BMRMM: An R Package for Bayesian Markov (Renewal) Mixed Models

by Yutong Wu and Abhra Sarkar

Abstract We introduce the BMRMM package implementing Bayesian inference for a class of Markov renewal mixed models which can characterize the stochastic dynamics of a collection of sequences, each comprising alternative instances of categorical states and associated continuous duration times, while being influenced by a set of exogenous factors as well as a 'random' individual. The default setting flexibly models the state transition probabilities using mixtures of Dirichlet distributions and the duration times using mixtures of gamma kernels while also allowing variable selection for both. Modeling such data using simpler Markov mixed models also remains an option, either by ignoring the duration times altogether or by replacing them with instances of an additional category obtained by discretizing them by a user-specified unit. The option is also useful when data on duration times may not be available in the first place. We demonstrate the package's utility using two data sets.

1 Introduction

Markov models (MMs) are widely used for modeling the transition dynamics of categorical state sequences. Classical Markov renewal models (MRMs) additionally model the state duration times, when available, where the state transitions follow Markov dynamics and the state duration times follow a continuous distribution that depends on the immediately preceding and following states (Figure 1). M(R)Ms have been widely used in different variations (Phelan, 1990; Eichelsbacher and Ganesh, 2002; Muliere et al., 2003; Alvarez, 2005; Diaconis and Rolles, 2006; Bulla and Muliere, 2007; Etterson et al., 2007; Bacallado et al., 2009; Li, 2009; Epifani et al., 2014; Siebert and Söding, 2016; Holsclaw et al., 2017; Sesia et al., 2019). There are also some sparse works on covariate-dependent Markov models (Muenz and Rubinstein, 1985; Gradner, 1990; Alioum et al., 1998; Islam and Chowdhury, 2006).

The existing literature however focuses very heavily on modeling single sequences. Sarkar et al. (2018) developed a highly flexible and computationally efficient class of Bayesian Markov mixed models (BMMMs) for jointly modeling a collection of categorical sequences, each one associated with an individual as well as a set of time-invariant external covariates (e.g., the sex and genotype of the associated individual, an experimental condition under which the sequence was generated, etc.). BMMMs characterize the state transition probabilities using a convex combination of a fixed covariate-dependent component and a random individual-level component, both being Dirichlet distributed. They further allow covariate levels with similar effects to be probabilistically clustered together, allowing automatic selection of the significant covariates, and providing a sophisticated framework for analyzing data sets having the aforementioned structure.

BMMMs however do not model duration times of the states which are often additionally available in real-world applications. Recently, Wu et al. (2023) extended BMMMs to Bayesian Markov renewal mixed models (BMRMMs), allowing for the additional analysis of continuous duration times which, depending on the application, can either be the duration for which a state persists or the duration between two consecutive states, i.e., inter-state intervals (ISIs). Specifically, they modeled the duration times using mixtures of gamma kernels with mixture probabilities being a convex combination a covariate-dependent effect and an individual-level effect, similar to BMMMs. Covariate levels with similar influences on mixture probabilities are clustered together in BMRMMs as well, allowing the selection of the significant covariates. BMRMMs thus holistically model both state transitions and continuous duration times, painting a comprehensive picture of the underlying stochastic dynamics.

In this article, we describe the R package BMRMM which implements BMMMs and BMRMMs, collectively referred to henceforth as BM(R)MMs. The **BMRMM** package runs posterior inference for categorical state transitions and continuous duration times, if available, via a Markov chain Monte Carlo (MCMC) algorithm, returning an object containing comprehensive inference results. The package also includes a suite of plotting functions to display the results graphically. Specifically for continuous duration times, when available, the package provides users with three different options: (i) ignore the duration times and model the state transitions alone as a BMMM; (ii) introduce an additional category by discretizing the continuous duration times by a user-specified unit, and analyze the appended state transitions as a BMMM; (iii) model the duration as a mixture of gamma kernels using a BMRMM, as proposed in Wu et al. (2023). Additionally, users can choose to turn off one or both of the fixed covariate effects and the random individual effects. Overall, the **BMRMM** package thus gives users a lot of flexibility in handling their data sets, providing inferences for both Bayesian Markov renewal or non-renewal models as needed.

The **BMRMM** package conveniently includes a synthetic `foxp2` data set that describes the laboratory study on the role of the FoxP2 gene implicated in speech deficiencies for adult mice, which is also the motivating application for the methodology of BMMMs and BMRMMs. Mutations in the FoxP2 gene have long been associated with severe deficits in vocal communication for mammals (MacDermot et al., 2005). Mice with and without the mutation singing under various "social contexts" have thus been studied in many experiments (Fujita et al., 2008; Castellucci et al., 2016; Gaub et al., 2016; Chabout et al., 2016). The FoxP2 data set (Chabout et al., 2016), e.g., comprises the sequences of syllables making up the songs as well as the lengths of inter-syllable intervals (ISIs). The data set `foxp2` included in the **BMRMM** package is taken from the simulation study of Wu et al. (2023). It is much shorter than the real FoxP2 data set but closely mimics its other aspects and is used in this paper to demonstrate how to obtain detailed inferences for both syllable transitions and ISI dynamics for a comprehensive analysis of the vocal repertoire in mice with and without the FoxP2 mutation.

The utility of the **BMRMM** package goes well beyond the FoxP2 data set. As described above, the package is designed for scenarios where the data set consists of categorical state sequences associated with an individual as well as a number of observed covariates where additional data on continuous duration times may or may not be available. Data sets with such structures are frequently observed in different areas of scientific research and can potentially benefit from the **BMRMM** package. For example, Islam and Chowdhury (2006) analyzed the transitions of different rainfall orders in three districts of Bangladesh under three covariates, wind speed, humidity, and maximum temperature. In an education assessment study, Zhang et al. (2019) recorded sequences of writing states, characterized by keystroke logs, for 257 eighth graders of various genders, races, and socioeconomic statuses. Combescure et al. (2003) estimated the control states of 371 asthma patients with different body mass indices (BMIs) and disease severity over a four-year-long study and produced the asthma control data set, which we will use as an additional example to demonstrate the utility of our package in this paper. For such data sets, the **BMRMM** package provides a flexible, sophisticated, and principled way to model fixed effects of the covariates and random effects of the individuals in both the state transition dynamics and the distribution of the ISIs.

Other computer programs for Markov models with covariates include MARKOV (Marshall et al., 1995) and MKVPCI (Alioum and Commenges, 2001). R packages for analyzing discrete Markov models include **markovchain** (Spedicato, 2017) and **msm** (Jackson, 2011). **SemiMarkov** (Listwon and Saint-Pierre, 2015) and **SMM** (Barbu et al., 2018) provide functions for the simulation and estimation of traditional semi-Markov models. Some R packages provide the inference of hidden semi-Markov models, such as **mhsmm** (O'Connell and Højsgaard, 2011) and **hhsmm** (Amini et al., 2022). Ferguson et al. (2012) built the **msSurv** package which provides a nonparametric estimation of semi-Markov models but does not consider covariates. There are also R packages implementing MRPs in specific application areas. For example, Kharrat et al. (2019) introduced the **Countr** package for flexible regression models based on MRPs. Pustejovsky (2021) developed the **ARPobservation** for simulating behavior streams based on alternating renewal processes. Other R packages for categorical data analysis include **catdap** (Katsura, 1980) and **vcd** (Meyer et al., 2022). To our knowledge, there has not been an R package that implements flexible Bayesian M(R)MMs.

In the following section, we summarize the technical details of BMRMMs. Documentation for the functions of the **BMRMM** package is then provided. Next, we demonstrate the usage of our package in analyzing two different data sets. The final section contains some concluding remarks.

2 The Bayesian Markov (renewal) mixed models

We briefly describe the BM(R)MM methodologies here – more details can be found in Sarkar et al. (2018) and Wu et al. (2023). Consider specifically a sequence s comprising T_s state instances and let $y_{s,t}$ denote the state at time t in sequence s. The states $y_{s,t}$'s come from a set $\mathcal{Y} = \{1, 2, \ldots, d_0\}$. Within a sequence s, there are $T_s - 1$ duration times (state persistence times or inter-state intervals), denoted by $\{\tau_{s,t}\}_{s=1,t=2}^{s_0,T_s}$, where $\tau_{s,t}$ is the duration between the $(t-1)^{th}$ and t^{th} states in sequence s, and s_0 represents the total number of sequences. Figure 1 presents a graphical summary of the data structure. Each sequence s is associated with p categorical covariates or factors, denoted by $x_{s,j} \in \mathcal{X}_j = \{1, 2, \ldots, d_j\}$, and an individual, denoted by i_s. Without loss of generality, we assume that the analyses of the transition probabilities and the duration times distributions both include all p covariates. Moreover, the analysis of duration times counts the previous state $y_{s,t-1}$ as an additional $(p+1)^{th}$ covariate. In the **BMRMM** package, users have the flexibility to select particular covariates for each analysis and exclude the previous state from the analysis of duration times. In their original definition in Pyke (1961), the duration time $\tau_{s,t}$ in an MRP was allowed to depend on both the preceding state $y_{s,t-1}$ and the following state $y_{s,t}$. To keep the notation simple and the methodology easy to understand for a broad audience, we however only include the preceding state $y_{s,t-1}$ as a predictor of $\tau_{s,t}$ in this paper. This analysis can be easily modified to have the pair $(y_{s,t-1}, y_{s,t})$ as a predictor instead of just $y_{s,t-1}$, as was actually done in Wu et al. (2023).

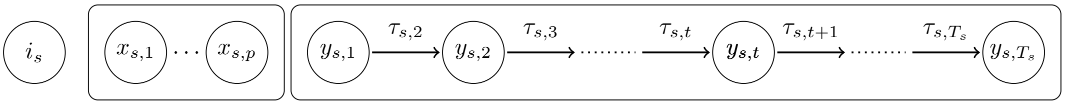


Figure 1: Graphical model showing the data structure: $y_{s,t}$ denotes the observed state at the t^{th} time location in the s^{th} sequence; $\tau_{s,t}$ denotes the observed duration times (either state persistence times or inter-state intervals) between the states $y_{s,t-1}$ and $y_{s,t}$; each sequence s is also associated with an individual i_s and a set of exogenous time-invariant covariates $x_{s,1}, \ldots, x_{s,p}$. The Markov mixed model considered in this article analyzes the state transitions $y_{s,t}$ in a collection of sequences; the Markov renewal mixed model additionally analyzes the duration times $\tau_{s,t}$; both models accommodate fixed effects of the covariates $x_{s,1}, \ldots, x_{s,p}$ and random effects of the individuals i_s.

