PostPM Manual

January 25, 2010

Acknowledgement: The program PostPM.exe was written in Fortran by Olav Laudy. The first version of this manual was written by Olav Laudy, thereafter Herbert Hoijtink has updated the manual.

1 Introduction

This manual explains how to use the program POSTMP.exe. This program uses a contingency table as input, and calculates the Bayes factors and posterior model probabilities for a set of models. Models are defined by a set of constraints on the cell probabilities. For details of the method, please read

Klugkist, I., Laudy, O. and Hoijtink, H. (submitted manuscript). Bayesian evaluation of inequality and equality constrained hypotheses for contingency tables.

This program is free. However, you are obliged to refer to this paper when publishing results obtained with this program.

2 General

The folder Executables contains two subfolders (unzip without loosing the folder structure, e.g. by first saving the zip-file to your computer and then using “extract here”). In the folder “Program”, you will find the program called POSTPM.exe and a file DFORRT.dll, which is needed to run the fortran program. The latter should be placed in each folder where the POSTPM.exe program is run from. The second folder ”Examples” contains data (compulsory named DATA.txt) and initialization files (compulsory named INI.txt) for the examples that are used in Klugkist, I., Laudy, O. and Hoijtink, H. (submitted manuscript).

When POSTPM.exe, DFORRT.dll, DATA.txt and INI.txt are located in the same folder, POSTPM.exe can be run after it is double clicked. During the run, it will show its progress.

3 Data

The data file DATA.txt is a plain text file, and contains frequencies. If a two way table is used, the first two columns, separated by spaces, refer to the cells in the contingency table. This is not used in the program, but serves as a reference to the user. After the first two identification columns, and a space, the cell frequency is denoted. Only spaces can be used, not tabs, to separate the columns.

Consider the two by two table in Table 1 The data format that is used in the file DATA.txt is shown in Table 2. Other examples can be found in the subfolder "Examples".

Table 1: An example of a two way contingency table

		B	
		1	2
A	1	100	200
	2	300	400

Table 2: An example of the file DATA.txt

1 1 100
1 2 200
2 1 300
2 2 400

4 Initialization file

The initialization file is called INI.txt and should be in the same directory as the program POSTPM.exe. The file consists of items preceded by ♯, a carriage return and then the value for that item. The settings are discussed below. Items are displayed in Italics type, comment in regular type.

♯♯

The double ♯♯ may be used to describe the model and constraints that are tested. It will not be read by the program, but serves as a reference to the user. It is placed at the top of the initialization file and can contain multiple lines, that are all preceded with ♯♯.

♯*dim_crostab*
2

The number of variables. The program does not use this in the calculations, but it is used to correctly read the data file.

♯*ncel_crostab*
15

The number of cells in the contingency table, and the number of entries in the data file.

♯*n_models*
3

The number of different models to be evaluated.

♯*model 1*
"restriction set 1 (will be elaborated below)"

♯*model 2*
"restriction set 2"

♯*model 3*
"restriction set 3"

♯*eof*

Indicates end of the initialization file. The INI.txt files for the examples analyzed in Klugkist, I., Laudy, O. and Hoijtink, H. (submitted manuscript), can be found in the subfolder "Examples".

4.1 Restrictions

Consider the data in Table 1. Suppose, it is to be investigated whether the odds ratio is larger than 1, that is, $p_{11} * p_{22}/(p_{12} * p_{21}) > 1$. The program uses consecutive numbers to indicate the frequencies in the file DATA.txt, that is, "$c1 * c4/(c2 * c3) > 1$". Thus, $c4$ is used to reference the 4th frequency in file DATA.txt. As the file DATA.txt shows, the 4th frequency refers to cell $(2, 2)$.

The following rules apply to writing restrictions:

- Restrictions are surrounded by quotes "". Note that no additional space is allowed between the quote and the beginning of the restriction (i.e., " $c1 * c4/(c2 * c3) > 1$" will not be read correctly)
- A frequency is denoted by c and a number. Denote l as the number of cells in the contingency table, with $l = 1, ..., L$, then c_l refers to the l^{th} number in the data file. Thus, the order of the data does not matter, as long as the c_1,...,c_L refers to the correct frequency.
- A restriction can consists of more than one expression. Each expression is on a different line in INI.txt and is surrounded by quotes.
- Both $>$ and $=$ may occur more often in a restriction or expression, but can not be used simultaneously in an expression. Not possible is "$c1/c2 > c3/c4 = 1$", but possible is two lines containing subsequently "$c1/c2 > c3/c4$" and "$c3/c4 = 1$".
- Note that $<$ can not be used, but can simply be replaced by $>$ and a multiplication of -1 of both sides of the equation.
- The following operators can be used: c_l, $+, -, *, /$, and ().

