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 source files, `.cpp` and `.h` files are provided in the `CProg` directory and the sub-directory `lib`. The pre-compiled windows executable files (`.exe`) are also provided.

2. Compile Program

To run a program, you need a compiled copy of the program. At this time, only the pre-compiled program under Windows 10 operating system is provided and the pre-compiled programs only works for the data in the same directory.

The `.cpp` and `.h` files can be compiled with any standard g++ compiler and GNU scientific library (https://www.gnu.org/software/gsl/). There are three main programs corresponding to three methods developed and compared in the manuscript. Compile each `*-main.cpp` and all other `.cpp` and `.h` files separately.

- `PX-GS`: Parameter-expanded algorithm based on the 2nd non-identifiable model and using the Gibbs sampling algorithm for sampling the correlation matrix
- `PX-GSM`: Parameter-expanded algorithm with marginalization based on the 2nd non-identifiable model and using the Gibbs sampling algorithm for sampling the correlation matrix
- `PX-MH`: Parameter-expanded algorithm based on the 1st non-identifiable model and using the Metropolis-Hastings algorithm for sampling the correlation matrix

3. Run Program with Example Data and Real Data Sets

A simulated data and two real data sets are provided in `DataSets` directory:

- PainStudy-X.dat and PainStudy-Y.dat data from the pain score study
- SchStudy-X.dat and Sch-Y.dat data from the Schizophrenia study
- Simu-X.dat, Simu-Y.dat: a simulated data

To run the program under windows 10,

(1) Create a working directory and copy the input data files and the pre-compiled windows executable file to the working directory.
(2) Type `CMD` to start Windows Commands.
(3) Use `CD` to take you to the working directory that contains the input data files and pre-compiled programs.
(4) Type the following command, for example, the press `ENTER` to run the program:

```
PX-GS -N 437 -P 4 -K 4 -S 20000 -m 20 -DP 1
-xFile SchStudy-X.dat -yFile SchStudy-Y.dat
```

A number of files containing the results will be generated and stored in a sub-directory. You can put the data files and the executable file to different directories and use `-dirName` to specify the directory that contains the data files.

The arguments used in the program are:

```
-N: number of samples
-P: number of covariates
-K: dimension of outcome (number of repeated measures)
-m: degrees of freedom for the proposed distribution of covariance
-m0: degrees of freedom for prior used in PX-MH method
-S: number of iterations in MCMC sampling
-PD: the prior structure for the covariance matrix, 1 for identify matrix and 2 for compound symmetry.
```

4. Post Process

A R program in `RProg` directory is provided to summarize the results. You need to do the following changes:

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

- Change the names of files for results if the different names are used.
- Change the corresponding parameters used at the beginning of the program, including the sample size, the number of covariates, the dimension of outcome (number of repeated measures), the number of iterations.

5. Input Files

Two input files are needed and Several example files are in `DataSets` directory.

`Y.dat`: contains the response vector. This file contains `N * K` 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 `K` rows are the outcomes of the first sample, and the second `K` rows are the outcomes of the second sample, etc.

`X.dat`: contains the covariates. This file contains `N * K` rows and `P` columns. Here `N` is the sample size, `K` is the dimension of outcome (number of repeated measures), and `P` is the number of covariates. First `K` rows are for the first sample, and the second `K` 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.

The following table list the basic information of three data sets provided here:

Data Set	Sample Size (N)	Number of Covariates (P)	Number of Repeats (K)
Pain Score Study	41	5	6
Schizophrenia Study	437	4	4
Simulated Data	500	2	5

6. Output Files

A number of files containing the results will be generated:

`CorrOrdBeta.dat`: coefficients

`CorrOrdR.dat`: correlation matrix of latent variables

`CorrOrdSig`: covariance matrix of latent variables

`CorrOrdGama.dat`: cut points for ordinal outcomes

`CorrOrdZ.dat`: latent variables for the response.

7. Contact Information

If you have any questions, please contact me at:

Xiao Zhang

Department of Mathematical Sciences

Michigan Technological University

Email: xzhang35@mtu.edu