2.1 Model for state transitions

For a sequence s associated with individual i and covariate levels $x_1, \ldots, x_p$, the transition probabilities $\Pr(y_{s,t} = y_t \mid i_s = i, x_{s,1} = x_1, \ldots, x_{s,p} = x_p, y_{s,t-1} = y_{t-1}) = P^{(i)}_{trans,x_1,\ldots,x_p}(y_t \mid y_{t-1})$ are defined as a convex combination of a fixed covariate effect component $\boldsymbol{\lambda}_{trans,x_1,\ldots,x_p}(\cdot \mid y_{t-1})$ and a random effect component $\boldsymbol{\lambda}^{(i)}_{trans}$:

$$P^{(i)}_{trans,x_1,\ldots,x_p}(y_t \mid y_{t-1}) = \pi^{(i)}_{trans,0}(y_{t-1})\lambda_{trans,x_1,\ldots,x_p}(y_t \mid y_{t-1}) + \pi^{(i)}_{trans,1}(y_{t-1})\lambda^{(i)}_{trans}(y_t \mid y_{t-1}). \tag{1}$$

The coefficients of the convex combination, namely, $\{\pi^{(i)}_{trans,0}(y_{t-1}), \pi^{(i)}_{trans,1}(y_{t-1})\}$ are individual-specific and satisfy $\pi^{(i)}_{trans,1}(y_{t-1}) = 1 - \pi^{(i)}_{trans,0}(y_{t-1})$.

For each covariate $j = 1, \ldots, p$, it is possible that some covariate levels exert a similar effect on the transition dynamics. For example, the components $\boldsymbol{\lambda}_{trans,x_1=1,x_2,\ldots,x_p}(\cdot \mid y_{t-1})$ and $\boldsymbol{\lambda}_{trans,x_1=2,x_2,\ldots,x_p}(\cdot \mid y_{t-1})$ would be equal if levels 1 and 2 of covariate 1 have similar influences on transition dynamics for fixed levels for covariates $2, \ldots, p$. A clustering mechanism for covariate levels allows the fixed component $\boldsymbol{\lambda}_{trans,x_1,\ldots,x_p}(\cdot \mid y_{t-1})$ to be the same for all levels with a similar influence. In particular, for covariate j, we construct the partition $\mathcal{C}^{(j)}_{trans} = \{\mathcal{C}^{(j)}_{trans,h_j}\}^{k_{trans,j}}_{h_j=1}$ of its levels, where $k_{trans,j}$ is the number of clusters for covariate j and h_j represents the cluster index. We introduce latent variables $\{z_{trans,j,\ell}\}^{p,d_j}_{j=1,\ell=1}$ that indicate the cluster index for the ℓ^{th} label of covariate j. Two levels of the covariate j, $\ell_1, \ell_2 \in \mathcal{X}_j = \{1, \ldots, d_j\}$, are clustered together if and only if $z_{trans,j,\ell_1} = z_{trans,j,\ell_2}$. For the fixed effects, we then replace the covariate levels $x_1, \ldots, x_p$'s with cluster indices $h_1, \ldots, h_p$'s and present the fixed effect as $\boldsymbol{\lambda}_{trans,h_1,\ldots,h_p}(\cdot \mid y_{t-1})$.

We set Dirichlet priors for both fixed and individual effect components and let them center around the same mean vector $\boldsymbol{\lambda}_{trans,0}$ to facilitate posterior computation. The probability vector $\boldsymbol{\lambda}_{trans,0}$ is also given a Dirichlet prior with mean $\boldsymbol{\lambda}_{trans,00}$ to capture the natural preferences of certain states in $\mathcal{Y}$:

$$\begin{aligned}
&\boldsymbol{\lambda}_{trans,h_1,\ldots,h_p}(\cdot \mid y_{t-1}) \sim \text{Dir}\left\{\alpha_{trans,0}\lambda_{trans,0}(1 \mid y_{t-1}), \ldots, \alpha_{trans,0}\lambda_{trans,0}(d_0 \mid y_{t-1})\right\}, \\
&\boldsymbol{\lambda}^{(i)}_{trans}(\cdot \mid y_{t-1}) \sim \text{Dir}\left\{\alpha^{(0)}_{trans}\lambda_{trans,0}(1 \mid y_{t-1}), \ldots, \alpha^{(0)}_{trans}\lambda_{trans,0}(d_0 \mid y_{t-1})\right\}, \\
&\boldsymbol{\lambda}_{trans,0}(\cdot \mid y_{t-1}) \sim \text{Dir}\left\{\alpha_{trans,00}\lambda_{trans,00}(1), \ldots, \alpha_{trans,00}\lambda_{trans,00}(d_0)\right\}.
\end{aligned}$$

We present the complete Bayesian hierarchical model for the transition dynamics as

$$\begin{aligned}
&(y_{s,t} \mid y_{s,t-1} = y_{t-1}, i_s = i, z_{trans,1,x_{s,1}} = h_1, \ldots, z_{trans,p,x_{s,p}} = h_p) \sim \\
&\qquad \text{Mult}\left\{P^{(i)}_{trans,h_1,\ldots,h_p}(1 \mid y_{t-1}), \ldots, P^{(i)}_{trans,h_1,\ldots,h_p}(d_0 \mid y_{t-1})\right\}, \text{ where} \\
&\mathbf{P}^{(i)}_{trans,h_1,\ldots,h_p}(\cdot \mid y_{t-1}) = \pi^{(i)}_{trans,0}(y_{t-1})\boldsymbol{\lambda}_{trans,h_1,\ldots,h_p}(\cdot \mid y_{t-1}) + \pi^{(i)}_{trans,1}(y_{t-1})\boldsymbol{\lambda}^{(i)}_{trans}(\cdot \mid y_{t-1}), \\
&z_{trans,j,\ell} \sim \text{Mult}\left\{\mu_{trans,j}(1), \ldots, \mu_{trans,j}(d_j)\right\}, \quad \boldsymbol{\mu}_{trans,j} \sim \text{Dir}(\alpha_{trans,j}, \ldots, \alpha_{trans,j}), \\
&\boldsymbol{\lambda}_{trans,h_1,\ldots,h_p}(\cdot \mid y_{t-1}) \sim \text{Dir}\left\{\alpha_{trans,0}\lambda_{trans,0}(1 \mid y_{t-1}), \ldots, \alpha_{trans,0}\lambda_{trans,0}(d_0 \mid y_{t-1})\right\}, \\
&\boldsymbol{\lambda}^{(i)}_{trans}(\cdot \mid y_{t-1}) \sim \text{Dir}\left\{\alpha^{(0)}_{trans}\lambda_{trans,0}(1 \mid y_{t-1}), \ldots, \alpha^{(0)}_{trans}\lambda_{trans,0}(d_0 \mid y_{t-1})\right\}, \\
&\boldsymbol{\lambda}_{trans,0}(\cdot \mid y_{t-1}) \sim \text{Dir}\left\{\alpha_{trans,00}\lambda_{trans,00}(1), \ldots, \alpha_{trans,00}\lambda_{trans,00}(d_0)\right\}, \\
&\pi^{(i)}_{trans,0}(y_{t-1}) \sim \text{Beta}(a_{trans,0}, a_{trans,1}), \\
&\alpha_{trans,0} \sim \text{Ga}(a_{trans,0}, b_{trans,0}), \qquad \alpha^{(0)}_{trans} \sim \text{Ga}(a^{(0)}_{trans}, b^{(0)}_{trans}).
\end{aligned}$$

2.2 Model for continuous duration times

The **BMRMM** package provides three options for analyzing the duration times: (i) ignore the durations altogether and only model the transition probabilities of the existing states, (ii) treat the durations as blocks of a new special category, with a discretization unit specified by users, (iii) model the durations as a continuous random variable with a flexible mixture of gamma distributions. For the first two options, we only need to apply the model described in the previous subsection. For the third option, we need to conduct a separate analysis of the duration times as described below.

Let K denote the number of gamma mixture components in the model for the duration times. We let $\mathbf{P}^{(i)}_{dur}(\cdot \mid x_1, \ldots, x_p, y_{s,t-1})$ denote the mixture probability vector given individual i, covariate levels $x_1, \ldots, x_p$, and preceding syllable $y_{s,t-1}$. The preceding state $y_{s,t-1}$ can be removed from the formula if the users do not wish to consider its influence in the inference of the duration times. The distribution of the continuous duration times, $\{\tau_{s,t}\}_{s=1,t=2}^{s_0,T_s}$, is then modeled as

$$\begin{aligned} &f(\tau_{s,t} \mid i_s = i, x_{s,1} = x_1, \ldots, x_{s,p} = x_p, y_{s,t-1} = y_{t-1}) \\ &= \textstyle\sum_{k=1}^{K} P^{(i)}_{dur}(k \mid x_1, \ldots, x_p, y_{t-1}) \mathrm{Ga}(\tau_{s,t} \mid \alpha_k, \beta_k), \end{aligned}$$

where α_k and β_k denote the shape and rate parameters of the k^{th} gamma mixture component, respectively. We introduce a set of latent variables $\{z_{dur,s,t}\}_{s=1,t=2}^{s_0,T_s}$ that represents the index of the mixture component. If $z_{dur,s,t}$ equals to k, then $\tau_{s,t}$ follows $\mathrm{Ga}(\alpha_k, \beta_k)$ distribution, i.e.,

$$\begin{aligned} f(\tau_{s,t} \mid z_{dur,s,t} = k) &\sim \mathrm{Ga}(\tau_{s,t} \mid \alpha_k, \beta_k), \\ \Pr(z_{dur,s,t} = k \mid i_s = i, x_{s,1} = x_1, \ldots, x_{s,p} = x_p, y_{s,t-1} = y_{t-1}) &= P^{(i)}_{dur}(k \mid x_1, \ldots, x_p, y_{t-1}). \end{aligned}$$

Similar to the model for the transition probabilities, the mixture probabilities are a convex combination of a fixed population-level effect and a random individual-level effect:

$$\mathbf{P}^{(i)}_{dur}(\cdot \mid x_1, \ldots, x_p, y_{t-1}) = \pi^{(i)}_{dur,0}(\cdot) \boldsymbol{\lambda}_{dur,x_1,\ldots,x_p,y_{t-1}}(\cdot) + \pi^{(i)}_{dur,1}(\cdot) \boldsymbol{\lambda}^{(i)}_{dur}(\cdot),$$

where $\boldsymbol{\lambda}_{dur,x_1,\ldots,x_p,y_{t-1}}(\cdot)$ is the baseline component, $\boldsymbol{\lambda}^{(i)}_{dur}(\cdot)$ is the random individual effect, and $\{\pi^{(i)}_{dur,0}(k), \pi^{(i)}_{dur,1}(k)\}_{k=1}^{K}$ are individual-specific coefficients such that $\pi^{(i)}_{dur,1}(k) = 1 - \pi^{(i)}_{dur,0}(k)$. Again, for each covariate $r = 1, \ldots, p, p+1$ (where the $(p+1)^{th}$ covariate is the preceding state y_{t-1}), we construct the partition $\mathcal{C}^{(r)}_{dur} = \{\mathcal{C}^{(r)}_{dur,g_r}\}_{g_r=1}^{k_{dur,r}}$ of its levels, where $k_{dur,r}$ is the number of clusters for covariate r and g_r represents the cluster index. We introduce latent variables $\{z_{dur,r,w}\}_{r=1,w=1}^{p+1,d_r}$ that indicate the cluster index for the w^{th} label of the r^{th} covariate. We now replace the population-level effect $\boldsymbol{\lambda}_{dur,x_1,\ldots,x_p,y_{t-1}}(\cdot)$ with $\boldsymbol{\lambda}_{dur,g_1,\ldots,g_p,g_{p+1}}(\cdot)$.

