

JASA ACS 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.

As of Sept. 1, 2016, the ACC Form must be included with all new submissions to JASA ACS.

This document is the initial version of the template that will be provided to authors. The JASA Associate Editors for Reproducibility will update this document with more detailed instructions and information about best practices for many of the listed requirements over time.

Data

Abstract (Mandatory)

Two raw dataset were used for our analysis. The first is daily average PM2.5 in micrograms per cubic meter from 2000 to 2012 inclusive for all zip codes in the mainland USA. The second is all hospitalizations among US Medicare beneficiaries aged 65+. The medicare dataset includes fields for date of hospitalization and the primary diagnosis on discharge, as denoted by codes from International Classification of Diseases, Revision 9 (ICD9).

Availability (Mandatory)

The raw Medicare data cannot be made publicly available due to restrictions laid out in a Data Use Agreement (DUA) with the Centers for Medicare & Medicaid Services, US Department for Health and Human Services.

Description (Mandatory if data available)

Permissions: Medicare data were provided by the Centers for Medicare & Medicaid Services under the DUA mentioned above. More information can be provided upon request.

Optional Information (complete as necessary)

N/A

Code

Abstract (Mandatory)

We provide code for: (1) fitting the spike and slab MORETreeS model (ssMORETreeS) via the variational inference algorithm described in our manuscript; (2) re-running the simulation study; (3) reproducing the results of the data example (although as data is not provided, this code cannot be run), and; (4) recreating all tables and figures in the manuscript (other than figures created using LaTeX code).

Description (Mandatory)

How delivered: R code available on GitHub

Licensing information (default is MIT License):

Link to code/repository: <https://github.com/emgthomas/moretrees>

Version information (e.g., for a Git repository, the number or branch+commit): Master branch, commit 99c8b2fd6b588243fff78c93e9969509c6b350ff.

Optional Information (complete as necessary)

Hardware requirements: Access to a cluster is recommended.

Supporting software requirements: The following R packages are required:

BoomSpikeSlab (version 1.1.0)
circlize (version 0.4.1)
collapsibleTree (version 0.1.6)
data.table (version 1.10.4-3)
doParallel version (1.0.11)
ggplot2 (version 3.2.0)
glue (version 1.3.1)
icd (version 2.3.1)
igraph version (1.1.2)
Matrix (version 1.2-11)
mclust (version 5.4)
plotly (version 4.7.1)
RColorBrewer (version 1.1-2)
reshape2 (version 1.4.3)
xtable (version 1.8-2)

Instructions for Use

Reproducibility (Mandatory)

All simulations can be re-run. The following tables and figures from the manuscript can be fully reproduced: Figure 3, Figures A.2 and A.6, Tables A.1 and A.2. All simulations can be reproduced. However, figures and tables related to the simulations can also be reproduced without re-running the simulations as full simulation results have been provided.

Simulations and relevant tables/figures can be reproduced by running the R scripts simulations.R followed by simulations_figures_and_tables.R. Simulation scenarios are denoted by the parameter simidx (which takes integer values 1 through 8) in the simulations.R script. All 8 scenarios (represented by the 8 panels in Figure 3) must be run to reproduce all data needed for the tables and figures. Use of a cluster is recommended.

The data example cannot be re-run as the data cannot be shared. However, results from the data example are saved in the data_example_results directory, and the following figures and table can be reproduced: Figure 4, Table 1, Figures A.3, A.4, and A.5, and the large table in Section C of the supplementary material. All these figures and tables related can be reproduced by running the script data_example_figures_and_tables.R.

Replication (Optional)

The script VI_functions.R provides all functions needed for fitting ssMORETreeS to case-crossover or case-control data via the variational inference algorithm described in the manuscript. The necessary input data types are described in the file. The script data_example_full.R can be used as a template.