Myriads Manual v 1.1

Antonio Carvajal-Rodríguez
Departamento de Bioquímica, Genética e Inmunología
Universidad de Vigo, Vigo 36310, Spain
Email: myriads@uvigo.es
Web: http://myriads.webs.uvigo.es
# TABLE OF CONTENTS

Introduction .................................................................................................................. 3

The Program ............................................................................................................... 3
  Download .................................................................................................................. 3
  Ready-to-use executables ......................................................................................... 3
  Compiling the program manually (linux/MAC) ....................................................... 3

Input ............................................................................................................................. 3

Program Usage ......................................................................................................... 4
  Interactive menu options ......................................................................................... 4
  Command line options for multiple testing correction .......................................... 5
  Interactive menu options for $p$-value simulation ................................................. 6
  Command line options for $p$-value simulation ..................................................... 7
  Run in MacOsX or Linux ......................................................................................... 9
  Run in Windows ....................................................................................................... 9

Dependence test ....................................................................................................... 10

Simulation ................................................................................................................. 10

Output ....................................................................................................................... 10
  Multiple testing ...................................................................................................... 10
  Dependence test .................................................................................................... 11
  Light output ........................................................................................................... 11
  Simulation output .................................................................................................. 12

References ................................................................................................................. 13
INTRODUCTION

The Myriads software performs three different tasks, namely, multiple testing adjustment, detection of dependency and $p$-value simulation.

THE PROGAM

Download

The program can be downloaded from

http://myriads.webs.uvigo.es

Ready-to-use executables

Windows: Myriads.exe
Linux: Myriads

Compiling the program manually (linux/MAC)

Go to the source folder that should include a file called Makefile, and just type make.

INPUT

Under the non-simulation mode, the program requires a SGoF-like input file (Fig. 1). The input file is by default called PvalMyriads.dat but any other name can be given by the user. The input file should have an integer number indicating the total number of tests and after it two columns with pairs of identifiers and $p$-values. The identifier can be a number or a character string. The list of $p$-values does not need to be sorted, indeed, is required to be not for the dependency test. There is no limit on the number of $p$-values other than computer memory. The simulation mode does not require input file.
Interactive menu options

When calling without arguments, the program offers a menu (Fig. 2)

Fig. 2. Myriads general menu.
Choosing the option 1 from the menu executes the default parameter values (see below).

If the program is called with any argument, the menu is skipped and the command line execution mode is performed.

**Command line options for multiple testing correction**

We also can skip the menu and run the default options from the command line by passing the tag -default. Thus, the calling

```
Myriads -default
```

is equivalent to the first option of the menu and to the line with arguments:

```
Myriads -simulation 0 -inpath . -outpath ./Myriads_output -inputfile PvalMyriads.dat -outputfile Myriads -SL 0.05 -epi0 A -grid 5 --robust 0 -dependence 0 -lag 1 -minblock 50 -light 0 -AdjustLarge 0
```

Note that if the tag -default is utilized, every argument after it will be ignored. This means that the line

```
-dependence 1 -default
```

is the same as

```
-dependence 1
```

but the line

```
-default -dependence 1
```

is the same as

```
-default (i.e. -default -dependence 0)
```

The utility of -default is to skip the menu without requiring the specification of any argument.

The explanation of each argument is as follows:

- `simulation <BOOLEAN>` Simulation (1|2) or analysis (0) mode. By default is 0.

- `inpath <STRING>` Specifies the path to the input file (default is the working directory represented by a dot).

- `outpath <STRING>` Specifies the path to the output file (default is ./Myriads_output).

- `input <STRING>` Specifies the input file name (default is PvalMyriads.dat).
- **output** <STRING> Specifies the output file name.
- **SL** <DOUBLE> The significance level (default 0.05).
- **epi0** <DOUBLE> Choose the Pi0 estimation method (option 8 from the menu, default is A).
- **grid** <DOUBLE> Change lambda in bootstrap and smooth (option 9 from the menu).
- **robust** <BOOLEAN> Change (1) q-value estimation to robust (option 7 from menu).
- **dependence** <BOOLEAN> Include the dependence test in the analysis (1) or not (0, default).
- **lag** <INTEGER> The lag for the autoregression in the dependence analysis (default is 1).
- **minblock** <INTEGER> The minimum block size for the dependence test (default is 50).
- **light** <BOOLEAN> Activates (1) the light output mode (default is 0).
- **AdjustLarge** <BOOLEAN> Activates (1) the adjusted p-values for SGoF when the number of tests is large (option 6 from menu).

**Interactive menu options for p-value simulation**

The simulation menu is displayed (Fig. 3) when the user chooses the options 10 or 11 from the general menu (Fig 2).
A similar menu is obtained when the option 1 is chosen from the general menu. From the simulation menu, choosing the option 0 without any change, executes the simulation default parameter values (see below).

**Command line options for p-value simulation**

The simulation model can also be executed from the command line simply by putting 1 (t-tests) or 2 (Cochran-Armitage tests) in the tag –simulation when running the command line option as explained above.