The mixture probability vectors are given Dirichlet priors with the mean vector $\boldsymbol{\lambda}_{dur,0}$, which itself centers around a global vector $\boldsymbol{\lambda}_{dur,00}$:

$$\begin{aligned} \boldsymbol{\lambda}_{dur,g_1,\ldots,g_{p+1}}(\cdot) &\sim \mathrm{Dir}\left\{\alpha_{dur,0}\lambda_{dur,0}(1), \ldots, \alpha_{dur,0}\lambda_{dur,0}(K)\right\}, \\ \boldsymbol{\lambda}^{(i)}_{dur}(\cdot) &\sim \mathrm{Dir}\left\{\alpha^{(0)}_{dur}\lambda_{dur,0}(1), \ldots, \alpha^{(0)}_{dur}\lambda_{dur,0}(K)\right\}, \\ \boldsymbol{\lambda}_{dur,0}(\cdot) &\sim \mathrm{Dir}\left\{\alpha_{dur,00}\lambda_{dur,00}(1), \ldots, \alpha_{dur,00}\lambda_{dur,00}(K)\right\}. \end{aligned}$$

We present the complete Bayesian hierarchical model for the continuous duration times as

$$\begin{aligned} &(\tau_{s,t} \mid z_{dur,s,t} = k) \sim \mathrm{Ga}(\tau_{s,t} \mid \alpha_k, \beta_k), \\ &(z_{dur,s,t} \mid i_s = i, z_{dur,1,x_{s,1}} = g_1, \ldots, z_{dur,p,x_{s,p}} = g_p, z_{dur,p+1,y_{s,t-1}} = g_{p+1}) \sim \\ &\qquad \mathrm{Mult}\left\{P^{(i)}_{dur,g_1,\ldots,g_{p+1}}(1), \ldots, P^{(i)}_{dur,g_1,\ldots,g_{p+1}}(K)\right\}, \text{ where} \\ &P^{(i)}_{dur,g_1,\ldots,g_{p+1}}(k) = \pi^{(i)}_{dur,0}(k)\lambda_{dur,g_1,\ldots,g_{p+1}}(k) + \pi^{(i)}_{dur,1}(k)\lambda^{(i)}_{dur}(k), \\ &\boldsymbol{\lambda}_{dur,g_1,\ldots,g_{p+1}}(\cdot) \sim \mathrm{Dir}\left\{\alpha_{dur,0}\lambda_{dur,0}(1), \ldots, \alpha_{dur,0}\lambda_{dur,0}(K)\right\}, \quad \alpha_{dur,0} \sim \mathrm{Ga}(a_{dur,0}, b_{dur,0}), \\ &\boldsymbol{\lambda}^{(i)}_{dur}(\cdot) \sim \mathrm{Dir}\left\{\alpha^{(0)}_{dur}\lambda_{dur,0}(1), \ldots, \alpha^{(0)}_{dur}\lambda_{dur,0}(K)\right\}, \quad \alpha^{(0)}_{dur} \sim \mathrm{Ga}(a^{(0)}_{dur}, b^{(0)}_{dur}), \\ &\boldsymbol{\lambda}_{dur,0}(\cdot) \sim \mathrm{Dir}\left\{\alpha_{dur,00}\lambda_{dur,00}(1), \ldots, \alpha_{dur,00}\lambda_{dur,00}(K)\right\}, \\ &\pi^{(i)}_{dur,0}(k) \sim \mathrm{Beta}(a_{dur,0}, a_{dur,1}), \\ &\alpha_k \sim \mathrm{Ga}(a_{dur,0}, b_{dur,0}), \quad \beta_k \sim \mathrm{Ga}(a_{dur,0}, b_{dur,0}). \end{aligned}$$

Inference is based on samples drawn from the posterior using a Metropolis-Hastings-within-Gibbs MCMC algorithm. Most full conditionals are available in closed form and can be directly sampled from. A Metropolis-Hastings step is however used for updating the discrete valued cluster configurations. There is, however, no conjugate prior for gamma distributions with unknown shape parameters (Damsleth, 1975). Recently, Miller (2019) designed a procedure that efficiently approximates the posterior full conditionals of gamma shape parameters under a gamma prior with another gamma density. We adopt this approximation in our MCMC algorithm.

3 The BMRMM R package

3.1 Package description

The **BMRMM** package is developed to implement Bayesian Markov (renewal) mixed models. The main function BMRMM of the package carries out detailed analyses of the state transitions and their duration times (if applicable) as described in the previous section. Moreover, the package includes a number of supplementary functions that use the results of the main function to produce numerical summaries, visualizations, and diagnostics. Table 1 provides a brief description of all functions.

Function	Description
BMRMM	Creates a BMRMM object.
summary.BMRMM	Summary for an object of class BMRMM and create a BMRMMsummary object.
plot.BMRMMsummary	Visualization of a BMRMMsummary object.
hist.BMRMM	Returns histograms of duration times for a BMRMM object.
diag.BMRMM	Provides MCMC diagnostic plots for a BMRMM object.
model.selection.scores	Returns the LPML and WAIC scores of the mixture gamma model.

Table 1: Summary of functions in the **BMRMM** package.

3.2 The main function BMRMM

The main function is BMRMM which implements the inference for both the state transition probabilities and the duration times. We summarize the parameters in Table 3 and present the function as follows.

```
BMRMM(data, num.cov, cov.labels = NULL, state.labels = NULL,
      random.effect = TRUE, fixed.effect = TRUE,
      trans.cov.index = 1:num.cov, duration.cov.index = 1:num.cov,
      duration.distr = NULL, duration.incl.prev.state = TRUE,
      simsize = 10000, burnin = simsize/2)
```

The parameter data specifies the target data set and needs to follow a certain structure. The first column should list the individual IDs i_s, followed by p columns for the values of the p associated covariates $x_{s,j}$, then two columns for the values of the previous state $y_{s,t-1}$, the current state $y_{s,t}$, and finally a column for duration times $\tau_{s,t}$. The package supports one to five categorical covariates that take on values $1, 2, \ldots$. The duration times column is optional if the user would like to use BMMM instead of BMRMM to analyze just the state transitions. This is shown in Table 2. The users can look at the included simulated data set foxp2 as an example.

Id	Covariate 1	...	Covariate p	Previous State	Current State	State Durations/ISI

Table 2: Columns of the desired input data set.

The number of covariates in the data set is specified by the argument num.cov. The argument cov.labels is a list of vectors giving the names of covariate levels in the covariate order that is presented in data while the parameter state.labels is a vector providing the names of the transition states. The default labels are Arabic numerals.

The random.effect parameter gives users the option to exclude the random individual effects. If random.effect is set to FALSE, the transition probabilities (and the mixture probabilities for duration times, if applicable) will only consider the influence of the covariate levels. Similarly, the fixed.effect parameter allows users to exclude the fixed population effects. The default values for random.effect and fixed.effect are both TRUE. The covariate indices for the two analyses can be specified by setting `trans.cov.index` and `duration.cov.index`. We note that indices specified by `trans.cov.index` and `duration.cov.index` refer to the index of the covariate when the first covariate is given index 1, thus different from its index in `data`.

Users can define `duration.distr` in the following three ways.

1. If users set `duration.distr` to be `NULL`, which is the default setting, then the duration times will be ignored and not modeled at all. The BMMM described will be implemented to analyze the existing state transitions alone.
2. If `duration.distr` is set as ``list(`mixDirichlet',unit)``, the duration times will be used to construct a new state ``` `dur.state' ```, which will be analyzed along with the original set of states. The additional argument `unit` must be defined and acts both as a threshold and as a block size for duration times. For example, if the `unit` is set to 5, then for each duration value greater than 5 units, each block of 5 unit in it will be treated as an instance of a new `'dur.state'` state. If there is a state transition from state ``` `a' ``` to ``` `b' ``` with a duration time of 15 seconds and the `unit` is specified at 5 seconds, then the updated Markov sequence will contain three consecutive ``` `dur.state' ``` states, i.e., ``(`a',`dur.state',`dur.state',`dur.state',`b')``. Since we adopt the floor operation, a duration time of say 17.68 seconds will also be replaced by three consecutive instances of `'dur.state'` states in this example. The BMMM model will then be implemented to analyze the resulting appended state transitions.

   These first two options may naturally result in loss of information and is therefore not recommended when a detailed analysis of the distribution of the duration times is warranted.
3. If `duration.distr` is set to be ``list(`mixgamma',shape,rate)``, the duration times are modeled as a continuous random variable using a flexible mixture of gamma kernels, as described for a BMRMM model. In this case, users can specify the prior shape and rate parameters with the `shape` and `rate` arguments within the definition of `duration.distr`. We note that `shape` and `rate` must be numeric vectors of the same length.

By default, we consider the previous state $y_{s,t-1}$ as a covariate when we model the duration times as continuous variables, i.e., `duration.incl.prev.state` is set to `TRUE`. Users can set this parameter to `FALSE` if they wish to exclude the previous state when analyzing the duration times. The remaining parameters `simsize` and `burnin` denote the total number of MCMC iterations and the number of burn-ins, respectively.

Argument	Explanation	Default value
`data`	the data set to be used following the required format	
`num.cov`	an integer giving the number of observed covariates in `data`	
`cov.labels`	a list of vectors giving names of all covariate levels	`NULL`
`state.labels`	a vector giving names of the states	`NULL`
`random.effect`	`TRUE` if random individual effects are included	`TRUE`
`fixed.effect`	`TRUE` if fixed population effects are included	`TRUE`
`trans.cov.index`	selects the covariates to analyze for transition probabilities	`1:num.cov`
`duration.cov.index`	selects the covariates to analyze for duration times	`1:num.cov`
`duration.distr`	specifies the distribution for duration times	`NULL`
`duration.incl.prev.state`	`TRUE` if y_{t-1} acts as a covariate for the analysis of duration times	`TRUE`
`simsize`	number of MCMC iterations	10000
`burnin`	number of burnins of the MCMC algorithm	`simsize/2`

Table 3: Arguments to the `BMRMM` function.

The `BMRMM` function returns an object of class `BMRMM`, which either contains only `results.trans` or both of `results.trans` and `results.duration` if duration times follow a mixture gamma distribution. For the state transitions, the posterior mean transition probability matrices for each combination of the covariate levels and each individual are given by `results.trans$tp.exgns.post.mean` and `results.trans$tp.anmls.post.mean`, respectively. Additionally, `results.trans$clusters` stores cluster configurations for each covariate from each MCMC iteration. As for duration times, the fields `results.duration$shape.samples` and `results.duration$rate.samples` record the shape and rate parameters, for each mixture component in every MCMC iteration, respectively. Meanwhile, `results.duration$comp.assignment` gives the assignment of the mixture component for each data point in the last MCMC iteration. Similar to transition probabilities, `results.duration$clusters` gives the cluster configurations of the covariates. Other elements of `results.trans` and `results.duration` can be found in the detailed R function description.

3.3 Summarizing BMRMM results

The **BMRMM** package provides an S3 method for summarizing results of a `BMRMM` object as follows.

```
summary.BMRMM(object, delta = 0.02, digits = 2, ...)
```

The object must be of class BMRMM. The argument delta is associated with local tests for transition probabilities, which we will explain further. The digit parameter is an integer used for number formatting, as in the general summary function. The summary.BMRMM function returns an object of class BMRMMsummary with the following fields.

