Bayesian Multiple Changepoint Detection for Stochastic Models in Continuous Time

Lu Shaochuan*

Abstract. A multiple changepoint model in continuous time is formulated as a continuous-time hidden Markov model, defined on a countable infinite state space. The new formulation of the multiple changepoint model allows the model complexities, i.e. the number of changepoints, to accrue unboundedly upon the arrivals of new data. Inference on the number of changepoints and their locations is based on a collapsed Gibbs sampler. We suggest a new version of forward-filtering backward-sampling (FFBS) algorithm in continuous time for simulating the full trajectory of the latent Markov chain, i.e. the changepoints. The FFBS algorithm is based on a randomized time-discretization for the latent Markov chain through uniformization schemes, combined with a discrete-time version of FFBS algorithm. It is shown that, desirably, both the computational cost and the memory cost of the FFBS algorithm are only quadratic to the number of changepoints. The new formulation of the multiple changepoint models allows varying scale of run lengths of changepoints to be characterized. We demonstrate the methods through simulations and a real data example for earthquakes.

Keywords: continuous-time forward-filtering backward-sampling algorithms, uniformization, infinite hidden Markov Models, Markov jump processes, Poisson processes.

1 Introduction

Changepoint models arise in many practical applications, e.g. signal processing, system control, bioinformatics, finance, among many others. Typically, chronologically ordered data streams are collected over a period during which a (random) number of structural breaks happen at unknown locations, so that observations within the same segment, or equivalently, observations between two consecutive changepoints, are regarded as homogeneous and observations across different segments are heterogeneous. The interest lies in making inference for the number and the locations of changepoints and other model parameters. Among many literatures, recently, pruned exact linear time PELT (Killick et al. 2012), wild binary segmentation WBS (Fryzlewicz 2014) and simultaneous multiscale change point estimator SMUCE (Frick et al. 2014) are popular.

In Bayesian context of retrospective changepoint detection, Monte Carlo strategies are often applied for posterior inference, see e.g. Stephens (1994), Green (1995), Chib (1998), Lavielle and Lebarbier (2001), Fearnhead (2006) and Ruggieri and Lawrence (2014), among many others. Online inference for multiple changepoint problems by use of particle filtering is discussed by Chopin (2007), Fearnhead and Liu (2007) and

---
*School of Statistics, Beijing Normal University, Xin Jie Kou Wai Da Jie 19, Beijing, P. R. China, lvshaochuan@bnu.edu.cn

Yildirim et al. (2013) and some others. Ko et al. (2015) proposes a Bayesian multiple changepoint model based on the Dirichlet process hidden Markov model, in which the number of changepoints is not required to be specified. Martinez and Mena (2014) proposes another Bayesian nonparametric approach to changepoint detection based on random partition distributions (Barry and Hartigan 1992) and a split-merge Markov Chain Monte Carlo (MCMC) algorithm is applied for the posterior inference. Most of these studies are suitable only for detecting changepoints appearing in discrete time series or discretely observed random processes. Although some continuous time random processes may be approximated ideally by their discrete time counterparts, it is necessary to develop multiple changepoint models in continuous time in their own right. So, the approximation error introduced by using the discrete-time counterpart, which is often difficult to quantify, can be avoided.

We consider Bayesian multiple changepoint problems for some continuous-time random processes. Multiple changepoint problems in discrete time are often treated in the framework of hierarchical Bayesian models, e.g. latent Markov models, due to some efficient posterior simulation strategies such as forward-backward recursions (Chib 1998, Scott 2002, Fearnhead and Liu 2007) and Viterbi dynamics can be utilized for the inference of changepoints. Similarly, we formulate multiple changepoint problems of random processes into continuous-time hidden Markov models. Recently, some multiple changepoint models for continuous-time stochastic models, such as (marked) Poisson process and finite Markov chain, are formulated as continuous-time hidden Markov models (Rao and Teh 2013, Lu 2019, Lu 2020). However, these models are only suitable for multiple changepoint problems with a fixed number of changepoints. Although the number of changepoints may be determined by use of some model selection criteria after a series of multiple changepoint models with diverse numbers of changepoints fitted, the definition of the model complexities in different criteria can be sharply different. In some cases, different numbers of changepoints may be identified by a range of model selection criteria. The uncertainties in the number of changepoints are difficult to quantify.

A new multiple changepoint model for stochastic models in continuous time is formulated, in which the changepoint locations are associated with a left-to-right latent Markov chain defined on a countable infinite state space. In a full Bayesian approach, the joint posterior of the number of changepoints and their locations is simulated via efficient Markov chain Monte Carlo strategies. We suggest a continuous-time version of forward-filtering backward-sampling (FFBS) algorithm within the Gibbs sampling scheme, which is an extension of the discrete-time forward-filtering backward-sampling algorithm of Fearnhead and Liu (2007). It is based on a randomized time-discretization for the latent Markov process on a countable infinite state space through uniformization schemes. Therefore, there are no time-discretization errors required to be accounted for. For multiple changepoint problems in discrete time, it is well-known that the computational cost of the FFBS algorithm is proportional to the number of observations and the memory cost increases with respect to the number of observations quadratically, which is prohibitive when the data set is large and the changepoints are only sparsely distributed. Often, particle filters or pruning techniques have been applied to deal with these problems: see Fearnhead and Liu (2007). Nevertheless, with the uniformization

scheme, it is possible to deal with the issue of computational cost and storage cost of a Bayesian multiple changepoint problem in continuous time directly at the scale of changepoint numbers. It is shown that both the computational cost and the memory cost of the FFBS algorithm are only quadratic to the number of uniformization times generated from the uniformization schemes, which can be scaled down to the number of changepoints. This is desirable in comparison to its discrete-time counterpart. Furthermore, By integrating out the model parameters and targeting the marginal posteriors of the number of changepoints and their locations, the efficiency of the posterior simulation scheme is improved.

In Section 2, a multiple changepoint model for a stochastic process is formulated as a continuous-time hidden Markov model. We then introduce a continuous-time FFBS algorithm for the simulation of the trajectories of the latent Markov chain defined on a countable infinite state space. The posteriors of the number of changepoints and their locations are approximated via a Gibbs sampling scheme. In some cases, it might be desirable to obtain the MAP estimates of changepoints. In section 3, we discuss the continuous-time version of the Viterbi algorithm, particularly for multiple changepoint problems of Markov jump processes and (marked) Poisson processes. We demonstrate the methods through 5 numerical examples. The first two numerical examples are simulation studies for the multiple changepoint problem of a Markov jump process. The last three numerical examples include simulation studies for Poisson process and a real data example of New Zealand deep earthquakes.

2   Model Formulation and Bayesian Multiple Changepoints Detection

2.1   Model Formulation

Let $\mathbf{Y}_\theta(t)$ be a continuous-time random process on a time interval $[0, T]$. Assume it is subject to abrupt changes at unknown locations $0 \triangleq \tau_0 < \tau_1 < \cdots < \tau_m < \tau_{m+1} \triangleq T$, such that the observed random process $\mathbf{Y}_\theta(t)$ is partitioned into $m + 1$ segments by m changepoints $\tau = (\tau_1, \cdots, \tau_m)$. The number of changepoints m is often unknown. The main interest lies in making inference for the number of changepoints m, their locations τ and model parameters $\theta = (\theta_1, \cdots, \theta_{m+1})$ in each segment of $\mathbf{Y}_\theta(t)$.

In a Bayesian context, the priors for the number of changepoints m, their locations τ and θ need to be specified. For computational reasons illuminated later, we implicitly define priors for the number of changepoints and their locations by defining priors for the run lengths, i.e. the segment lengths, between consecutive changepoints. The priors for the model parameters θ in $\mathbf{Y}_\theta(t)$ will be given in Section 4 after $\mathbf{Y}_\theta(t)$ is defined. We assume *a priori* run length of the i-th segment is exponentially distributed with rate q_i. This type of priors for the changepoints can be defined as a latent Markov chain $\mathbf{X}(t)$ on a numerable infinite state space, with a left-to-right transition rate matrix $\mathbf{Q}$ parameterized exactly in terms of a multiple changepoint model. We define the transition

rate matrix of a left-to-right Markov chain $\mathbf{X}(t)$ by

$$\mathbf{Q} = \begin{pmatrix} -q_1 & q_1 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ 0 & -q_2 & q_2 & \cdots & 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -q_m & q_m & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

When $\mathbf{X}(t)$ is in the i-th state, $\mathbf{Y}_\theta(t)$ is parameterized by θ_i. This type of model formulation is a particular type of continuous time hidden Markov models (CTHMMs) and also a type of product partition models (Barry and Hartigan, 1992).

In the following, we define the brackets $[j, i]$ after a matrix A as the (j, i)-th entry of A. The sample path of a random process $\mathbf{Z}(t)$ on a time interval $[a, b]$ or $[a, b)$ is denoted by $\mathbf{Z}[a, b]$ or $\mathbf{Z}[a, b)$ respectively. Generically, $p(.)$ denotes a probability density (mass) function. For convenience, $\mathbf{Y}_\theta(t)$ is abbreviated to $\mathbf{Y}(t)$. Throughout the discussion, we denote a vector $(z_1, \cdots, z_n)$ by $z_{1:n}$.

