Supplementary Notes: Segment Parameter Labelling in MCMC Change Detection

Alireza Ahrabian

Abstract

This work addresses the problem of segmentation in time series data with respect to a statistical parameter of interest in Bayesian models. It is common to assume that the parameters are distinct within each segment. As such, many Bayesian change point detection models do not exploit the segment parameter patterns, which can improve performance. This work proposes a Bayesian change point detection algorithm that makes use of repetition in segment parameters, by introducing segment class labels that utilise a Dirichlet process prior.

Index Terms— Change Detection, Markov Chain Monte Carlo, Dirichlet Process, Autoregressive Models.

I. INTRODUCTION

Change detection algorithms are an important tool in the exploratory analysis of real world data. Namely, such algorithms seek to partition time series data with respect to a statistical parameter of interest, where examples include, the mean, variance and autoregressive model weights, to name but a few. Change point detection algorithms have found applications in fields ranging from financial to bio-medical data analysis.

Accordingly, many approaches have been proposed to address the problem of segmenting time series data. Namely, the work in [1] proposed an online method, based on the likelihood ratio test that both detects and estimates a change in the statistical parameter of interest. More recently, the work in [2] proposed a computationally efficient algorithm that segments data by minimising a cost function using dynamic programming, while [3] utilises a classification algorithm (kernel-SVM) in order to estimate the change point locations. Bayesian approaches to time series segmentation have also proven to be useful. In particular, by deriving the posterior distribution of the parameters interest (that includes, the number of segments, as well as, the change point locations), one can then use suitable methodologies (for evaluating the posterior distribution) in order to both estimate the change point locations. Bayesian approaches to time series segmentation have also proven to be useful. In particular, by deriving the posterior distribution of the parameters interest (that includes, the number of segments, as well as, the change point locations), one can then use suitable methodologies (for evaluating the posterior distribution) in order to both infer as well as predict. Examples of such work includes, a fully hierarchical Bayesian model proposed in [4], that utilised a Markov Chain Monte Carlo (MCMC) sampler in order to evaluate the posterior distribution of the target parameters. While an exact (that is, avoiding the use MCMC sampling algorithms) segmentation algorithm was developed in [5], by evaluating the posterior distribution using an recursive algorithm.

The change point algorithms mentioned in the previous section, assume that statistical parameters of interest from different segments are distinct and thus independent. However, many real world processes can often be modeled by parameters generated from a fixed number of states, where such parameters can be re-assigned more than once (parameter repetition). In particular, hidden Markov models (HMM) and their extensions [6], assign to each data point a state label (that evolves according to Markov chain) corresponding to a set of parameters that govern the emission probability of generating the data point. While, mixture models [7][8] and their extensions (e.g. Dirichlet processes [7]) assume that each data point is generated from a probability distribution; with the parameters of the distribution belong to state drawn from a discrete distribution (that captures the clustering of the data points, with respect to the parameter of interest). However, it should be noted that such methods assign a parameter belonging to a particular state to each time point, and not to the parameters corresponding to a given segment.

This work is based on the method outlined in [9] that incorporates segment parameter repetition in the estimation process of a change point detection algorithm, when estimating changes in autoregressive processes. Namely, the work proposed to extend the Bayesian change point detection algorithm proposed in [4], by incorporating a parameter class variable that utilises a Dirichlet process prior for identifying the number of distinct segment parameters. By including the parameter class variable, segment parameter repetition is captured during the estimation process of the transition times (change point locations), resulting in more robust segmentation.

II. BACKGROUND

A. MCMC Change Point Detection

In this work, we consider the change point detection algorithm proposed in [4]. That is, given a set of transition times \( \tau_K = [\tau_1, ..., \tau_K] \) where \( \tau_0 = 1 \) and \( \tau_K + 1 = N \), that partition a data set \( x \) into \( K + 1 \) segments, where for each segment (consisting of data points between the time indices \( \tau_i + 1 \leq \tau \leq \tau_{i+1} \)) there exists the following functional relationship between the data points \( x_{\tau_i+1:\tau_{i+1}} \) and the statistical parameter \( \phi_i \in \mathbb{R}^D \), that is

\[
x_{\tau_i+1:\tau_{i+1}} = f_d(x_{\tau_i+1:\tau_{i+1}}, \phi_i) + n_{\tau_i+1:\tau_{i+1}}
\]

