both $t$ and $\theta$ and use the posterior probability law $p(t, \theta|x)$ as a tool for doing any inference. Choosing a prior pdf for $t$ is also usual in classical approach. A simple model is the following:

$$t_n = t_{n-1} + \epsilon_n \quad \text{with} \quad \epsilon_n \sim P(\lambda),$$

(6)

where $\epsilon_n$ are assumed iid and $\lambda$ is the a priori mean value of time intervals $(t_n - t_{n-1})$. If $N$ is the number of change point we can take $\lambda = \frac{T}{N+1}$. With this modeling we have:

$$p(t|\lambda) = \prod_{n=1}^{N+1} P(t_n - t_{n-1}|\lambda) = \prod_{n=1}^{N+1} e^{-\lambda \frac{(t_n - t_{n-1})}{(t_n - t_{n-1})!}}$$

$$\ln p(t|\lambda) = -(N+1)\lambda + \ln(\lambda) \sum_{n=1}^{N+1} (t_n - t_{n-1}) - \sum_{n=1}^{N+1} \ln((t_n - t_{n-1})!))$$

(7)

With this prior selection, we have

$$p(x, t|\theta, N) = p(x|t, \theta, N) p(t|\lambda, N)$$

(8)

and

$$p(t|x, \theta, N) \propto p(x|t, \theta, N) p(t|\lambda, N)$$

(9)

In Bayesian approach, one goes one step further with assigning prior probability laws to the hyperparameters $\theta$, i.e., $p(\theta)$ and then one writes the joint a posteriori:

$$p(t, \theta|x, \lambda, N) \propto p(x|t, \theta, N) p(t|\lambda, N) p(\theta|N)$$

(10)

where here we noted $\theta = \{\mu_n, \sigma_n^2, \rho_n, n = 1, \ldots, N\}$. 

A classical choice for \( p(\theta) \) is the conjugate priors which, in general, results in
- Gaussian pdfs \( p(\mu_n) = \mathcal{N}(\mu_0, \sigma^2_0) \) for position parameters \( \mu_n \),
- Inverse Gamma (IG) pdfs \( p(\sigma^2_n) = \mathcal{IG}(\alpha_0, \beta_0) \) for variances \( \sigma^2_n \) and
- Inverse Wishart (IW) pdfs \( p(\Sigma_n) = \mathcal{IW}(\Lambda_0, \beta_0) \) for the covariance matrices \( \Sigma_n \).

When the likelihood \( p(x|\mathbf{t}, \theta) \) and the priors \( p(t|\theta) \) and \( p(\theta) \) are chosen and
the expression of the posterior probability law \( p(t, \theta|x) \) obtained, one can do
any inference on the on the unknown parameters of the problem \( t \) and \( \theta \) separately
or jointly. Two main approaches for the estimation are:
- the methods which are based on the the computation of the modes (Maximum
  a posteriori MAP) of the different posterior probability laws, and
- the methods which are based on the the computation of the means of the
  different posterior probability laws.

### 3.1 MAP optimization based methods

The methods which are based on the the computation of the modes of the
different posterior probability laws result in the following optimization problems:

1. **Learning:** Infer on \( \theta \) given a training set \( x \) and \( t \);
   \[
   \hat{\theta} = \arg \max_{\theta} \left\{ p(\theta|x, t) \right\} = \arg \max_{\theta} \left\{ p(x|t, \theta) p(\theta) \right\} \tag{11}
   \]

2. **Supervised estimation:** Infer on \( t \) given \( x \) and \( \theta \):
   \[
   \hat{t} = \arg \max_{t} \left\{ p(t|x, \theta) \right\} = \arg \max_{t} \left\{ p(x|t, \theta) p(t|\theta) \right\} \tag{12}
   \]