With m changepoints τ on $[0, T]$, $\mathbf{Y}(t)$ is partitioned into $m + 1$ segments. The i-th segment of $\mathbf{Y}(t)$ is associated with the model parameter θ_i. Given τ and θ, we assume conditional independence of $\mathbf{Y}(t)$ across segments. The likelihood is written as

$$\mathbb{L}(m, \tau, \theta) = p\left(\mathbf{Y}[0, T] | m, \tau, \theta\right) = \prod_{i=1}^{m+1} p\left(\mathbf{Y}[\tau_{i-1}, \tau_i) | \theta_i\right).$$

After defining priors for θ, the marginal likelihood of $\mathbf{Y}(t)$ on an interval $[s, t)$ is given by

$$\mathcal{M}(s, t) = \int_\Theta p\left(\mathbf{Y}[s, t) | \theta\right) p(\theta) \, d\theta.$$

We assume the marginal likelihood of $\mathbf{Y}(t)$ within an interval can be exactly evaluated, which is achieved either by choosing a conjugate prior $p(\theta)$ for θ or carrying out a numerical integration. The priors for m changepoints located at $\tau_1, \cdots, \tau_m$ are given by

$$p(m, \tau) = \prod_{i=1}^{m} q_i e^{-q_i(\tau_i - \tau_{i-1})} e^{-q_{m+1}(\tau_{m+1} - \tau_m)}.$$

Thus, the marginal posterior of the number of changepoints and their locations is written as

$$p(m, \tau | \mathbf{Y}[0, T]) \propto \prod_{i=1}^{m} \{q_i e^{-q_i(\tau_i - \tau_{i-1})} \mathcal{M}(\tau_{i-1}, \tau_i)\} e^{-q_{m+1}(\tau_{m+1} - \tau_m)} \mathcal{M}(\tau_m, \tau_{m+1}), \quad (1)$$

which is known up to a normalizing constant.

Generally, reversible jump MCMC (RJMCMC) methods can be applied to simulate approximately from the marginal posterior of the number of changepoints and their positions. However, these methods can be computationally inefficient, leading to problems

such as slow convergence and poor mixing. Instead, we may consider a time discretization of $\mathbf{Y}(t)$ by assuming that the locations of changepoints are only at the endpoints of a set of consecutive time intervals, partitioned evenly at $0 < t_1 < t_2 < \cdots < t_n < T$. Then, efficient posterior simulation strategies such as forward-backward recursions (Chib 1998, Scott 2002, Fearnhead and Liu 2007) can be applied. Nevertheless, the approximation error by time-discretization techniques is often difficult to quantify. We suggest a randomized time-discretization by use of the uniformization scheme for a continuous-time Markov jump process on a countable infinite state space, so that quantifying the time-discretization error is bypassed.

2.2 Uniformization Scheme

Let $\mathbf{Z}(t)$ be a time-inhomogeneous Markov chain on a countable infinite state space with transition rates

$$q_{ij}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(\mathbf{Z}(t + \Delta t) = j | \mathbf{Z}(t) = i), j \neq i.$$

Assume for some constant $\lambda > 0$,

$$q_i(t) = \sum_{j \neq i} q_{ij}(t) < \lambda. \tag{2}$$

Let $\mathbb{P}_{s,t}[i,j]$ denote the transition probability of observing $\mathbf{Z}(t)$ in state j at time t given that it was in state i at time s. Define a transition probability matrix $\mathbb{P}_t$ by

$$\mathbb{P}_t[i,j] = \begin{cases} \frac{q_{ij}(t)}{\lambda}, & if\ j \neq i; \\ 1 - \sum_{k \neq i} \frac{q_{ik}(t)}{\lambda}, & if\ j = i. \end{cases} \tag{3}$$

Then, for all $s \leq t$ and all i, j,

$$\mathbb{P}_{s,t}[i,j] = \sum_{k=0}^{\infty} \frac{(t-s)^k \lambda^k}{k!} e^{-\lambda(t-s)} \int_s^t \int_s^t \cdots \int_s^t \mathbb{P}_{t_1} \mathbb{P}_{t_2} \cdots \mathbb{P}_{t_k}[i,j] d\mathbb{U}(t_1, t_2, \cdots, t_k), \tag{4}$$

where $\mathbb{U}(t_1, t_2, \cdots, t_k)$ is the k-dimensional uniform distribution of the order statistics $t_1 < t_2 < \cdots < t_k$ and $\mathbb{P}_{t_1} \mathbb{P}_{t_2} \cdots \mathbb{P}_{t_k}$ is the matrix product of the transition matrices in (3), see Van Dijk et al. (2018) and Jensen (1953).

In this case, we say that the Markov chain is subordinated to the Poisson process. The above equation has a natural interpretation as Poisson randomization. During $[s, t]$, jumps occur according to a Poisson process with rate λ. Given that k jumps have occurred, the jump times are uniformly distributed on $[s, t]$. At these Poisson epochs, the state transitions happen in terms of a discrete-time Markov chain with the transition probability matrix defined by (3). It is noted that the uniformization scheme is also viable for a Markov jump process on a countable infinite state space, which can be used in multiple changepoint models with a potentially unbounded number of changepoints.

2.3 Gibbs Samplers with a Continuous-Time FFBS Algorithm

Assume $\{t_1, t_2, \cdots, t_n\}$ is a sequence of uniformization times of $\mathbf{X}(t)$ generated from the uniformization schemes, such that $\mathbf{X}(t_1), \cdots, \mathbf{X}(t_n)$ is a discrete-time Markov chain with the transition probability matrix

$$\mathbb{P} = \begin{pmatrix} p_1 & 1-p_1 & 0 & \cdots & 0 & 0 & \cdots \\ 0 & p_2 & 1-p_2 & \cdots & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & p_m & 1-p_m & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}. \tag{5}$$

Given $t_{1:n}$, the locations of changepoints of $\mathbf{Y}(t)$ are choosen from the set of discrete times $\{t_1, t_2, \cdots, t_n\}$. We then treat $\mathbf{Y}_1 \triangleq \mathbf{Y}[0, t_1), \mathbf{Y}_2 \triangleq \mathbf{Y}[t_1, t_2), \cdots, \mathbf{Y}_{n+1} \triangleq \mathbf{Y}[t_n, T)$ as a sequence of granularized observations. Let s_i denote $\mathbf{X}(t_i)$ for $i = 1, \ldots, n$ and let $g(i, k)$ denote the probability mass function of the run length of a segment, beginning from t_i and ending at t_k, so that

$$g(i, k) \triangleq (\prod_{j=i}^{k-1} p_{s_j})(1 - p_{s_k}), \qquad k > i. \tag{6}$$

The associated cumulative distribution function of the run length for this segment is written as

$$\mathbb{G}(i, k) \triangleq \sum_{j=i+1}^{k} g(i, j).$$

Let C_t be a state variable taking values in $0, 1, 2, \cdots, t - 1$, which is defined to be the location of the most recent changepoint before time t. If there is no changepoint before t, $C_t = 0$. $\{C_t\}$ is a discrete-time Markov chain on $0, 1, 2, \cdots, t - 1$. In Fearnhead and Liu (2007), the transition probability of C_t is given by

$$\mathbb{P}(C_{t+1} = j | C_t = i) = \begin{cases} \frac{1 - \mathbb{G}(i, t+1)}{1 - \mathbb{G}(i, t)}, & if \ j = i; \\ \frac{g(i, t+1)}{1 - \mathbb{G}(i, t)}, & if \ j = t; \\ 0, & otherwise. \end{cases} \tag{7}$$

The forward filtering algorithm gives the filtering probabilities

$$\mathbb{P}(C_{t+1} = i | \mathbf{Y}_{1:t+1})$$
$$\propto \begin{cases} \mathbb{P}(C_t = i | \mathbf{Y}_{1:t}) \left(\frac{1 - \mathbb{G}(i, t+1)}{1 - \mathbb{G}(i, t)} \right) \left(\frac{\mathcal{M}(i, t+1)}{\mathcal{M}(i, t)} \right), & for \ i = 0, 1, \cdots, t - 1; \\ \mathcal{M}(t, t+1) \sum_{j=0}^{t-1} \mathbb{P}(C_t = j | \mathbf{Y}_{1:t}) \left(\frac{g(j, t+1)}{1 - \mathbb{G}(j, t)} \right), & for \ i = t. \end{cases} \tag{8}$$

When the filtering probabilities are stored, the backward sampling step generates samples from the posteriors of the number of changepoints and their locations. So, the

last changepoint C_n is generated first according to the filtering probabilities $\mathbb{P}(C_n|\mathbf{Y}[0,T])$. Given the latest changepoint $C_n = t$, the next changepoint is generated recursively and backwards according to

$$\mathbb{P}(C_t = i|\mathbf{Y}_{1:n}, C_n = t) \propto \mathbb{P}(C_t = i|\mathbf{Y}_{1:t}) \left(\frac{g(i, t+1)}{1 - \mathbb{G}(i, t)} \right). \tag{9}$$

The recursion terminates when $C_t = 0$. This is another version of the continuous-time forward-filtering backward-sampling (FFBS) algorithm: see Rao and Teh (2013) for another version of the FFBS algorithm for Markov-modulated Poisson processes with a finite number of states and continuous-time Bayesian network.