(1)
for segments indexed by \( i = \{0, \ldots, K \} \), where \( n_{\tau_{i-1}:\tau_i} \) is a set of i.i.d. Gaussian noise samples with zero mean and variance \( \sigma_i^2 \). In particular, an example of the functional relationship \( f_d(x_{\tau_i+1:\tau_{i+1}}, \phi_i) \) is given by the autoregressive model of order \( D-1 \), that is

\[
f_d(x_{\tau_i+1:\tau_{i+1}}, \phi_i) = \begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
1 & x_{\tau_{i+1}} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{\tau_{i-1}} & x_{\tau_{i-2}} & \cdots & x_{\tau_{i-\theta_i+1}}
\end{bmatrix} \begin{bmatrix}
\phi_{\theta_i}^0 \\
\phi_{\theta_i}^1 \\
\vdots \\
\phi_{\theta_i}^{D-1}
\end{bmatrix}
\]

where \( \phi_i = [\phi_0^i, \ldots, \phi_{D-1}^i]^T \). Accordingly, one can define the posterior distribution for the target parameters of interest; namely, the number of transition times \( K \) and set of transition time points \( \tau_K \), that is

\[
p(K, \tau_K | x) \propto \int p(x|\Phi, K, \tau_K) p(\Phi, K, \tau_K)\,d\Phi
\]

(2)

where \( \Phi = [\phi_0, \ldots, \phi_K] \) corresponds to the vector of segment parameters that is treated as a nuisance parameter and thus integrated out of the posterior distribution. Finally, it should be noted that the likelihood function in the posterior distribution (2) assumes that the parameters \( \phi_i \) are distinct for each segment \( i \), that is

\[
p(x|\Phi, K, \tau_K) = \prod_{i=0}^{K} p(x_{\tau_i+1:\tau_{i+1}} | \phi_i)
\]

(3)

B. Dirichlet Process Mixture Model

Consider a set of \( N \) exchangeable data points \( x \), such that probability distribution of the data points \( p(x) \) can be represented by a set of \( V \) class distributions, that is

\[
p(x) = \sum_{v=1}^{V} \pi_v f(x|\theta_v)
\]

(4)

where \( \pi_v \) is the mixing coefficient and \( \theta_v \) corresponds to the class parameter/s of the probability distribution \( f(.) \). The Dirichlet process mixture model (DPMM) can be seen as the limiting case of the mixture model (MM) specified in (4). This can be seen by first considering the re-formulation of the mixture model shown in (4), by introducing the class indicator random variable \( c_i \) for the \( i^{th} \) data point

\[
x_i|c_i, \theta \sim f(x_i|\theta_{c_i})
\]

\( c_i | \pi \sim \text{Discrete}(\pi_1, \ldots, \pi_V) \)

\( \theta_v \sim G_0 \)

\( \pi | \alpha \sim \text{Dir}(\alpha/V, \ldots, \alpha/V) \)

(5)

where \( \theta = [\theta_1, \ldots, \theta_V] \), \( \pi = [\pi_1, \ldots, \pi_V] \) and \( G_0 \) denotes the prior distribution on the parameters \( \theta_v \). The probability of selecting a given class \( c_i \) is determined by the mixing coefficients \( \pi \). In particular, the the joint distribution of the class indicator random variables is given by

\[
p(c_1, \ldots, c_N | \pi) = \prod_{v=1}^{V} \pi_v^{n_v}
\]

(6)

where \( n_v \) corresponds to the number of data points assigned to class \( v \). The distribution of the the \( i^{th} \) class indicator variable given all other class variables \( c_{-i} \) that excludes the \( c_i \), is given by the following

\[
p(c_i = v | c_{-i}, \alpha) = \frac{\int p(c_i = v, c_{-i} | \pi) p(\pi | \alpha) \, d\pi}{\int p(c_{-i} | \pi) p(\pi | \alpha) \, d\pi}
\]

(3)

\[
= \frac{n_{-i,v} + \alpha/V}{N - 1 + \alpha}
\]

where \( n_{-i,v} \) is the number of data points (excluding \( x_i \)) assigned to class \( v \) and \( p(\pi | \alpha) \) is a symmetric Dirichlet distribution with parameter \( \alpha/V \).