3. **Unsupervised estimation:** Infer on \( t, \theta \) or jointly on \( t \) and \( \theta \) given \( x \).
   \[
   (\hat{t}, \hat{\theta}) = \arg \max_{(t, \theta)} \left\{ p(t, \theta|x) \right\} = \arg \max_{(t, \theta)} \left\{ p(x|t, \theta) p(t|\theta) p(\theta) \right\} \tag{13}
   \]

We can also first focus on the estimation of \( \theta \) by integrating out \( t \) from
\( p(t, \theta|x) \) to obtain
\[
p(\theta|x) = \int p(t, \theta|x) \, dt = \int p(x|t, \theta) p(t|\theta) \, dt p(\theta) = p(x|\theta) p(\theta) \tag{14}
\]
and then estimate \( \theta \) by
\[
\hat{\theta} = \arg \max_{\theta} \left\{ p(\theta|x) \right\} = \arg \max_{(t, \theta)} \left\{ p(x|t, \theta) p(t|\theta) p(\theta) \right\} \tag{15}
\]
and then use it as in the supervised estimation case.
\[
\hat{t} = \arg \max_{t} \left\{ p(t|x, \hat{\theta}) \right\} = \arg \max_{t} \left\{ p(x|t, \hat{\theta}) p(t|\hat{\theta}) \right\}
\]
## 3.2 Posterior mean and MCMC methods

The methods which are based on the computation of the posterior means result in integration computation. Indeed, rarely these integrations can be done analytically, and often, they are done by the MCMC methods.

Here, we propose the following Gibbs sampling MCMC algorithm:

Iterate until convergence

- Sample \( t \) using \( p(t|x, \theta, N) \)
- Sample \( \theta_n \):
  - \( \mu_n \) using \( p(\mu_n|x, t, N) \)
  - \( \sigma^2_{\theta_n} \) using \( p(\sigma^2_{\theta_n}|x, t, N) \)
  - \( \rho_n \) using \( p(\rho_n|x, t, N) \)

In the following, we give some details on the expressions of these posterior laws and the sampling algorithms which we implemented.

- First, note that, thanks to the conjugacy, we have:
  \[
  p(\mu_n|x, t) = \mathcal{N}(\hat{\mu}_n, \hat{\sigma}^2_n) \quad \text{with} \quad \hat{\mu}_n = \hat{\sigma}^2_n \left[ \frac{\alpha_0}{\sigma^2_n} + 1' \Sigma_n^{-1} x_n \right] \\
  \hat{\sigma}^2_n = \left( 1' \Sigma_n^{-1} 1 + \frac{1}{\sigma^2_n} \right)^{-1}
  \]
  \[
  p(\sigma^2_{\theta_n}|x, t) = IG(\alpha_n, \beta_n) \quad \text{with} \quad \alpha_n = \alpha_0 + \frac{n}{2} \\
  \beta_n = \beta_0 + \frac{1}{2} (x_n - \mu_n)' R_n^{-1} (x_n - \mu_n) 
  \]
  where \( R_n = \text{Toeplitz}(1, \rho_n, \rho^2_n, \cdots, \rho^2_n) \). Thus, these posterior laws are classical ones and generating samples from them is quite simple and easy.

- \( p(\rho_n|x, t) \) is not a classical law, but we can write its expression which is given by:

  \[
  p(\rho_n|x, t, N) = \prod_{n=0}^{N} p(\rho_n|x_n, t, N) \\
  \propto \left( \frac{1}{\sigma^2_n (1-\rho^2_n)} \right)^{\frac{1}{2}} \exp \left\{ - \frac{1}{2 \sigma^2_n (1-\rho^2_n)} \sum_{i=1}^{n} (x_{n_i} - \mu_n)' R_n^{-1} (x_{n_i} - \mu_n) \right\}
  \]
  (16)

As we can see, this is not a classical probability density and we do not have a simple way to generate samples for this density. The solution we propose is to use, in this step, a Hastings-Metropolis algorithm for sampling this density.