The computational cost of this version of FFBS algorithm is quadratic in n, that is the actual number of simulated uniformization times. The memory cost of this version of FFBS algorithm is only quadratic in n. By computing directly on the scale of changepoints, the memory cost is significantly lowered, which is otherwise prohibitive as in the usual case when the filtering probabilities are computed and stored in an event-by-event scale, particularly for a very long sequence of observations in which the changepoints are only sparsely distributed. The proposed method is potentially suitable for the detection of sparse changepoints in a very long sequence of observations.

Gibbs sampling is completed by sampling the model parameters conditioned on the full path of the latent Markov chain. Let $\Gamma(a, b)$ be the prior distribution of q_i. So, the joint prior of $\mathbf{Q}$ is given by $\prod_{i=1}^{m} \frac{b^a}{\Gamma(a)} q_i^{a-1} \exp\{-bq_i\}$. The hyperparameters a and b in the priors are given subjectively in numerical examples. The likelihood of $\mathbf{X}(t), 0 \leq t \leq T$ is written as $\prod_{i=1}^{m} q_i \exp\{-q_i(\tau_i - \tau_{i-1})\}$. Therefore, the posterior distribution of $\mathbf{Q}$ is given by

$$p(\mathbf{Q}|\mathbf{X}[0,T]) \propto \prod_{i=1}^{m} \frac{b^a}{\Gamma(a)} q_i^{(a+1)-1} \exp\{-[b + (\tau_i - \tau_{i-1})]q_i\}. \tag{10}$$

The full conditional of q_i is distributed according to $\Gamma(a+1, b+(\tau_i - \tau_{i-1}))$. A two-block Gibbs samper is given in Algorithm 1.

Algorithm 1 Block Gibbs sampler 1.

---

1. Initialize $\mathbf{Q}_0$, the number of changepoints m_0 and their locations $\tau_1, \tau_2, \cdots, \tau_{m_0}$;
2. Uniformization scheme: Sample a sequence of Poisson times $\{t_1, t_2, \cdots, t_n\}$ with a constant intensity rate λ satisfying (2) and define a discrete-time Markov chain with the transition probability matrix $\mathbb{P}$ given in (3);
3. Sample a set of changepoints $\tau_1, \tau_2, \cdots, \tau_{m^*}$ by use of a discrete-time FFBS algorithm as aforementioned;
4. Conditioned on the current trajectory of the latent Markov chain $\mathbf{X}(t)$, sample the transition rate matrix $\mathbf{Q}$ according to (10);
5. Repeat steps (2) $\sim$ (4) until the maximum number of iterations reached.

One strength of this model formulation is that it allows a varying scale of segment lengths to be characterized in the transition rate (changepoint recurrence rate) matrix

Q. When the rate parameters in **Q** are highly variable, the Poisson rate λ utilized in the uniformization scheme will be dominated by the largest q_i in **Q**, potentially leading to an excessive number of Poisson events generated by the uniformization procedure, which will significantly inflate the memory cost and computational cost and slow the algorithm. Instead, we consider a modification of the previous uniformization scheme by using a nonhomogeneous Poisson process with the Poisson rate $\lambda(t)$ adaptively tuned with respect to the hazard rate (changepoint recurrence rate) $q(t)$. We choose a piecewise constant intensity function $\lambda(t)$ to implement the randomized time-discretization for the latent Markov chain piecewisely (Van Dijk 1992). To simulate the subordinated Poisson process with a piecewise constant intensity function $\lambda(t)$, the standard thinning methods (Lewis and Shedler 1979) can be employed. Algorithm 1 is replaced by the following Algorithm 2.

Algorithm 2 Block Gibbs sampler 2.

---

1. Initialize $\mathbf{Q}_0$, the number of changepoints m_0 and their locations $\tau_1, \tau_2, \cdots, \tau_{m_0}$;
2. Sample a sequence of Poisson times $\{t_1, t_2, \cdots, t_n\}$ from a nonhomogeneous Poisson process with a piecewise constant intensity function $\lambda(t)$ via the thinning methods. Calculate state transition probabilities of the discrete-time Markov chain according to (3), with λ in (2) replaced by $\lambda(t)$;
3. Sample a set of changepoints $\tau_1, \tau_2, \cdots, \tau_{m^*}$ by use of a discrete-time FFBS algorithm as aforementioned;
4. Conditioned on the current trajectory of the latent Markov chain $\mathbf{X}(t)$, sample the transition rate matrix **Q** according to (10);
5. Repeat steps (2) $\sim$ (4) until the maximum number of iterations reached.

In many cases, the posterior of the model parameter θ in each segment is also of interest. We may sample from $p(\theta|\mathbf{Y}[0, T])$ by a collapsed Gibbs sampler, in which the model parameter θ is integrated out from the joint posterior and the block Gibbs sampler (Algorithm 2) is applied to the remaining components until convergence. Then, the model parameter θ in each segment is sampled according to the full conditionals, see Liu (1994).

In simulation studies, we observed that the simulated trajectories of the latent Markov chain, the locations of changepoints, typically show some "knots". These "knots" happens when the filtering probabilities of a changepoint location are not concentrated in a single uniformization time, but instead spread across several consecutive uniformization times. In this case, the backward sampling step will sometimes treat part of them or all of them as individual changepoints. This sort of "knot" will inflate the memory and computational costs if kept into the next iteration. A direct approach to pruning these "knots" is to slightly modify the backward sampling step by resampling the most recent changepoint before t several times according to (9) and choosing the most distant one from t as the most recent changepoint before t. Our numerical studies show that this is a very effective approach to prune the extra changepoints cheaply.

In step 2 of Algorithm 3, the changepoints themselves are treated as uniformization times in the next iteration, so that likely changepoint locations can be kept in following iterations.

Algorithm 3 Collapsed Gibbs sampler 3.

---

1. Set an initial value m_0 for the number of changepoints , their locations
$\tau = \{\tau_1, \cdots, \tau_{m_0}\}$ and the changepoint recurrence rate matrix $\mathbf{Q}_0$ with $m + 1$ states;
2. Let the Poisson intensity rate $\lambda(t)$ be k times of the changepoint recurrence rate,
$\lambda(t) = kq_i$ when $t \in [\tau_{i-1}, \tau_i), i = 1, \cdots, m + 1$. Simulate a nonhomogeneous Poisson
process $\mathcal{U} = \{u_1, u_2, \cdots, u_K\}$ with piecewise constant intensity function by thinning
methods. Combine $\mathcal{U}$ and τ to form a sequence of uniformization times $\{t_1, \cdots, t_n\}$.
Calculate the state transition probabilities $\{p_{s_1}, \cdots, p_{s_n}\}$ of the discrete-time Markov
chain $\{\mathbf{X}(\mathbf{t_1}), \cdots, \mathbf{X}(\mathbf{t_n})\}$ according to (3), with $\lambda(t)$ in (2) replaced by $(k + 1)q_i$ when
$t \in [\tau_{i-1}, \tau_i)$;
3. Conditioned on the current model parameters, recursively calculate the forward
probabilities according to (8) and sample a set of changepoints $\tau_1, \tau_2, \cdots, \tau_{m'}$
according to (9). In the backward sampling step, the "knots" are pruned by repeatedly
sampling the most recent changepoint before t and choosing the most distant one from
t as the most recent changepoint;
4. Given the current trajectory of the latent Markov chain $\mathbf{X}(t)$, sample the transition
rate matrix $\mathbf{Q}$ according to (10);
5. After the Markov chain converges, sample the model parameters θ_i in each segment
from the full conditionals $p(\theta_i|\mathbf{Y}[\tau_{i-1}, \tau_i)), i = 1, \cdots, m' + 1$;
6. Repeat steps (2)–(5) until the maximum number of iterations reached.

The current model formulation is more flexible than multiple changepoint models with the number of changepoints fixed or the maximum number of changepoints bounded. It is also noted that when all q_is are given in advance, we may run an exact Monte Carlo sampling algorithm for changepoints. However, this is unrealistic in practice. As q_is in $\mathbf{Q}$ might be different, varying scale of run lengths between changepoints can be incorporated into the model, which is desirable in practice to avoid model bias. This approach is applicable for multiple changepoint models for any random process when the marginal likelihood can be exactly evaluated, which is of course evident when conjugate priors are introduced, for example for the exponential family of random processes, or the marginal likelihood can be numerically computed.

We call the Poisson rate in the uniformization scheme a *resolution parameter*. The use of a high intensity rate will lead to the generation of a large number of uniformization times, so that the positions of changepoints can be located in a finer time scale. However, both the memory cost and computational cost will inflate significantly, making the algorithm less efficient. Nevertheless, choosing a relatively small Poisson rate will reduce both the memory cost and computational cost on average, but the accuracy of estimation for the locations of changepoints will be lowered: see numerical examples in Section 4. The resolution parameter needs to be carefully tuned for a tradeoff between the costs and accuracies of the estimation.