Important: Not al restrictions are admissible. Always check the admissibility of a constraint using the following steps:

- Step 1: Write down the constraint. Example 1: "$c1 * c4/(c2 * c3) = 1$". Example 2: "$c1 * c4 = c2 * c3$". Example 3: "$c1 * c4 - c2 * c3 = 0$". This is the form in which you intend to enter the constraint in INI.txt.
- Step 2: Define *elements of a restriction* as combinations of c's that are separated from other combinations by a $+$, $-$, or $>$. Example 1 has

one element: $c1 * c4/(c2 * c3)$. Example 2 and 3 have two elements: $c1 * c4$ and $c2 * c3$.

- Step 3: Multiply each c in each element by d. For Example 1 this renders: $d * c1 * d * c4/(d * c2 * d * c3)$. For Examples 2 and 3 this renders $d * c1 * d * c4$ and $d * c2 * d * c3$.

- Step 4: A restriction is admissible if the d's cancel in each element. As can be seen for Example 1 the d's in the numerator and denominator cancel. Stated otherwise, the restriction formulated in Example 1 is admissible. As can also be seen, the d's in both elements of the restrictions in Example 2 and 3 do not cancel. Stated otherwise, the restrictions formulated in Examples 2 and 3 are not admissible.

Examples of admissible constraints are: "$c1/(c1 + c2) > c3/(c3 + c4)$" and "$c1/c2 = 1$". Another example of an inadmissible constraint is: "$c1 = c2$". Table 3 shows examples of the restrictions that are used in the article, in consecutive order.

Table 3: An example of restrictions used in Klugkist, I., Laudy, O. and Hoijtink, H. (submitted manuscript)

Example 1	
M_1	"c1*c4/(c2*c3)=1"
M_2	"c1*c4/(c2*c3)>1"
Example 2	
M_1	"c1*c4/(c2*c3)=c5*c8/(c6*c7)=1"
M_2	"c1*c4/(c2*c3)=c5*c8/(c6*c7)"
M_3	"c1*c4/(c2*c3)>1"
	"c5*c8/(c6*c7)>1"
M_4	"c1*c4/(c2*c3)>c5*c8/(c6*c7)"
M_5	"c1*c4/(c2*c3)>c5*c8/(c6*c7)>1"

Example 1.

Model M_1 restricts the odds ratio to be equal to 1, i.e. no association in the contingency table. Model M_2 restricts the odds ratio to be larger than 1, i.e., there is a positive association in the contingency table.

Example 2.

Model M_1 restricts two odds ratios to be equal to 1. Model M_2 restricts two odds ratios to be equal to each other. Model M_3 consists of two lines. This model states that both odds ratios are larger than 1. Model M_4 restricts the first odds ratio to be larger than the second odds ratio, and model M_5 restricts the first odds ratio to be larger than the second odds ratio, and both are restricted to be later than 1. Other examples can be found in the INI.txt files in the subfolder "Examples".

5 Running the Program

The program is started with two mouse-clicks on PostPM.exe. Take notice of the following pointers:

- The program is not fast. The example in the folder ex-1odds will run within 10 minutes. The example in the folder ex-2odds will run within 60 minutes. The example in the folder ex-theory will take about 5 hours. For the latter example you are well advised to start the program in the evening and look at the result in the morning.

- The program is faster if restrictions are spread out over multiple lines. That is, do not use

  $$"c1/(c1+c2+c3) > c2/(c1+c2+c3) > c3/(c1+c2+c3)" \tag{1}$$

  but use

  $$"c1/(c1+c2+c3) > c2/(c1+c2+c3)" \tag{2}$$

  and

  $$"c2/(c1+c2+c3) > c3/(c1+c2+c3)". \tag{3}$$

- The worse the support in the data for a model, the longer the computation time. This is caused by the fact that a precise estimate of fit for a badly fitting model requires a large number of iterations. If the situation is really grim (in terms of computation time) the program will announce this. In such a case you are well-advised to abort the program and delete the model from the set of models under investigation. The Bayes factor of such a model is close to zero.

6 Output file

First, per model the Bayes factor w.r.t the unconstrained model is printed. Furthermore, the program prints the Expected APosteriori estimate (EAP) and the 95% central credibility intervals for the cell probabilities.

In the last part of the output, posterior model probabilities are printed. First, for each model the posterior model probability against the unconstrained model is printed. If there is more than 1 model to fit, the last section of the output also contains a simultaneous comparison of all models using posterior model probabilities. Two comparisons are made, the first including the unconstrained model, the last without the unconstrained model.