As an instrumental density we propose to use a Gaussian approximation of the posterior density, i.e., we estimate the mean \( m_{\rho_n} \), and the variance \( \sigma^2_{\rho_n} \) of \( p(\rho_n|x, t, N) \) and we use a Gaussian law \( \mathcal{N}(m_{\rho_n}, \sigma^2_{\rho_n}) \) to obtain a sample. This sample is accepted or rejected following \( p(\rho_n|x, t, N) \). In practice we compute \( m_{\rho_n} \) and \( \sigma^2_{\rho_n} \) calculating by approximation of their definition:

\[
 m_{\rho_n} \quad \rightarrow \quad \int_{0}^{1} \rho_n \ p(\rho_n|x, t, N) \\
 \sigma^2_{\rho_n} \quad \rightarrow \quad \int_{0}^{1} \rho_n^2 \ p(\rho_n|x, t, N) - m^2_{\rho_n}
\]
To generate samples from $p(t|x, \theta, N)$ can be obtained by a method based on recursion on the change points. An approximation of this method is possible to obtain an algorithm whose computational cost is linear in the number of observations [10].

The main idea behind this algorithm is to compute the conditional probability laws $p(t|t_{j-1}, x)$ which permits to generate recursively samples for $t_j$. To give main relations to achieve this, first we note by $x_{t,s} = [x(t), x(t + 1), \ldots, x(s)]$ and define the the following probabilities:

\[
R(t, s) = p(x_{t,s}|t, s \text{ in the same segment}) \\
Q(t) = P(x_{t,s}| \text{change point at } t-1), \quad Q(1) = P(x)
\]

Let also note $g(t_j - t_{j-1})$ the a priori density of the interval between two change points, and $G(t)$ the associated distribution function. Then one can show that the posterior distribution of $t_j$ given $t_{j-1}$ is

\[
p(t_j|t_{j-1}, x) = \frac{R(t_{j-1}, t_j)Q(t_j + 1)g(t_j - t_{j-1})}{Q(t_{j-1})}
\tag{17}
\]

and the posterior distribution of no further change point is given by

\[
p(t_j = T|t_{j-1}, x) = \frac{p(t_{j-1}, T)(1 - G(T - t_{j-1} - 1))}{Q(t_{j-1})}
\tag{18}
\]

Thus, we have all the necessary expressions for generating samples $(t, \theta)^{(1)}, (t, \theta)^{(2)}, \ldots$, from the joint posterior law $p(t, \theta|x)$. Note however that, we need to generate a great number of those samples to achieve the convergence of the Markov chain. When this convergence is achieved, we can use those final samples to compute any statistics such as the mean, the median or take as the final output the most frequently generated samples. The main advantage however is to use these samples to generate their histograms which are good representative of their marginal posterior probabilities.

4 Other formulations

Other formulation can also exist. We introduce two sets of indicator variables $z = [z(t_0), \ldots, z(t_0 + T)]'$ and $q = [q(t_0), \ldots, q(t_0 + T)]'$

where

\[
q(t) = \begin{cases} 
1 & \text{if } z(t) \neq z(t - 1) \\
0 & \text{elsewhere}
\end{cases} = \begin{cases} 
1 & \text{if } t = t_n, n = 0, \ldots, N \\
0 & \text{elsewhere}
\end{cases}
\tag{19}
\]

Thus, $q$ can be modeled by a Bernoulli process

\[
P(Q = q) = \lambda^{\sum_i q_i}(1 - \lambda)^{N - \sum_i q_i} = \lambda^{\sum_i q_i}(1 - \lambda)^{N - \sum_i q_i}
\]
and \( z \) can be modeled by a Markov chain, i.e., \( \{z(t), t = 1, \cdots , T\} \) forms a Markov chain:

\[
P(z(t) = k | z(t - 1) = l) = p_{kl}, \quad k = 1, \cdots , K, \quad P(z(t) = k) = p_k, \quad \text{with} \quad \sum_k p_{kl} = 1. \tag{20}
\]