However, even using Algorithm 2–3, it is still possible the uniformization scheme occasionally produces too many Poisson events (uniformization times), causing issues of inflated computational costs and memory costs. In this case, we may set an upper bound on the number of Poisson events. The truncation error can be quantified as

follows:

$$\|\bar{\mathbb{P}}_{s,t}^L - \mathbb{P}_{s,t}\| \leq \sum_{k=L+1}^{\infty} \frac{\lambda^k(t-s)^k}{k!} e^{-\lambda(t-s)} \leq \frac{\lambda^L(t-s)^L}{L!}, \tag{11}$$

where $\|.\|$ denotes the standard supremum norm and $\bar{\mathbb{P}}_{s,t}^L$ is the truncated version of (4) at level $k = L$: see Van Dijk (2018).

3   Continuous-Time Viterbi Algorithm

In some cases, it is desirable to obtain *maximum a posteriori* (MAP) estimates of the set of changepoints. The MAP estimates of the set of changepoints, the optimal path of the latent Markov chain $\mathbf{X}(t)$, is selected from potentially an infinite number of trajectories, which is inaccessible in general. Nevertheless, when $\mathbf{Y}(t)$ is Markov jump process or (marked) Poisson process as illustrated in the next section. The following algorithm is able to retrieve exactly the most probable sample path by a continuous-time version of the Viterbi algorithm. Let $u_1 < u_2 < \cdots < u_n$ be the set of event times, the arrival times of a (marked) Poisson process or the transition times of a Markov jump process respectively. Define

$$M_i(u_k) = \max_{\mathbf{X}[0,u_k)} \log p\bigg(\mathbf{X}[0,u_k), \mathbf{X}(u_k) = i, \mathbf{Y}[0,u_k]\bigg). \tag{12}$$

For these Markov renewal processes, Viterbi recursion shows that

$$M_j(u) = \max_i \left\{ M_i(u_k) + \log p\big(\mathbf{X}(u) = j, \mathbf{Y}(u_k, u] \big| \mathbf{X}(u_k) = i\big) \right\}, \tag{13}$$

for $u_k < u \leq u_{k+1}$ and $j = i$ or $j = i + 1$, in which $\log p\big(\mathbf{X}(u) = j, \mathbf{Y}(u_k, u] \big| \mathbf{X}(u_k) = i\big)$ needs to be maximized over the path space from u_k to u. When $\mathbf{Y}(t)$ is Markov jump process or (marked) Poisson process, $\log p\big(\mathbf{X}(u) = j, \mathbf{Y}(u_k, u] \big| \mathbf{X}(u_k) = i\big)$ can be explicitly maximized: see Bebbington (2007), Lu (2019) and Lu (2020). It turns out that the most probable path of the latent Markov chain $\mathbf{X}(t)$ has state transitions exactly at the event times, the arrival times of a (marked) Poisson process $\mathbf{Y}(u)$ or the state transition epochs of the observed Markov jump process $\mathbf{Y}(u)$ respectively. As a result, the optimal set of changepoint locations is chosen exactly from these event times: see Lu (2019) and Lu (2020) for nearly the same argument. The search for the set of most probable changepoints can be equivalently implemented by a discrete-time version of the Viterbi algorithm (Viterbi, 1967).

Let $f_k(j)$ be the probability density of the most probable path of $x(t)$ up to time u_k upon reaching state j, and $\phi_k(j)$ be the optimal state at time u_{k-1} for the path reaching state j at the next time step u_k.

The memory cost in this case is quadratic in n, which is prohibitive for a long sequence of event times. Instead, we may first obtain the MAP estimates of the number of changepoints from the marginal posteriors. Conditioned on the number of changepoints, then, the MAP estimates of the locations of changepoints can be obtained from

---

Algorithm 4 Continuous-time Viterbi Algorithm.

1. Initialize $f_1(j) = \pi_1 p(\mathbf{X}(u_1) = j, \mathbf{Y}(0, u_1]|\mathbf{X}(0) = 1)$ and set $\phi_1(j) = 0$ for $j = 1$.

2. For $k = 2, \cdots, n$ and $j = i\,or\,i + 1$, recursively compute

$$f_k(j) = \max_i \left\{ f_{k-1}(i) p(\mathbf{X}(u_k) = j, \mathbf{Y}(u_{k-1}, u_k]|\mathbf{X}(u_{k-1}) = i) \right\}$$

and

$$\phi_k(j) = \arg\max_i \left\{ f_{k-1}(i) p(\mathbf{X}(u_k) = j, \mathbf{Y}(u_{k-1}, u_k]|\mathbf{X}(u_{k-1}) = i) \right\}.$$

3. Let $x_n = \arg\max_j f_n(j)$ and backtrack the state sequence as follows:

For $k = n - 1, \cdots, 1, x_k = \phi_{k+1}(x_{k+1})$.

---

the continuous-time version of a Viterbi algorithm for a finite-state hidden Markov model (HMM): see Bebbington (2007), Lu (2019) and Lu (2020). In this case, both the memory cost and the computational cost are only proportional to mn, where m is the MAP estimate of the number of changepoints.

Algorithm 4 is only suitable for special cases as demonstrated in the above examples. In general, we may obtain approximately the MAP estimates of changepoints via the uniformization scheme as follows. First, we obtain MAP estimates of the number of changepoints directly from the marginal posteriors. Conditioned on the number of changepoints, we may perform a randomized time-discretization for both $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ through the uniformization scheme. Conditioned on a particular time-discretization of $\mathbf{X}(t)$ and $\mathbf{Y}(t)$, we carry out the discrete-time Viterbi algorithm to obtain a "MAP" estimate of the locations of changepoints. This randomized time-discretization for the latent Markov chain $\mathbf{X}(t)$ by the uniformization scheme and the following discrete-time Viterbi algorithm can be performed many times. We may choose the one which gives the highest posterior density of θ as the "MAP" estimate among the potentially many different segmentations. We call it a "randomized Viterbi algorithm".

4 Numerical Examples

4.1 Multiple Changepoint Models for Markov Jump Processes

Markov chains are widely used in modelling weather, bioinformatics, speech recognition and modern credit management, see Polansky (2007), Arnesen et al. (2016) and Xing et al. (2012). Let $\mathbf{Y}(t)$ be a r-state continuous-time finite Markov chain observed on $[0, T]$, specified by the transition rate matrix $\mathbf{D}_{\mathbf{X}(t)}$. $\mathbf{Y}(t)$ is subject to abrupt changes at unknown locations corresponding to a random number of changepoints. We consider multiple changepoint models of Markov jump processes with the changepoint locations associated with a latent Markov chain $\mathbf{X}(t)$. When $\mathbf{X}(t)$ is in state k, $\mathbf{Y}(t)$ is parameterized by $\mathbf{D}_k = (d_{ij}(k))_{r \times r}$. The likelihood of the trajectory of $\mathbf{Y}(t)$, with the transition

	$\mathbf{D}_1$			$\mathbf{D}_2$			$\mathbf{D}_3$		
(a)	−1	0.5	0.5	−6	2	4	−0.2	0.1	0.1
Model	0.1	−0.2	0.1	0.5	−1	0.5	2	−4	2
for	1	2	−3	0.04	0.06	−0.1	0.1	0.2	−0.3
4	$\mathbf{D}_4$			τs					
Segments	−3	1	2	$\tau_1 = 678.7$					
	0.2	−0.7	0.5	$\tau_2 = 1753.7$					
	2	2	−4	$\tau_3 = 2668.3$					
	$\mathbf{D}_1$			$\mathbf{D}_2$			$\mathbf{D}_3$		
(b)	−1.19	0.67	0.52	−4.9	1.62	3.28	−0.2	0.09	0.11
Model	0.09	−0.20	0.11	0.63	−1.05	0.42	2.33	−4.47	2.14
for	1.13	1.95	−3.08	0.05	0.06	−0.11	0.11	0.15	−0.26
4	$\mathbf{D}_4$			τs			MAP of τs		
Segments	−3.12	1.02	2.10	$\tau_1 = 668.6$			$\tau_1 = 678.7$		
	0.2	−0.77	0.57	$\tau_2 = 1727.3$			$\tau_2 = 1761$		
	1.93	1.88	−3.81	$\tau_3 = 2654.6$			$\tau_3 = 2666.6$		

Table 1: The table lists the estimated $\mathbf{D}$, the locations of changepoint τs. The priors for the hyper parameters in $\mathbf{Q}$ and D are $\Gamma(1,1)$ and $\Gamma(2,2)$ respectively.

rate matrix $\mathbf{D} = (d_{ij})$, is written as

$$L = \prod_{i=1}^{r} \left(\prod_{\substack{1 \leq j \leq r \\ j \neq i}} d_{ij}^{N_{ij}} \right) e^{-d_i T_i},$$

where the sufficient statistics N_{ij} and T_i are the total transition numbers of $\mathbf{Y}(t)$ from i to j and the overall sojourn time of $\mathbf{Y}(t)$ in state i respectively. By choosing priors of $\mathbf{D}$ from the Gamma distribution $\Gamma(\alpha, \beta)$, which is a conjugate prior, the marginal likelihood of $\mathbf{Y}(t)$ can be explicitly written as:

$$\prod_{i=1}^{r} \prod_{\substack{1 \leq j \leq r \\ j \neq i}} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha + N_{ij})}{(\beta + T_i)^{(\alpha + N_{ij})}}. \tag{14}$$

