

Documentation for Bayesian Joint Deconvolution Model

Huayu Liu and Nichole E Carlson

1 General information

The code in this folder is for the joint Bayesian deconvolution model as investigated in the paper [Liu, et al. (2017) Modeling Associations between Latent Event Processes Governing Time Series of Jointly Secreted Pituitary Hormones]. 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. In addition, there are still several parameters that are being hard coded in the jt_deconvolution_main.c file. These include the priors on ρ and ν on lines 273-278 and the starting values for ρ and ν on lines 333 and 336 respectively. Finally the parameters in the Strauss are set on lines 339-339 and 277-278.

Two data files are needed, one for triggers hormone and the other for response hormone. 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. two data files need to have exactly the same number of columns.

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 two data file for the subject being analyzed. The first one is the trigger hormone series and the second one is the response hormone series.
2. The name of the root of the output file for the common parameters (number of pulses, 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 joint_mcmc.c file and the screen output is hard coded at NNN=5000 in mcmc.c).
4. Mean ($\mathbf{m}_s^{x,y}$) of the prior on the population pulse mass and width mean vectors. The order is: trigger pulse mean, response pulse mass mean, trigger pulse width mean, response pulse width mean. with correlation 0.
5. First three numbers: prior distribution of $\Sigma_{\alpha,s}$: mean of the subject-to-subject variation for mean mass in trigger hormone, mean of the subject-to-subject variation for mean mass in response hormone, subject-level correlation between trigger and response hormone mass.
Fourth to fifth number: upper bound on the uniform prior on the $\Sigma_{\omega,s}$: subject-to-subject variation for mean width in trigger hormone, subject-to-subject variation for mean width in response hormone. The correlation between trigger and response hormone width is assumed to be 0.
Sixth to seventh number: Variance of the prior on the population pulse mass for trigger and response hormone.

Eighth to eleventh number: upper bound on the uniform prior on the pulse-to-pulse variation for mass and width for trigger hormone (Υ_s^x), and pulse-to-pulse variation for mass and width for response hormone (Υ_s^y).

6. Population Mean (m_b^x) and variance (v_b^x) of the prior **on the baseline** (μ_b^x) and upper bound on the uniform prior on the SD of the subject-to-subject variation (σ_b^x) for trigger hormone, mean (m_b^y) and variance (v_b^y) of the prior on the baseline (μ_b^y) and upper bound on the uniform prior on the SD of the subject-to-subject variation (σ_b^y) for response hormone.

7. Population Mean (m_h^x) and variance (v_h^x) of the prior **on the half-life** (μ_h^x) and upper bound on the uniform prior on the SD of the subject-to-subject variation (σ_h^x) for trigger hormone, mean (m_h^y) and variance (v_h^y) of the prior on the half-life (μ_h^y) and upper bound on the uniform prior on the SD of the subject-to-subject variation (σ_h^y) for response hormone.

8. The two parameters in the Inverse-Gamma prior on the model error variance parameter for trigger hormone (Σ_e^x), the two parameters in the Inverse-Gamma prior on the model error variance parameter (Σ_e^y) for response hormone.

9. The starting values of $\boldsymbol{\mu}_s^{x,y}$: the mean pulse mass for trigger and response hormone, the starting values of the mean pulse width for trigger and response hormone.

10. First three numbers: the starting values for entries in $\Sigma_{\alpha,s}$: subject-to-subject variation for mean mass in trigger hormone, subject-to-subject variation for mean mass in response hormone, subject-level correlation between trigger and response hormone mass.

Fourth to fifth number: the starting values for $\Sigma_{\omega,s}$: subject-to-subject variation for mean width in trigger hormone, subject-to-subject variation for mean width in response hormone.

Sixth to ninth number: the starting values for Υ_s^x and Υ_s^y .

