This article is a tutorial about how to compile and to use the programs in the paper "Bayesian Inferences of Latent Class Models with An Unknown Number of Classes via Reversible Jump Markov Chain Monte Carlo" (the RLCA-RJ paper). The object of these programs is to perform regression latent class analysis (RLCA) by adopting reversible jump Markov Chain Monte Carlo (RJMCMC) on a dataset with the specific format describe in section 2.

# 1 How to compile `RLCA_RJMCMC.cpp`

The file `RLCA_RJMCMC.cpp` contained in the folder `supplemental` is a `C++` program. One can compile these program by using `g++` compiler in Linux-like system or in Cygwin software under windows system.

## 1.1 For Linux-like system user

Linux-like system users can use the file `makefile_RLCA` to compile these programs. The compiling steps are as follows.

1. Change the working directory to the folder supplemental.
   `$ cd /YOURPATH/supplemental`
2. Run the following commands:
   `$ make -f makefile_RLCA`
   If it compiles successfully, it will produce a executable file `RLCA_RJMCMC` in the current directory.

## 1.2 For Windows system user

Windows system user need to download the software Cygwin and then install it. Cygwin is a Linux-like environment for Windows, once Cygwin have been installed, one can compile `RLCA_RJMCMC.cpp` by using `g++` compiler in Cygwin. The followings are the steps for installation of Cygwin and compilation.

1. Download and install setup.exe in the website of Cygwin (by clicking "install or update now!").
2. Click "Next" → choose "Install from internet" → type in your root directory (eg: D:/cygwin) → type in the directory that will storage the downloaded materials → select your internet connection as "Direct Connection" → choose a ftp as your download site → select "Base", "Devel" and "Editor" packages to be installed and click "Next" → installation completes.
3. Copy/save the supplemental folder to the home directory of Cygwin (eg: D:/cygwin/home/USER/).
4. Open Cygwin terminal and change the working directory to the folder supplemental.
   `$ cd supplemental`
5. Convert file format from Windows to Linux:
   `$ dos2unix *`
6. Run the following commands:
   `$ make -f makefile_RLCA`
   If it compiles successfully, it will produce a executable file `RLCA_RJMCMC.exe` in the current directory.

# 2 The format of dataset

To perform `RLCA_RJMCMC`, three files are required as inputs of `RLCA_RJMCMC`: `Items_response_file`, `Latent_class_covariate` and `Condition_covariates_file`.

- Items response file: The `Items_response_file` has specific format, recording the response variables with multiple items in RLCA. Suppose that there are 2 items to be analyzed, the first item has three categories and the second has two. The most often coded format is shown in Table 1. To use our

| obs. index | item 1 | item 2 |
|---|---|---|
| 1 | 3 | 2 |
| 2 | 1 | 1 |
| 3 | 2 | 2 |

Table 1: 2 items with often coded format.

program, we need to change the items response file to 0-1 coding as that in Table 2. The variable in item 1 has 3 categories. Table 2 summarize this variable with 3 column to represent for the item 1. The subsequent two columns represent for the item 2. Notice that, the first column is necessarily

| | | | | | |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 1 | 0 | 1 |
| 2 | 1 | 0 | 0 | 1 | 0 |
| 3 | 0 | 1 | 0 | 0 | 1 |

Table 2: 2 items with 0-1 format.

required in `Items_response_file` for our program standing for observation index, and others columns are 0-1 values to summary the items values. One could use the `R` program `transY.R` in `supplemental` folder to help transforming the format from Table 1 to Table 2. Save these items values in the format of Table 2 with filename `itemsFile.txt`.

- Latent class covariates file: `Latent_class_covariates_file` is a file contains covariates associated with latent class probabilities (the $\mathbf{x}$'s in the RLCA-RJ paper). The first column of this file must be the observation index. The other columns are covariate variables and, basically, one column represents a covariate variable. Therefore, suppose $p$ covariates are of interest, there are totally $p+1$ columns in latent class covariates file, and save filename as `LCXFile.txt`. If a $C$-level nominal covariate is of interest, $C-1$ dummy variables with $C-1$ columns should be created by user in `LCXFile.txt`.
- Condition covariates file: `Condition_covariates_file` is the file contains covariates associated with conditional probabilities (the $\mathbf{z}$'s in the RLCA-RJ paper). This file is in the same format as latent class covariates file, and save with a filename (eg:`conZFile.txt`).

