Journal of Statistical Software

June 2017, Volume 78, Book Review 4. *doi: 10.18637/jss.v078.b04*

Reviewer: Hakan Demirtas, Rachel Nordgren
University of Illinois at Chicago

Introduction to Scientific Programming and Simulation Using R (2nd Edition)

> Owen Jones, Robert Maillardet, Andrew Robinson
> Chapman &Hall/CRC, Boca Raton, 2014.
> ISBN 978-14665699997. xxiv + 582 pp. USD 67.16 (P).
> `http://www.ms.unimelb.edu.au/~apro@unimelb/spuRs/`

The Introduction to Scientific Programming and Simulation Using R (2nd Edition) is a useful and well-organized book. The writing is orderly, logical, consistent, intriguing, and engaging. We have read many programming and simulation oriented books that vary in context, scope, and difficulty level. This one turned out to be one of our favorites. It stands out in the sense that a decent dose of theory is given in addition to the programming related aspects. It covers an immense amount of material, yet manages to do so both thoroughly and clearly. The book has 24 chapters, divided into four parts, and each part contains enough material for a nine week class. The preface indicates that calculus is a co-requisite, and that prior knowledge of probability or programming is not assumed. With some care in teaching, particularly for numerical integration and the random variable theory, that is a reasonable background for what the authors list as the core content for a course on scientific programming and simulation (Chapters 1–6, 9–11, 14–16, 20.1–20.2, 21, 23.1–23.2, 24). To more fully benefit from this book, the mathematical level and quantitative comprehension of the readers should go a little beyond what is stated in the book (for example, some linear algebra knowledge would be beneficial). While there is something for everyone ranging from novice programmers to mathematically or statistically sophisticated researchers, a decent level of mathematical aptitude is needed. It could simultaneously be a reference book, a self-study book, and a textbook, which is hard to achieve by any standards. End-of-chapter exercises are mostly interesting and thought provoking, and the first author of this manuscript occasionally uses them in his doctoral level computational statistics and masters' level R programming classes. Most R books out there are designed for data analysis, whereas this one exclusively concentrates on computation and simulation, which can (and does!) lead to a better understanding on how to turn algorithms into code.

The book comes with an R package **spuRs** that is available from the Comprehensive R Archive Network at `https://CRAN.R-project.org/package=spuRs`. The functionality is satisfactory, however, the reference manual for **spuRs** and other documentation are outdated, and need to be updated. The manual does not mention/include all data sets and scripts in the book. One has to separately download some of these from a zip file. The authors properly acknowledge this, and provide guidance on how to get access to all resources in the book and on the first author's website (http://www.ms.unimelb.edu.au/~apro@unimelb/spuRs/). The number of typographic errors is within acceptable limits, and the authors make an errata file available for the First Edition. While the book scores high on readability, some intricate details (perhaps too many to grasp in the first reading!) are occasionally given. It was easier to grasp the material on the second reading, however, we still thought that some of the notation was confusing.

We now give a general picture and summary of the coverage across the book. As mentioned before, there are four major parts and 24 chapters. Part 1 (Chapters 1–8) is concerned core knowledge of R and programming concepts. Chapter 1 describes how to obtain and install R, and the package **spuRs**, which complements the book. Chapter 2 shows how to use R to do arithmetic calculations; create and manipulate variables, vectors, and matrices; work with logical expressions and the built-in functions. Chapter 3 introduces a set of basic programming structures that are the building blocks of many programs. Chapter 4 describes some of the infrastructure that R provides for importing data for subsequent analysis and for displaying results in the form of graphics. Chapter 5 covers the creation of functions, the rules that they should follow, and how they relate to the environments from which they are called. Chapter 6 gives information about more sophisticated data structures such as lists and dataframes. Chapter 7 provides further exposition of the graphical capabilities of R. Chapter 8 discusses more advanced aspects of programming in R.

Part 2 (Chapters 9–13) focuses on numerical techniques. Chapter 9 is about numerical accuracy and program efficiency, it gives details concerning how computers operate, and their implications for programming practice. Chapter 10 presents a range of different techniques for root-finding. It covers fixed-point iteration, the Newton-Raphson method, the secant method, and the bisection method. Chapter 11 introduces numerical integration, it covers the trapezoid rule, Simpson's rule, and adaptive quadrature. Chapter 12 covers the optimization tools in the multivariate case. Chapter 13 is pertinent to systems of ordinary differential equations.

