Documentation for Single Subject Bayesian Deconvolution

Kenneth Horton and Nichole E Carlson

1 General information

The code in this folder is for the single subject Bayesian deconvolution model as investigated in this paper (i.e., the exact same priors). We also have an R package (https://github.com/XXX); however the priors for the pulse masses and widths differ from the log normal. The package uses a truncated t-distributions. To run this code, the user needs to compile the 8 *.c files in this folder. Then to run the executable, the user needs to create an input file setting parameters in the priors, starting values for the MCMC algorithm and file names for output and data input. An example with numbers is provided in the simdata_example folder. We also provide an example with words used to describe the numbers needed. Below we also provide a dictionary for the values that need to be in the input file.

The data files need to be a space delimited *.dat file with 3 columns and a header row. The column names are OBS (integer), TIME (in min), hormone concentration (conc. units).

In addition, in the directory a seed file ("seed.dat") is provided and needs to be in the directory with the complied executable file. An example seed file is in this directory for ease of copying. This file is the input for the random number generators.

2 Creating the input file

In the input.dat file the following are the values needed by row (space delimited). We first set the file names, then needs for the MCMC run length, then the parameters for the priors are specified, followed by starting values and initial proposal variances.

1. The name of the data file for the subject being analyzed.

2. The name of the output file for the common parameters (baseline, half-life, variance parameters, etc.) and the name of the output file for the pulse specific parameters (pulse locations, mass and width).

3. Number of MCMC iterations to run (thinning is hard coded at NN=50 in the mcmc.c file and the screen output is hard coded at NNN=5000 in mcmc.c).

4. Mean (m_α) and variance (v_α^2) of the prior on the log pulse mass mean (μ_α).

5. Mean (m_ω) and variance (v_ω^2) of the prior on the log pulse width mean (μ_ω).

6. Mean (m_b) and variance (v_b) of the prior on the baseline (θ_b).

7. Mean (m_h) and variance (v_h) of the prior on the half-life (θ_h).

8. The two parameters in the Inverse-Gamma prior on the model error variance parameter (σ_e^2).

9. The maximum value for the Uniform prior on the pulse-to-pulse standard deviation of log pulse mass (ν_α) and log pulse width (ν_ω).

10. The mean of the Poisson distribution for the prior on the number of pulses (r).

11. The starting values of the mean log pulse mass (μ_α) and the mean log pulse width (μ_ω).

12. The starting values of the baseline (θ_b) and half-life (θ_h).

13. The starting value for the model error variance (it is inverted in the algorithm).

14. The starting value for the pulse-to-pulse SD of the log pulse masses (ν_α) and log pulse widths (ν_ω).

15. The initial proposal variances for baseline (θ_b) and half-life (θ_h). The correlation is assumed -0.9, which sets the covariance for the matrix. The hard coding is in mcmc.c.

16. The initial proposal variances for the pulse to pulse SD of log pulse mass (ν_α) and log pulse width (ν_ω).

17. The initial proposal variances for the individual log pulse masses ($\log \alpha_k$), log pulse widths ($\log \omega_k$), and pulse locations (in minutes, τ_k).

3 Interpreting the output files

1. For each subject there is common hormone parameters (*.out) and pulse parameters file (*.out), both named by the user in the input file.

2. Common parameters output file: These are the parameters that are common across pulses. The columns are: number of pulses (N_s), baseline (θ_b), mean log pulse mass μ_α, mean log pulse width μ_ω, half-life (θ_h), model error (σ_e^2), pulse-to-pulse SD for log pulse mass (ν_α), pulse-to-pulse SD for log pulse width (ν_ω).

3. Pulse parameters output file: These are the pulse specific parameters. The columns are: iteration number after thinning, number of pulses (N_s), pulse number in this iteration (a counter), pulse location (τ_k), pulse mass (α_k), pulse width (ω_k).

4 Interpreting the screen output

Every 5000 iterations the following output information is written to the screen. The purpose of the output is generally to monitor acceptance rates prior to a full run being complete.

Each write to the screen has the following form:

1. The iteration number, the current value of the log likelihood.

2. The current parameter value for A (mean log pulse mass hormone concentration scale), s (mean log pulse width min^2 scale), d (half-life in min), v (model error variance).

3. The pulse location (in hours), mass (conc. units), width ($hours^2$) for each pulse.

4. Current acceptance rates in the birth-death algorithms: rem = acceptance % of individual pulse masses, rew = acceptance % of individual pulse widths, time = acceptance % of individual pulse locations, md = acceptance %, revm = acceptance % of SD for pulse masses, revw = acceptance % of SD for pulse widths).

5 Compiling the code

A makefile exists in the file for compiling the c code in this directory.

1. deconvolution_main.c: This file reads in the data. Sets the initial values, the priors and the proposals for the MH algorithms. It initializes the pulse parameter lists and the cluster lists. It calls the MCMC algorithm.

2. format_data.c: This file is called in deconvolution_main.c and is the algorithm for reading in the data.

3. mcmc.c: This file is the birth-death MCMC algorithm (function mcmc) and controls screen and file output. It is called from deconvolution_main.c.

4. birthdeath.c: This file is the birth-death algorithm for the pulse locations. The function birthdeath is called from the mcmc.c file.

5. The other files in the directory are supporting subroutines for the random number generators, etc.