



REVIEW OF MATHSTATICA (V.1): AN ADD-ON TO MATHEMATICA

H. D. VINOD*

Economics Department, Fordham University, Bronx, NY 10458, USA

1. OVERVIEW

mathStatica is an add-on package for Mathematica. I used Version 1.0 in conjunction with Version 4.2 of Mathematica for Windows. The latter sells for \$1880, (academic \$895, student \$135) at www.wolfram.com. mathStatica comes with a nicely produced book, Rose and Smith (2002), which sells for about \$80 and includes two CDs inside the back cover. The first contains a trial version of Mathematica (which needs a password), and the second contains mathStatica. A gold version of mathStatica sells for an additional \$89 (if you own the book) or \$159 (if you do not). The software can be ordered on-line at: www.mathstatica.com. Support is available by E-mail. Both Mathematica and mathStatica are license protected with their own independent licenses.

The attraction of this software for econometricians is that it makes it fun to learn and relearn some conceptually relevant and important but analytically tedious results from mathematical statistics. No other software offers anything even close. For example, it is fun to get moments from characteristic functions of economically realistic densities such as Pareto, or for transformed versions of known densities. One can invert the characteristic function to uniquely determine the density or its quantiles by simple commands, either numerically or analytically. As it develops further, this software promises to make seemingly esoteric econometrics and finance easier to understand and use. We hope that the authors strive to make these tools user-friendly and provide additional econometrics examples. A second attraction is its ability to achieve a near perfect numerical accuracy. However, since mathStatica does not generally provide pre-programmed routines for numerical results, it is not intended to be a substitute for packages like GAUSS or S-Plus.

* Correspondence to: H. D. Vinod, Economics Department, Fordham University, Bronx, NY 10458, USA.
E-mail: Vinod@fordham.edu

2. SPECIFIC COMMENTS ON MATHSTATICA

2.1. Overall Philosophy and Objectives

The Preface of Rose and Smith (2002) notes that it is a symbolic engine of tools for doing mathematical statistics with an emphasis on problem solving rather than proofs, with the potential to enliven the experience. Since the preface claims to piggyback on the advantages of Mathematica, we indicate some disadvantages to alert readers unfamiliar with it. The authors provide a computer tool for finding moments, transformations, Cramer-Rao bounds, maximum likelihood (ML) estimators, and many more manipulations for nearly arbitrarily specified random variables. For example, multivariate, piecewise, discrete, censored, mixture and many other sophisticated random variables are available. One has the option to do the calculations analytically or numerically. The program is multi-purpose, flexible, and can solve new problems in seconds.

Although this software is for mathematical statistics, not just econometrics, it clearly fills a need of many econometricians. It is recommended for those who want to quickly refresh their skills in math stat concepts, without the drudgery of tedious mathematical derivations. Even experts may find that it saves time in their work as referees for journals, and in their research, for checking mathematical derivations with a variety of alternative probability models. It can indeed enliven the educational experience if used in a classroom.

2.2. Implementation

mathStatica runs on Windows, Mac OS, Linux, and UNIX. My Mathematica license permitted me to test it only on 'Windows ME,' and it did not crash. On a 1 to 10 scale of user-friendliness (10 = most friendly), Mathematica-mathStatica is about 6. On 'ease of learning' it is 6, as context sensitive help is available by highlighting and pressing F1. An introductory slide show and a 'Getting Started' booklet are helpful. I would have avoided considerable frustration if only (i) I had read the first sixty pages of the Mathematica book carefully, and (ii) I had exploited the access to the electronic version of the Rose and Smith book in the form of notebook files and started with Appendices A1 to A9. This is the first book I have seen where every input equation and every diagram in the book can be verified, recreated, and changed if appropriate.

I installed Mathematica-mathStatica on my HP Pavilion PC Model 8756C with Intel Pentium III processor running at 850 MHz, using the recommended file names and the license number and password included in the package. The installation was a bit daunting due to slightly unclear directions on page 3 of the book where a back slash is denoted by a triangle. I am told that this is a standard indicator of structure used by operating systems other than Windows.

The mathStatica book suggests loading it by typing '<<mathStatica.m'. Since I am used to GAUSS (Vinod, 2000a), which has the prompt '>>', I looked in vain for '<<' as a prompt. Another *unlearning* for me involved Mathematica's use of the semi-colon to suppress screen output. I now appreciate the semicolon facility for suppressing displaying of very large data matrices or other long outputs, since the output remains available by simply typing the % symbol. Multiplication can be stated by a space (instead of the usual *), which I thought was elegant. However, if one forgets the space, diagnostics do not help. Hence I continue to use a * for multiplications.

