Documentation for Population Bayesian Deconvolution

Kenneth Horton and Nichole E Carlson

1   General information

The code in this folder is for the population Bayesian deconvolution model as investigated in this paper (i.e., the exact same priors). To run this code, the user needs to compile the 8 *.c files in the src 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 1 column for time of observations in minutes (first column) plus a column of observations for each subject. There is no header row.

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 root 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). A subject number is added at the end of each file with the *.out extension.

3. Number of subjects in the analysis and the number of MCMC iterations to run (thinning is hard coded at NN=50 in the pop_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 population log pulse mass mean (μ_α) and the maximum of the uniform prior of the SD on the pulse-to-pulse variation in the log masses (σ_α in Σ_s) and the maximum of the uniform prior of the SD on subject-to-subject variation in mean log pulse masses (v_α in Υ_s).

5. Mean (m_ω) and variance (v_ω^2) of the prior on the log pulse width mean (μ_ω) and the maximum of the uniform prior of the SD on the pulse-to-pulse variation in the log widths (σ_ω in Σ_s) and the maximum of the uniform prior of the SD on subject-to-subject variation in mean log pulse widths (v_ω in Υ_s).

6. Mean (m_b) and variance (v_b) of the prior on the baseline (θ_b) and upper bound on the uniform prior on the SD of the subject-to-subject variation (σ_b).

7. Mean (m_h) and variance (v_h) of the prior on the half-life (θ_h) and upper bound on the uniform prior on the SD of the subject-to-subject variation (σ_h).

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

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

10. The starting values of the mean log pulse mass (μ_α), SD's of the pulse-to-pulse variation of log pulse mass (σ_α) and the subject-to-subject variation of the mean log pulse mass (υ_α).

11. The starting values of the baseline (θ_b) and SD of the subject-to-subject variation of the baselines (σ_b).

12. The starting values of the half-life (θ_h) and the SD of the subject-to-subject variation of the half-lives (σ_h).

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

14. The initial proposal variances for subject level mean log pulse mass, population mean log pulse mass, and individual log pulse masses.

15. The initial proposal variances for subject level mean log pulse width, population mean log pulse width, and individual log pulse widths.

16. The initial proposal variances for the population mean baseline and the subject level baselines.

17. The initial proposal variances for the population mean half-life and the subject level half-lives.

18. The initial proposal variances for the pulse locations.

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. There is also a population parameters file (fileroot from above.out. Same name as the subject common parameter file without subject number).

2. Population parameters output file: These are the parameters that are common across subjects. The columns are: mean log pulse mass (μ_α), mean log pulse width (μ_ω). Subject-to-subject SD of mea log pulse mass and mean log pulse width. Pulse-to-pulse SD of pulse mass and Width. Mean baseline and SD of baselines between subjects. Mean half-life and SD of half-lives between subjects. Model error variance.

3. Subject level common parameters file: These are the parameters that are common across pulses in a subject. The columns are: number of pulses (N_s), mean log pulse mass μ_α, mean log pulse width μ_ω, baseline (θ_b), half-life (θ_h).

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

2. The current parameter value for mean baseline, mean half-life, mean of mean log pulse masses, mean of mean log pulse width (min^2 scale), model error variance.

3. Current acceptance rates in the birth-death algorithms: SD of mean log pulse masses (s_ma) and SD of log pulse masses (s_a), individual log pulse masses ($A_k i$).

4. Acceptance rates for SD of mean log pulse widths, SD of log pulse widths, individual log pulse widths.

5. Acceptance rates for SD of baseline, SD of half-life, mean baseline/half-life and individual pulse locations.

5 Compiling the code

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

1. pop_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. pop_format_data.c: This file is called in deconvolution_main.c and is the algorithm for reading in the data.

3. pop_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. pop_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.