- trans.global and dur.global

  These two fields give the global test results from the inference of transition probabilities and duration times. Global tests show the significance of the covariates in affecting the state transitions and duration times. Specifically, for each covariate, the empirical distribution of the size of the clusters in the stored MCMC iterations is calculated. The null hypothesis that a covariate is not important is equivalent to the event that all its levels are in the same cluster, or, in other words, that the cluster size for the covariate is just one.
- trans.probs.mean and trans.probs.sd

  The two fields provide the mean and standard deviation for the posterior mean of each transition type under all combinations of covariate levels, respectively.
- trans.local.mean.diff and trans.local.null.test

  The BMRMMsummary object also contains local test results for transition probabilities. Local tests analyze the differences between the transition probabilities associated with two different levels of a covariate *j*, fixing the levels of the other covariates. For every pair of levels of covariate *j*, trans.local.mean.diff gives the absolute differences in transition probabilities for each transition type in the MCMC iterations. The local null hypothesis we test for each transition type is that this difference is at least the pre-specified value delta. Meanwhile, trans.local.null.test gives the probability of the null hypothesis that the difference between two covariate levels is not significant under each transition type.
- dur.mix.params and dur.mix.probs

  For each mixture component, dur.mix.params provides the estimates of the gamma shape and rate parameters from the last MCMC iteration. For every covariate level, users can obtain the mixture probabilities by calling the field dur.mix.probs, which can be further used to estimate the length of the duration times.

3.4 Visualizing results with BMRMM plotting functions

The main plotting function of the package, plot.BMRMMsummary, is an S3 method for class BMRMMsummary. It gives the barplots for global tests as well as heatmaps for the posterior mean and standard deviation for transition probabilities, local tests for transition probabilities, mixture parameters and probabilities for duration times. The parameters of plot.BMRMMsummary include x, which must be an object of class BMRMMsummary and type, which is a single string representing the field of x that needs to be plotted. The function also takes general plotting arguments such as xlab, ylab, etc.

```
plot.BMRMMsummary(x, type, xlab = NULL, ylab = NULL, main = NULL, col = NULL, ...)
```

When duration times are analyzed as continuous variables using mixture gamma distributions, the users can use the S3 method hist.BMRMM to generate histograms for duration times along with the estimated posterior distribution. The parameter x is an object of class BMRMM. The argument comp gives the specific mixture component that the user would like to investigate. When comp is NULL, which is the default setting, the histogram of all observed duration times is plotted and superimposed with the posterior mean of the fitted mixture gamma distribution. When comp is a specific integer, we will be looking at the last MCMC iteration. The histogram for duration times assigned with component comp will be presented alongside the mixture gamma distribution with the shape and rate parameters from the last MCMC iteration. Users can refer to the documentation of the general hist function to see the interpretation for the rest of the parameters.

```
hist.BMRMM(x, comp = NULL, xlim = NULL, breaks = NULL, main = NULL,
           col = 'gray', xlab = 'Duration times', ylab = 'Density', ...)
```

Finally, users can check the MCMC diagnostics with the traceplots and autocorrelation plots produced by the function diag.BMRMM. The object parameter should be an object of class BMRMM. For transition probabilities, users can specify the covariate levels as well as the state transitions they are interested in by defining cov.combs and transitions, respectively. For duration times, users can define components, a numeric vector, to obtain the diagnostic plots for shape and rate parameters of the specific component kernels.

```
diag.BMRMM(object, cov.combs = NULL, transitions = NULL, components = NULL)
```

3.5 Model selection scores for continuous duration times

When the duration times are modeled using mixtures of gamma distributions, model selection can be performed on the number of mixture components using the function `model.selection.scores`.

```
model.selection.scores(object)
```

The function takes an `object` as its input, which must be an object of class `BMRMM`. It returns a list consisting of the log pseudo marginal likelihood (LPML) (Geisser and Eddy, 1979) and the widely applicable information criterion (WAIC) (Watanabe and Opper, 2010) scores of the model. Larger values of LPML and smaller values of WAIC indicate better model fits. They are particularly suitable for complex Bayesian hierarchical models as they can be easily computed from the MCMC samples.

4 Illustrations on the synthetic FoxP2 data set

The FoxP2 data set records the songs sung by adult male mice of two genotypes, wild type or FoxP2, denoted by W and F, respectively (Chabout et al., 2016). The mice sang under three social contexts, U (fresh female urine on a cotton tip placed inside the male's cage), L (an awake and behaving adult female placed inside the cage), and A (an anesthetized female placed on the lid of the cage). Each song comprises a sequence of syllables and continuous inter-syllable intervals (ISIs). The data set can be used to analyze the effect of the FoxP2 gene on the vocal syntax of mice, in turn providing insights into the effects of the gene on human vocal communication abilities and related deficiencies. The real FoxP2 data set originates from the study by Chabout et al. (2016) and requires permission to use. Wu et al. (2023) simulated a data set that closely mimics the real one. For demonstrating the **BMRMM** package, we included in it a shortened version of this synthetic data set which we refer to as the `foxp2` data set.

The `foxp2` synthetic data set has 17391 rows and 6 columns, which are Id, Genotype, Context, Prev_State, Cur_State, and Transformed_ISI. The original FoxP2 data set records ISIs in seconds. In the simulated data set `foxp2`, following Wu et al. (2023), log(1+ISI) values are used which give a better model fit.

Id	Genotype	Context	Prev_State	Cur_State	Transformed_ISI
1	2	2	3	3	0.20197711
1	2	2	3	3	0.06972753
1	2	2	3	3	0.07211320
1	2	2	3	3	0.15790932
1	2	2	3	3	0.06781471
1	2	2	3	3	0.09426236

Table 4: Part of the simulated FoxP2 data set `foxp2`.

If we are only interested in analyzing the transition probabilities with the covariates genotype and social contexts, we would use the main function as follows.

```
R> res.fp2 <- BMRMM(foxp2, num.cov = 2)
```

If we would like to pick specific covariates for our analyses, we can define `trans.cov.index` and `duration.cov.index` accordingly. For example, if we only want to use context for transition probabilities and genotype for ISIs, we would run the following.

```
R> res.fp2 <- BMRMM(foxp2, num.cov = 2,
                    trans.cov.index = c(2), duration.cov.index = c(1))
```

If we would like to analyze the ISIs as part of the original state sequence following a mixture Dirichlet distribution, as was done by Sarkar et al. (2018), the ISIs are replaced by (possibly consecutive) "silent" states by dividing them into blocks of 250 milliseconds. The `BMRMM` function can do this by setting `duration.distr` as a list with the string `'mixDirichlet'` and the argument `unit` as $\log(0.25 + 1)$, based on the log transformation.

```
R> res.fp2 <- BMRMM(foxp2, num.cov = 2,
                    duration.distr = list('mixDirichlet', unit = log(0.25+1)))
```

In the next example, we would like to analyze the ISIs as continuous variables following a mixture gamma distribution. For syllable transitions, we use both genotype and context as covariates. For the ISIs, in addition to these two, we also use the preceding syllable as a covariate.

```
R> res.fp2 <- BMRMM(data = foxp2, num.cov = 2, state.labels = c('d', 'm', 's', 'u'),
                    cov.labels = list(c('F', 'W'), c('U', 'L', 'A')),
                    duration.distr = list('mixgamma', shape = rep(1, 4), rate = rep(1, 4)))
```

In what follows, we show the results for the last function call. The returned `res.fp2` have two parts, which are named `res.fp2$results.trans` and `res.fp2$results.duration`. Now we demonstrate how we print and visualize the results.

First, we obtain a `BMRMMsummary` object, `sm.fp2`, by calling the `summary.BMRMM` function on the returned results `res.fp2`. The global test results for identifying the significant covariates can be found by calling the fields `trans.global` and `dur.global`. The function `plot.BMRMMsummary` is called to visualize the global tests using barplots, as presented in Figure 2. We recall that a covariate is significant when its levels formed more than one cluster with very high posterior probability (the bar heights). Figure 2 and the printed results suggest that every covariate is significant for the ISIs but only the social context is significant for the transition probabilities.

```
R> sm.fp2 <- summary(res.fp2)
R> sm.fp2$trans.global
            label_data
cluster_data Context Genotype
           1       0        1
           3       1        0
R> sm.fp2$dur.global
            label_data
cluster_data Context Genotype prev_state
           2    0.00     1.00       0.10
           3    1.00     0.00       0.90
           4    0.00     0.00       0.01
R> plot(sm.fp2, 'trans.global')
R> plot(sm.fp2, 'dur.global')
```

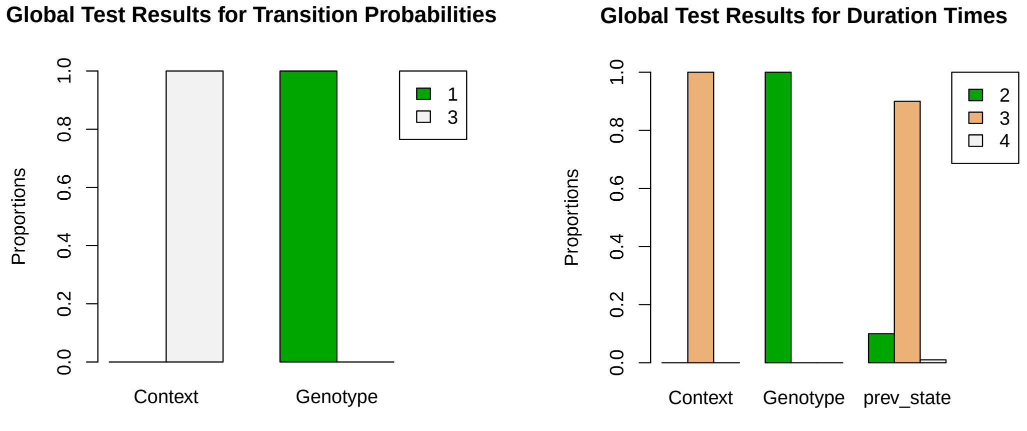


Figure 2: Results for the simulated `foxp2` data set showing the global tests of significance of the covariates for the state transitions (left) and the ISIs (right). The bars represent the estimated posterior probabilities of the number of clusters formed by the levels of each covariate.