We demonstrate two numerical examples for Markov jump processes. In the first numerical example, a 3-state Markov jump process is observed, which is partitioned by several changepoints at unknown locations. The trajectory of the observed Markov jump process and the exact locations of changepoints are displayed in the top of Figure 1. The parameters of the simulated Markov jump process are specified in Table 1. In this numerical example, each segment consists of 300 state transitions. We run the Gibbs sampler in Algorithm 3 a total of 6000 iterations, with the last 2800 iterations collected as independent draws. The posterior of the number of changepoints is indicated in the left top of Figure 2. In the figure, the dominant mode of the posterior of the number of changepoints is 3, which is consistent with the true model. The MAP estimate of changepoint locations given by the continuous-time Viterbi algorithm is listed in Table 1.
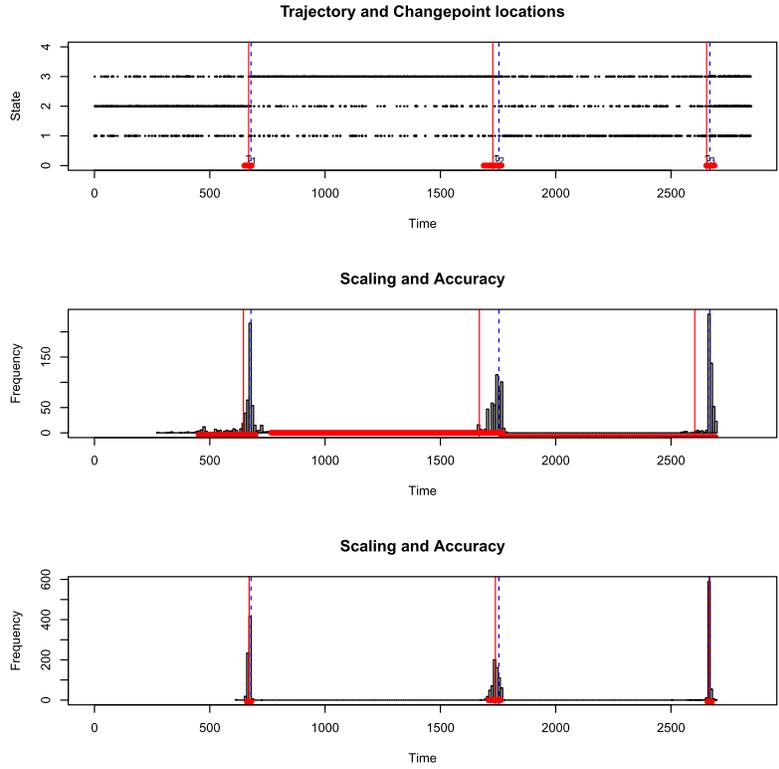
Figure 1: The top of the figure displays the trajectory of the simulated Markov jump process with 3 changepoints. The three plots indicate the effect of different values of k on the accuracies of the estimation at three levels, $k = 6, 3$ and 11 in the top, the middle and the bottom of the figure respectively. In all the plots, the posterior means and the 95% HPD credible intervals are demonstrated by solid red lines, with the exact positions of 3 changepoints indicated by dashed blue lines.

In the uniformization scheme, the Poisson intensity rate is k times the changepoint recurrence rate. In this numerical example, we also demonstrate the scaling effects of k on the efficiencies and accuracies of the estimates. The right of Figure 2 indicates the central processing unit (CPU) time and the effective sample size per iteration with 3 different scalings $k = 6, 3$ and 11. From the figure, it is indicated that the computational cost per iteration is quadratic to the scaling coefficient k. However, the effective sample size is only roughly linear with respect to the scaling parameter. In all cases, we observed a bias of the estimates for the locations of changepoints by posterior means (see Figure 1), which is a price paid for pruning nearly repeated changepoints. In Figure 1, the blue dash lines mark the exact locations of changepoints and the solid red lines mark the posterior means of changepoint locations. It is observed that all the posterior means of changepoints are located to the left of the exact changepoint locations. The degree of bias is obviously dependent on the scaling: see Figure 1. The bias is significant when
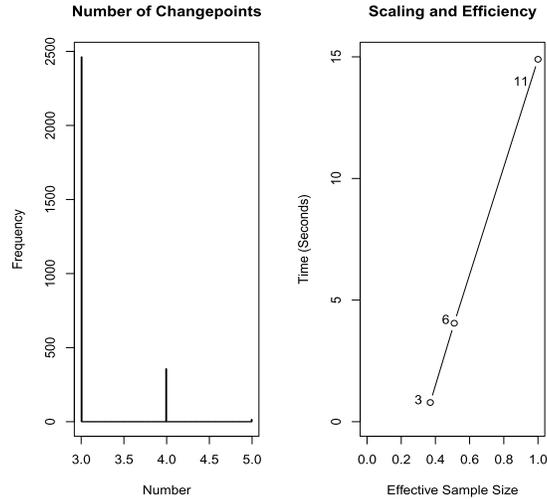
Figure 2: The left of the figure displays the posterior distribution of the number of changepoints. The right of the figure demonstrates the effect of the scaling parameter k on the efficiency of the estimation.

$k = 3$ in the middle of Figure 1, but negligible for $k = 11$ in the bottom of Figure 1. Further exploration shows that the 95% highest posterior density (HPD) intervals of changepoint locations, which are demonstrated in Figure 1 by red horizontal bold line segments, are also affected by the scaling parameter. For $k = 3$, the 95% HPD intervals of changepoint locations are so wide that the upper (lower) bounds of the posterior credible intervals are partially superimposed on adjacent one: see the middle of Figure 1. This is a poor estimate of the uncertainties of changepoint locations due to low resolution. With increasing k, the 95% HPD intervals of changepoint locations are narrowed. Further increasing k above 11 is unnecessary as it won't improve the accuracies of the credible intervals much. The posterior means of the transition rate matrix of the observed Markov jump processes are listed in Table 1. The estimates are close to the true values. Some of the posteriors of the transition intensity rate and their 95% highest posterior density (HPD) intervals are shown in Figure 3. From the figure, it is apparent that the estimation is reasonably accurate. In this numerical example, to alleviate the memory cost and the computational cost, the number of Poisson events generated from the uniformization scheme is upper-bounded by 200. The truncation error can be evaluated by (10). In practice, this upper bound is rarely reached and the truncation error is negligible.

The second numerical example is still a multiple changepoint model for Markov jump processes. The observed two-state Markov jump process is partitioned into 5 segments by 4 changepoints. In this case, each segment consists of only 70 state transitions. The transition intensity rate matrices of the observed Markov chain in adjacent segments are close to each other. The true parameters of the observed Markov jump processes are listed in Table 2. This numerical example is designed to make the detection of change-points more difficult. For this numerical example, the Gibbs sampler in Algorithm 3
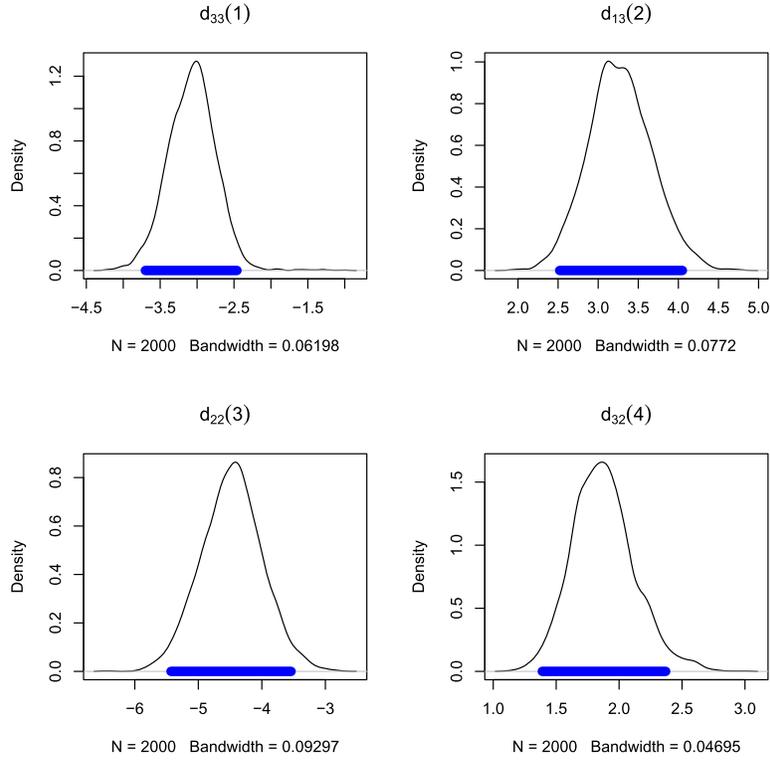
Figure 3: Kernel density estimates for some of the model parameters. Line segments in bold beneath each kernel density estimate indicate the 95% highest posterior density for the model parameters. The left top, right top, left bottom and right bottom of the figure show the posterior distributions of $d_{33}(1), d_{13}(2), d_{22}(3)$ and $d_{32}(4)$ respectively. The number of posterior samples and the bandwidth used in the kernel density estimation are given.

