USABILITY OF MATHEMATICAL SOFTWARES IN PROSTATE SCREENING

Mária Molnárné Nagy
University of Debrecen,
Doctoral School of Information Science and Technology
nagymaria1@gmail.com

Intoduction

This poster discusses 4 mathematical software regarding the usage in the research phase of the PRO-FILTER, which will be a prostate screening computer service. This service is based on the PSA-level changes of the individual, the goal is to provide a prediction on the PSA-level, and from this future value propose a medical proceeding according to the state of the cancer. The states are specified in the section about the model family of the cancerous process. The first implementable method of the screening algorithm is described on the poster: Discussion of mathematical approach of psa level test during prostate cancer analysis by Péter Fésüs.

The four mathematical software which where examined as possible tools are: *Mathematica*, *Maxima*, *R* and *Scilab*.

These tools where chosen because Maxima, R and Scilab are counted as powerful free solutions [5]. The most well known proprietary solutios are Maple, Mathematica and MATLAB. Maple has no trial version, and MATLAB has its own distribution network also for the trial version, which takes time to proceed, so it does not arrive in time. So the only proprietary program which was examined, is the Mathmatica.

Prostate cancer facts [1]

Prostate cancer is one of the most frequent cancer diseases in men, but usually it isn't an agressive tumor, it grows slowly and it is quite well treatable, but in Hungary the most cases are diagnosed in late stages, when only the symtoms can be treated. The diagnosing of the prostate cancer based on 3 methods

- *Prostate-Specific Antigen* (PSA) level test, quite cheap easily usable blood test
- *Digital Rectal Examination* (DRE), quite cheap easily usable test
- *Multiple Transrectal Ultrasound Guided Biopsy,* expensive test with higher complication rate

The PSA and DRE tests are theoretically usable for screening purposes. The PSA-level is naturally increases with age, so the critical level is age specific, but this level is not specific enough, because the rate of the false negative and false positive cases is higher then desired.

Model family of the cancerous process

In [4] the progress of cancer cell increasing is assigned to 5 types, which describe the progress quite universally.

- Type I. Not savable state: The cancer is highly progressive, it is not curable. Rapid intervention can extend life.
- Type II. Risky state: The cancer progresses not too fast, there is no need for immediate intervention, but intervention is neccesary. It is at least temporary curable
- Type III. Lucky state: The cancer cell count is low and slowly increasing. Nintervention not always neccesary. Probable that not the cancer will be the cause of death.
- Type IV. Favorable state: The cancer cell count increases, than stagnate. The cancer is detectable but not malignant
- Type V. Normal state: There is no cancer detectable.

Description of the programs

- **Mathematica**[11] Wolfram reffers it as „the world's ultimate application for computations”. It has definitely more capabilities than a general purpose CAS, there is possible to build a presentation or even a whole application with it. It is a sophisticated tool. It handles cells, which can be not only input cells for computation, but other elements for example a presentations text part.
- **Maxima** [9]It is a Computer Algebra System (CAS), for manipulate symbolic and numeric expressions. It is able to plot 2d and 3d functions. The „Maxima is a descendant of Macsyma, the legendary computer algebra system developed in the late 1960s at the MIT”
- **R** [8] R can compete with the commercial statistical computing and analisys softwares. It is a program language and an environment, with a wide functionallity. It has the capability to draw all statistic related graphics
- **Scilab** [10] It is an “open source software for numerical computations”. It is a close match to MATLAB, even has its own converter from MATLAB.

Requirements and preferences of the tools

- Batch running natively or with simple script available
- GNU/GPL licence preferred
- Good quality documentation available
- Living community site
- Generally used operating system support (Windows, OS-X, Unix)
- „Pretty” basic plotting
- Basic numerical functionality
- Statistical module

Test case descriptions

The test case for each tool is a non-linear (quadratic) least square function fit. Each program is able to do it, some have multiple algoritmhs for it. The choosen function is from [3] the polinom (2).

$$p(t)=b+vt+\frac{a}{2}t^2$$

The test data set is:

t	0	0,5	1	10	15
p	5,6	6,4	7,2	25,6	39,4

As a first test the I did the estimation using the least square solution described in the manuscript [3]. It has simple steps, so each program is able to do it. The next step was to discover each tools statistical package, to find the function which uses non-linear least square datafit.