The plotting function can be called to visualize the posterior transition probabilities under different combinations of the covariate levels. We show in Figure 3 the heatmaps for the posterior mean and standard deviation of the transition probabilities for each transition type for the following combinations of covariates: (F, A) and (W, L).

```
R> plot(sm.fp2, 'trans.probs.mean')
R> plot(sm.fp2, 'trans.probs.sd')
```

We also perform the local test to assess the influence of genotype on the transition probabilities by computing the absolute difference of the transition probabilities between F and W among the thinned MCMC samples after burn-ins, i.e., $|\Delta\lambda_{trans,\cdot,x_2}(y_t \mid y_{t-1})| = |\lambda_{trans,1,x_2}(y_t \mid y_{t-1}) - \lambda_{trans,2,x_2}(y_t \mid$

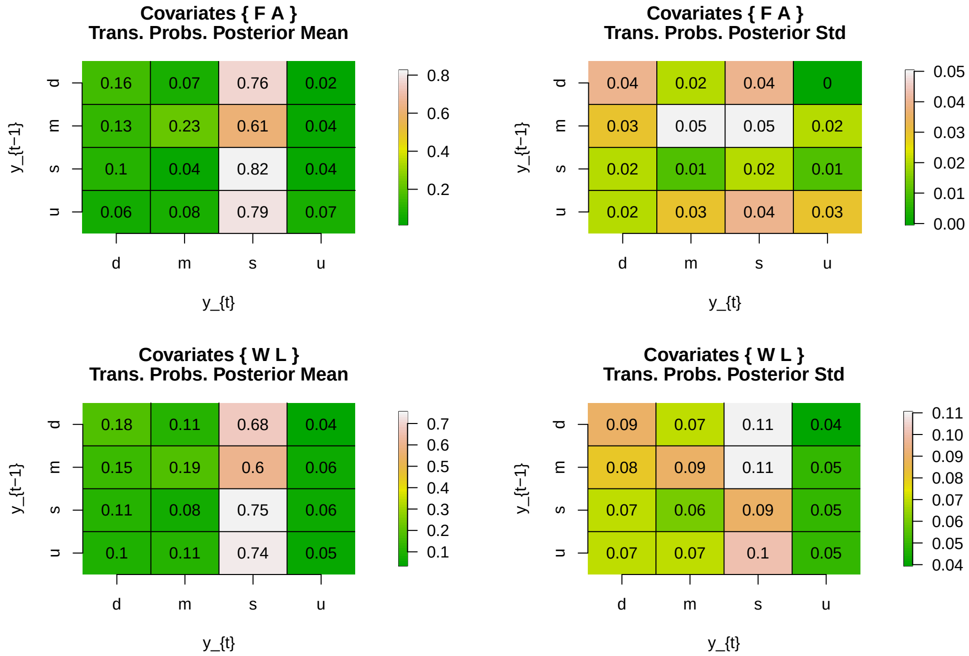


Figure 3: Results for the simulated foxp2 data set showing the posterior mean and standard deviation for each transition type for selected covariate combinations, (F, A) (top) and (W, L) (bottom).

$y_{t-1})|$. The estimated posterior probability for the null hypothesis is therefore the proportion of times $|\Delta\lambda_{trans,\cdot,x_2}(y_t \mid y_{t-1}) \leq \delta|$ is observed in the MCMC samples, where x_2 is the social context and δ is the user-specific difference threshold delta. The plotting function plot.BMRMMsummary gives the plots for all local test results if we set the type to be `trans.local.mean.diff' or `trans.local.null.test'. Here, we show the results of local tests for the covariate 1 (i.e., genotype) with delta equaling the default value of 0.02, and present the plots in Figure 4. From the figure, we see that the posterior probabilities of the null hypotheses are generally large for most transition types (e.g., transitions to the syllable u) regardless of the social context, indicating that genotype does not have a strong influence on transition probabilities with a fixed context under these transition types.

```
R> plot(sm.fp2, 'trans.local.mean.diff')
R> plot(sm.fp2, 'trans.local.null.test')
```

Next, we turn our attention to the ISIs. We first check the fit of our estimated mixture gamma distribution presented in Figure 5a. We then look further into the shape of each mixture component in Figure 5b. From the histogram for each component, we see that components 2 and 4 represent longer ISIs while components 1 and 3 represent shorter ISIs.

```
R> hist(res.fp2, xlim = c(0,1))
R> for(comp in 1:4) {
     hist(res.fp2, comp = comp)
   }
```

We examine the values of mixture parameters and mixture probabilities for each covariate level in the last MCMC iteration, which provides insights into the influence of the covariate on ISI lengths.

```
R> sm.fp2$dur.mix.params
       shape.k rate.k
Comp 1   29.07 394.30
Comp 2    1.23   1.49
Comp 3    8.46 465.66
Comp 4    3.03  20.13

R> sm.fp2$dur.mix.probs
$Genotype
          F    W
Comp 1 0.46 0.48
Comp 2 0.19 0.13
```

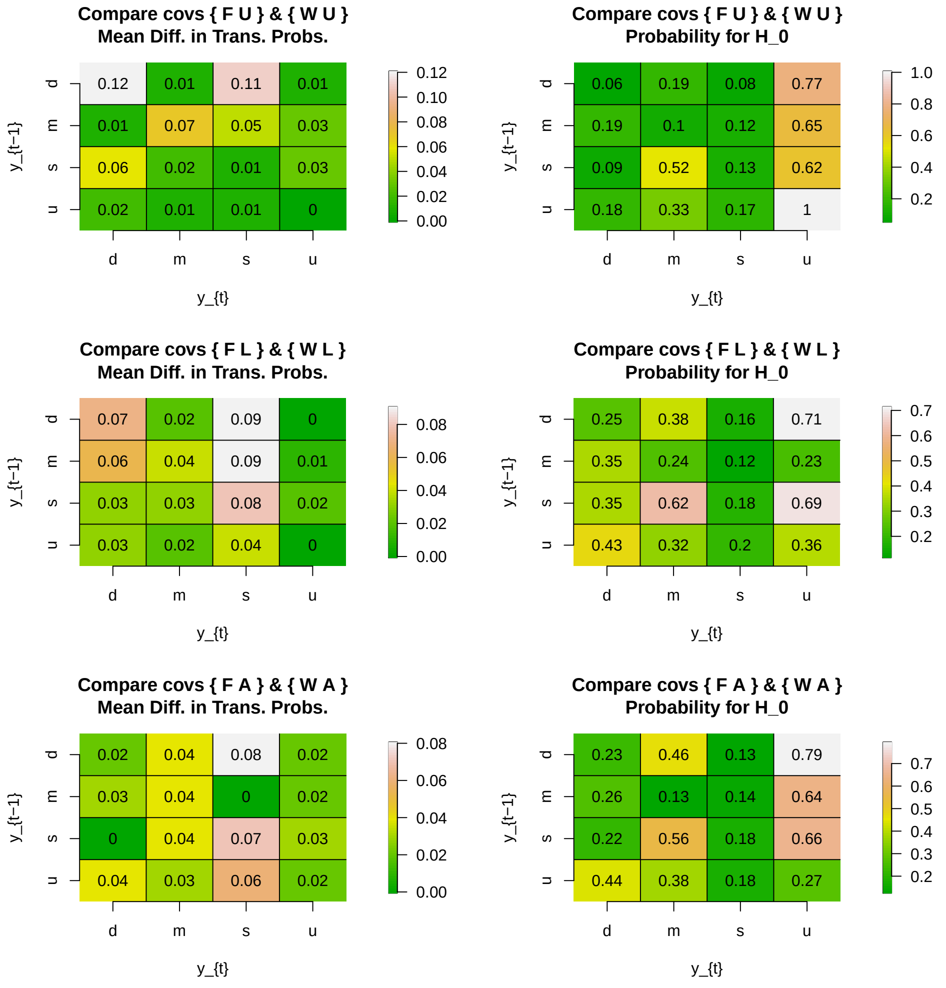
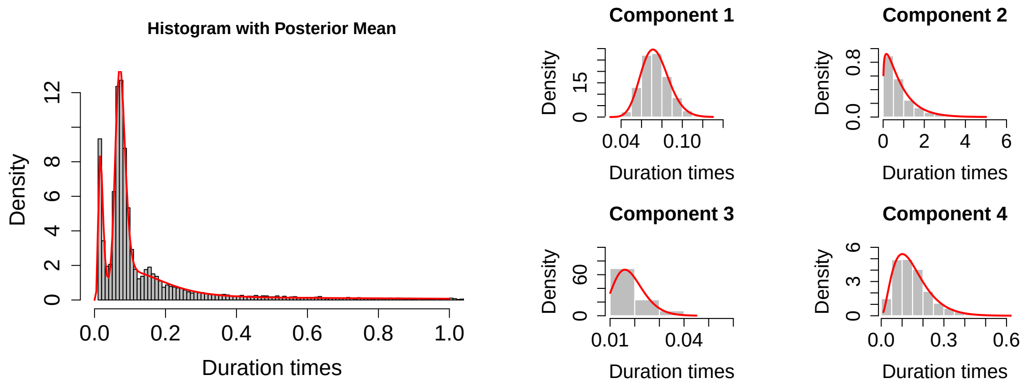


Figure 4: Results for the simulated `foxp2` data set showing local test results for genotypes fixing the social context, U (top), L (middle), and A (bottom). The averaged absolute difference in transition probabilities between F and W is presented on the left. The posterior probabilities of the corresponding null hypotheses are on the right.



(a) Histogram of ISIs with the estimated posterior mean (red line) of their marginal gamma mixture density averaged from recorded MCMC samples.

(b) Histograms of ISIs for each component of the gamma mixture model along with the component density (red lines) from the last MCMC iteration.

Figure 5: Results for the simulated `foxp2` data set showing the histograms of the ISIs with the estimated posterior gamma mixture density (left) and the histograms of the ISIs for each mixture component (right).

```
Comp 3 0.10 0.15
Comp 4 0.25 0.24

$Context
          U    L    A
Comp 1 0.58 0.35 0.48
Comp 2 0.14 0.16 0.18
Comp 3 0.08 0.21 0.08
Comp 4 0.20 0.28 0.25

$prev_state
          d    m    s    u
Comp 1 0.52 0.52 0.42 0.42
Comp 2 0.13 0.13 0.22 0.17
Comp 3 0.11 0.11 0.10 0.18
Comp 4 0.24 0.24 0.26 0.23
```

From the mixture probabilities, we see that mice with genotype F have a much higher mixture probability in component 2 compared to genotype W, which indicates mice with the FoxP2 mutation require a longer ISI between pronouncing two syllables, a reflection of vocal impairment.

