JASA A&CS Reproducibility Initiative - Author Contributions Checklist Form

The purpose of the Author Contributions Checklist (ACC) Form is to document the code and data supporting a manuscript, and describe how to reproduce its main results.

Data

Abstract

We have included a synthetic data set, Data_Set.mat, which was simulated by closely mimicking the real Foxp2 data set analyzed in the paper. The simulation design is described as 'simulation scenario D' in Section 6 of the main paper. The simulated data set itself is also analyzed in the main paper. The data set comprises a matrix Data=[MouseId Genotype Context Ytminus1 Yt] with 148778 rows and 5 columns. The column contents are given below.

- **MouseId** = The mouse identifying labels *within* a genotype. Takes values in [1,2,…,8] for 8 Foxp2 mutant mice mice and in [1,2,…,6] for 6 wild type control mice.
- **Genotype** = The genotype the mouse comes from. Takes values in [F,W]=[1,2]. Together with MouseId, Genotype uniquely identifies a mouse.
- **Context** = The context under which the mouse is singing. Takes values in [U,L,A]=[1,2,3].
- **Ytminus1** = The syllable that precedes Yt. Takes values in [d,m,s,u,x]=[1,2,3,4,5].
- **Yt** = The syllables that make up the mouse songs. Takes values in [d,m,s,u,x]=[1,2,3,4,5]. Collectively, the entries in column Yt with [MouseId,Genotype,Context]=[m,g,c] comprise the song sung by mouse m from genotype g under context c.

MouseId	Genotype	Context	Ytminus1	Yt
1	1	1	3	3
1	1	1	3	5
1	1	1	5	1
1	1	1	1	3
…	…	…	…	…

Availability

The original data could not made available as they are being used for other projects and grants.

Code

Abstract

The codes implement Bayesian semiparametric mixed effects Markov models developed in the main paper. The codes are written in Matlab. The codes comprise a main program and a few utility functions. The codes also utilize the Matlab Tensor toolbox, see below. The implementation is fully automated, taking in only the data matrix (as described above) as a single argument. Additional descriptions and instructions are included as detailed comments in the body of the codes.

Description
The codes included in a zipped file. The main file is BSPMEMM.m. The main file calls some external functions that are also included in the zipped file. To run the codes, all m. Files should be included the working directory.
Licensing information = MIT License.
Version information = 1.1

Optional Information
Supporting software requirements = The codes depend on the (freely available) Tensor toolbox for Matlab. The toolbox can be freely downloaded from the url given below.
Brett W. Bader, Tamara G. Kolda and others. MATLAB Tensor Toolbox Version 2.6, Available online, February 2015. URL: http://www.sandia.gov/~tgkolda/TensorToolbox/.
Instructions on how to include the toolbox in the Matlab working path are included as comments at the beginning of the main file BSPMEMM.m.

Instructions for Use

Reproducibility
All results for simulation scenario D described in Section 6 of the main paper can be reproduced using the codes and the data set. These include Figure 8 and Figure 9 in the main paper and Figure S.4 and Figure S.6 in the Supplementary Materials.