Journal of Statistical Software

January 2008, Volume 29, Book Review 2. *http://www.jstatsoft.org/*

Reviewer: Juana Sanchez
University of California at Los Angeles

Probability with R – An Introduction with Computer Science Applications

Jane M. Horgan
John Wiley & Sons, Hoboken, New Jersey, 2008.
ISBN 978-0-470-28073-7. 394 pp. USD 99.95.
http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470280735.html

---

Probability with R is a self-contained book with no need for ancillary material, other than the free software R. It is a book that will interest researchers using probability and R in their research, teachers of probability and statistics, the curious person seeking an understanding of this field, and those who need a user-friendly introduction to R before tackling the online manuals.

The book grew from a one-semester course in introductory probability for undergraduate students in computer science and reflects a very organized instructor who is very knowledgeable of, and concerned with, her students' motivations and background and the teaching of probability. Nothing in the lectures conveyed in the chapters is left to improvisation. Everything the student needs in order to do the exercises at the end of the chapters is in the book. However, Professor Horgan is developing a Web page to serve as a forum for teachers and students. At the moment of the writing of this book review, the most valuable item in the Web page is a set of four tutorials for students. The tutorials contain theoretical problems that the students can see solved step by step. The Web site also contains some of the chapters converted into lecture presentations that teachers can use. None of that is really needed if the teacher and students commit to reading the book well. There is no other book in the market like this at the moment.

The book follows the traditional division of topics in introductory probability curricula. It has five parts that discuss, respectively: the basics of R for summarizing data and doing graphical displays (Part I, three chapters); basic probability rules and counting methods (Part II, five chapters); discrete univariate random variables and their probability distributions (Part III, six chapters), continuous univariate random variables and their distributions (Part IV, five chapters) and Markov and Chebyshev's bounds (Part V, one chapter). For each distribution, R is used to compute quantiles, to obtain random numbers, to illustrate graphically probabilities and cumulative distributions and to double check the theoretical results. Formulas are kept to a minimum, but the theory is presented rigorously. Mathematics plays an important role in the chapters, but it is simulation and experimentation that illustrates the concepts best. Two appendices contain additional proofs and derivations of results.

Each chapter in the book is full of examples and applications of the methods to computer science; the examples are solved to the smallest level of detail, using theory and R. All of the R code needed is provided in the context of the example. The author teaches students how to use the functions of R, but at the same time shows them how to use the most basic R to build their own functions to do the same routines. Some nice plots illustrating outcomes under different conditions are also given, for example, showing how the reliability changes as the number of components in a system changes (Chapter 124, page 124). The exercises at the end of the chapters build on what has been done in that chapter, and help strengthen both the concepts and the skills in using R to solve problems. R code is not distracting, rather, it is blended very harmoniously within the discussion of the theory.

Unique to this book, and of very high interest to any teacher interested in showing the applications of probability, are: Chapter 8 in Part II dedicated exclusively to reliability; Chapter 14 in Part III dedicated to sampling inspection schemes; Chapter 19 in Part IV dedicated to process control, and Chapter 17 of Part IV on applications of the exponential distribution to reliability and to queueing modeling. Noteworthy too is Chapter 16 on the exponential distribution where a clear connection is made between exponential waiting times and Poisson outcomes. By the time the student reaches those chapters, s/he has already acquired enough skill in R, simulation and the theory needed to work on these nontrivial applications of probability on his/her own. It is rare in probability books to see this hands-on approach to learning the topic. The small and the big picture is always given within the book. Through its interactive, experimental, and simulation-based style, complemented with detailed instructions on how to proceed, the book constantly challenges the reader to experiment on his/her own and to see the concepts come through with R.

Professor Horgan's book could be used to teach any introductory probability course in science, engineering and computer science, with no additional effort needed from the instructor, since the book makes a perfect set of lecture notes and homework exercises. However, if the majority of students are from other fields, say sociology or economics, law or business, the instructor will need to supplement the book with examples from those areas and make an effort to show how queues, reliability, and process control might appear in those other areas. Those teaching a calculus-based course in introductory statistics at the upper-division level would benefit from using this book, too, assuming that the instructor would go on with the teaching of inference also using R. For those instructors teaching more mathematically-oriented introductory probability, the chapters in the book can be used as introductions to the topics, perhaps adding extra notes on the theorems and the mathematical solutions to keep their mathematics-oriented students motivated. All the applied chapters are a perfect set of activities for an applied probability seminar.

Most importantly, a great quality of this book is that it could be used by anybody with a reasonable level of education to self-teach themselves probability and its applications. Those who already know probability but do not know how to use R to do many fancy things, like showing graphically the area of probability under a density curve or simulating random variables from distributions, would benefit immensely from reading this book. It is more user-friendly than the R manuals and, because the code is used within the examples, it teaches R for probability much more effectively.

The only drawback I can see in this book is the omission of the bivariate distributions and expectation results for multiple random variables. Any instructor would like to see a similar book written to teach that topic. But the book is not intended to do that, so this comment

should be taken more as a wish list than a criticism. For the audience this book is intended, the book is the best lecture notes a teacher can give to students and to anybody wishing to self-teach themselves, with that caveat.

As a final note, there are some typos in the book: page 9, in the R code after `names(results)` the command `arch[5]` should be `arch1[5]`; page 87, one of the capital E in the top two formulas should not have a bar; page 290, Example 16.2, X should be T; page 314, 15 lines from the bottom, a λ is missing.

Reviewer:

Juana Sanchez
University of California at Los Angeles
Department of Statitistics
8125 Math Sciences Building, Box 951554
Los Angeles, CA 90095-1554, United States of America
E-mail: jsanchez@stat.ucla.edu
URL: http://www.stat.ucla.edu/~jsanchez/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 29, Book Review 2 *Published:* 2009-01-02
January 2008