Notice that, first, the first column in each file must be the observation index, and their values must coincide with each other. Second, the input files should not contain any header names in each file. Three demonstrated files are in the folder supplemental: `itemsFile.txt`, `LCXFile.txt` and `conZFile.txt` for illustration.

# 3 Parameters setting

`parametersfile.txt` is a file specifies required parameters in program `RLCA_RJMCMC`. `RLCA_RJMCMC` automatically recognizes filename "`parametersfile.txt`", and reads the content of `parametersfile.txt` to change the default parameters setting in the program. Parameters can be classified into two categories. One is associated with data (data parameters); the other is associated with program (program parameters). The data parameters must be set depending on your data characteristics, which includes:

totalsweep
An integer value. The number of sweeps in MCMC runs.

N
An integer value. The number of observations in analyzed data.

M
An integer value. The number of items in data.

K1, ...
An integer value. K1 is the number of categories of item 1. K2 is the number of item two, and so on.

xprev
An integer value. The number of covariate columns (except for the column of observation index) in `LCXFile.txt`.

L
An integer value. The number of covariate columns (except for the column of observation index) in `conZFile.txt`.

J
An integer value. The initial value of number of classes in RJMCMC for RLCA model.

Program parameters having the default setting according to the RLCA-RJ paper in the program. and the values are not necessary to change. However, if the default setting is not appropriate for the data to be analyzed, the change of program parameters are recommended. The principle of how to adjust to a proper setting is shown in RLCA-RJ paper. Program parameters include:

BDjump_prob
A logical value. If BDjump_prob=1, the birth and the death step is engaged in RJMCMC; else BDjump_prob=0, the birth and the death step is not allowed to perform.

MSjump_prob
A logical value. If MSjump_prob=1, the split and the merge step is engaged in RJMCMC; else MSjump_prob=0, the split and the merge step is not allowed to perform.

prostd_beta
A numeric value greater than 0. prostd_beta is the standard deviation (SD) value of the proposals ($\sigma_{BD}$) in the reversible jump steps for $\beta$s. In the RLCA-RJ paper, we set prostd_beta = prostd_gamma.

prostd_gamma
A numeric value greater than 0. prostd_gamma is the SD value of the proposals ($\sigma_{BD}$) in the reversible jump steps for $\gamma$s. In the RLCA-RJ paper, we set prostd_beta = prostd_gamma.

priorstd_beta
A numeric value greater than 0. priorstd_beta is the SD value of the prior ($\sigma_P$) for $\beta$s. In the RLCA-RJ paper, we set priorstd_beta = priorstd_alpha = priorstd_gamma.

priorstd_gamma
A numeric value greater than 0. priorstd_gamma is the SD value of the prior ($\sigma_P$) for $\gamma$s. In the RLCA-RJ paper, we set priorstd_beta = priorstd_alpha = priorstd_gamma.

priorstd_alpha
A numeric value greater than 0. priorstd_alpha is the SD value of the prior ($\sigma_P$) for $\alpha$s. In the RLCA-RJ paper, we set priorstd_beta = priorstd_alpha = priorstd_gamma.

c_beta
A numeric value less than 0. c_beta is the scale constant value ($c^*$) of the Gibbs sampling for generating $\beta$s.

c_gamma
A numeric value less than 0. c_gamma is the scale constant value ($c^*$) of the Gibbs sampling for generating $\gamma$s.

c_alpha
A numeric value less than 0. c_alpha is the scale constant value ($c^*$) of the Gibbs sampling for generating $\alpha$s.

# 4 The usage of RLCA_RJMCMC

SYNOPSIS

```
RLCA_RJMCMC Items_response_file [OPTIONS] [FILE...]
```

OPTIONS

`-h, ---help`
Print a usage message briefly summarizing these command-line options.

`-r, ---regression`
This option follows a latent class covariates filename. If this option is specified, a regression latent class analysis is invoked.

`-e, ---extended`
This option follows a conditional covariates filename. If this option is specified, a regression extension latent class analysis is invoked.