**t-tests**
The calling

Myriads -simulation 1

is equivalent to choosing the option 0 of the menu and is also equivalent to the line with arguments:

Myriads -simulation 1 --twotail 1 -numfiles 1 -numtests 1000 -n1 20 -n2 20 -blocksize 1000 -numblocks 1 -SL 0.05 -pi0 1 -sd1 1 -sd2 1 -rho1 0 -rho2 0 -power 0.8 -randepos 0 -dir Myriads_Sims_t_effcorr 0 -fixseed 0

The explanation of the new arguments is as follows:

- **twotail** <BOOLEAN> Two tail (1) or one tail (0) t test (default is 1).
- **numfiles** <INTEGER> The number of p-value files to generate (default is 1).
- **numtests** <INTEGER> The number of p-values in each file (default is 1000).
- **shift** <INTEGER> Default is 0. The output file name is MyriadsPval_number.txt. The number is obtained as number=number of file + shift;
- **n1** <INTEGER> Sample size 1 for the two sample t-test (default is 20).
- **n2** <INTEGER> Sample size 2 for the two sample t-test (default is 20).
- **blocksize** <INTEGER> The size of each correlated block if any (default is 1000).
- **numblocks** <INTEGER> The number of blocks \( N_b \). If \( N_b \) is not passed as argument, the program automatically computes \( N_b = \left( \frac{\text{numtests}}{\text{blocksize}} \right) + 1_{(\text{numtests} \% \text{blocksize} \geq \text{minblock})} \), that defines the maximum possible number of blocks with size blocksize, plus one more block, with size the residue of \( \text{numtests} / \text{blocksize} \), in the case that the residue have the minimum acceptable block size (50 by default).
- pi0 <DOUBLE> Proportion of true nulls (default is 1, the complete null).
- sd1 <DOUBLE> Standard deviation from the distribution of sample 1 (default is 1).
- sd2 <DOUBLE> Standard deviation from the distribution of sample 2 (default is 1).
- rho1 <DOUBLE> Within-block correlation in sample 1 (default is 0).
- rho2 <DOUBLE> Within-block correlation in sample 2 (default is 0).
- power <DOUBLE> Power for the two-sample t-test (default is 0.8).
- randepos <BOOLEAN> Positions with effects (in sample 2) are randomly distributed (1) or (0) they are consecutive within the block (default is 0).
- effcorr <BOOLEAN> The correlation in sample 2 affects only to the positions with effects (1) or (0) the correlation affects the consecutive positions within block independently of the effects (default is 0).
- dir <STRING> Specifies the output directory name (default is Myriads_Sims).
- fixseed <INTEGER> If positive it fixes the seed for the Monte Carlo simulations to that value, so that the result for different runs with equal arguments should be the same (default is 0, so that the seed is randomly generated at each program execution).

Cochran-Armitage tests

The calling

Myriads -simulation 2

is equivalent to choosing the option 0 of the simulation menu and is also equivalent to the line with arguments:

Myriads -simulation 2 -twotail 1 -numfiles 1 -numtests 1000 -blocksize 1000 -numblocks 1 -SL 0.05 -pi0 1 -rho1 0 -power 0.8 -randepos 0 -dir Myriads_Sims_CA -fixseed 0 -prevalence 0.1 -risk2 3 -phi 0.5 -diseasemodel 0.5

The meaning of most arguments is the same as for the other simulation case. However, note that sample sizes and standard deviations are not considered when the simulation=2 mode is called. This is because sample sizes are automatically computed based on the desired power. Also, only one correlation value is necessary. Additionally, new arguments are needed for the case-control simulation. The explanation of the new arguments is as follows:

- maf <DOUBLE> Minor allele frequency (default is 0.1).
- prevalence <DOUBLE> Prevalence of disease (default is 0.1).
- risk2 <DOUBLE> Homozygote genotype relative risk (default is 3).
- phi <DOUBLE> Proportion of cases (default is 0.5, i.e. same sample size for cases and controls).
- diseasemodel <DOUBLE> Recessive (0), additive (0.5) or dominant (1) disease model (default is 0.5).
- rho1 <DOUBLE> Within-block correlation in the SNPs (default is 0).
- power <DOUBLE> Power for the two-sample Cochran-Armitage test (default is 0.8).
- randepos <BOOLEAN> Positions with effects are randomly distributed (1) or (0) they are consecutive within the block (default is 0).

Run in MacOsX or Linux

From the console or terminal type

./Myriads

jointly with the desired arguments. If there are no arguments the menu will appear. If we want to skip the menu and run the default arguments we just call

./Myriads -default

which run the Myriads default options for multiple testing correction as already explained.

Run in Windows

Double click or go to the command prompt (cmd.exe) and type

Myriads

for running the program with the interactive menu. If you want to skip the menu and run the default arguments just type

Myriads -default

You can also access the Run command by pressing the Windows logo key +r then drag and drop the .exe file from your folder and add the desired arguments, e.g. if Myriads.exe is in the folder Myriads then after drag and drop you will have

C:\ Myriads \ Myriads.exe

now add the desired arguments and then hit the Intro key. For example:

C:\ Myriads \ Myriads.exe -inputfile pval1.txt -dependence 1
DEPENDENCE TEST

Myriads incorporate an autocorrelation test based on the generalized Durbin-Watson (D-W) statistic (Ali, 1987; Vinod, 1973). The test permits to identify strong dependencies in the p-values and estimate the minimum detectable block size of correlated values. The lowest the detected block size the strongest the correlation in the data.

