



Journal of Statistical Software

September 2010, Volume 36, Book Review 4.

<http://www.jstatsoft.org/>

Reviewer: Joseph M. Hilbe
Arizona State University

Introduction to Scientific Programming and Scientific Simulation Using R

Owen Jones, Robert Maillardet, Andrew Robinson
Chapman & Hall/CRC, Boca Raton, Florida, 2009.
ISBN 978-1-4200-6872-6. 472 pp. USD 79.95.

<http://www.crcpress.com/product/isbn/9781420068726>

It is not often that I think that a statistics text is one that most scientific statisticians should have in their personal libraries. The 2008 text by Owen Jones, Robert Maillardet, and Andrew Robinson titled, *Introduction to Scientific Programming and Simulation Using R* is such a text. I am assuming, however, that R is the primary statistical package that is used among statisticians who are evaluating biological, geological, astronomical, environmental, and other scientific research areas. But even if a researcher does not use R as their preferred statistical tool, authors of research articles in scientific journals now appear to overwhelmingly employ R for executing and displaying published statistical results. In order to fully understand – or to referee – these articles, an understanding of R is desirable. This text provides scientific researchers with a working knowledge of R for both reviewing and for engaging in the statistical evaluation of scientific data.

The text is partitioned into four sections: “Programming”, “Numerical techniques”, “Probability and statistics”, and “Simulation”. Appendices provide a highly useful glossary of R commands as well as a listing of all programs and functions developed in the book, together with the page on which each is first used or defined. Exercises are provided for all chapters except the first.

In this review, I shall provide a brief overview of the chapters and discussion presented in each of the above mentioned sections.

Programming

1. Setting up

Provides directions on installing R, setting up a working directory, and using the help system. Guidance on getting R ready for work.

2. R as a calculating environment

Explains how to use R for basic calculating. Discusses foremost aspects of the R environment including variables, functions, vectors, handling missing data, assignments and expressions, and matrices. Fully worked out examples are given for defining and using each of the above.

3. Basic programming

Explanation with numerous examples of branching, looping, vector-based programming, program flow, debugging, and proper programming style.

4. Input and Output

Demonstrates input of data from keyboard or from file, and output to various formats.

5. Programming with functions

Discussion of how to construct properly working functions including vector and recursive programming.

6. Sophisticated data structures

The chapter presents examples on constructing and manipulating dataframes and lists, and how to define the scope of functions.

7. Better graphics

A basic introduction to graphics. A special section on the use of the **lattice** package is provided.

8. Pointers to further programming techniques

The final chapter in this section provides examples on object-oriented programming and the use of compiled C code in R functions.

Numerical techniques

For those who already have some proficiency in R, the important part of the book begins with this section.

9. Numerical accuracy and program efficiency

Significance of digits, determining the time functions take to run, and looping and memory control are all addressed with examples in this chapter.

10. Root-finding

The basics of estimation. Constructing Newton-Raphson code and estimation using secant and bisections methods.

11. Numerical integration

Examples using both the trapezoidal and Simpson's rule, and adaptive quadrature.

12. Optimization

Discussion and examples are given on how to optimize. Examples using Newton's method, the golden rule method, steepest ascent, and multivariate optimization are examined in detail. A fully developed function is presented for curve fitting.

Probability and statistics

13. Probability

An introduction to probability theory, including Bayes theorem. No R code is used in this chapter.

14. Random variables

A general overview of the nature of random and pseudo-random number variables; R code is developed for their construction. PDF and CDF theory, with the gamma PDF and an example.

15. Discrete random variables

Focus on details of the binomial, Poisson, and negative binomial distributions.

16. Continuous random variables

Focus on continuous variables including the normal, and χ^2 . Simulations are developed. An example of the Poisson process and gamma distribution is presented with a fully worked out example of a discrete queue simulation.

17. Parameter estimation

A major chapter providing theory and examples of the central limit theorem, development of confidence intervals using various methods, and point estimation.

Simulation

18. Simulation

Methods are given for constructing simulated data sets with specified distributions and parameter. Rejection method is examined with code for constructing a variety of simulations based on the method.

19. Monte Carlo integration

Methods of Monte Carlo integration are examined.

20. Variance reduction

Importance and reduction sampling techniques are provided. Methods of variance reduction examined.

21. Case studies

Several fully worked out examples are given, with complete R code provided. Examples are given for the SIR (susceptible, infected, and removed) epidemic models, inventory analysis, and seed dispersal from forestry management. Simulations are demonstrated with R code provided and explained.

22. Student projects

Projects are specified for reader/student evaluation. The subjects are: (1) level of a dam, (2) roulette, (3) Buffon's needle and cross, (4) insurance risk, (5) squash, and (6) stock prices. Code is provided on the book's web site for the analysis of each of these example problems.