- **Scilab**: The non-linear least square problem is solvable with scilab using datafit, leastsq or lsrsolve fumction. Each need a function definiton and measurement data. The lsrsolve uses starting values.
- **Maxima**: The leastsquares collection contains multiple functions to estimate parameters. For the test the lsuare_estimates function solution was implemented, which first seeks an exact solution then an approximate.
- **R**: There multiple possibilities to solve the problem. Function lsfit is the matrix form solution. The basic non-linear least squares model is implemented in the nlm function, it needs a starting values. The stats package contains extensive facilities, as nls function. The result of these function are more detailed than other tools, for example return correlations, residual values, etc...
- **Mathematica**: The least square function is in the Optimalization package. LeastSquares function is the implementation of the least square problem in matrix form. There is also a general finction FindFit, which performs “optimal nonlinear unconstrained or constrained fit to data”. It is a simple to use tool, with it the test case is solvable in two lines.

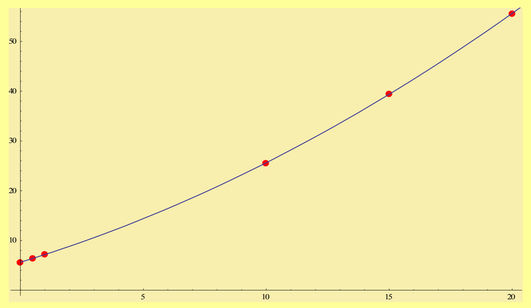
Comparison table

	Scilab	R	Maxima	Mathe-matica
Basic numerical functionality	yes	yes	yes	yes
Statistical module	yes	yes	yes	yes
Scripts	yes	yes	yes	yes
GNU/GPL	yes	yes (GPL-compatible)	yes	no (proprietary)
OS support	U/M/W	U/O/W	U/O/W	U/O/W

Plots

- **Scilab**: The plot function can be easily parametized,. The result can be zoomed and exported to multiple formats including vectorial formats too. The output is acceplable for presentation purposes too.
- **Maxima**: It has a built in graphical tool, but also can use 3rd party tools as gnuplot. It was not possible to create a plot with the result function and the input data in the same graph, because of a bug in the program. The output is parametarizable, but without parameters it is not too nice
- **R**: It contains all the neccesary graphical functionality for statistical purposes. The graphs can be parametized easily. The result is resizable, but not zoomable, export is possible to multiple formats, icluding postcript.
- **Mathematica**: The built in graphical tool is really powerful, the created plot can be manipulated with menu points, even a drawing tool is usable to add extra elements. Professional as expected from a commertial system.

The graph generated by Mathematica



Conclusion

Each tools fulfill almost all the requirements. The plotting functionality of the Maxima is not has issues. The Mathematica is properiary, but all other aspect of the program are ideal. That leaves the Scilab and R system as best opportunities. The Scilab is better for statistical analysis, it contains more special functions. But for quick numerical computations the Scilab is the better tool.

References

[1] Dr. Nagy Károly: A prosztata rosszindulatú betegségének rövid összefoglalása (Budapest, 31/03/2012, Manuscript)
[2] Dobó Andor: Sztochasztikus rákos folyamatok (Budapest, 24/05/2012, Manuscript)
[3] Dobó Andor: Prosztatarák terjedésének matematikai leírása (Budapest, 07/03/2012, Manuscript)
[4] Dobó Andor: A rákos folyamat modell-családjának diszkutálása (Budapest, 07/06/2012, Manuscript)
[5]Antonio Cangiano :3 awesome free Math programs (02/06/2007) http://math-blog.com/2007/06/02/3-awesome-free-math-programs/
[6] List of numerical analysis software – Wikipedia (version modified on 09/01/2014) http://en.wikipedia.org/wiki/List_of_numerical_analysis_software
[7] List of computer algebra systems – Wikipedia (version modified on 30/12/2013) http://en.wikipedia.org/wiki/List_of_computer_algebra_systems
[8] The R Project for Statistical Computing http://www.r-project.org/
[9] Maxima, a Computer Algebra System http://maxima.sourceforge.net/
[10] Scilab homepage http://www.scilab.org/
[11]Mathematica homepage http://www.wolfram.com/mathematica/
[12] Mathematica Stack Exchange http://mathematica.stackexchange.com/

Aknowledgement

This poster would not be possible without
Mr Andor Dobó , who described the process of cancer from the mathematical viewpoint,
Dr Károly Nagy who provided the medical backgroud of the work and
Péter Fésüs, who encouraged the publication