Taking the number of classes \( V \to \infty \) and assuming that there exists a finite number of represented classes \( V' \), such that the number of data points assigned to each class is greater than zero; accordingly the conditional probability in (3) for represented classes is given by

\[
p(c_i = v | c_{-i}, \alpha) = \frac{n_{-i,v}}{N - 1 + \alpha}
\]

(4)
That is, \( P(c_i = v | c_{-i}, \alpha) \) is the probability of assigning the class variable \( c_i \) to the represented class \( v \). Furthermore, as \( V \rightarrow \infty \) there exists an countably infinite number of classes that excludes the represented classes such that, \( c_i \neq c_l \), for all \( l \neq i \). In order to calculate the probability of assigning \( c_i \) to a new class, consider the following

\[
\sum_{v' = 1}^{\infty} p(c_i = v' | c_{-i}, \alpha) = 1
\]

\[
\sum_{v' = 1}^{V'} p(c_i = v' | c_{-i}, \alpha) + \sum_{v' = V' + 1}^{\infty} p(c_i = v' | c_{-i}, \alpha) = 1
\]

where \( \sum_{v' = 1}^{\infty} p(c_i = v | c_{-i}, \alpha) \) corresponds to the sum of the probabilities of the represented classes (that is, there exists a data point assigned to the class) and \( \sum_{v' = V' + 1}^{\infty} p(c_i = v | c_{-i}, \alpha) \) the sum of the probabilities of all other classes. Accordingly the probability of assigning \( c_i \) to a new class is given by the following

\[
p(c_i \neq c_l \text{ for all } i \neq l | c_{-i}) = \sum_{v' = V' + 1}^{\infty} p(c_i = v' | c_{-i}, \alpha)
\]

\[
= 1 - \sum_{v' = 1}^{V'} p(c_i = v' | c_{-i}, \alpha)
\]

\[
= \frac{\alpha}{N + \alpha}
\]

The conditional posterior distribution of the class variable \( c_i \), given \( \theta \), can then be determined as follows

\[
p(c_i = v | c_{-i}, x_i, \theta) \propto \frac{n_{i,v}}{N + \alpha} L(x_i | \theta_v)
\]

for a class where \( n_{i,v} > 0 \) and corresponds to the likelihood \( L(x_i | \theta_v) \). The conditional posterior probability for assigning a data point to a new class is given by

\[
p(c_i \neq c_l \text{ for all } i \neq l | c_{-i}, x_i) \propto \frac{\alpha}{N + \alpha} \int L(x_i | \theta) dG_0(\theta)
\]

Finally, it should be noted that the Dirichlet process mixture model can be written as follows

\[
x_i | \theta_i \sim f(x_i | \theta_i)
\]

\[
\theta_i \sim G
\]

\[
G \sim \text{DP}(G_0, \alpha)
\]

where \( G \) is drawn from the Dirichlet process with base measure \( G_0 \).

### III. Proposed Method

In this work we propose to extend the model developed in [4], by assigning a class variable \( c_i \) for the parameters \( \phi_i \) in each segment (for \( i = 0, ..., K \)) thereby capturing the dependencies between segment parameters for improved change point estimation. This performance improvement is achieved by concatenating data points of segments with the same class labels; thereby providing more degrees of freedom when assessing if a change point exists. In the sections that follow, we will provide a detailed description of the proposed method.

#### A. Bayesian Model

In particular, we modify the set of distinct segment parameters, \( \Phi = [\phi_0, ..., \phi_K] \), by introducing the class variable \( c_i \) such that, \( \Phi = [\phi_{c_0}, ..., \phi_{c_K}] \), where the parameters in each segment are effectively then being drawn from a set of class parameters, \( \Phi^c = [\phi^c_1, ..., \phi^c_V] \), where \( V \leq K + 1 \). That is, each segment parameter can be formulated as a multivariate Gaussian mixture model (the Gaussian assumption enables tractable posterior distributions) of the class parameters \( \phi^c_v \):

\[
p(\Phi | \Phi^c, \Sigma^c, \pi) = \sum_{v=1}^{V} \pi_v MN(\Phi | \phi^c_v, \Sigma^c_v)
\]