The mathStatica book suggests learning Mathematica conventions by unconscious osmosis, but this needs many days of trial and error. At the very start, the reader should use the sequence: 'File menu, Palettes' and choose 'Basic Typesetting'. Shortcuts for entering mathematical operations are

nice to know early on. For example, the sequence 'Escape | Escape' enters λ , whereas 'Ctrl-' enters a subscript box. Since most software does not come with a detailed electronic book with hyperlinks, a typical reader may not notice the power offered by this feature. I recommend its extensive use to avoid frustration. If one buys the 'gold' version, one can e-mail the '.nb' or notebook file to technical help people in Australia and get an answer fairly quickly subject to time differences.

mathStatica is not meant for simple regression. However, highlighting the word 'Regress' and pressing the F1 key in Mathematica gives information about how to run it with two or three lines of code. Its output has all the basics, including adjusted R^2 and p values. To get information about how to get more diagnostic goodies, the user has to highlight the word Durbin or 'DurbinWatson' and hit the F1 key. The user who highlights Durbin-Watson with a hyphen will be out of luck. Many sophisticated tools for outlier detection available here are missing in other econometric packages.

2.3. Data Input and Data Management

It is explained in the Rose and Smith book that all inputs follow the Mathematica conventions, especially capitalizations (first caps even for mixed words). One of the available datasets is brought in by the command `xdata=ReadList['nerve.dat', Number]`. It needs the word 'Number' to identify that it is numerical data, with the capital N. I wish future versions would accept some misspellings to enhance user-friendliness. If I mistype 'Tale' instead of 'Table,' Mathematica already points out my typing error and asks whether I mean 'Table' or 'Take'. I would like to see more such smart behavior. Of course, in some cases we would need the option to turn off the smart behavior. After some trial and error I found that my test data are best input by the command `Import['c:\\temp\\test.dat']`, where double back-slashes and quotes are needed in my Windows environment as in GAUSS. Alternatively, `SetDirectory['c:\\temp']` is better if more files from the same directory are needed.

Mathematica-mathStatica is interactive, but commands can be stored anywhere in a text file and called with '`<filename`'. This needs no quotes provided the batch input file is in the directory set by the `SetDirectory` command earlier in the session. I needed help from support@wolfram.com, in this context. They gave me an ID number by return mail and I got a reply in about 48 hours, which asked me to go to the FAQ's on their web page. The FAQs are not numbered and I had to search for the right one.

Highlighting the word 'Import' and pressing the F1 key is a quick way to find permissible Import formats. Since Excel workbooks can be saved as files with the CSV extension, they can be readily imported. Due to its symbolic nature, Mathematica-mathStatica can readily implement any data transformation by simply defining it. However, it does not offer ready-made time-series filters or a graphical user interface (GUI) for such tasks. The mathStatica book comes with ten trial data sets and some artificial data. Bank of Melbourne stock price data are used for ARCH, whereas a biometric dataset is used to explain maximum likelihood (ML). Very large data sets can be manipulated and stored given adequate memory and disk space.

2.4. Estimation

Given data, a probability distribution may be fit in mathStatica with a wide range of options described in chapters 1 to 6 of the book. For almost any estimator, chapter 7 describes estimation

of sampling distributions. The electronic book can be a template for the following estimation techniques: admissible, asymptotic unbiased, best unbiased, density estimation, moments or cumulants, Hessian estimator of variance, van Beek's bound for the Method of Moments estimator, non-parametric kernel density, and many others. Chapters 10 and 11 describe the theory of unbiased and ML estimation with a wide range of distributional choices. Chapter 12 has ML estimation with data, where the ARCH example nicely illustrates how to find the causes of a crash and how to impose restrictions on parameters to recover from it. MathStatica is not intended for implementing econometric estimators such as generalized method of moments (GMM), State Space, ARMA, or GARCH. Some templates are offered by another add-on Mathematica package called Time Series.