In the multivariate case, or more precisely in bivariate case (image processing), \( q \) may represent the contours and \( z \) the labels for the regions in the image. Then, we may also give a Markov model for them. For example, if we note by

\[
q_r
\]

for \( r \in S \) the position of a pixel, \( S \) the set of pixels positions and by \( V(r) \) the set of pixels in the neighborhood of the pixel position \( r \), we may use an Ising model for \( q \)

\[
P(Q = q) \propto \exp \left\{ -\rho \sum_{r \in S} \sum_{s \in V(r)} \delta(z(r) - z(s)) \right\} \tag{21}
\]

or a Potts model for \( z \):

\[
P(z) \propto \exp \left\{ -\rho \sum_{r \in S} \sum_{s \in V(r)} \delta(z(r) - z(s)) \right\}. \tag{22}
\]

Other more complex modelings are also possible.

With these auxiliary variables, we can write

\[
p(x | z, \theta) = \sum_{n=1}^{N} P(z_j = n)N(\mu_n,1,\Sigma_n) = \sum_{n=1}^{N} p_k N(\mu_n,1,\Sigma_n) \tag{23}
\]

if we choose \( K = N \).

Here, \( \theta = \{ N, \{ \mu_n, \sigma_n, p_n, \text{ } n = 1, \cdots, N \}, \{ p_{kl}, k, l = 1, \cdots, N \} \} \) and the model is a mixture of Gaussians.

We can again assign appropriate prior law on \( \theta \) and give the expression of \( p(z, \theta | x) \) and do any inference on \( z, \theta \).

Finally, we can also use \( q \) as the auxiliary variable and write

\[
p(x | q, \theta) = (2\pi)^{-N/2} \left( \prod_{n=1}^{N} \frac{1}{\sigma_n} \right) \exp \left\{ -\frac{1}{2\sigma_n^2} \sum_{n=1}^{N} (x(t_n) - \mu_n)^2 \right\} \\
+(2\pi)^{-T-N/2} \left( \prod_{n=1}^{N} \frac{1}{\sigma_n} \right)^{(l_n-1)} \exp \left\{ -\frac{1}{2\sigma_n^2} \sum_{j=1}^{T} (1 - q_j) (x_j - x_{j-1})^2 \right\} \\
= (2\pi)^{-T/2} \left( \prod_{n=1}^{N} \frac{1}{\sigma_n} \right)^{(l_n)} \exp \left\{ -\frac{1}{2\sigma_n^2} \sum_{j=1}^{T} \left[ (1 - q_j) (x_j - x_{j-1})^2 + q_j (x_j - \mu_n) \right] \right\} \tag{24}
\]

and again assign appropriate prior law on \( \theta \) and give the expression of \( p(q, \theta | x) \) and do any inference on \( q, \theta \). We are still working on using these auxiliary hidden variables particularly for applications in data fusion in image processing and we will report on these works very soon.
5 Simulation results

To test the feasibility and to measure the performances of the proposed algorithms, we generated a few simple cases corresponding to only changes of one of the three parameters $\mu_n$, $\sigma^2_n$ and $\rho_n$.

In each case we present the data, the histogram of the a posteriori samples of $t$ during the first and the last iterations of the MCMC algorithm. For each case we also give the value of the parameters used to simulate the data, the estimated values when the change points are known and the estimated values by the proposed method.

5.1 Change of the means

As we can see in Fig. 2, we obtain precise results on the position of the change points. In the case of change of means, the algorithm is very fast to converge to the good solution. In fact it needs only few iterations (about 5). The main cause of this results is the importance of the means in the likelihood $p(x|t, \Theta, N)$.

We can also see in Table 1 that the estimations of the means are very precise, particularly when the size of the segment is long.

