1. General Descriptions

This document describes how to use the programs to analyze multivariate ordinal data using multivariate probit models and the Markov chain Monte Carlo (MCMC) methods developed by our group.

The `CProg` directory contains the corresponding `.cpp` files and the pre-compiled windows executable files (Please refer to Detailed Descriptions of Program section for more details)

2. Compile Program

To run a program, you need a compiled copy of program. At this time, only the pre-compiled program under Windows 10 system is provided.

The `.cpp` files can be compiled with any standard g++ compiler and GNU scientific library

(https://www.gnu.org/software/gsl/)

3. Run Program with Example Data

An example is provided in `DataResults` directory. Two input files, "`outY.dat`" and "`outX.dat`" are provided.

To run the program under windows 10,

 (1) Type `CMD` to start Windows Commands.
 (2) Use `CD` to take you to the directory that contains the pre-compiled programs.
 (3) Type the name of program, e.g., `MH-ID-PID`, the press `ENTER` to run the program.

A number of files containing the results will be generated and stored in `DataResults` directory.

4. Post Process

A R program in `RProg` directory is provided to summarize the results. You need to change the working directory to the directory that contains your results files.

5. Run Program with New Data

The program is still under development so the pre-compiled programs can only be used for two files provided.

To use new data, you need to do the following:

(1) Prepare files for new data. Please refer to Input Files section for more details.

(2) Open the corresponding `.cpp` file and make the following changes:

    a. Change folder for input files and output files: search `"strcpy(dirName, "../DataResults/")"` change `"../DataResults/"` to the direcotry that contains your data.

    b. Change the name of input files: search `"strcat(fileName, "outY.dat");"` and `"strcat(fileName, "outX.dat");"` and change `"outY.dat"` and `"outX.dat"` to the name of your input data files, respectively.

    c. Change parameters used in the program, around lines 58-64, find and change:

        `"int N = 50;"` - sample size

        `"int K = 2;"` - number of covariates

        `"int rep = 5;"` - dimension of outcome (number of repeated measures)

        `"int m0 = 10;"` - degrees of freedom for prior

        `"int m = 100;"` - degrees of freedom of proposed distribution in MH sampling

        `"int Inum = 10000;"` - number of iterations

        `"int ordcut = 4;"` - number of levels for ordinal variable

    d. change the name of output files: search `"strcat(outBetaFile, "CorrOrdBeta.dat");"`,`"strcat(outZFile, "CorrOrdZ.dat");"`, `"strcat(outRFile, "CorrOrdR.dat");"`, `"strcat(outWRFile, "CorrOrdWR.dat");"`, `"strcat(outGamaFile, "CorrOrdGama.dat");"` and change to those file names to the name of your output files.

e. Recompile the program and run (please refer to **Run Program with Example Data** section for more details).

(3) Post process of results: a R program is provided to summarize the results. You need to do the following changes:

a. Change the working directory to the directory that contains your results files.

b. Change the names of files for results if the different names are used.

c. Change the corresponding parameters used at the beginning of the program, including:

> `N` - sample size
>
> `K` - number of covariate
>
> `Rep` - dimension of outcome (number of repeated measures)
>
> `Inum` - number of iterations
>
> `Ordcut` - number of levels for ordinal variable

6. Input Files

Two input files are needed and two example files are in `DataResults` directory.

> `OutY.dat`: contains the response vector. This file contains `N * rep` rows and each row contains one observation, where `N` is the sample size and `rep` is the dimension of outcome (number of repeated measures). First `rep` rows are the outcomes of the first sample, and the second `rep` rows are the outcomes of the second sample, etc.

> `OutX.dat`: contains the covariates. This file contains `N * rep` rows and `K` columns. Here `N` is the sample size, `rep` is the dimension of outcome (number of repeated measures), and `K` is the number of covariates. First `rep` rows are for the first sample, and the second `rep` rows are for the second sample, etc. Each column represents a covariate. In general, the first column is all 1 to represent the design vector for the intercept.

7. Output Files

A number of files containing the results will be stored in `DataResults` directory.

`CorrOrdBeta.dat`: coefficients

`CorrOrdR.dat`: correlation matrix of latent variables

`CorrOrdWR`: covariance matrix of latent variables

`CorrOrdGama.dat`: cut points for ordinal outcomes

`CorrOrdZ.dat`: empty, currently not used.

8. Detailed Descriptions of Program

Ten `.cpp` files and `.exe` files are provided. A brief description of these program can be found below:

`MH-ID`: the MCMC sampling using Metropolis-Hastings algorithm for sampling the correlation matrix based on the identifiable model

   `MH-ID-PID`: the `MH-ID` algorithm with non-informative prior (`N-ID`)

   `MH-ID-PAR`: the `MH-ID` algorithm with the informative prior (`I-AR`)

`PX-MH`: Parameter-expanded algorithm based on the 1st non-identifiable model and using the Metropolis-Hastings algorithm for sampling the correlation matrix

   `PX-MH-PID`: the `PX-MH` algorithm with non-informative prior (`N-ID`)

   `PX-MH-PAR`: The `PX-MH` algorithm with the informative prior (`I-AR`)

`PX-MH-M`: Parameter-expanded algorithm with marginalization based on the 1st non-identifiable model and using the Metropolis-Hastings algorithm for sampling the correlation matrix

   `PX-MH-M-PID`: the `PX-MH-M` algorithm with non-informative prior (`N-ID`)