	$\mathbf{D}_1$		$\mathbf{D}_2$		$\mathbf{D}_3$		$\mathbf{D}_4$	
(a) True	-1	1	-0.5	0.5	-0.8	0.8	-2	2
	1	-1	0.2	-0.2	0.4	-0.4	1	-1
	$\mathbf{D}_5$		τs		τs			
Model	-1.5	1.5	$\tau_1 = 84.5$		$\tau_3 = 433.6$			
	0.8	-0.8	$\tau_2 = 312.2$		$\tau_4 = 484$			

Table 2: The transition rate matrices of the observed Markov jump process in 5 segments and the exact locations of changepoints. The priors for the hyper parameters in $\mathbf{Q}$ and D are $\Gamma(1,1)$ and $\Gamma(2,2)$ respectively.
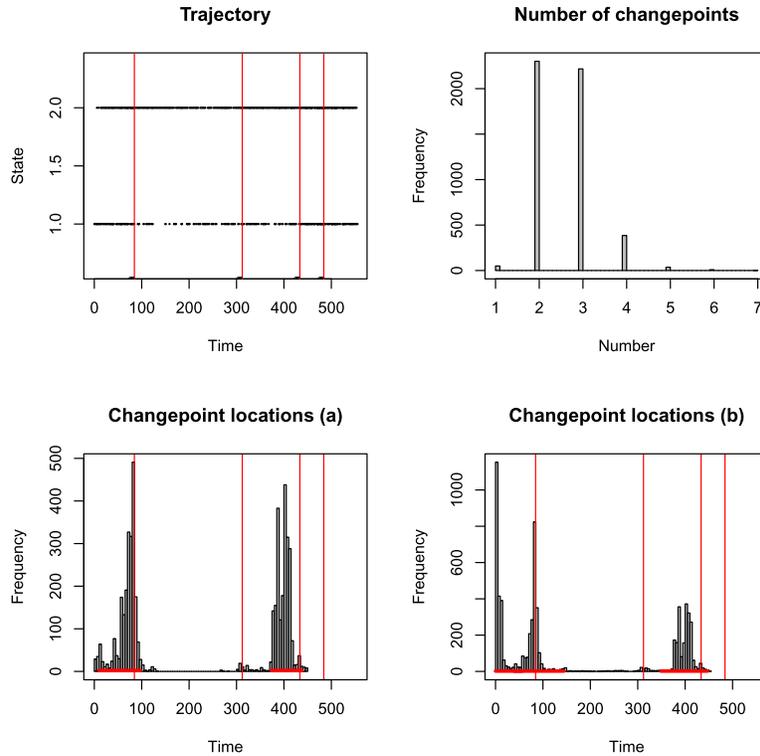
Figure 4: The left top of the figure shows the trajectory of the simulated Markov jump process with 4 changepoints marked by solid red lines. The right top of the figure indicates the posterior of the number of changepoints. The two plots in the bottom of the figure demonstrate the 95% HPD credible intervals of changepoint locations by solid red lines beneath the histograms for a two-changepoint model and a three-changepoint model respectively. The exact positions of the changepoints are indicated by solid red lines.

was run for 1000 burn-in iterations followed by 5000 iterations, which are collected as posterior samples. In the uniformization scheme, the Poisson intensity rate is 6 times the changepoint recurrence rate. We set an upper bound of 300 for the number of uniformization times. In all the simulations, the number of uniformization times rarely reaches this bound. The truncation error is negligible. The left top of Figure 4 indicates the trajectory of the observed Markov jump process and the exact locations of changepoints. The posterior of the number of changepoints is indicated in the right top of Figure 4. A bimodal pattern in the posterior suggests large uncertainties in the number of changepoints. In this case, the number of changepoints might be 2 or 3, which is less than the true values. We also demonstrate the posteriors of changepoint locations for a two-changepoint model and a three-changepoint model in the left bottom and right bottom of Figure 4 respectively, along with 95% HPD credible intervals. It is observed

from the bottom of Figure 4 that only the first and the third changepoints can be identified and other changepoints are simply missed because the observations are not informative about their locations. The MAP estimates of the changepoint locations for a two-changepoint model given by the continuous-time Viterbi algorithm are 84.5 and 410.7. The proposed model formulation is more flexible than the multiple changepoint model with fixed number of changepoints in Lu (2020).

In the third numerical example, we consider multiple changepoint models for a Poisson process $\mathbf{Y}(t)$ with the changepoint locations associated with a latent Markov chain $\mathbf{X}(t)$, so that when $\mathbf{X}(t)$ is in the i-th state, the intensity rate of $\mathbf{Y}(t)$ is parameterized by λ_i. The stochastic intensity rate of the doubly stochastic point process $\mathbf{Y}(t)$ is specified by $\lambda_{\mathbf{X}(t)}$. The likelihood of a Poisson process with intensity rate λ is given by $\lambda^N e^{-\lambda T}$, where N is the total number of Poisson arrivals and T is the overall observation time. After taking a conjugate prior $\Gamma(\alpha, \beta)$ and integrating out λ from the posterior, the marginal likelihood of $\mathbf{Y}(t)$ is exactly written as:

$$\frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + N)}{(\beta + T)^{(\alpha+N)}}. \tag{15}$$

A total of 10,000 Poisson events are partitioned into 5 segments by 4 changepoints, with 2,000 Poisson events in each segment. The Poisson intensity rates specified in the 5 segments are 4,1,3,5 and 2. For this relatively long sequence of observations, it would be very expensive to detect the sparsely distributed changepoints on an event-by-event scale. One may discretize the Poisson process into a sequence of discrete Poisson observations by splitting the time intervals evenly and applying extant methods to detect changepoints in a discrete-valued time series framework. To control the discretization error, one may choose a relatively small bin width. However, with increasing bin numbers, the computational costs and memory costs inflate quickly: see Fearnhead (2006). To lower the computational costs and memory costs, one may decrease the number of bins by choosing a large bin width. In this case, however, the discretization error will inflate and the estimation of the locations of changepoints will be at low resolution. The advantage of randomized time discretization via uniformization is that there is no time-discretization error and the changepoints can be detected directly on the scale of the number of changepoints.

For this numerical example, the Gibbs sampler in Algorithm 3 was run for 500 burn-in iterations followed by 500 iterations, which are collected as posterior samples. In the uniformization scheme, the Poisson intensity rate is 7 times the changepoint recurrence rate. We set an upper bound of 300 for the number of uniformization times. The number of uniformization times rarely reaches this bound. With this uniformization scheme, the FFBS algorithm is efficient to simulate changepoints from a moderate number of uniformization times, which is often less than 100. The left top of Figure 5 demonstrates the cumulative time of 10,000 simulated Poisson events. The posterior of the number of changepoints is indicated in the right top of Figure 5. The dominant mode in the posterior of the number of changepoints is 4. The posterior of the locations of changepoints is demonstrated in the bottom of Figure 5, with 95% HPD intervals of the changepoint locations and their posterior mean estimates indicated in the figure. The changepoint
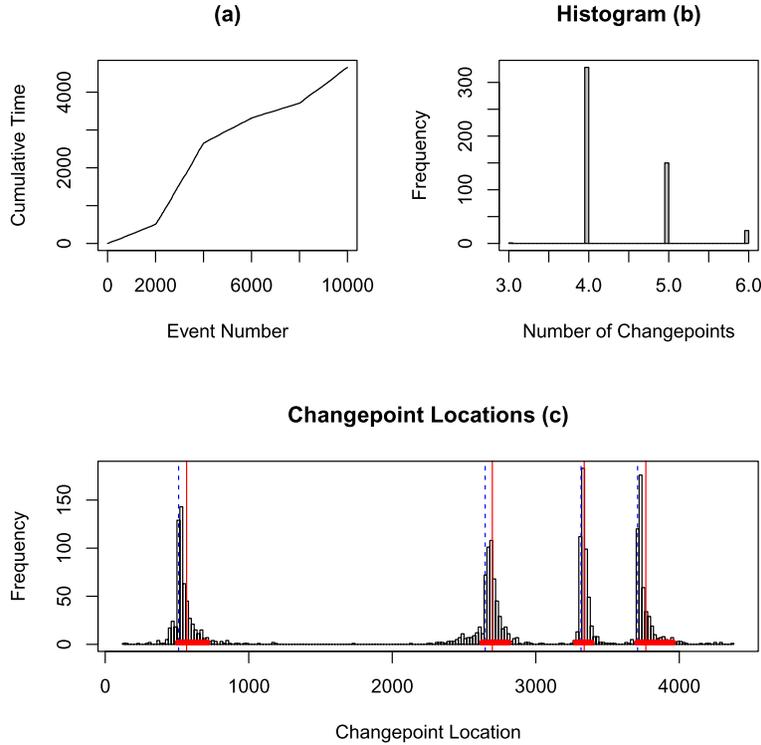
Figure 5: The left top of the figure displays the cumulative event times of 10,000 simulated Poisson observations with 4 changepoints. The right top of the figure indicates the posterior of the number of changepoints. The bottom of the figure shows the posterior means and the 95% HPD credible intervals of changepoint locations as solid red lines, with the histograms of changepoint locations displayed and exact positions of changepoints marked by dashed lines. The priors for the hyper parameters in $\mathbf{Q}$ and λ are $\Gamma(2, 2)$ and $\Gamma(2, 2)$ respectively.

locations given by the MAP estimation via the Viterbi algorithm are at event times 2002, 4001, 6001 and 8012.

