This folder contains several executable files for Windows 11 Operating System, two input files, and several output files. The executable files are used to analyze the multinomial probit model.

Please contact Prof. Xiao Zhang at xzhang35@mtu.edu if you need executable files for other operating systems such as Linux and Unix.

Methods Implemented

There are four executable files. The following describes the method implemented in each executable file.

- PX-GFN-U.exe: the PX-GFN algorithm in the manuscript (parameter-expanded Gibbs sampling with fully marginalization for the MNP model)
- PX-GMN-U.exe: the PX-GMN algorithm in the manuscript (parameter-expanded Gibbs sampling with marginalization for the MNP model)
- PX-GS.exe: the PX-GS algorithm in the manuscript (parameter-expanded Gibbs sampling)
- PX-MH.exe: the PX-MH algorithm in the manuscript (parameter-extended Metropolis-Hastings)

Run Program

The use of four executable files is similar. Here we use PX-MH.exe as an example to illustrate how to run the program in Windows 11 Operating System. The following steps are needed:

1. Download all files to a folder.
2. Open Command Prompt: you need to (1) Open the Start menu; (2) Type "cmd" in the search box; and (3) Click Command Prompt to open it.
3. Use command “cd” to change the folder to the folder that contains the executable files and input files.
4. Type the following command:

PX-MH -N 2657 -P 6 -K 5 -m0 10 -m 4000 -RS 0 -S 20000 -PD 3 -SP 1 -NN 1 6 -NO 0 -xFile Input_X.dat -yFile Input_Y.dat -vt 1,1,1,1,1 -outPre res-

Arguments Used

Each argument is passed to the program through the following way: the symbol hyphen ("-"), then the key word, the space, and the value. Different arguments are separated by spaces.

- -N: the number of samples in the data.
    - For this example, "-N 2657" means that there are 2647 samples.
- -P: the number of covariates
    - For this example, "-P 6" means that there are 6 covariates.
- -K: the number of latent variables which is the number of levels of the nominal response minus 1.
    - For this example. "-K 5" means that there are 5 latent variables. Therefore, the number of levels for this nominal response is 6.
- -m0: the degrees of freedom for the prior distribution. 10 is usually used.
    - For this example, "-m0 10" means that the degrees of freedom for the prior distribution.
- -m: the degrees of freedom for the proposed distributions. It is usually adjusted so that the acceptance rate is around 15.
    - For this example, "-m 4000" means that the degrees of freedom for the proposed distributions.
    - Note that this is only needed for PX-MH.exe. The other three programs (PX-GFN-U.exe, PX-GMN-U.exe, and PX-GS.exe) do not need this.
- -RS: the seed for the random number generator.
    - For this example. "-RS 0" means that the seed for the random number generator is 0.
- -S: the number of iterations for Gibbs sampling.
    - For this example, "-S 20000" means that the number of iterations for Gibbs sampling is 20000.
- -PD: the different types of prior distributions. 1 to 4 represents different prior distributions.
    - For this example, "-PD 3" means that the prior distribution 3 is used.
- -SP: the initial value.
    - For this example, "-SP 1" means the initial value is 1.

- -NN: the number of nominal reponses and their levels.
    - For this example, "-NN 1 6" means that there is one nominal response with 6 levels.
- -NO: the number of ordinal responses. This is for future use. At this stage, "0" should be used.
    - For this example, "-NO 0" means that there are no ordinal responses.
- -xFile: the file name for covariates.
    - For this example, "-xFile Input_X.dat" means that the name of the file for covariates is "Input_X.dat".
- -yFile: the file name for responses
    - For this example, "-yFile Input_Y.dat" means the name of the file for responses is "Input_Y.dat".
- -vt: At the current stage, it should be K number of 1 and is separated by ",".
- -outPre: the prefix of the names of the output files.
    - For this example, "-outPre res-" means that the prefix of the names of all the output files "res-".

Input Files

The program needs two files as input: one file contains all covariates and its name is specified by "-xFile" and the other file contains all responses and its name is specified by "-yFile". Assume that (1) There are N samples; (2) There are K latent variables (thus K + 1 levels) for the nominal response; and (3) There are M variables as the covariates. Note that the total number of covariates P will be K + M since there is an intercept for each latent variable.

File for Covariates

As we have mentioned, the total number of covariates P will be K + M since there is an intercept for each latent variable. Therefore, the file contains K * N lines. There are K lines for the design matrix of each sample and each line has P numbers. For K lines of each sample, A K by K identity represents the design matrix for the intercept. For K lines of each sample, the last M columns represent the design matrix for the other covariates. For this example, K = 5, M = 1, N = 2657, P = K + M = 6. So there are 13285 = K * N = 5 * 2647 lines and each line has 6 numbers.

File for Responses

The file contains K * N lines. There are K lines for each sample and each line has one number – the level of the response of that variable. The levels are coded from 0 to K. For this example, K = 5 and the first two samples have levels 5 and 5, so the first 5 lines are 5 and the last 5 lines are 5 too.

Output Files

Assume that (1) The number of iterations is S; (2) The number of latent variables is K; and (3) The number of covariates is P. There are 4 output files: one for the coefficients, one for correlation matrix, one for unidentified covariance matrix, and one for identified covariance matrix.

File for Coefficients

The file contains P * S lines, and each line has one number. Therefore, P lines are the coefficients for each iteration.

Files for Correlation Matrix, Unidentified Covariance Matrix, and Identified Covariance Matrix

Each file contains K * K * K lines, and each line has one number. Therefore, P lines are the correlation (or covariance matrix) for each iteration.