Finally, we check the MCMC diagnostic plots and see if we had good mixing for the parameters. Here we focus on a specific transition type, $u \to m$, for covariate combination $\{F, U\}$ and a specific mixture component (component 2). We show these plots in Figure 6.

```
R> diag.BMRMM(res.fp2, cov.combs = list(c(1, 1)),
              transitions = list(c(4, 2)), components = c(2))
```

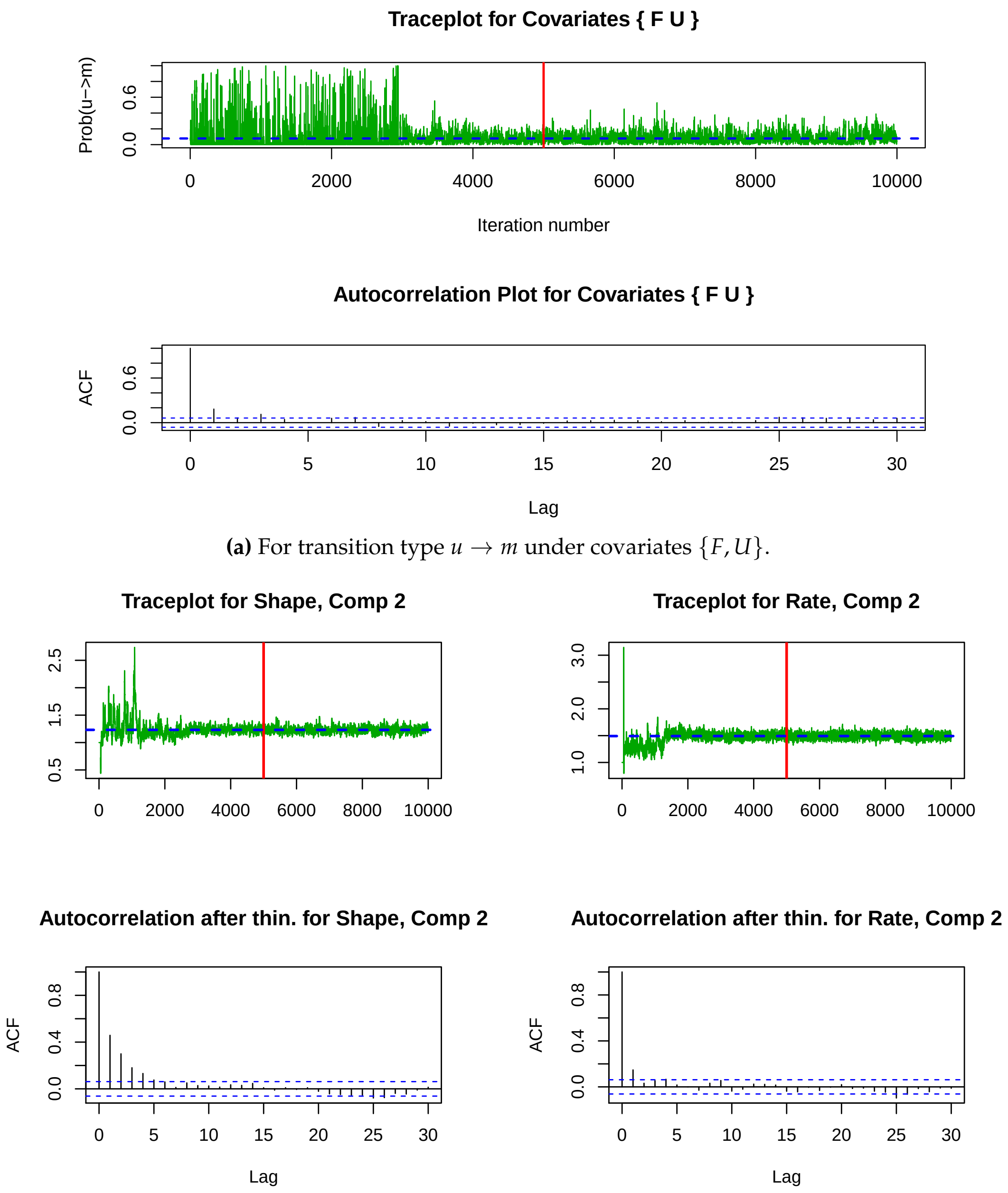
5 Illustrations on the asthma control data set

The **BMRMM** package is able to analyze duration times in detail which could either be the ISIs, as seen in the synthetic `foxp2` data set, or the state persistence times, as in a traditional semi-Markov model. To demonstrate the usage of our package in analyzing the state persistence times, we use the asthma control data set from the ARIA (Association pour la Recherche en Intelligence Artificielle) study of severe asthmatic patients (Combescure et al., 2003) in France between 1997 and 2001. At each visit, a chest physician graded the asthma control status of the patient using control scores (Juniper et al., 1999). The data set contains the sojourn time of the control states as well as three covariates: Asthma severity, sex, and the body mass index (BMI) of the patients. Saint-Pierre et al. (2003) used a Markov model with piece-wise constant intensities to model the asthma control evolution and proposed a regression model for analyzing the effect of covariates. Combescure et al. (2003) used the data set to assess the relationship between asthma severity and control of asthma. Listwon and Saint-Pierre (2015) fitted a semi-Markov model for the sojourn times using exponential and Weibull distributions and analyzed the effect of covariates individually due to complexity. Our **BMRMM** package is able to analyze the effect of the three covariates while also incorporating random individual effects exhibited by different patients on transition dynamics and state duration times.

The `asthma` data set we use here is from the **SemiMarkov** package (Listwon and Saint-Pierre, 2015). We have renamed and reordered the columns such that the data set fits the required format. Specifically, the data set has 928 rows, recording the asthma control states of 371 patients, which is one of the following three transient states: Optimal control (State 1), sub-optimal control (State 2), and unacceptable control (State 3). Each state can transit to any other two states and the state duration times are recorded. The data set also contains three binary covariates of the asthma patients, including the disease severity (1 if mild-moderate and 2 if severe), BMI (body mass index, 1 if BMI $<$ 25 and 2 otherwise), and sex (1 if women and 2 if men). We display part of the processed data in Table 5, where `Duration` is the sojourn time in `Prev_State`.

We investigate the transition dynamics and state persistence times of the `asthma` data set using the `BMRMM` function. We consider $K = 4$ mixture components for state persistence times. The choice of K is derived from running the BMRMM model several times with different K's and comparing the fitness of the models using the LPML and WAIC scores.

```
R> res.asm <- BMRMM(data = asthma, num.cov = 3, state.labels = c(1, 2, 3),
                    cov.labels = list(c('Mild-Moderate','Severe'),
                                      c('BMI<25','BMI>=25'),
                                      c('Women','Men')),
```



(a) For transition type $u \to m$ under covariates $\{F, U\}$.

(b) For gamma mixture component 2 shape and rate parameters.

Figure 6: Results for the simulated `foxp2` data set showing the MCMC diagnostic plots, including traceplots and autocorrelation plots.

```
duration.distr = list('mixgamma', shape = rep(1, 4),
                                  rate = rep(1, 4)))
```

We name the returned `BMRMM` object `res.asm` and obtain the `BMRMMsummary` object `sm.asm` by calling the `summary.BMRMM` function. As in the FoxP2 application, we first plot the global test results for both transition probabilities and state persistence times in Figure 7. For the transition probabilities, only the severity of asthma is significant while for duration times only the preceding state is significant. The BMI value and the sex are not significant for either transition dynamics or state durations.

```
R> sm.asm <- summary(res.asm)
R> sm.asm$trans.global
            label_data
cluster_data  BMI Severity  Sex
           1 0.96     0.00 1.00
           2 0.04     1.00 0.00
R> sm.asm$dur.global
```

Id	Severity	BMI	Sex	Prev_State	Cur_State	Duration
2	2	2	1	3	2	0.1533
2	2	2	1	2	2	4.1232
3	2	2	2	3	1	0.0958
3	2	2	2	1	3	0.2300
3	2	2	2	3	1	0.2656
3	2	2	2	1	1	5.4073

Table 5: Part of the asthma data set from the ARIA study of severe asthmatic patients.

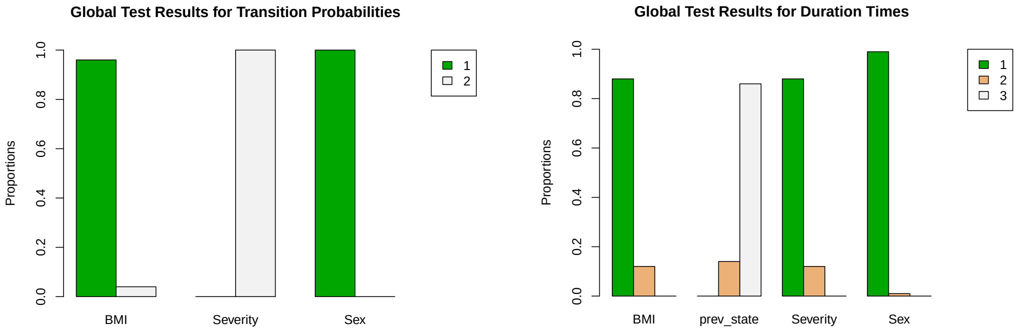


Figure 7: Results for the asthma data set showing the global tests of significance of the covariates for the state transitions (left) and the state persistence times (right). The bars represent the estimated posterior probabilities of the number of clusters formed by the levels of each covariate.

```
                      label_data
cluster_data  BMI prev_state Severity  Sex
           1 0.88       0.00     0.88 0.99
           2 0.12       0.14     0.12 0.01
           3 0.00       0.86     0.00 0.00
R> plot(sm.asm, 'trans.global')
R> plot(sm.asm, 'dur.global')
```

We show the posterior mean and standard deviations of the state transition probabilities for men and women with severe conditions and BMI ≥ 25 in Figure 8. We see that for severe patients with BMI ≥ 25, the transition probabilities are similar for men and women. We also take a look at the local test results for the BMI values fixing the severity of the patients in Figure 9. Though the absolute differences between the two covariate levels for BMI are small, the probabilities for the null hypotheses are also small, especially for transitions to state 1 and state 2. This suggests that even though the influence of BMI on state transitions is not significant globally, it is significant given the severity of asthma condition regardless of sex.

```
R> plot(sm.asm, 'trans.probs.mean')
R> plot(sm.asm, 'trans.probs.sd')
```

Figure 10a presents the histograms of the entire asthma data set, superimposed with the posterior mean of the mixture gamma distribution. With four components, the estimated mixture gamma fits the asthma data well. From the histogram for each component, we see from Figure 10b that component 1 and 2 represents shorter state persistence times while components 3 and 4 represent longer durations.

```
R> hist(res.asm, xlim = c(0,1))
R> for(comp in 1:4) {
     hist(res.asm, comp = comp)
   }
```

We investigate the covariates' influence by examining the mixture probabilities from the last MCMC iteration. An interesting discovery is that the mixture probabilities for both sexes, BMI levels, and severity levels are the same, indicating that the levels of these three covariates do not strongly influence the distributions of state durations. This matches the global test results in Figure 7. If the preceding state is state 1, which is optimal control for asthma, the state duration time is longer than that if the previous state is 2 or 3, as there is a lower weight in component 1 and higher weight in

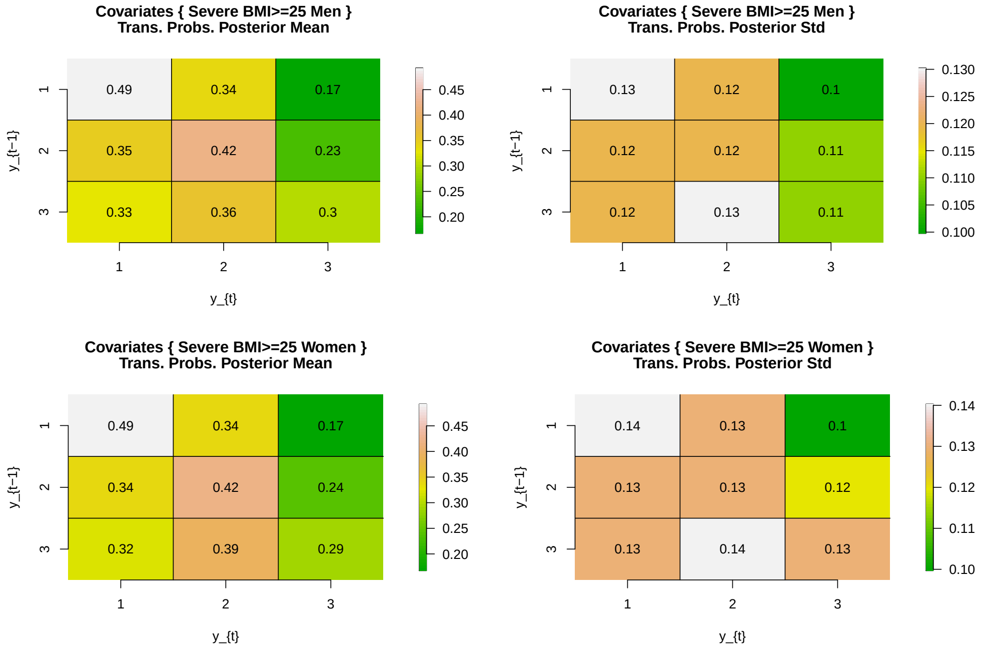


Figure 8: Results for the asthma data set showing the posterior mean and standard deviation for each transition type for selected covariate combinations, $\{\text{Severe}, \text{BMI} \geq 25, \text{Men}\}$ (top) and $\{\text{Severe}, \text{BMI} \geq 25, \text{Women}\}$ (bottom).