The fourth numerical example is a mimic of the real data analysis in the next numerical example. A small number of Poisson events are partitioned by 2 changepoints, with 230 observations in each segment. The Poisson intensity rates in the 3 segments are 15, 5 and 15. For this relatively small number of observations, the Gibbs sampler in Algorithm 3 was run for 3000 iterations, with the last 2500 iterations collected as posterior samples. The Poisson rate used in the uniformization scheme is 7 times the changepoint recurrence rate. Often, only a dozen of uniformization times are generated and the upper bound of 300 for the number of uniformization times is never reached. The left top of Figure 6 displays the cumulative time of 690 Poisson events. The right top of Figure 6 indicates the posterior of the number of changepoints. The dominant mode
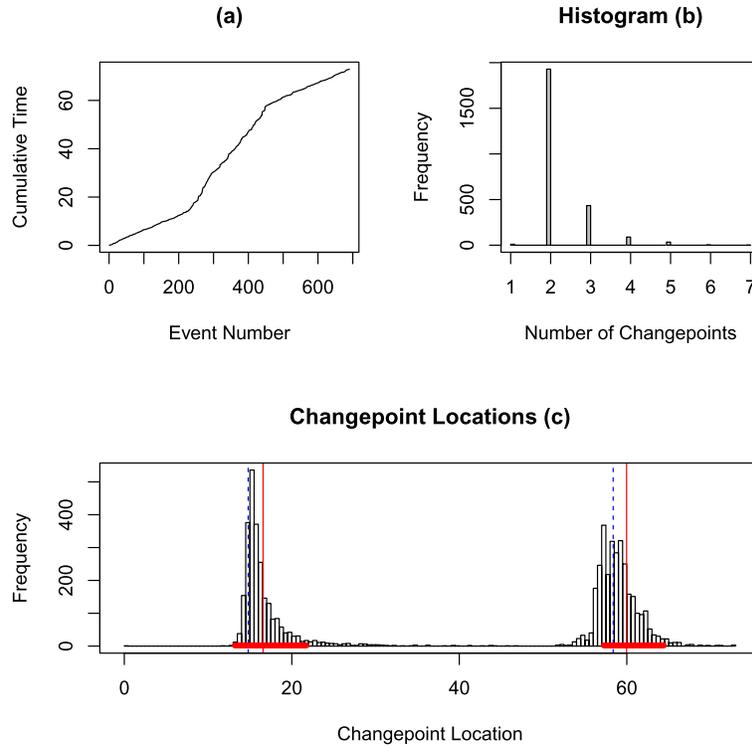
Figure 6: The left top of the figure displays the cumulative event times of 690 Poisson observations with 2 changepoints. The right top of the figure indicates the posterior of the number of changepoints. The bottom of the figure shows the posterior means and the 95% HPD credible intervals of changepoint locations as solid red lines, with the histograms of changepoint locations displayed and exact positions of changepoints marked by dashed lines. The priors for the hyper parameters in **Q** and λ are $\Gamma(2, 2)$ and $\Gamma(2, 2)$ respectively.

in the posterior of the number of changepoints is 2. The posterior of the locations of changepoints is shown in the bottom of Figure 6, with 95% HPD intervals of the changepoint locations and their posterior mean estimates indicated in the figure. The MAP estimates of changepoint locations via the Viterbi algorithm are eventtimes 228 and 448.

4.2 Time-Varying Patterns of Deep Earthquakes via Multiple Changepoint Models

We apply the method to model the occurrence patterns of New Zealand deep earthquakes. The data set studied in this analysis is the New Zealand earthquake catalogue, freely available from GNS Science of New Zealand via Geonet (www.geonet.org.nz). We choose those events from the New Zealand catalogue within a poly-
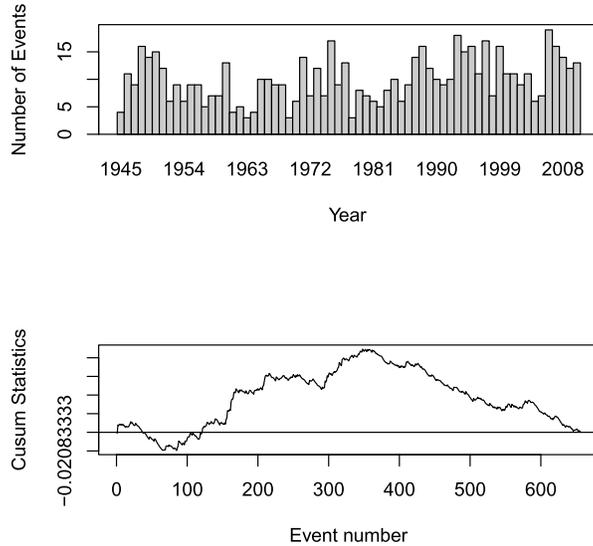
Figure 7: The top of the figure is the yearly counts of deep earthquakes. The bottom of the figure is the centralized and normalized cumulative sum of the inter-event times for deep earthquakes.

gon with vertices $(171°E, 43°S)$, $(172.5°E, 39.3°S)$, $(176°E, 36.25°S)$, $(177°E, 36.25°S)$, $(178.7°E, 37.6°S)$, $(178°E, 40°S)$ and $(174°E, 43°S)$, at depth greater than 45 km with magnitude above 5 in the New Zealand version of local magnitude scale, from 1945 until 2011. The total number of recorded earthquakes is 646. Nearly the same data set is analysed in Lu (2018). The occurrence time distribution of New Zealand deep earthquakes is displayed in the top of Figure 7. From Figure 7, it seems that the yearly counts of New Zealand deep earthquakes varies form time to time. Whether the time-varying behavior of the earthquake occurrence rate is associated with structural breaks or it is just random fluctuation is of importance for seismic hazard evaluation and seismic risk forecasts. The centralized and normalized cumulative sum of inter-event times Δt_i, which is given by $\frac{\sum_{i=1}^{j} \Delta t_i}{\sum_{i=1}^{n} \Delta t_i} - \frac{j}{n}$, is indicated in the bottom of Figure 7. This statistic behaves as a Brownian bridge on $[0,1]$ if the observations are from a time-homogeneous Poisson process. Otherwise, if the statistic deviates from the line segment: $y = 0, 0 \leq x \leq 1$ significantly, the observations should not be regarded as a homogeneous Poisson process: see Galeano (2007) and Yang and Kuo (2001) for the detection of changepoints of Poisson processes. From the bottom of Figure 7, it appears that a single-changepoint or a two-changepoint model might be suitable to characterize the time-varying behavior of the earthquake occurrence rate.

In this numerical example, the Gibbs sampling scheme iterates 8000 times, with the last 2,000 samples treated as the posterior samples. We set an upper bound of 200 for the number of Poisson event numbers produced by the uniformization scheme. The truncation error is negligible. The left top of Figure 8 shows the magnitude versus
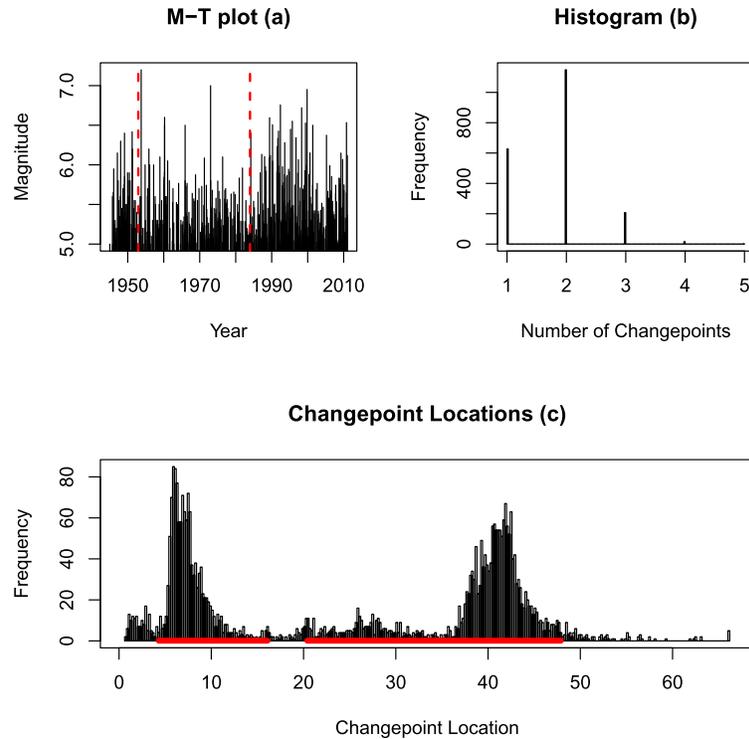
Figure 8: The left top of the figure is the magnitude versus time plot of the deep earthquakes, with the posterior means of changepoint locations marked by red dash lines. The right top of the figure indicates the posterior of the number of changepoints. The bottom of the figure demonstrates the 95% HPD credible intervals of changepoint locations by solid red lines and their histograms. The priors for the hyper parameters in **Q** and λ are $\Gamma(2,2)$ and $\Gamma(2,2)$ respectively.