The numerical superiority of Mathematica arises from the user-specified number of digits of accuracy. For example, it can compute π up to, say, a million digits. Also, the Mathematica command 'Rationalize' works with rational numbers directly and can avoid decimal approximations tied to machine accuracy altogether. These advantages are not automatic and usually need simple, but additional programming steps; see McCullough (2000).

mathStatica offers several techniques that are not available elsewhere. For example, (i) given a probability distribution, the sampling distribution of its order statistics can be computed by using the mathStatica command OrderStat. This is useful in Finance, where the sampling distribution of value at risk (VaR) is needed. (ii) Vinod (2000b) argues that to properly measure risk (of loss), the standard deviation should be restricted to the 'loss range' of excess returns. mathStatica permits estimation and inference for a statistic restricted to any such range extremely easily. (iii) The ideas of sufficient statistics and the Rao-Blackwell theorem come alive with the use of mathStatica. (iv) Chapter 11 has three-dimensional graphs, which make the likelihood function and ML estimation fun to understand and explain. (v) Statistical inference, which involves analytical evaluation of expectations, can use mathStatica, for example, to develop 'estimating functions', as in Vinod (1998, 2000b), or to create a small sample pivotal, as in Davidson and MacKinnon (2002).

By opening the electronic version of Chapter 12 of the Rose and Smith book, it was rather fun to see the frequency plots for the raw data and the fitted data side by side. McCullough and Vinod (1999, 2003), Nerlove (1999), and others argue the need to explore the likelihood surface before accepting the solution produced by any software. The gradient at the solution should be zero and the eigenvalues of the Hessian should be negative to make sure that the observed log-likelihood is concave. This is done by the convenient mathStatica commands 'Grad' and 'Hessian', in conjunction with the Mathematica command 'Eigenvalues'. Such exploration and profiling of the likelihood function is feasible with mathStatica. However, the authors need to provide more examples of such operations, since the syntax is not easy to learn. I expect to use it to teach the subtle differences between BFGS, Newton-Raphson, and other solvers. There are wonderful tools for computing moment-generating functions, characteristic functions, and asymptotic limits. Since Mathematica alone does not yield tractable likelihood functions, the 'SuperLog[On]' and 'Grad[f, β]' commands are a critical contribution of mathStatica.

2.5. Documentation

The Mathematica book with 1450 pages is all documentation. Rose and Smith's 500-page book contains both documentation and detailed explanations. The new user of Mathematica should start with a very careful reading of the 'Getting Started' booklet and should actually issue all the

commands in it. There are many unintuitive notations, which must be learned first. I might have saved considerable grief if I had learned the replacement operator `'/.'`, the distinction between `'x'` and `'x_.'` with underscores, arrows, colons, double brackets, etc. The *mathStatica* book has a large index but did not have any entry for sampling distribution (although the topic is covered) or singular value decomposition (since the topic is not covered). The manuals are attractively printed, easy to read, and will stand up to long and repeated use.

Extensive on-line help can be obtained by many methods: (i) If one knows the name or part of the name of a command, one can type `'?command-name.'` For example, `'?Central*'` provided help for central moments and conversion to raw moments, cumulants, etc. However, when I typed `'central'` without the initial capital letter, it did not find anything. (ii) Highlighting an expression and pressing the F1 key may be the best method. (iii) When one starts *mathStatica*, the computer displays on-line links to chapter numbers in the book and electronic index. I strongly recommend the electronic index rather than the index of the book, since the former has easy hyperlinks for clicking. Some links are in the form of 'closed group' cells or what I call 'blue right brackets' on the right ends of various groups of lines in the electronic manual. Clicking on these brackets provides a nice access to all commands and explanatory text. More importantly, one can highlight and paste them in one's own program to avoid frustrations due to typing or syntax errors.

2.6. Model Diagnostics and Quality of Output

Mathematica and *mathStatica* can provide a comprehensive set of diagnostic statistics for many statistical models, provided one knows how to program them. It provides very few statistical diagnostics automatically, as do most standard software packages.

I like the graphical output and the fact that it can be easily cut and pasted into MS-Word and TeX documents directly. This feature is often not as convenient with other software. An annoying feature of the transfer to MS-Word is that the entire set becomes a 'picture', which does not line up unless one uses: 'Format menu, Font, Character and Spacing', 'Position: Lowered' and 'By: 4pt'. The results of each session are stored in 'Mathematica notebook files,' denoted by (*.nb) and designed for later use. One can save a notebook as text, TeX, XML, or HTML, among others. I found the HTML option the most convenient for transferring to MS-Word after I downloaded the *MathReader* fonts from Mathematica's website.

Most chapters in the *mathStatica* book end with well-designed exercises, which illustrate the power of the software as a learning tool. For example, economists can use Azzalini's (1985) skew-Normal distribution to represent asymmetric density functions; see Figure 1, which also illustrates Mathematica's graphics capabilities. Its extension to skew-t, which has the advantage of allowing for fat tails, appears to be even more useful. I have used this density on mutual fund return data and found that it gives a more realistic representation. There is considerable evidence that the behavior of financial markets during boom years and recessions is not symmetrical, and that different industry groups react differently during the ups and downs of business cycles. Adding one parameter λ to the usual normal theory has a considerable potential in financial economics. No other software can readily give the score function for λ , the parameter of interest.