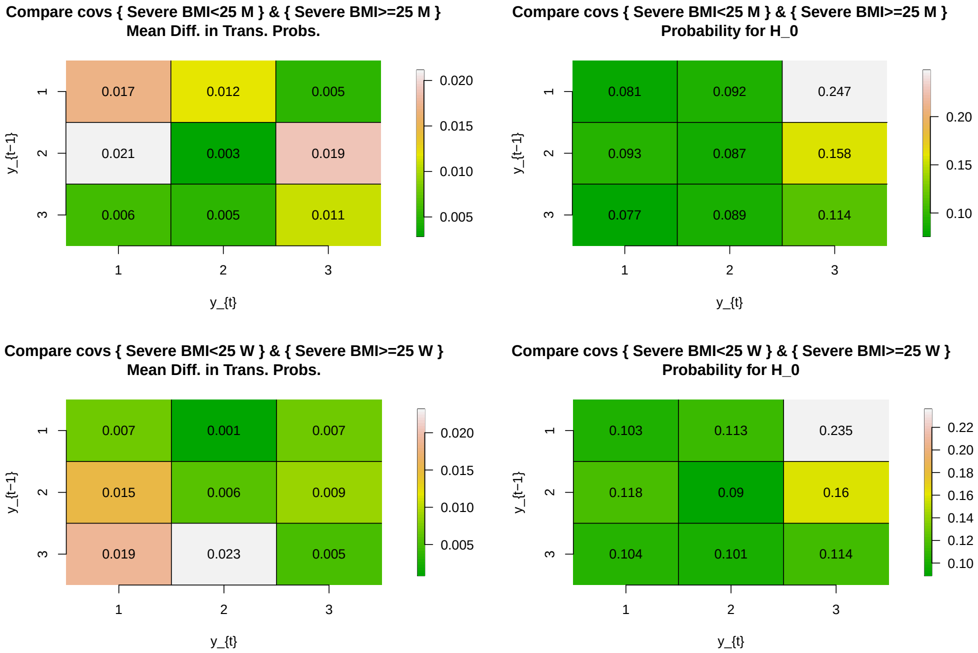
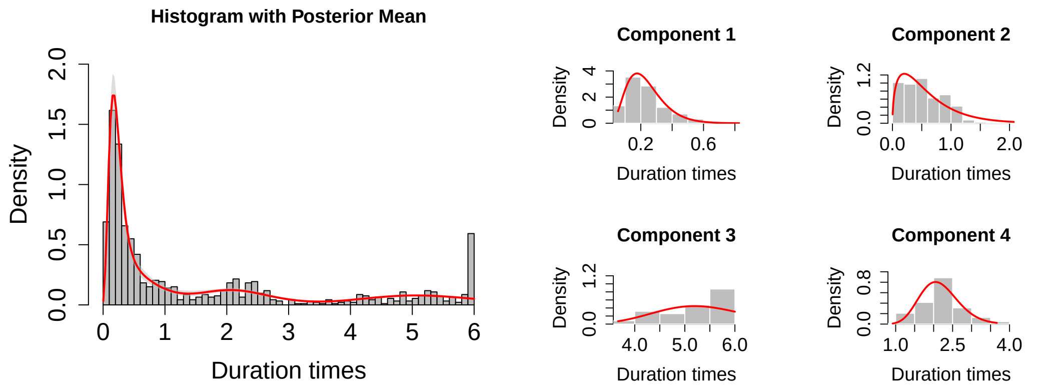


Figure 9: Results for the asthma data set showing local test results for BMI fixing asthma severity condition and sex, men (top) and women (bottom). The averaged absolute difference in transition probabilities between BMI < 25 and BMI ≥ 25 is presented on the left. The posterior probability of the null hypothesis is on the right.

components 3 and 4. On the other hand, if the preceding state is 3, which is unacceptable control, then the state duration time is much shorter, as the mixture probability in component 1 is much higher when the preceding state is state 3. We present some examples of the diagnostic plots for the asthma data set in Figure 11.

```
R> sm.asm$dur.mix.probs
$Severity
       Mild-Moderate Severe
Comp 1          0.37   0.37
```



(a) Histogram of ISIs with the estimated posterior mean (red line) of their marginal gamma mixture density averaged from recorded MCMC samples.

(b) Histograms of ISIs for each of the three components of the gamma mixture model along with the component density (red lines) from the last MCMC iteration.

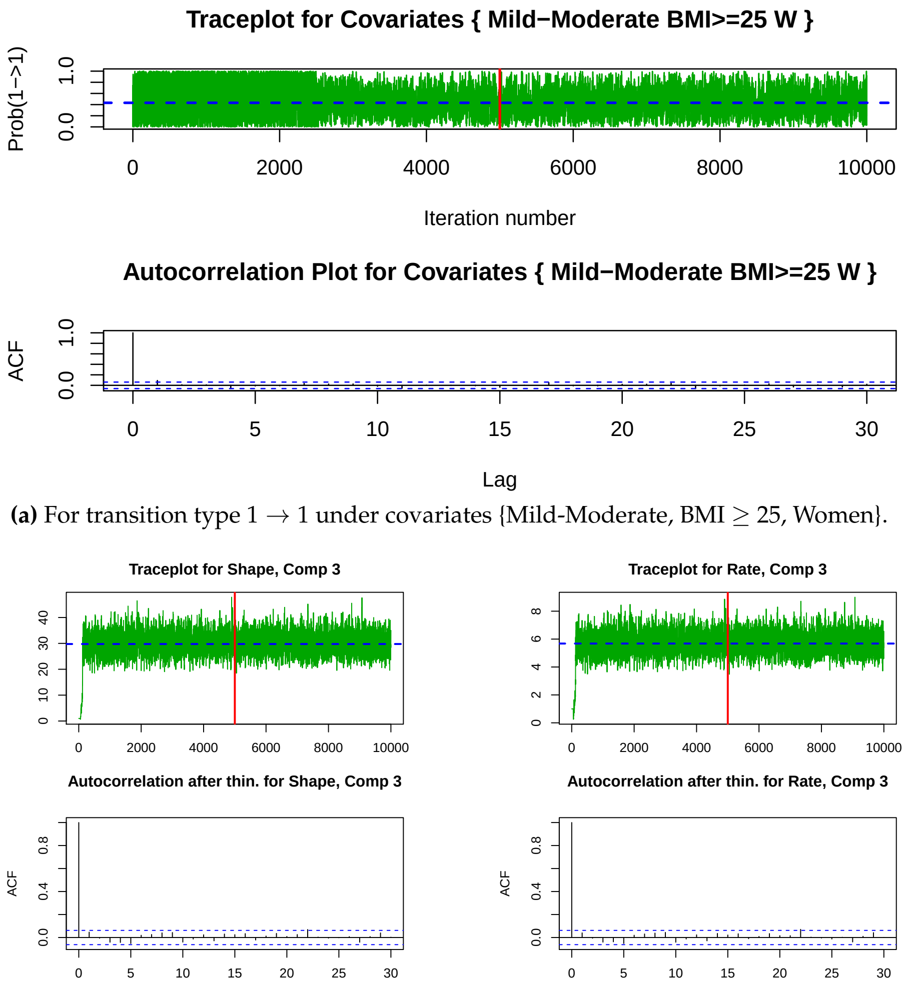
Figure 10: Results for the `asthma` data set showing the histograms of ISIs with the estimated posterior gamma mixture density (left) and histograms of ISIs for each mixture component (right).

```
Comp 2          0.26   0.26
Comp 3          0.20   0.20
Comp 4          0.17   0.17

$BMI
       BMI<25 BMI>=25
Comp 1   0.37    0.37
Comp 2   0.26    0.26
Comp 3   0.20    0.20
Comp 4   0.17    0.17

$Sex
       Women  Men
Comp 1  0.37 0.37
Comp 2  0.26 0.26
Comp 3  0.20 0.20
Comp 4  0.17 0.17

$prev_state
          1    2    3
Comp 1 0.27 0.33 0.51
Comp 2 0.23 0.33 0.23
Comp 3 0.31 0.20 0.09
Comp 4 0.19 0.15 0.17
```

```
R> diag.BMRMM(res.fp2, cov.combs = list(c(1, 2, 1)),
              transitions = list(c(1, 1)), components = c(3))
```

6 Conclusion

We presented the **BMRMM** package which implements both Bayesian Markov mixed models (BMMM) for analyzing the state transitions and Bayesian Markov renewal mixed models (BMRMM) for additionally analyzing the duration times (being either state persistence times or inter-state intervals) in a collection of categorical sequences, using flexible Dirichlet and gamma mixtures, respectively. The BMRMM takes into account fixed effects of the associated covariates as well as random effects of the associated individuals while simultaneously selecting the significant covariates separately for the state transitions and the duration times. The package includes a synthetic `foxp2` data set to demonstrate the data framework and function usages. The package also provides a series of plotting functions for visualizing the results of the analyses, including various global and local hypotheses tests, MCMC diagnostics, etc. We are committed to maintaining and further developing the package in the future.



(a) For transition type 1 → 1 under covariates {Mild-Moderate, BMI $\geq$ 25, Women}.

(b) For gamma mixture component 3 shape and rate parameters.

Figure 11: Results for the `asthma` data set showing the MCMC diagnostic plots, including traceplots and autocorrelation plots.

Future improvements to the package may include more options for the distribution types of transition probabilities and duration times beyond the currently available mixture Dirichlet and mixture gamma distributions, respectively.

7 Acknowledgements

We thank two anonymous reviewers very much for their careful review of our work and their constructive comments that led to significant improvements to both the package and this paper.