   `PX-MH-M-PAR`: the `PX-MH-M` algorithm with the informative prior (`I-AR`)

`PX-GS`: Parameter-expanded algorithm based on the 2nd non-identifiable model and using the Gibbs sampling algorithm for sampling the correlation matrix

   `PX-GS-PID`: the `PX-GS` algorithm with the non-informative prior (`N-ID`)

   `PX-GS-PAR`: the `PX-GS` algorithm with the informative prior (`I-AR`)

`PX-GS-M`: Parameter-expanded algorithm with marginalization based on the 2nd non-identifiable model and using the Gibbs sampling algorithm for sampling the correlation matrix

> `PX-GS-M-PID`: the `PX-GS-M` algorithm with the non-informative prior (`N-ID`)

> `PX-GS-M-PAR`: the `PX-GS-M` algorithm with the informative prior (`I-AR`)

9. Reproduction of Results in Manuscript

`OutY`.dat and `OutX`.dat are for a simulated data that contains `N = 50` samples, `rep = 5` repeated ordinal outcomes and two covariates including the intercept. Here we illustrate how to use our program to reproduce Tables 1 and 2 and Figures 2 corresponding to MH-ID method with N-ID prior in the manuscript.

(1) Run program `MH-ID-PID` in `CProg` folder.

(2) Run R program `PostProc.R` in `RProg` folder (please remember to change the working folder first).

The figure is attached below.
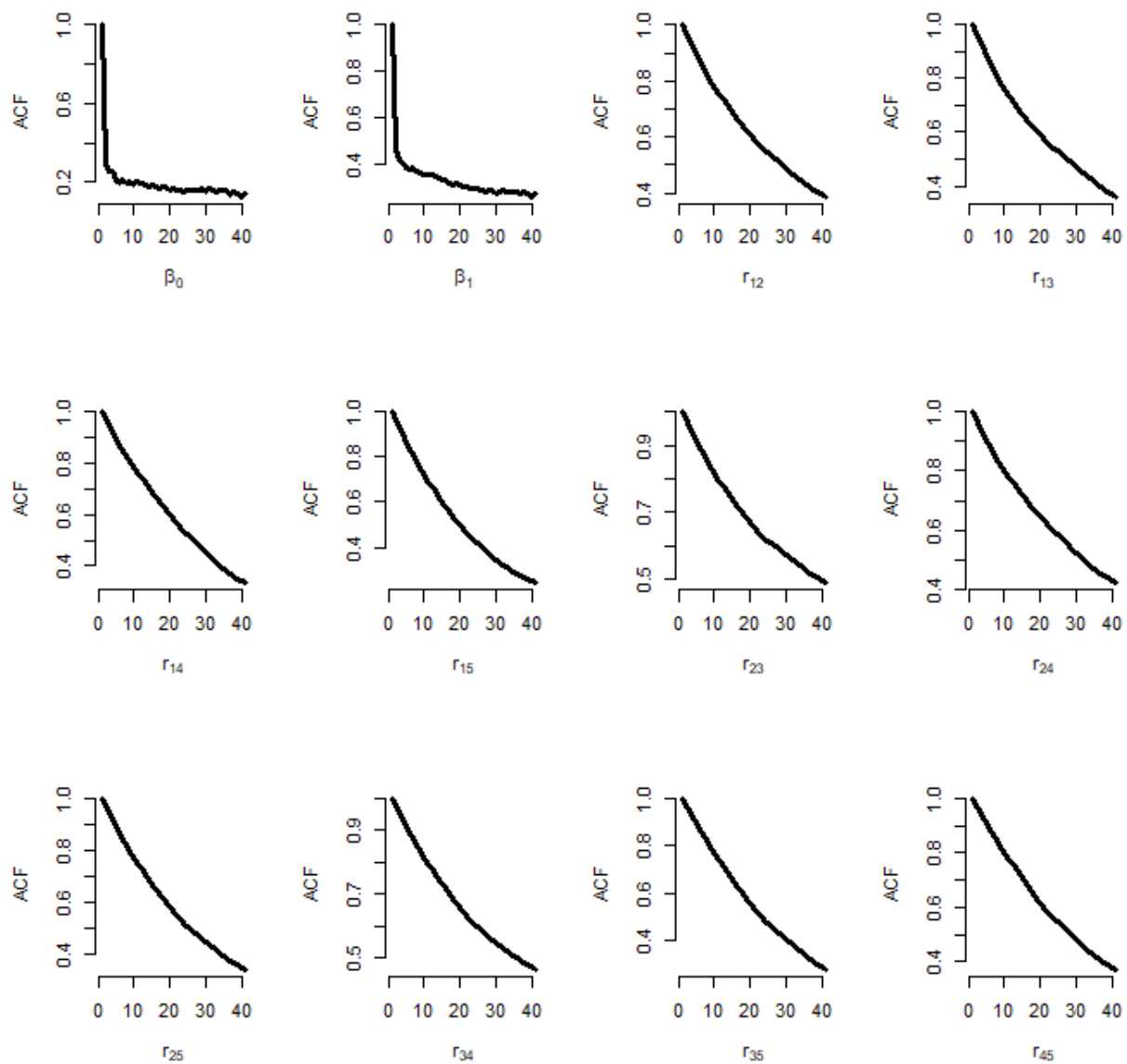
10. Contact Information

If you have any questions, please contact me at:

    Xiao Zhang

    Department of Mathematical Sciences

    Michigan Technological University

    Email: xzhang35@mtu.edu