\(^1\text{The classes that exclude the represented classes.}\)
where $\Sigma^c = [\Sigma^c_1, \ldots, \Sigma^c_K]$ and $\Sigma^c_i \in \mathbb{R}^{D \times D}$. Accordingly, we present a change point estimation model that incorporates the class variable $c_i$ and a Dirichlet process prior on the likelihood on the class probabilities, that is

$$x_{\tau_i + 1: \tau_{i+1}} | \phi^c_i, \sigma^2_i \sim f_j(x_{\tau_i + 1: \tau_{i+1}} | \phi^c_i, \sigma^2_i)$$

$$\sigma^2_i \sim G_0$$

$$\phi^c_i | \phi^c, \Sigma^c_i \sim \mathcal{MN}(\phi^c_i | \phi^c, \Sigma^c_i)$$

$$(\phi^c_i, \Sigma^c_i) \sim G$$

$$G \sim DP(G_0, \alpha)$$

$$x_{\tau_i + 1: \tau_{i+1}} | \tau_K, \phi^c_i \sim f_j(x_{\tau_i + 1: \tau_{i+1}} | \phi^c_i)$$

$$\tau_K, K \sim \text{Bin}(\tau_K, K | \lambda)$$

for $i = 0, \ldots, K$. Furthermore, $\text{Bin}(.)$ corresponds to a Binomial distribution, $G_0$ is the joint prior distribution of both the class parameter $\phi^c_i$ and the variance of the class parameter $\Sigma^c_i$, the prior distribution of the variance of the data points with the same class label is given by $G_0$ and $f_j(.)$ corresponds to the joint Normal distribution.

The posterior distribution of the model in (11), consists of the following parameters, $\{\tau_K, K, \phi^c, \Sigma^c, \sigma^2\}$, where $\sigma^2 = [\sigma^2_1, \ldots, \sigma^2_K]$. Inference of the parameters is carried out by using a Metropolis-Hastings-within-Gibbs sampling scheme. The Gibbs moves are performed on each parameter in the set, $\{c, \phi^c, \Sigma^c, \sigma^2\}$, while a variation of the Metropolis-Hastings algorithm is used to obtain samples for the parameters $\{\tau_K, K\}$. The marginal posterior distribution of parameters $\{c, \phi^c, \Sigma^c, \sigma^2\}$ are given by the following; namely, the marginal posterior distribution of the class parameters $\phi^c_i$, assuming the conjugate prior distribution, $p(\phi^c_i | \lambda, \phi, \sigma^2_i) \sim \mathcal{N}(\lambda, \delta \sigma^2_i \delta^2 I_D)$, where $\lambda \phi \in \mathbb{R}^D$ and $I_D$ is an identity matrix of dimension $D$, is given by

$$p(\phi^c_i | c, \tau_K, K, \Sigma^c, x) \sim \mathcal{MN}(\mu^c_i, \Sigma^c_\phi) \quad v = 1, \ldots, V$$

(12)

where $\mu^c_i = \Sigma^{-1}_\phi \left( n_v \phi^c_i (\Sigma^c_i)^{-1} + \lambda^c_i \phi^c_i \right)$ and $\Sigma^c_\phi = (n_v (\Sigma^c_i)^{-1} + 1)^{-1} \Sigma^c_i - 1$ for $\phi^c_i = \frac{1}{n_v} \sum_{i : c_i = v} \phi^c_i$ where $n_v$ corresponds to the number of segment parameters $\phi^c_i$ assigned to class $v$. The marginal posterior distribution (Inverse Wishart distributed) of the class covariance matrix $\Sigma^c_i$, given the conjugate prior distribution, $p(\Sigma^c_i | \beta, \omega) \sim \mathcal{IW}(\beta, \Omega)$ where $\beta \in \mathbb{R}$ and $\Omega \in \mathbb{R}^{D \times D}$, is shown by the following

$$p(\Sigma^c_i | c, \tau_K, K, \phi^c_i, x) \sim \mathcal{IW}(\alpha^c_i, B^c_i) \quad v = 1, \ldots, V$$