As a demonstration of how the book can be used to construct an optimization, I used the methods discussed in the text to develop code that can be used to estimate an NB-C model using R. It is a simple optimization function in which the model response and predictors are specified in the code and the parameter estimates standard errors, z value, and confidence intervals are displayed upon running the script. I later enhanced the model so that the code could accept any data and not have it specified in the function itself. It is included in my book, *Negative Binomial Regression* (Hilbe 2011). It should be noted that, before this, R did not have any functions to estimate an NB-C model. First, the data is loaded

```
R> load("medpar.rda")
R> medpar$type <- factor(medpar$type)
```

Then the function `nbc.reg.ml()` for NB-C optimization is defined

```
R> nbc.reg.ml <- function(b.hat, X, y) {
+   a.hat <- b.hat[1]
+   xb.hat <- X %*% b.hat[-1]
+   mu.hat <- 1 / ((exp(-xb.hat)-1)*a.hat)
+   p.hat <- 1 / (1 + a.hat*mu.hat)
+   r.hat <- 1 / a.hat
+   sum(dnbinom(y, size = r.hat, prob = p.hat, log = TRUE))
+ }
```

The design matrix is

```
R> nbcX <- model.matrix(~ hmo + white + type, data = medpar)
```

with starting points (discovered by trial and error!)

```
R> p.0 <- c(alpha = 0.5, cons = -1, hmo = 0, white = 0, type2 = 0, type3 = 0)
```

The joint conditional log-likelihood is maximized by

```
R> nbc.fit <- optim(p.0, nbc.reg.ml, X = nbcX, y = medpar$los,
+   control = list(fnscale = -1, maxit = 10000), hessian = TRUE)
```

and the parameter estimates and asymptotic standard errors are obtained by

```
R> (nbc.beta.hat <- nbc.fit$par)

      alpha      cons      hmo      white      type2      type3
0.44516214 -0.20128564 -0.01399781 -0.02491594  0.04101961  0.10723427

R> (nbc.se.beta.hat <- sqrt(diag(solve(-nbc.fit$hessian))))

      alpha      cons      hmo      white      type2      type3
0.019775949 0.013496886 0.010729402 0.010756500 0.008932639 0.010062008

R> nbc.results <- data.frame(Estimate = nbc.beta.hat,
+   SE = nbc.se.beta.hat, Z = nbc.beta.hat / nbc.se.beta.hat,
+   LCL = nbc.beta.hat - 1.96 * nbc.se.beta.hat,
+   UCL = nbc.beta.hat + 1.96 * nbc.se.beta.hat)
R> rownames(nbc.results) <- c("alpha", colnames(nbcX))
R> nbc.results <- nbc.results[c(2:nrow(nbc.results), 1),]
R> nbc.results
```

	Estimate	SE	Z	LCL	UCL
(Intercept)	-0.20128564	0.013496886	-14.913488	-0.22773954	-0.174831744
hmo	-0.01399781	0.010729402	-1.304622	-0.03502744	0.007031817
white	-0.02491594	0.010756500	-2.316361	-0.04599868	-0.003833201
type2	0.04101961	0.008932639	4.592105	0.02351164	0.058527585
type3	0.10723427	0.010062008	10.657343	0.08751273	0.126955803
alpha	0.44516214	0.019775949	22.510280	0.40640128	0.483922999

For comparative purposes, I show the maximum likelihood results of a NB-C model using Stata. The results are nearly identical to the thousandths place. The Stata code is provided by [Hilbe \(2007\)](#).

```
. cnbreg los white hmo type2 type3
```

```
Canonical Negative Binomial Regression          Number of obs =      1495
                                                Wald chi2(4)    =      131.15
Log likelihood = -4796.6997                    Prob > chi2     =      0.0000
```

los	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
white	-.0248636	.0107598	-2.31	0.021	-.0459524 -.0037749
hmo	-.0141484	.0107354	-1.32	0.188	-.0351894 .0068926
type2	.0409448	.0089345	4.58	0.000	.0234335 .0584561
type3	.1072065	.0100628	10.65	0.000	.0874837 .1269293
_cons	-.2012769	.0135036	-14.91	0.000	-.2277436 -.1748103

```

-----+-----
      /lnalpha |  -.8091036   .0444641  -18.20   0.000   -.8962516   -.7219556
-----+-----
      alpha |    .445257   .0197979
-----+-----
AIC Statistic =          6.424

```

Scientific Programming and Simulation Using R can be used to develop a host of similar models as well as functions for a variety of analytic needs. It is particularly useful for understanding and developing modeling and simulation software. I highly recommend the text, finding it to be one of the most useful books I have read on the subject.

References

- Hilbe J (2007). “hcnbreg: Stata Module to Estimate Heterogeneous Canonical Negative Binomial Regression.” Statistical Software Components, Boston College Department of Economics. URL <http://ideas.repec.org/c/boc/bocode/s456870.html>.
- Hilbe JM (2011). *Negative Binomial Regression*. 2nd edition. Cambridge University Press, Cambridge.

Reviewer:

Joseph M. Hilbe
 University of Hawaii, and
 Sociology and Statistics, Arizona State University
 Tempe, Arizona, United States of America
 E-mail: hilbe@asu.edu