EXAMPLE

- For adopting Bayesian RLCA model with RJMCMC, enter the commands:
  `$ ./RLCA_RJMCMC itemsFile.txt -r LCXFile.txt -e conZFile.txt`
- For adopting Bayesian RLCA model with RJMCMC without covariates $\mathbf{z}$, enter the commands:
  `$ ./RLCA_RJMCMC itemsFile.txt -r LCXFile.txt`
- For adopting Bayesian LCA model with RJMCMC, enter the commands:
  `$ ./RLCA_RJMCMC itemsFile.txt`
- If analysts want to adopting (R)LCA model without dimension jumping, the values of BDjump_prob and MSjump_prob in `parameterfile.txt` should be modified with zero.

# 5 The outputs of RLCA_RJMCMC

`RLCA_RJMCMC` will produce `out*.txt` files while it is executing. The `out*.txt` contains `outJ.txt`, `outbeta.txt`, `outgamma.txt` and `outalpha[1-M].txt`, where `M` is the number of items. One row in each file is the value of posterior sample(s) in the corresponding (RJ)MCMC sweeps. The length of rows is determined by the value of totalsweep in `parameterfile.txt`. In each file, the first column is the sweep index, and the other columns are the posterior samples. For example in `outJ.txt` there are two columns in `outJ.txt`. The first column is the sweep index; the second is the value of $J$ under each sweep.

# 6 Analysis of the outputs

In the folder supplemental, two programs `analyze_output.cpp` and `separate_output.cpp` are the programs to analyze the output files of `RLCA_RJMCMC`. Both files will recognize `parametersfile.txt` and read the data parameters as input parameters. Please be sure that the data parameters is correct for your data and it is also the data parameters for your `out*.txt` files.

`analyze_output.cpp` is the program for the analysis of the output files. It will return two parts of informations: probability information and coefficient information. To use `analyze_output.cpp`, first compile it with

```
$ make -f makefile_analyze
```

and then execute it with

$ ./analyze_output -L *lower_bound* -U *upper_bound* -a *alpha*

where *lower_bound* and *upper_bound* are the lower bound and upper bound of the value of $J$ and *alpha* is the significant level. If we specify `-L` 4 `-U` 6, it will return the analyzed results under every value of $J$ ranging between 4 and 6. The specified `-a` *alpha* will give the significance remark on (1-*alpha*)×100% credible interval for each posterior sample under each value of $J$ when it is significant. After executing this command, an output file [1]`analyze_output.txt` will be produced. The file `analyze_output.txt` contains the analyzed results shown on the terminal.

`separate_output.cpp` is the program for separating the output files according to the value of $J$. That is, this program will collect the sweeps belong to the same value of $J$, so that we can do further analysis for the posterior samples under the specific value of $J$. To use `separate_output.cpp`, type the command

```
$ g++ separate_output.cpp -o separate_output
```

for compilation and

$ ./separate_output -L *lower_bound* -U *upper_bound*

for execution. In the command, we use *lower_bound* and *upper_bound* to specify which values (ranging from *lower_bound* to *upper_bound*) of $J$ being selected for separation.

# 7 Example

In the folder supplement, we give simulated file `itemsFile.txt`, `LCXFile.txt` and `conZFile.txt` for illustrating how to use these programs.

```
$ # change to the directory supplemental
$ cd supplemental
$ dos2unix *

$ # For compilation, enter the commands:
$ make -f makefile_RLCA
$ g++ separate_output.cpp -o separate_output
$ make -f makefile_analyze

$ # For adopting Bayesian RLCA model with RJMCMC, enter the commands:
$ ./RLCA_RJMCMC itemsFile.txt -r LCXFile.txt -e conZFile.txt
```

[1]The interpretation of statics is summarized in Section 8

```
$ # When the out*.txt produced from the above command,
$ # type the following command to analyze the results of the posterior samples of J=6.
$ ./analyze_output -L 6 -U 6 -a 0.05

$ # Or type the following command to obtain the posterior samples
$ # of J=6 (save as out*_6.txt).
$ ./seperate_output -L 6 -U 6
```

# 8 The statistics of analyze_output.txt

The summary of data parameters was printed on the screen:

```
M=5
N=1500
L=2
xprev=2
```

The frequency of posterior samples of $J$.

```
==================
J frequency
------------------
1 0
2 344
3 1060
4 1273
5 13456
6 64478
7 17295
8 1594
9 470
10 14
11 13
12 3
13 0
14 1
==================
```