(13)

where $\alpha^c_i = n_v + \beta$ and $B^c_i = \beta \Omega + \sum_{l : c_l = v} (\phi^c_i - \phi^c_l)(\phi^c_l - \phi^c_i)^T$. Furthermore, the posterior distribution of the variance $\sigma^2_v$ for the data points from segments with the same class label $v$, along with the inverse Gamma prior distribution, $p(\sigma^2_v | \nu, \gamma) \sim \mathcal{IG}(\nu, \gamma)$, is given by (for $v = 1, \ldots, V$)

$$p(\sigma^2_v | c, \tau_K, K, x) \sim \mathcal{IG}(\nu^c + d_v, \gamma + Y^T_v P_v Y_v)$$

(14)

where $d_v$ is the number of data points with label $v$, $Y_v$ is the concatenated vector of all data points with the same segment label $v$, $P_v = (I_D - G_e M_v M_v^T)$, with $M_v = (G_v^T G_v + \delta^2 I_D)^{-1}$. Finally, the marginal posterior distribution for the class labels $c_i$ are given by: $p(c_i = v | c_{i-1}, \phi^c_i, \Sigma^c_i)$ for $n_{i-1,v} \geq 0$ (shown in (12) with likelihood $L(\phi^c_i | \phi^c, \Sigma^c_i) \sim \mathcal{MN}(\phi^c_i | \phi^c, \Sigma^c_i)$; while for the posterior probability for a new class $p(c_i \neq c_l$ for all $i \neq l | c_{i-1}, \phi^c_i$) is given by (13) (see [1] for more details).

The conditional posterior distribution of the parameters $\{\tau_K, K\}$ can be obtained by first considering the following marginal posterior distribution, $p(\tau_K, K | \lambda, c, \phi^c, \Sigma^c, x)$ where $\lambda$ is an hyperparameter of the following prior distribution, $p(\lambda) = \lambda^K (1 - \lambda)^{T - K - 1}$. Having selected the appropriate conjugate priors, we can integrate out the nuisance parameters $\{\phi^c, \Sigma^c, \lambda\}$, thereby significantly reducing the number of parameters required to specify the posterior distribution for $\{\tau_K, K\}$. To this end, we first obtain the following posterior distribution (that incorporates the prior distributions of the nuisance parameters)

$$p(\Phi^c, \sigma^2, \tau_K, K | c, x) \propto p(x | \Phi^c, \sigma^2, K, \tau_K, c_K) \times p(K | \lambda) p(\lambda) \prod_{v=1}^{V} p(\phi^c_v | \lambda, \phi, \delta) p(\sigma^2_v | \nu, \gamma)$$

(15)

where $p(\lambda)$ has uniform probability over the interval $[0, 1]$ and the likelihood function is given by

$$p(x | \Phi^c, \sigma^2, K, \tau_K, c_K) = \prod_{v=1}^{V} \prod_{i : c_i = v} p(x_{\tau_i + 1: \tau_{i+1}} | \phi^c_v, \sigma^2_v)$$

Furthermore, $G_v$ is concatenation of input data points such that, $Y_v = G_v \phi_v^c$ is satisfied.
where by combining data points from the same segment class label \( v \), we can potentially obtain more accurate parameter estimation owing to the increased number of number available for estimating \( \{ \tau_K, K \} \). Integration of (15) with respect to the parameters \( \{ \Phi_v, \sigma_v^2, \lambda \} \) results in the following expression for the conditional posterior distribution of the parameters \( \{ \tau_K, K \} \)

\[
p(\tau_K, K | \mathbf{c}_K, \mathbf{x}) \propto \prod_{v=1}^{V} \frac{2\pi}{\Gamma(K + 1)} \Gamma(N - K + 1) \left( \frac{\gamma}{2} \right)^{\frac{1}{2}}_\tau \times \Gamma \left( \frac{d_v + \nu}{2} \right) \pi^{-d_v} [\gamma + Y_v^T P_v Y_v]^{-\frac{d_v + \nu}{2}} |\mathbf{M}_v|^{-\frac{1}{2}} \tag{16}
\]

Finally, we note that there are some challenges from drawing samples from (16) due to the dependence on \( \mathbf{c}_K \) that we have addressed in the next section.