For performing the dependence test the user just need to choose the option 5 in the menu or add the corresponding arguments in the command line:

    Myriads -dependence 1

See the Supplementary Document (Carvajal-Rodriguez, 2017) for detailed information about the algorithm.

SIMULATION

Myriads incorporates the option for simulating p-values by conducting a two sample t-test or a Cochran-Armitage case-control test.

For performing the simulation the user just need to choose the options 10 (t-test) or 11 (CA test) in the menu or add the corresponding arguments in the command line:

    Myriads -simulation 1 (or Myriads -simulation 2)

See the Supplementary Document (Carvajal-Rodriguez, 2017) for detailed information about the simulation algorithms.

OUTPUT

Multiple testing

The program provides with various output files in the folder ‘Myriads_output’ or any other folder defined by the user.

1) Myriads.txt which includes only the tests remaining significant after the different adjustments.
2) *Myriads_Fulloutput.html* and *Myriads_Fulloutput.xls* which provide with the full list of a priori significant *p*-values, the adjusted *p* after each method and the estimated *q*-values. The .html file returns just the output coming from the last program execution while the .xls accumulate the output through different executions.

The adjusted B-H, BonSEV and SB *p*-values were computed following Yekutieli & Benjamini (1999). When the number of tests is below 10,000 (a myriad), the adjusted *p*-values for SGoF were computed following (Castro-Conde and de Uña-Álvarez, 2015). For higher number of tests, the metatest *p*-values are given by default, the adjusted *p*-values when computed, follow the approximation developed in (Carvajal-Rodriguez, 2017). Values under the SLIM column correspond to the *q*-value obtained using the SLIM estimate for the proportion of true nulls.

**Dependence test**

If the dependence test is performed, another file called *MyriadsDepRes.txt* is added to the previous ones. If dependence was detected, the content of this file includes the rejection level, the minimum block size for which the test was significant and the number of significant blocks at that size. An example of the output is as follows:

```
Input file: PvalMyriads.dat
Myriads dependence analysis output:
---------------------
Minimum allowed block size = 20
REJECTION LEVEL = 0.00066
Zcrit = 3.40655
Number of p-values = 3170
Block size = 160
Number of blocks = 19
Number of significant blocks = 1
Lag = 1
Test = 3.58074

Sat Aug 12 16:36:43 2017
```

**Light output**

This is the output obtained when the –light argument is set to 1. In this case the only file obtained is by default called *MyriadsFast.txt*, although, both the folder and the file
name, can be changed by the corresponding arguments as we have seen in the command line options section. The light mode is only available from the command line and only under the non-simulation mode. It is useful when we desire to analyze many files so that the output of each is accumulated in a line as follows:

<table>
<thead>
<tr>
<th>FILE</th>
<th>#m</th>
<th>m0</th>
<th>#SIGT</th>
<th>P0</th>
<th>BH</th>
<th>BSEV</th>
<th>SB</th>
<th>SGOF</th>
<th>SLIM</th>
<th>QVAL</th>
<th>Qmin</th>
<th>DEP</th>
<th>BlockSize</th>
</tr>
</thead>
<tbody>
<tr>
<td>file.txt</td>
<td>3170</td>
<td>3170</td>
<td>606</td>
<td>0.69</td>
<td>94</td>
<td>102</td>
<td>2</td>
<td>423</td>
<td>124</td>
<td>158</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>file2.txt</td>
<td>3170</td>
<td>3170</td>
<td>306</td>
<td>0.99</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.61</td>
<td>1</td>
<td>160</td>
</tr>
</tbody>
</table>

Most of the headers are self-explanatory. The number under $m_0$ is a guess for the number of true nulls computed as $m \times true pi0$ where $true pi0$ is 1 by default but can be user-defined by means of the argument -truePI0. The later can be useful if we are analyzing simulated files under known conditions, for benchmarking purposes. The number below the different methods refers to the number of $p$-values detected as significant after the adjustment, or in the SLIM and QVAL cases the number of $q$-values that are below the pre-defined $\alpha$-level. $Qmin$ is the lowest $q$-value divide by the PI0 estimate. If the dependence test is not performed or no dependence was detected (DEP 0 in any case) the block size is 0, on the contrary, if dependence is detected (DEP 1), the minimum significant block size is given.

**Simulation output**

If the simulation mode is executed Myriads produce as many files as indicated by the argument –numfiles. Each file contents as many $p$-values as indicated by the argument –numtests. The default output folder name is Myriads_Sims_ this name can be changed by the argument -dir. The format of the output files coincide with the input file format for the analysis of $p$-values (Fig. 1). The name of each file is MyriadsPval_1.txt, MyriadsPval_2.txt etc until the number of files indicated in –numfiles is reached. See (Carvajal-Rodriguez, 2017) for details on the simulation procedure.
REFERENCES