![Fig. 2: Change in the means. up to down: simulated data, histogram in the 50th iteration, histogram in the first iteration, real position of the change points.](image)

| $m$ | $\hat{m}_{x,t}$ | $\hat{\sigma}^2|x, t$ | $\hat{m}_{x}$ | $\hat{\sigma}^2|x$ |
|-----|-----------------|----------------------|----------------|-------------------|
| 1.5 | 1.4966          | 0.0015               | 1.4969         | 0.0013            |
| 1.7 | 1.7084          | 0.0017               | 1.7013         | 0.0038            |
| 1.5 | 1.4912          | 0.0020               | 1.5015         | 0.0045            |
| 1.7 | 1.6940          | 0.0014               | 1.6929         | 0.0016            |
| 1.9 | 1.9012          | 0.0015               | 1.8915         | 0.0039            |
5.2 Change in the variances

We can see in Fig. 3 that we have again good results on the position of the change points. However, for little difference of variances, the algorithm give an uncertainty on the exact position of the change point. This can be justified by the fact that the simulated data give itself this uncertainty.

In Table 2 we can see again good estimations on the variances on each segments.

|  0.01 | 0.0083 | 0.0081 |
|  1   | 0.9918 | 0.9508 |
| 0.001 | 0.0007 | 0.0026 |
| 0.1  | 0.0945 | 0.0940 |
| 0.01 | 0.0079 | 0.0107 |

5.3 Change in the correlation coefficient

The results showed in Fig. 4 are worse than in the two first cases. The position of the change points are less precise, and we can see that another change point appears. This affects the estimation of the correlation coefficient in the third segment because the algorithm alternates between two positions of change point. This problem can be justified by the fact that a value of the correlation coefficient near 1 implies locally a change of the mean, which can be considered by the algorithm as a change point. Also this problem appears when the size of the segments are far from the *a priori* size $\lambda$. 
Different correlation coefficient

| $\alpha$ | $\hat{\alpha}|x$ |
|---------|----------------|
| 0       | 0.0988         |
| 0.9     | 0.7875         |
| 0.1     | 0.3737         |
| 0.8     | 0.8071         |
| 0.2     | 0.1710         |

5.4 Influence of the prior law

In this section we study the influence of the \textit{a priori} on $\lambda$, i.e., the size of the segments. In the following we fix the number of change points as before and we change the \textit{a priori} size of the segments by $\lambda_0 = \frac{1}{2}$ and $\lambda_1 = 2\lambda$. We apply then our algorithm on the change of the correlation coefficient.
In figure Fig. 5, we can see that the algorithm has detected other change points, forming segments whose size is near $\lambda_0$. This result shows the importance
of the \textit{a priori} when the data are not enough significant. We can also see this conclusion in Fig. 6 where only three change points are detected, forming segments whose size is again near $\lambda_1$. We can also remark that fixing \textit{a priori} a size $\lambda$ comes down to fix the number of change points. Our algorithm give then good results for instance if we have a good \textit{a priori} on the number of change points.

6 Conclusions

In this paper, first we presented a Bayesian approach for estimating change points in time series. The main advantage of this approach is to give, at each time instant $t$, the probability that a change point has been occurred at that point. Then, based on posterior probabilities, we can not only give the time instants with highest probabilities, but also give indications on the amount of those changes via the values of the estimated parameters. In the second part, we focused on a piecewise Gaussian model and studied the changes in the means, in the variances and in the correlation coefficients of the different segments. As a conclusion, we could show that the detection of change points due to changes in the mean is easier than those due to changes in variances or changes in correlation coefficient. In this work, first we assumed to know the number $N$ of change points, but the proposed Bayesian approach can be extended to estimate this number too. We are investigating the estimation of the number of change points in the same framework. We also studied the role of the \textit{a priori} parameter $\lambda$ on the results. Finally, we showed that other modeling using other hidden variables than change point time instants are also possible and are under investigation. We are also investigating the extension of this work to image processing (2D signals) where the change points are contours.

References


