STooDs configuration files

July 2, 2021

1 Introduction

The configuration files aim at specifying the key ingredients of a STooDs case study:

1. A *dataset* containing values taken by one of several predictand variables Y varying in space, time or other dimensions, along with the values taken by potential covariates X (aka predictors).
2. A *model* making assumptions on the probabilistic mechanism that generated the data. Typically, a STooDs model can be schematized as follows:
   (a) each variable in the dataset follows a distribution.
   (b) the parameters of this distribution may vary in space, time or other dimensions.
   (c) this variability is specified by formulas that combine parameters, covariates and processes.
3. The *prior distribution* specified for each parameter (if any).
4. The *hyperdistribution* specified for each process (if any).
5. the properties of the *MCMC sampler* used to explore the posterior distribution associated with the model.

Typical configuration files can be found in the folder 'Examples'.

2 Dataset configuration file

The data file should be structured as follows:

1. (compulsory) A single column containing all data values for the predictand Y.
2. (compulsory) A single column containing the variable index. If there is a single variable, the whole column should be equal to 1. If there are K variables, integers 1...K should be used to indicate the variable associated with each row.
3. (optional) One or several columns containing the covariates values X.

4. (optional) One or several columns containing the dimension indices values. For instance, if 'time' and 'space' dimensions are used, a first column should contain the time step and a second column the site number associated with each row. The names of these columns should correspond to (or at least contain) the names of the dimensions.

5. (optional) A single column containing the censoring type of the data. 0 corresponds to interval censoring, any negative number corresponds to a 'less than' censoring (true value is smaller than Y[i]), any positive number corresponds to 'more than' censoring (true value is larger than Y[i]).

6. (optional) A single column containing the width of the censoring interval for each data. True value is supposed to be in Y[i]+/- width[i] width (0 thus leads to no censoring)

The dataset configuration specifies how the data file should be interpreted and it contains the following lines:

1. (string) A descriptive name for the dataset.

2. (string) The file where the dataset is stored. It is recommended to use quotes and to write the full path to the data file.

3. (integer) The number of header lines in the data file.

4. (integer) The number of rows in the data file (excluding header lines).

5. (integer) The number of columns in the data file

6. (integer) The column containing the values of the predictand.

7. (integer) The column containing the variable index.

8. (integer) The columns containing the values of the covariates (comma-separated if several covariates, 0 if no covariate).

9. (integer) The columns containing the dimension index (comma-separated if several dimensions, 0 if no dimension is used).

10. (integer) The column containing the censoring type (0 for no censoring).

11. (integer) The column containing the censoring interval width when interval censoring is used (0 if not used).

3 Model configuration file

The model configuration file should be named 'model.config' and it contains the following lines:

1. (string) A descriptive name for the model.

2. (integer) The number of variables nVar.

3. (string) The name of each variable (size nVar, comma-separated).

4. (string) The parent distribution of each variable (size nVar, comma-separated).
5. (integer) The number of parameters for each parent distribution (size nVar, comma-separated).
6. (string) The name of each parent parameter (size sum(nParentPar), comma-separated).
7. (integer) The number of covariates nCov.
8. (string) The name of each covariate (size nCov, comma-separated).
9. (integer) The number of model parameters nPar.
10. (string) The name of each model parameter (size nPar, comma-separated).
11. (string) The configuration file for model parameters (priors)
12. (integer) The number of dimensions nDim.
13. (string) The configuration file for each dimension (size nDim, comma-separated).
14. (integer) The number of processes nPro.
15. (string) The name of each process (size nPro, comma-separated).
16. (string) The configuration file for each process (size nPro, comma-separated).
17. ...
18. (string) The formula for deriving the last parent parameter from model parameters, processes and covariates.
19. (string) The dataset configuration file.

4 Parameter configuration file

5 Dimension and process configuration files

5.1 Dimension

5.2 Process

6 MCMC configuration file