Part 3 (Chapters 14–19) is about the salient aspects probability, random variables, and expectation, which are central to the simulation paradigm. Chapter 14 introduces mathematical probability. Chapter 15 defines discrete and continuous random variables, along with different routes of describing their distributions (probability density and mass functions, expectation, variance, covariance, and independence). Chapters 16 and 17 present the theory and applications of discrete and continuous variables, respectively. Chapter 18 covers point and interval estimation. Chapter 19 describes discrete and continuous time Markov chains, transition matrices, and limiting behavior, among other things.

Part 4 (Chapters 20–24) is concerned with stochastic modeling, simulation, random number generation, Monte carlo integration, case studies, and projects. Chapter 20 simulates uniformly distributed random variables and discrete random variables, and describes the inversion and rejection methods for simulating continuous random variables. Chapter 21 covers simulation-based approaches to integration such as the hit-and-miss and Monte Carlo integration methods. Chapter 22 describes several sampling-based procedures that are designed to improve the efficiency such as antithetic sampling, control variates, and importance sampling. Chapter 23 presents a few case studies from diverse areas. Chapter 24 includes a set of problems that can be solved by motivated students.

Two chapters (13 and 19) are new to the 2nd edition: Chapter 13 and 19, whose more detailed descriptions follow below. The introduction to Chapter 13 gives a good explanation of what ordinary differential equations (ODEs) are, and why we might want to solve systems of them. Although the methods only use first order ODEs, the authors explain that any higher orders can be handled with the substitution of new variables for the higher order derivatives. The analysis and methods are specifically limited to initial value problems, which makes sense, since those are both the simplest and the most common form. They provide an example based on predator-prey equations, which has a two-dimensional output (referred to as a trajectory) which can be interpreted as a graph. The first three sections cover three methods for solving: Euler's method, Midpoint method, and Fourth-order Runge-Kutta method. Each section contains code and equations, as well as graphs for various step sizes showing the degree to which accumulated errors will cause the trajectory of the example problem to diverge. Section 13.4 compares the efficiency of the three methods. A simple ODE with an analytical solution is used, allowing the error to be calculated. Code is created to halve the step size whenever a specific tolerance is not achieved, and the number of calls per method is tabulated. The results are displayed in a table and a log-log plot, making it clear that the Fourth-order Runge-Kutta is vastly superior. The final Section 13.5 discusses Adaptive step size. The idea is similar to adaptive quadrature, with the step size adjusted based on how quickly the function is changing. Code is provided, and a new example is introduced, involving a safe landing for a lunar landing module. Depending on the problem, the ability to adjust step size could be very useful and considerably more efficient.

Chapter 19 is divided between discrete time chains (19.1–19.5), continuous time chains (19.6–19.9), and applications (19.10–19.12). The authors give three requirements for a discrete Markov chain: 1) A memoryless process, that is, the current step only depends on the immediately preceding one, 2) The time points are a discrete set, and 3) The possible states are also discrete. The discussion is limited to cases where the transition probabilities are unchanged over time. Examples are given, including a gambling game and a mouse in a maze. State transition diagrams are discussed, and the various equations are derived. The applications provided are easy to understand and could be expanded to fit a variety of problems. Continuous Markov chains extend the concept of discrete Markov chains to a continuous version of time. Requirement (3) still holds, and requirement (1) is modified such that a memoryless process is still required, but now the future depends only on the present and not on the past. The discussion is again limited to homogenous transition probabilities. A continuous chain can be thought of as a Poisson process, or as a number of competing events. The equations and diagrams become more complicated and less intuitive, and full understanding would require a lot of time spent studying the material, which we did not invest. For those who do, we believe this would a useful chapter with its insightful applications.

All in all, this book contains an incredible amount of material, and does a great job of explaining not only how to follow their examples, but also to extrapolate and create new tools. Anyone interested in learning more about programming in R, basic probability, numerical techniques, simulation, or any of the other topics covered in this book, can substantially benefit from reading this book. It concisely presents fundamental ideas on simulation and computation techniques, with many examples, in an intuitively appealing manner. It has the potential to be an important reference, consulting source, and a classic text.

Reviewer:

Hakan Demirtas, Rachel Nordgren
Division of Epidemiology and Biostatistics (MC 923)
University of Illinois at Chicago, School of Public Health
1603 West Taylor Street
Chicago, IL, 60612-4336, United States of America
E-mail: demirtas@uic.edu, rnordg2@uic.edu
URL: http://www.healthstats.org/members/hdemirtas.html

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
June 2017, Volume 78, Book Review 4 *Published:* 2017-06-21
doi:10.18637/jss.v078.b04