### B. Gibbs Sampling

A summary of the Gibbs sampling scheme for drawing samples for the parameters \( \{ \tau_K, K, \mathbf{c}_K, \Phi, \Sigma, \sigma^2 \} \), is provided in Algorithm 1. Observe that the parameters \( \{ \tau_K, K \} \), are dependent on the segment class variables \( \{ \mathbf{c}_K \} \), and in turn, the segment class variables are dependent on the parameters \( \{ \tau_K, K, \Phi, \Sigma, \sigma^2 \} \). Furthermore, the marginal posterior distributions for both \( \{ \tau_K, K \} \) and \( \{ \mathbf{c}_K \} \) are intractable and therefore require sampling schemes; in particular, a nested Gibbs sampling scheme was used in order to draw samples for the class variables \( \{ \mathbf{c}_K \} \), due to the dependence on the parameters \( \{ \Phi, \Sigma, \sigma^2 \} \) (as shown in Algorithm 1). While, A modification of the Metropolis-Hastings algorithm (having integrated out the nuisance parameters) outlined in [4] was used in order to draw samples from the conditional posterior distribution, \( p(\tau_K, K | \mathbf{c}_K, \mathbf{x}) \); in particular, a variation was developed that incorporates the segment labels \( \mathbf{c}_K \).

Given the \( j^{th} \) samples, \( \{ \tau_K, K \} \), we first select with a certain probability, one the following:

- \( K \to K + 1 \): create a new change point (birth), with probability, \( b \)
- \( K \to K - 1 \): remove an existing change (death) with probability, \( d \)
- \( K \to K \): update of change point positions with probability, \( u \)

where \( b = d = u \) for \( 0 < K < K_{\text{max}} \), and \( b + d + u = 1 \) for \( 0 \leq K \leq K_{\text{max}} \). Furthermore, for \( K = 0, d = 0 \) and \( b = u \), while for \( K = K_{\text{max}} \), \( b = 0 \) and \( d = u \).

A birth move consists of proposing a new transition time \( \tau_{\text{prop}} \), with the following proposal distribution, \( q(\tau_{K+1} | \tau_K) = q(\tau_{\text{prop}} | \tau_K) = \{ \frac{1}{N - K - 2} \} \) for \( \tau_{\text{prop}} \in S_{\text{prop}} \), where \( \tau_{K+1} \) corresponds to the set of change points that includes both \( \tau_{\text{prop}} \) and \( \tau_K \), while \( S_{\text{prop}} \) corresponds to the set of time indices \( \{2, N - 1\} \) excluding the time points \( \tau_K \). Conversely, the proposal distribution for removing \( \tau_{\text{prop}} \) from \( \tau_{K+1} \), is given by \( q(\tau_K | \tau_{K+1}) = \{ \frac{1}{K + 1} \} \) for \( \tau_{\text{prop}} \in \tau_{K+1} \). Accordingly, the proposed transition time \( \tau_{\text{prop}} \) is accepted with the following probability, \( \alpha_{\text{birth}} = \min\{1, r_{\text{birth}}\} \)

\[
r_{\text{birth}} = \frac{p(\tau_{K+1}, K + 1 | \mathbf{c}_{K+1}, \mathbf{x}) \times q(\tau_K | \tau_{K+1}) \times q(K | K + 1)}{p(\tau_K, K | \mathbf{c}_K, \mathbf{x})}
\]

with \( q(K + 1 | K) = b \) and \( q(K | K + 1) = d \), corresponding to the proposal distributions for the unit increment and decrement (respectively) of the parameter \( K \). It should be noted that in order to determine the acceptance ratio \( r_{\text{birth}} \), we need to evaluate \( p(\tau_{K+1}, K + 1 | \mathbf{c}_{K+1}, \mathbf{x}) \), where there is now a dependence on \( \mathbf{c}_{K+1} \). This dependence arises due to the proposed transition time \( \tau_{\text{prop}} \), splitting the segment between the time indices \( \{ \tau_i, \tau_{i+1} \} \) into \( \{ \tau_i, \tau_{\text{prop}}, \tau_{i+1} \} \), as well as, splitting the segment class variable \( \{ c_i \} \), into two new class variables \( \{ \hat{c}_i, \tilde{c}_{i+1} \} \). As we have not yet inferred the new class variables from the conditional class posterior distributions, we assume that the two classes \( \{ \hat{c}_i, \tilde{c}_{i+1} \} \) are distinct (that is, \( \hat{c}_j \neq \tilde{c}_k \) for all \( j \neq k \) and \( j = 1, 2 \)) and thus independent from all other segments, to circumvent the lack of information we have for assignment to an existing class (please refer to Figure ).