time (M-T) plot of deep earthquakes, with the locations of changepoints indicated by vertical lines. The posterior of the number of changepoints is shown in the right top of Figure 8. From the figure, the MAP estimation of the number of changepoints is 2. The posterior probability mass for the number of changepoints equal to or above three is almost negligible. For the two-changepoint model, the posterior means for the locations of changepoints are about 1953 and 1987, with 95% credible intervals indicated by the red lines in boldface in the bottom of Figure 8. The credible interval for the changepoint at 1987 looks a bit over-diffused, which might be attributed to the Poisson rate varying with some sort of trend. The MAP estimation of the changepoint locations is at 1957.5 and 1991.2. In this numerical example, we also run the RJMCMC (Green 1995) algorithm for a significantly longer time to obtain the posterior of the number of changepoints of deep earthquakes (Figure 9), which is very diffusive in comparison to the right top of Figure 8. Generally, these methods have problems of poor mixing and highly correlated posterior samples.
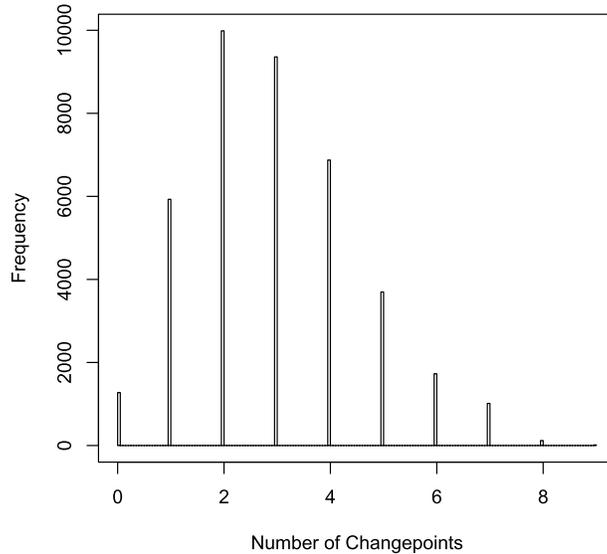
Figure 9: The histogram of changepoint numbers produced from 40,000 posterior samples collected after a significant longer burning time via the RJMCMC method. The prior distribution of the number of changepoints is uniformly distributed between 0 and $k_{\max} = 65$. The prior distribution for the intensity rate parameter λ is $\Gamma(2, 2)$.

References

Arnesen P., Holsclaw T. and Smyth P. (2016) "Bayesian Detection of Change points in Finite-State Markov Chains for Multiple Sequences." *Technometrics*, 58(2):205–213. MR3488299. doi: https://doi.org/10.1080/00401706.2015.1044118.   531

Barry D. and Hartigan J. A. (1992) "Product partition models for change point problems." *The Annals of Statistics*, 20(1):260–279. MR1150343. doi: https://doi.org/10.1214/aos/1176348521.   522, 524

Bebbington M. S. (2007) "Identifying volcanic regimes using Hidden Markov Models." *Geophysical Journal International*, 171(2):921–942.   530, 531

Chib S. (1998) "Estimation and Comparison of Multiple Change-Point Models." *Journal of Econometrics*, 86(2):221–241. MR1649222. doi: https://doi.org/10.1016/S0304-4076(97)00115-2.   521, 522, 525

Chopin N. (2007) "Dynamic detection of change points in long time series." *Annals of the Institute of Statistical Mathematics*, 59(2):349–366. MR2394170. doi: https://doi.org/10.1007/s10463-006-0053-9.   521

Fearnhead P. (2006) "Exact and Efficient Bayesian Inference for Multiple Changepoint Problems." *Statistics and Computing*, 16(2):203–213. MR2227396. doi: https://doi.org/10.1007/s11222-006-8450-8.   521, 537

Fearnhead P. and Liu Z. (2007) "On-line inference for multiple changepoint problems." *Journal of the Royal Statistical Society B*, 69(4):589–605. MR2370070. doi: https://doi.org/10.1111/j.1467-9868.2007.00601.x. 521, 522, 525, 526

Frick K., Munk A. and Sieling H. (2014) "Multiscale change point inference." *Journal of the Royal Statistical Society B*, 76(3):495–580. MR3210728. doi: https://doi.org/10.1111/rssb.12047. 521

Fryzlewicz P. (2014) "Wild binary segmentation for multiple change-point detection." *Annals of Statistics*, 42(6):2243–2281. MR3269979. doi: https://doi.org/10.1214/14-AOS1245. 521

Galeano P. (2007) "The use of cumulative sums for detection of changepoints in the rate parameter of a Poisson Process." *Computational Statistics and Data Analysis*, 51(12):6151–6165. MR2407705. doi: https://doi.org/10.1016/j.csda.2006.12.042. 540

Green P. J. (1995) "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination." *Biometrika*, 82(4):711–732. MR1380810. doi: https://doi.org/10.1093/biomet/82.4.711. 521, 541

Jensen A. (1953) "Markoff chains as an aid in the study of markoff processes." *Scandinavian Actuarial Journal*, 1953(Supp 1), 87–91. MR0057488. doi: https://doi.org/10.1080/03461238.1953.10419459. 525

Killick R., Fearnhead P. and Eckley I. A. (2012) "Optimal Detection of Changepoints With a Linear Computational Cost." *Journal of the American Statistical Association*, 107(500):1590–1598. MR3036418. doi: https://doi.org/10.1080/01621459.2012.737745. 521

Ko S., Chong T. and Ghosh P. (2015) "Dirichlet process hidden Markov multiple changepoint Model." *Bayesian Analysis*, 10(2):275–296. MR3420883. doi: https://doi.org/10.1214/14-BA910. 522

Lavielle M. and Lebarbier E. (2001) "An Application of MCMC Methods for the Multiple Change-Points Problem." *Signal Processing*, 81(1):39–53. 521

Lewis P. A. W. and Shedler G. S. (1979) "Simulation of nonhomogeneous Poisson processes by thinning." *Naval Research Logistics Quarterly*, 26(3):403–413. MR0546120. doi: https://doi.org/10.1002/nav.3800260304. 528

Liu J. S. (1994) "The collapsed gibbs sampler in Bayesian computations with applications to a gene regulation problem." *Journal of the American Statistical Association*, 89(427):958–966. MR1294740. 528

Lu S. (2019) "A Bayesian multiple changepoint model for marked Poisson processes with applications to deep earthquakes." *Stochastic Environmental Research and Risk Assessment*, 33(1):59–72. 522, 530, 531

Lu S. (2020) "Bayesian multiple changepoint detection for Markov jump processes." Computational Statistics. doi: https://doi.org/10.1007/s00180-020-00956-6. 522, 530, 531, 537

Martinez A. F. and Mena R. H. (2014) "On a nonparametric change point detection model in Markovian regimes." *Bayesian Analysis*, 9(4):823–858. MR3293958. doi: https://doi.org/10.1214/14-BA878. 522

Polansky A. M. (2007) "Detecting change-points in Markov chains." *Computational Statistics and Data Analysis*, 51(12):6013–6026. MR2407695. doi: https://doi.org/10.1016/j.csda.2006.11.040. 531

Rao V. and Teh Y. W. (2013) "Fast MCMC Sampling for Markov Jump Processes and Extensions." *Journal of Machine Learning Research*, 14(1):3295–3320. MR3144463. 522, 527

Ruggieri E. and Lawrence C. E. (2014) "The Bayesian Change Point and Variable Selection Algorithm: Application to the δ18O Proxy Record of the Plio-Pleistocene." *Journal of Computational and Graphical Statistics*, 23(1):87–110, MR3173762. doi: https://doi.org/10.1080/10618600.2012.707852. 521

Scott S. L. (2002) "Bayesian methods for hidden Markov models: Recursive computing in the 21st century." *Journal of the American Statistical Association*, 97(457):337–351. MR1963393. doi: https://doi.org/10.1198/016214502753479464. 522, 525

Stephens D. A. (1994) "Bayesian Retrospective Multiple-Changepoint Identification." *Applied Statistics*, 43(1):159–178. 521

Van Dijk N. M. (1992) "Uniformization for nonhomogeneous Markov chains." *Operations Research Letters*, 12(5): 283–291. MR1189730. doi: https://doi.org/10.1016/0167-6377(92)90086-I. 528

Van Dijk N. M., Van Brummelen S. P. and Boucherie R. J. (2018) "Uniformization: Basics, extensions and applications." *Performance Evaluation*, 118:8–32. 530

Viterbi A. J. (1967) "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm." *IEEE Transactions on Information Theory*, 13(2):260–269. 530

Xing H., Sun N. and Chen Y. (2012) "Credit Rating Dynamics in the Presence of Unknown Structural Breaks." *Journal of Banking and Finance*, 36(1):78–89. 531

Yang T. Y. and Kuo L. (2001) "Bayesian binary segmentation procedure for a Poisson Process with multiple changepoints." *Journal of Computational and Graphical Statistics*, 10(4):772–785. MR1938979. doi: https://doi.org/10.1198/106186001317243449. 540

Yildirim S., Singh S. S. and Doucet A. (2013) "An online expectation–maximization algorithm for changepoint models." *Journal of Computational and Graphical Statistics*, 22(4), 906–926. MR3173749. doi: https://doi.org/10.1080/10618600.2012.674653. 522