Prevalence and conditional probability of each class (statistics of the probability is calculated according to Huang, G. H. and Bandeen-Roche, K., 2004)

```
================================================================================
TOTAL CLASSES: 6
PROBABILITY INFORMATION
--------------------------------------------------------------------------------
LATENT CLASS PROBABILITY:
---------------------------------
prevalence
probability
---------------------------------
Class 1 :  0.449764
Class 2 :  0.028847
```

```
Class 3 :  0.137170
Class 4 :  0.013958
Class 5 :  0.163524
Class 6 :  0.206737
---------------------------------

--------------------------------------------------------------------------------
CONDITIONAL PROBABILITY:
item 1       class1 class2 class3 class4 class5 class6
item.level 1 0.9258 0.1006 0.0392 0.7806 0.0600 0.9934
item.level 2 0.0505 0.6125 0.3246 0.2104 0.5537 0.0042
item.level 3 0.0237 0.2869 0.6362 0.0090 0.3863 0.0025
--------------------
item 2
item.level 1 0.2455 0.0231 0.0002 0.1199 0.1038 0.1167
item.level 2 0.1818 0.1134 0.9794 0.1388 0.8814 0.2077
item.level 3 0.5726 0.8635 0.0204 0.7414 0.0148 0.6756
--------------------
item 3
item.level 1 0.0028 0.9558 0.7316 0.0102 0.0017 0.9576
item.level 2 0.9930 0.0334 0.2227 0.5750 0.9684 0.0236
item.level 3 0.0043 0.0109 0.0457 0.4147 0.0299 0.0188
--------------------
item 4
item.level 1 0.0243 0.0026 0.8329 0.8233 0.7749 0.7104
item.level 2 0.0045 0.9864 0.1486 0.0029 0.0043 0.2446
item.level 3 0.9712 0.0110 0.0185 0.1738 0.2208 0.0450
--------------------
item 5
item.level 1 0.0896 0.0765 0.0706 0.0169 0.0229 0.0025
item.level 2 0.4396 0.9055 0.8649 0.1131 0.0213 0.4177
item.level 3 0.4707 0.0180 0.0646 0.8700 0.9559 0.5798
--------------------------------------------------------------------------------
```

Latent prevalence regression from RLCA for the relationship between underlying subclass and external covariates. Coefficient of latent class covariates after taking exponential and information about significant of credible interval. COEFFICIENTS INFORMATION ———————————————————————— exp(coefficient) of latent class covariates (* represents for significant for 95———————————————————————— intercept 2.1755* 0.13954* 0.6635 0.067518* 0.79097 covariate 1 0.034386* 0.21715 3.3601* 0.39851* 3.8625* covariate 2 0.14193* 48.975* 6.1999* 0.0078593* 2.4074* ————————————————————————

Conditional probability regression from RLCA for the relationship between items and external covariates. Coefficient of covariates after taking exponential and information about significant of credible interval. ———————————————————————— exp(coefficient) of covariates (* represents for significant for 95———————————————————————— item 1 ———————————- item.level 1 covariate 1 0.0047886* covariate 2 4.8788* item.level 2 covariate 1 18.18* covariate 2 0.019361* ————————————- item 2 ————————————- item.level 1 covariate 1 14.963* covariate 2 15.804* item.level 2 covariate 1 0.7082 covariate 2 0.044823* ————————————- item 3 ———————————- item.level 1 covariate 1 0.15224* covariate 2 10.872* item.level 2 covariate 1 0.95343 covariate 2 0.030011* ————————————- item 4 ————————————- item.level 1 covariate 1 0.068827* covariate 2 87.161* item.level 2 covariate 1 0.014136* covariate 2 0.036679* ————————————- item 5 ———————————

————————- item.level 1 covariate 1 0.036454* covariate 2 6.7663* item.level 2 covariate 1 0.74366 covariate 2 0.0092295* ————————————————————————————————

# 9 Other files

The sensitivity analysis is written in R software. The file `sensitivity_analysis` contains the `R` program to perform the sensitivity analysis in the Section 5 of RLCA_RJ paper. Various plots Fig. 4-7, 9-11 and table 3 of your own data can be produced by using this file.

The file `transY.R` help recoding the items file from the format of Table 1 to 2. Please see the example in this file for understanding how to use it.