

MLOD : A Program for Maximized LOD score calculation for nuclear families

Yun Joo Yoo & Nancy R. Mendell

Updated Apr 2008

1. Introduction

MLOD is a program to calculate the LOD score maximized over the five-dimensional parameter space (four genetic model parameters and a recombination fraction parameter) for the data of nuclear families. Four genetic model parameters are the disease allele frequency, and three penetrance values for the two homozygous disease genotypes and a heterozygous disease genotype. MLOD does not require any assumption on genetic models like other LOD score computation programs and instead finds the highest LOD score within the parameter space.

2. Files included in the package

MLOD.exe : Executive file for MLOD
MLOD.pdf : The manual for MLOD
inputSample.txt : Sample input file
outputSample.txt : Sample output file

3. Overview of MLOD

MLOD will calculate the LOD score maximized over five parameters. Following is the definition of those parameters.

- (1) q denotes the frequency of the disease allele, D , assuming a biallelic disease locus with alleles D and d in Hardy-Weinberg equilibrium, i.e., $q = \Pr(D)$.
- (2) f_0 denotes the penetrance rate for those homozygous for the d allele, i.e., $f_0 = \Pr(\text{Affected} | dd)$; $0 \leq f_0 \leq 1.0$.
- (3) f_1 denotes the penetrance rate for those heterozygous for the D allele, i.e., $f_1 = \Pr(\text{Affected} | Dd)$, with $f_0 \leq f_1 \leq 1.0$.
- (4) f_2 denotes the penetrance rate for those homozygous for the D allele, i.e., $f_2 = \Pr(\text{Affected} | DD)$, $f_1 \leq f_2 \leq 1.0$.

Also, the recombination fraction between the marker and disease gene is denoted as θ (theta, $0 \leq \theta \leq 0.5$)

The maximization algorithm for MLOD uses Powell's direction set method (Press et al, 1993). This method finds a maximum of a function on n -dimensional space by finding a

maximum point on a straight line at each step and then moving to the next straight line which includes the optimum point from the previous step and has a new direction. Powell's algorithm may converge to a local maximum. To find the global maximum, we adapted several strategies such as repeating the search using various randomly selected starting points, additional searches through the direction of unit vectors after getting a local maximum and other searches based on twenty starting points. Since any other algorithm to find a global maximum, the results of MLOD program may fall into the local maximum. The number of iteration can be set in the program and we recommend 100 or at least 20 iterations.

4. How to run MLOD

MLOD.exe runs on Microsoft Windows 95, 98, NT, 2000, ME, XP and Vista (all 32bit system). MLOD can be run in MS-DOS command line mode, or in Windows mode using a batch file.

To run MLOD, run MLOD.exe file in command line followed by the input and output data file names:

```
>MLOD [input file] [output file]
```

Running MLOD in command line mode

- 1) Open a MSDOS command window. This can most likely be found on your computer under “Start → All Programs → Accessories → Command Prompt.
- 2) At the prompt, type the command “chdir” followed by the full path name of the folder in which you saved the MLOD program, and press enter
- 3) At the new prompt, which should now include the correct path name, type “MLOD inputfile.txt outputfile.txt” and press enter if input and output file names are inputfile.txt and outputfile.txt.
- 4) The program will take a few minutes to run and, upon completion, will save the output as “outputfile.txt” in the same folder in which the input file and MLOD program are located.

Running MLOD by creating a batch file

- 1) If you create a batch file containing the command line to run MLOD, you can run MLOD in the windows mode without opening MS-DOS command window.

- 2) Using a text editor program (e.g. Notepad), create a text file containing the command line to run MLOD (e.g. “MLOD inputfile.txt outputfile.txt”), then save the file as batch.bat. Make sure that the file extension is “.bat”.
- 3) Copy input data file in the directory where MLOD is saved.
- 4) Double click “batch.bat”. Upon completion, the output file “outputfile.txt” will be saved in the same directory.
- 5) Note that you the command window will be closed upon completion of batch mode. You cannot monitor the error message or screen output once it is closed.

*If you do not specify the output file name, it will use the name for default output file: “oufile.txt”. See next example command line.

Example 1. Command line example

```
>MLOD inputsamle.txt outputsample.txt
```

5. How to prepare input files

The input file of family data follows LINKAGE format except that it requires additional header line at the top of the file to specify the number of iterations, the number of families in the data, the maximum family size and the number of different marker alleles (Four numbers). Detailed format of [input file] is as follows;

*All data should be given in numeric values.

- First raw : [number of iterations] [number of families] [maximum family size] [number of different marker alleles]
- Second raw~: Data of each individual
 - column 1: Family ID
 - column 2: Individual ID
 - column 3: Father’s ID
 - column 4: Mother’s ID
 - column 5: Sex (1=Male, 2=Female)
 - column 6: Affection Status (1=Not affected, 2=Affected)
 - column 7 & 8 : Marker genotype

Example 2 : [input file]

```
20 40 6 4
1 1 0 0 1 2 2 1
1 2 0 0 2 1 3 4
1 3 1 2 1 2 4 1
1 4 1 2 2 2 4 1
1 5 1 2 1 1 2 4
1 6 1 2 2 2 3 2
2 1 0 0 1 2 2 1
2 2 0 0 2 1 3 4
2 3 1 2 1 2 4 1
2 4 1 2 2 2 4 1
2 5 1 2 1 1 1 4
2 6 1 2 2 1 2 4
3 1 0 0 1 1 1 2
3 2 0 0 2 2 3 4
3 3 1 2 1 2 1 4
3 4 1 2 2 2 2 3
3 5 1 2 1 2 2 3
3 6 1 2 2 2 1 3
```

*The marker allele symbols should be 1 to N (the number of different marker alleles).

5. The output files

The output file will contain the output values of the maximized LOD score and the values of five parameters at the maximum. The asymptotic p-values for maximized lod score are not given in this program. See Ulgen et al (2004) and Yoo & Mendell for the details of the null distribution of MLOD. However, we recommend to use critical values of MLOD as 1.6, 2.3 and 4.6 corresponding to the type I error of 0.05, 0.01 and 0.0001.

Example 3: [Output file]

```
q      f2      f1      f0      theta  lod
0.0001 1.0000 1.0000 0.0554 0.2106 0.7837
```

6. Missing data

In the updated version, MLOD can handle the missing data in phenotype and genotypes. Missing data for phenotype and genotype is marked as 0.

7. Reference

Yun Joo Yoo and Nancy R. Mendell (2008), The Power and Robustness of Maximum LOD Score Statistics, *Annals of Human Genetics* (In press)

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1993). *Numerical Recipes in C : The art of scientific computing, second Edition*. Cambridge University Press.

Ulgen, A., Yoo, Y. J., Gordon, D., Finch, S. and Mendell, N. R. (2004), Percentiles of the null distribution of 2 maximum lod scores tests. *Hum Hered* 57, 39-48.