11. The starting values of the baseline (u_b^x) and SD of the subject-to-subject variation of the baselines (σ_b^x) for trigger hormone, the starting values of the baseline (u_b^y) and SD of the subject-to-subject variation of the baselines (σ_b^y) for response hormone.
12. The starting values of the half-life (u_h^x) and SD of the subject-to-subject variation of the half-life (σ_h^x) for trigger hormone, the starting values of the half-life (u_h^y) and SD of the subject-to-subject variation of the baselines (σ_h^y) for response hormone.
13. The starting value for the model error variance.
14. The initial proposal variances for subject level pulse width variance for trigger and response hormone.
15. The initial proposal variances for variance of baseline and half-life for trigger and response hormone.
16. The initial proposal variances for subject mean mass for trigger and response pulse mass, and their correlation.
17. The initial proposal variances for subject level pulse width for trigger and response hormone.
18. The initial proposal variances for pulse level variance for mass and width for trigger and response hormone.
19. The initial proposal variances for subject level baseline and half-life variance for trigger and response hormone.
20. The initial proposal variances for pulse-level mass and width for trigger and response hormone.
21. The initial proposal variances for pulse locations for trigger and response hormone.

22. The initial proposal variances for η in the t-distribution for pulse level mass and width for trigger and response hormone.
23. The initial proposal variances for expected number of response pulses associated with each trigger (ρ) and spread parameter for each trigger-response pair (ν).

3 Interpreting the output files

1. For each subject there is common hormone parameters, and pulse parameters file (*.out, both named by the user in the input file for trigger and response hormone. There is also a population parameters file for trigger and response hormone (fileroot from above.out. Same name as the subject common parameter file without subject number). The files for trigger hormone contain x, and the files for response hormone contain y.
2. Population parameters output file: These are the parameters that are common across subjects. The columns are: square of the spread parameter for each trigger-response pair (ν^2), the expected number of response pulse for each trigger (ρ), mean pulse mass (μ_α), mean pulse width (μ_ω), four entries in the variance-covariance matrix of subject-to-subject level pulse mass between trigger and response hormone, variance of subject-to-subject pulse width, Mean baseline and SD of baselines between subjects. Mean half-life and SD of half-lives between subjects. Model error variance. The response hormone file dose not have first two columns.
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 pulse mass μ_α , mean pulse width μ_ω , baseline (θ_b), half-life (θ_h). The subject level common parameter file for trigger hormone also has two column of ν^2 and ρ as last two column.
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), and two η parameters in drawing the pulse level mass and width using t-distribution.

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 values for the trigger hormone: iteration, ν the response width in the Cox model, mean baseline, ρ the after number of responses for each trigger, mean mass and width, subject to subject variance and covariance of pulse masses between trigger and response pulse mass means, variance of the pulse widths, pulse to pulse variation in mass and width, mean), mean baseline, mean half-life, subject to subject variance of baseline and half-life, model error variance for trigger (labelled with LH) and response (labelled with FSH) hormone.
3. The current parameter values for the response hormone: iteration, mean mass and width, subject to subject variance and covariance of pulse masses between trigger and response pulse mass means, variance of the pulse widths, pulse to pulse variation in mass and width, mean), mean baseline, mean half-life, subject to subject variance of baseline and half-life, model error variance for trigger (labelled with LH) and response (labelled with FSH) hormone.
4. Acceptance rates for trigger and response pulse parameters.
5. Current listing of pulses for the trigger and response for each person in the study.

5 Compiling the code

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

1. `jt_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. `jt_format_data.c`: This file is called in `deconvolution_main.c` and is the algorithm for reading in the data.
3. `jt_mcmc.c`: This file is the birth-death MCMC and other MCMC algorithm (function `mcmc`) and controls screen and file output. It is called from `jt_deconvolution_main.c`.
4. `jt_birthdeath.c`: This file is the birth-death algorithm for the pulse locations. The function `birthdeath` is called from the `linklistv2.c` file.
5. The other files in the directory are supporting subroutines for the random number generators, etc.