References

A. Alioum and D. Commenges. MKVPCI: a computer program for Markov models with piecewise constant intensities and covariates. *Computer Methods and Programs in Biomedicine*, 64:109–119, 2001. URL `https://doi.org/10.1016/s0169-2607(00)00094-8`. [p193]

A. Alioum, V. Leroy, D. Commenges, F. Dabis, R. Salamon, G. d'Epidémiologie Clinique du SIDA en Aquitaine, et al. Effect of gender, age, transmission category, and antiretroviral therapy on the progression of human immunodeficiency virus infection using multistate Markov models. *Epidemiology*, 9:605–612, 1998. URL `https://www.jstor.org/stable/3702781`. [p192]

E. E. Alvarez. Estimation in stationary Markov renewal processes, with application to earthquake forecasting in Turkey. *Methodology and Computing in Applied Probability*, 7:119–130, 2005. URL `https://doi.org/10.1007/s11009-005-6658-2`. [p192]

M. Amini, A. Bayat, and R. Salehian. hhsmm: an R package for hidden hybrid Markov/semi-Markov models. *Computational Statistics*, 1:1–53, 2022. URL https://doi.org/10.1007/s00180-022-01248-x. [p193]

S. Bacallado, J. D. Chodera, and V. Pande. Bayesian comparison of Markov models of molecular dynamics with detailed balance constraint. *The Journal of Chemical Physics*, 131:1–10, 2009. URL https://doi.org/10.1063/1.3192309. [p192]

V. S. Barbu, C. Bérard, D. Cellier, M. Sautreuil, and N. Vergne. SMM: An R package for estimation and simulation of discrete-time semi-Markov models. *The R Journal*, 10:226–246, 2018. URL https://doi.org/10.32614/RJ-2018-050. [p193]

P. Bulla and P. Muliere. Bayesian nonparametric estimation for reinforced Markov renewal processes. *Statistical Inference for Stochastic Processes*, 10:283–303, 2007. URL https://doi.org/10.1007/s11203-006-9000-x. [p192]

G. A. Castellucci, M. J. McGinley, and D. A. McCormick. Knockout of Foxp2 disrupts vocal development in mice. *Nature Scientific Reports*, 6:1–14, 2016. URL https://doi.org/10.1038/srep23305. [p193]

J. Chabout, A. Sarkar, S. Patel, T. Raiden, D. B. Dunson, S. E. Fisher, and E. D. Jarvis. A Foxp2 mutation implicated in human speech deficits alters sequencing of ultrasonic vocalizations in adult male mice. *Frontiers in Behavioral Neuroscience*, 10:1–18, 2016. URL https://doi.org/10.3389/fnbeh.2016.00197. [p193, 199]

C. Combescure, P. Chanez, P. Saint-Pierre, J. P. Daures, H. Proudhon, P. Godard, et al. Assessment of variations in control of asthma over time. *European Respiratory Journal*, 22:298–304, 2003. URL https://doi.org/10.1183/09031936.03.00081102. [p193, 203]

E. Damsleth. Conjugate classes for gamma distributions. *Scandinavian Journal of Statistics*, 2:80–84, 1975. URL https://www.jstor.org/stable/4615580. [p196]

P. Diaconis and S. W. Rolles. Bayesian analysis for reversible Markov chains. *The Annals of Statistics*, 34:1270–1292, 2006. URL https://doi.org/10.1214/009053606000000290. [p192]

P. Eichelsbacher and A. Ganesh. Bayesian inference for Markov chains. *Journal of Applied Probability*, 39:91–99, 2002. URL https://www.jstor.org/stable/3215920. [p192]

I. Epifani, L. Ladelli, and A. Pievatolo. Bayesian estimation for a parametric Markov renewal model applied to seismic data. *Electronic Journal of Statistics*, 8:2264–2295, 2014. URL https://doi.org/10.1214/14-EJS952. [p192]

M. A. Etterson, B. Olsen, and R. S. Greenberg. The analysis of covariates in multi-fate Markov chain nest-failure models. *Studies in Avian Biology*, 34:55–64, 2007. URL https://sora.unm.edu/node/139709. [p192]

N. Ferguson, S. Datta, and G. Brock. msSurv: An R package for nonparametric estimation of multistate models. *Journal of Statistical Software*, 50:1–24, 2012. URL https://doi.org/10.18637/jss.v050.i14. [p193]

E. Fujita, Y. Tanabe, A. Shiota, M. Ueda, K. Suwa, M. Y. Momoi, and T. Momoi. Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. *Proceedings of the National Academy of Sciences*, 105:3117–3122, 2008. URL https://doi.org/10.1073/pnas.0712298105. [p193]

S. Gaub, S. E. Fisher, and G. Ehret. Ultrasonic vocalizations of adult male Foxp2-mutant mice: behavioral contexts of arousal and emotion. *Genes, Brain and Behavior*, 15:243–259, 2016. URL https://doi.org/10.1111/gbb.12274. [p193]

S. Geisser and W. F. Eddy. A predictive approach to model selection. *Journal of the American Statistical Association*, 74:153–160, 1979. URL https://doi.org/10.2307/2286745. [p199]

W. Gradner. Analyzing sequential categorical data: Individual variation in Markov chains. *Psychometrika*, 55:263–275, 1990. URL https://doi.org/10.1007/BF02295287. [p192]

T. Holsclaw, A. M. Greene, A. W. Robertson, and P. Smyth. Bayesian nonhomogeneous markov models via pólya-gamma data augmentation with applications to rainfall modeling. *The Annals of Applied Statistics*, 11:393–426, 2017. URL https://doi.org/10.1214/16-AOAS1009. [p192]

M. A. Islam and R. I. Chowdhury. A higher order Markov model for analyzing covariate dependence. *Applied Mathematical Modelling*, 30:477–488, 2006. URL https://doi.org/10.1016/j.apm.2005.05.006. [p192, 193]

C. Jackson. Multi-state models for panel data: the msm package for R. *Journal of Statistical Software*, 38:1–28, 2011. URL https://doi.org/10.18637/jss.v038.i08. [p193]

E. Juniper, P. O'byrne, G. Guyatt, P. Ferrie, and D. King. Development and validation of a questionnaire to measure asthma control. *European Respiratory Journal*, 14:902–907, 1999. URL https://doi.org/10.1034/j.1399-3003.1999.14d29.x. [p203]

K. Katsura. catdap, a categorical data analysis program package. *Computer Science Monograph*, 14, 1980. URL https://CRAN.R-project.org/package=catdap. [p193]

T. Kharrat, G. N. Boshnakov, I. McHale, and R. Baker. Flexible regression models for count data based on renewal processes: The Countr package. *Journal of Statistical Software*, 90:1–35, 2019. URL https://doi.org/10.18637/jss.v090.i13. [p193]

B. Li. Markov models for Bayesian analysis about transit route origin–destination matrices. *Transportation Research Part B: Methodological*, 43:301–310, 2009. URL https://doi.org/10.1016/j.trb.2008.07.001. [p192]

A. Listwon and P. Saint-Pierre. SemiMarkov: An R package for parametric estimation in multi-state semi-Markov models. *Journal of Statistical Software*, 66:1–16, 2015. URL https://doi.org/10.18637/jss.v066.i06. [p193, 203]

K. D. MacDermot, E. Bonora, N. Sykes, A.-M. Coupe, C. S. Lai, S. C. Vernes, F. Vargha-Khadem, F. McKenzie, R. L. Smith, A. P. Monaco, et al. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *The American Journal of Human Genetics*, 76:1074–1080, 2005. URL https://doi.org/10.1086/430841. [p193]

G. Marshall, W. Guo, and R. H. Jones. MARKOV: A computer program for multi-state Markov models with covariables. *Computer Methods and Programs in Biomedicine*, 47:147–156, 1995. URL https://doi.org/10.1016/0169-2607(95)01641-6. [p193]

D. Meyer, A. Zeileis, and K. Hornik. *vcd: Visualizing Categorical Data*, 2022. URL https://CRAN.R-project.org/package=vcd. R package version 1.4-10. [p193]

J. W. Miller. Fast and accurate approximation of the full conditional for gamma shape parameters. *Journal of Computational and Graphical Statistics*, 28:476–480, 2019. URL https://doi.org/10.1080/10618600.2018.1537929. [p196]

L. R. Muenz and L. V. Rubinstein. Markov models for covariate dependence of binary sequences. *Biometrics*, 41:91–101, 1985. URL https://doi.org/10.2307/2530646. [p192]

P. Muliere, P. Secchi, and S. G. Walker. Reinforced random processes in continuous time. *Stochastic Processes and their Applications*, 104:117–130, 2003. URL https://doi.org/10.1016/S0304-4149(02)00234-X. [p192]

J. O'Connell and S. Højsgaard. Hidden semi Markov models for multiple observation sequences: The mhsmm package for R. *Journal of Statistical Software*, 39:1–22, 2011. URL https://doi.org/10.18637/jss.v039.i04. [p193]

M. J. Phelan. Bayes estimation from a Markov renewal process. *The Annals of Statistics*, 18:603–616, 1990. URL https://doi.org/10.1214/aos/1176347618. [p192]

J. E. Pustejovsky. *ARPobservation: Simulating recording procedures for direct observation of behavior*, 2021. URL https://CRAN.R-project.org/package=ARPobservation. R package version 1.1. [p193]

R. Pyke. Markov renewal processes: definitions and preliminary properties. *The Annals of Mathematical Statistics*, 32:1231–1242, 1961. URL https://www.jstor.org/stable/2237923. [p193]

P. Saint-Pierre, C. Combescure, J. Daures, and P. Godard. The analysis of asthma control under a Markov assumption with use of covariates. *Statistics in Medicine*, 22:3755–3770, 2003. URL https://doi.org/10.1002/sim.1680. [p203]

A. Sarkar, J. Chabout, J. J. Macopson, E. D. Jarvis, and D. B. Dunson. Bayesian semiparametric mixed effects Markov models with application to vocalization syntax. *Journal of the American Statistical Association*, 113:1515–1527, 2018. URL https://doi.org/10.1080/01621459.2018.1423986. [p192, 193, 199]

M. Sesia, C. Sabatti, and E. J. Candès. Gene hunting with hidden Markov model knockoffs. *Biometrika*, 106:1–18, 2019. URL https://doi.org/10.1093/biomet/asy033. [p192]

M. Siebert and J. Söding. Bayesian Markov models consistently outperform PWMs at predicting motifs in nucleotide sequences. *Nucleic Acids Research*, 44:6055–6069, 2016. URL https://doi.org/10.1093/nar/gkw521. [p192]

G. A. Spedicato. Discrete time Markov chains with R. *The R Journal*, 9:84–104, 2017. doi: 10.32614/RJ-2017-036. URL https://doi.org/10.32614/RJ-2017-036. [p193]

S. Watanabe and M. Opper. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11:3571–3594, 2010. URL https://doi.org/10.48550/arXiv.1004.2316. [p199]

Y. Wu, E. D. Jarvis, and A. Sarkar. Bayesian semiparametric Markov renewal mixed models for vocalization syntax. *Biostatistics*, 2023. URL https://doi.org/10.1093/biostatistics/kxac050. To appear. [p192, 193, 199]

M. Zhang, P. W. van Rijn, P. Deane, and R. E. Bennett. Scenario-based assessments in writing: An experimental study. *Educational Assessment*, 24:73–90, 2019. URL https://doi.org/10.1080/10627197.2018.1557515. [p193]

Yutong Wu
Department of Mechanical Engineering
The University of Texas at Austin
204 E Dean Keeton St C2200, Austin, TX 78712-1591
United States
ORCID: 0000-0001-7828-9981
yutong.wu@utexas.edu

Abhra Sarkar
Department of Statistics and Data Sciences
The University of Texas at Austin
2317 Speedway D9800, Austin, TX 78712-1823
United States
ORCID: 0000-0002-6924-8464
abhra.sarkar@utexas.edu