To save space we cannot illustrate how *mathStatica* provides moments, moment generating function, order statistics, Fisher Information, first and second order conditions for maximizing a log likelihood function, and myriad other operations. Researchers worried about non-normal errors can have more options using this software.

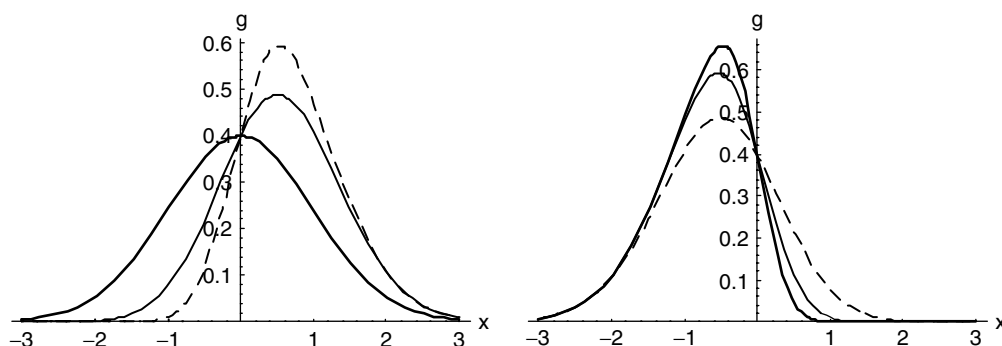


Figure 1. The left panel has SN density for $\lambda = 0$ (dark line, unit Normal), $\lambda = 1$ (plain) and $\lambda = 2$ (dashed line). The right panel has $\lambda = -3$ (dark line), $\lambda = -2$ (plain line), and $\lambda = -1$ (dashed line)

I have used mathStatica in my Econometrics class, and students seem to especially like the animations. For example, the animation of bivariate normal quantiles explains the idea of collinearity. Classroom use requires the use of the Format menu, which provides 'Presentation' as an option in 'screen style environments.' It is also helpful to save notebooks in HTML format for classroom use. If a student asks a math stat question, it is convenient to go to the electronic index and show the answer with graphics, if any, to the class with the zooming options located at the bottom left. Chapters in the mathStatica book end with well-designed exercises, which illustrate the power of the software as a learning tool. Unfortunately, some matrix algebra commonly used in mathematical statistics and hypothesis testing (Likelihood Ratio, Wald, LM, restricted least squares) are not included.

3. CONCLUSION

For the modest price, mathStatica is a wonderful new package with a new philosophy. Since mathStatica can be fun to use, it has tremendous potential to enhance the popularity and quality of research in econometrics. The graphics in Mathematica-mathStatica can generate plots of arbitrary functions, nearly arbitrary mixtures, and transformations of a very wide range of probability distributions for easy export to word processors. However, for standard econometrics computations, one should not expect to replace packages similar to GAUSS, (aptech.com) or S-Plus (insightful.com). I strongly recommend mathStatica as a great concept-learning and research tool, especially for rigorously evaluating the properties of many econometric estimators and in teaching.

REFERENCES

- Azzalini A. 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**: 171–178.
- Davidson R, MacKinnon JG. 2002. Bootstrap J tests of nonnested linear regression models. *Journal of Econometrics* **109**: 167–193.
- McCullough BD. 2000. The accuracy of Mathematica 4.0 as a statistical package. *Computational statistics* **15**: 279–299.
- McCullough BD, Vinod HD. 1999. The numerical reliability of econometric software. *Journal of Economic Literature* **37**: 633–665.

- McCullough BD, Vinod HD. 2003. Verifying the solution from a nonlinear solver: A case study. *American Economic Review* (in press).
- Nerlove M. 1999. Chapter one: *The likelihood principle*. Dept. of Agricultural and Resource Economics: University of Maryland.
- Rose C, Smith MD. 2002. *Mathematical Statistics with Mathematica*. Springer Verlag: New York.
- Vinod HD. 1998. Foundations of statistical inference based on numerical roots of robust pivot functions. *Journal of Econometrics* **86**: 387–396.
- Vinod HD. 2000a. Review of GAUSS for Windows, including its numerical accuracy. *Journal of Applied Econometrics* **15**: 211–220.
- Vinod HD. 2000b. Foundations of multivariate inference using modern computers. *Linear Algebra and Its Applications* **321**: 365–385.