**Algorithm 1**

**Require:**
- Set input parameters (discussed in Section IV).
- Initialize: \( \{ \tau_K, K, \mathbf{c}_K, \Phi, \Sigma, \sigma^2, V \} \)
- Set: \( N_{\text{iter}} \) and \( N_{\text{iter}} \).

**for** loop over main = \( \{ 1, \ldots, N_{\text{iter}} \} \) **do**

**for** loop over \( \{ 1, \ldots, N_{\text{iter}} \} \) **do**

- Sample \( p(\Phi | \mathbf{c}_K, \tau_K, K, \Sigma, \mathbf{x}) \) for \( v = 1, \ldots, V \), shown in (12).
- Sample \( p(\Sigma | \mathbf{c}_K, \tau_K, K, \Phi, \mathbf{x}) \) for \( v = 1, \ldots, V \), shown in (13).
- Sample \( p(\sigma^2 | \mathbf{c}_K, \tau_K, K, \mathbf{x}) \) for \( v = 1, \ldots, V \), shown in (14).
- Sample class variable \( c_i \) for \( i = 0, \ldots, K \), using (7) and (8).

**end for**

- Sample \( p(\tau_K, K | \mathbf{c}_K, \mathbf{x}) \), refer to Section III.B.

**end for**
The death move proposes to remove a transition time $\tau_{prop}$, by choosing with uniform probability from the set $\tau_K$; where the removal of $\tau_{prop}$ is accepted with probability $\alpha_{death} = \min\{1, r^{-1}_{birth}\}$. As in the previous case (birth move), we need to determine $r_{birth}$, however, now we need to evaluate $p(\tau_{K-1}, K-1|x)$. That is, the segments between the transition times, $\{\tau_i, \tau_{prop}\}$ and $\{\tau_{prop}, \tau_{i+2}\}$ where $\tau_{prop} = \tau_{i+1}$, are combined into one segment $\{\tau_i, \tau_{i+2}\}$, along with the segment class variables $\{c_i, c_{i+1}\}$ being combined into one segment with a new class variable $\{\hat{c}_i\}$. Using the argument utilised for the birth of a change point we assign a distinct value to the new class variable, that is, $\hat{c}_i \neq c_j$ for all $j \neq i$.

The update of the transitions times is carried by first removing the $j^{th}$ transition time index $\tau_j$ from $\tau_K$ and proposing a new change point at some new location, for all $j = \{1, \ldots, K\}$. That is, the death move is first applied followed by a birth move for all transition times in $\tau_K$.

**REFERENCES**

[1] A. Brandt, “Detecting and estimating parameter jumps using ladder algorithms and likelihood ratio tests,” IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1017–1020, 1983.

[2] R. Killick, P. Fearnhead, and I. A. Eckley, “Optimal detection of changepoints with a linear computational cost,” Journal of the American Statistical Association, vol. 500, no. 107, pp. 1590–1598, 2012.

[3] F. Desobry, M. Davy, and C. Doncarli, “An online kernel change detection algorithm,” IEEE Transactions on Signal Processing, vol. 53, no. 8, pp. 2961–2974, 2005.

[4] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, “Bayesian curve fitting using MCMC with applications to signal segmentation,” IEEE Transactions on Signal Processing, vol. 50, no. 3, pp. 747–758, 2002.

[5] P. Fearnhead, “Exact Bayesian curve fitting and signal segmentation,” IEEE Transactions on Signal Processing, vol. 53, no. 6, pp. 2160–2166, 2005.

[6] I. Rezek and S. Roberts, “Ensemble hidden Markov models with extended observation densities for biosignal analysis,” Springer London, pp. 419–450, 2005.

[7] R. M. Neal, “Markov chain sampling methods for Dirichlet process mixture models,” Journal of Computational and Graphical Statistics, vol. 2, no. 2, pp. 249–265, 2000.

[8] C. E. Rasmussen, “The infinite Gaussian mixture model,” Advances in Neural Information Processing Systems 12, pp. 554–560, 2000.

[9] A. Ahrabian, S. Enshaeifar, C. Cheong-Took, and P. Barnaghi, “Segment parameter labelling in MCMC mean-shift change detection,” IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4244–4248, 2018.