Journal of Statistical Software

June 2014, Volume 58, Book Review 1. http://www.jstatsoft.org/

Reviewer: Gary Evans
Purdue University

Analyzing Spatial Models of Choice and Judgment with R

David A. Armstrong III, Ryan Bakker, Royce Carroll, Christopher Hare, Keith T. Poole, Howard Rosenthal
CRC Press, Boca Raton, 2014.
ISBN 978-14665-1715-8. 336 pp. USD 66.45.
http://voteview.com/asmcjr.asp

Owing to its often elegant mathematics and sophisticated graphical output, latent space modeling has become a popular statistical method. From its origins in psychometrics, it has been effectively applied, in one form or another, to areas as diverse as machine learning, network analysis, and general categorical data analysis. Historically, one of its earliest applications outside of psychometrics was in political science, which remains a rich source of modeling innovation. Such applications and innovations are the subject of this excellent book which will take readers on a thoroughgoing tour bringing them right to the cutting edge of spatial modeling as it is used with political choice data and revealing to them the considerable capabilities of R for carrying out such procedures.

The book is aimed at practicing political science researchers and, to use the authors' phrase, "expert [statistical] methodologists." (Readers and instructors who intend to use the book should be aware that it is written at this advanced level.) For the latter audience in particular, the first chapter provides a brief, but lucid discussion on latent space theory as it applies to voting and political choices and, for the former, a well-considered argument for using such models as a standard analytical practice for such data.

The two key ideas underlying the modeling approach, ideological constraint and basic-space theory, are concisely presented. These explain our ingrained tendency to view the wide variety of political issues along a 1-dimensional scale or spectrum and provide the justification for modeling political data in 1 or 2 dimensions, with a second dimension to account for whatever non-ideological factors may be present in a particular vote or choice. Again in the authors' words, the book's "spatial models are not just useful abstractions, but also accurate depictions of decision-making processes." In other words, the methods that will be presented do not just produce interesting data visualizations, but psychologically sound models. That is, the computation of optimum positions in the latent space for voters and choices can be viewed as meaningful parameter estimation. This is a principle that is employed throughout the book in the authors' interpretive analyses of their results.

It should be said that this principle has, over the years, been the subject of some contention, with a school of thought holding that the increased mathematical complexity involved in stochastic distance modeling of voting or choice psychology does not add value to the visualization components of these spatial methods. The authors here are very persuasive, in Chapter 1 and throughout the book, in making the case in favor of such modeling. For this contribution to this important debate alone, and there are certainly others, this is a book that should be read by anyone interested in this subject. Chapter 1 also includes interesting discussions of directional versus proximity voting and of other theoretical political science findings that will enhance the reader's enjoyment and understanding of the later chapters.

Chapter 2 provides a sort of 2-part introduction to R. The first part, around 15 pages, discusses basic R commands while the second, around 10 pages, is on reading and writing data in R. The chapter is well-written but seems out of place (particularly the first section) in a book whose readers are likely to have some working knowledge of such topics as bootstrapping, Bayesian methods, and MCMC (all of which are employed in later chapters) and, therefore, of some fairly sophisticated statistical software. The R code in the modeling demonstrations from Chapter 3 on is well written and described so that such readers, merely working through these examples, are likely to pick up the basics.

Also, most significantly, all R code used in the book appears, in a form that can be directly copied into R, at the book website, `http://voteview.com/asmcjr.asp` (part of the excellent `voteview.com` site which readers of this book will definitely want to explore further). The code is nicely annotated there and just a bit more annotation would easily accomplish what the first part of this chapter does. Also at this website are easily-loaded versions of all datasets used in the text and exercises of the book. They load into R with just a double-click. Thus, the second part of Chapter 2, most useful to researchers whose own data may be in some other software format but not really for readers of book, could be moved to an appendix. Much more useful, perhaps, might be a discussion of the package system in R and a table, with their uses, of the packages discussed in the book and perhaps others involving spatial modeling. In short, most readers of the book can skip this chapter. This is not to be taken as a grave criticism of the material in the chapter but a formal one of the editorial decision to include the chapter in this manner. In such an indispensable work as, in our opinion, this will come to be regarded, such formalities should not be overlooked.

Chapters 3 through 7 present the spatial models. Chapter 3, on analyzing issue scales, is the best (though perhaps it would be better to say it is my favorite since they are all of high quality). We are introduced to Aldrich-McKelvey (A-M) scaling via the **basicspace** package (Poole, Rosenthal, Lewis, Lo, and Carroll 2013) and its `aldmck` function. The procedure is carried out in R on three interesting datasets (French political parties, Vietnam War views, and US Urban Unrest) all of which, as noted above, are available at the `voteview.com` website. An estimation of standard errors in the positions of the rated stimuli (since, again, these are viewed as parameter estimates) using bootstrapping is demonstrated as is a Bayesian approach to A-M scaling. The worked examples in these sections all feature diligently detailed and explained R code, including special functions that are written in R to carry out the analyses. Most notably, the plotting code, in this chapter and throughout the book, is quite sophisticated, a key strength of the work given the importance of effective graphics for analysts using spatial methods.

After A-M scaling, a generalization of the method developed by Professor Poole known as basic space scaling is discussed. Unlike A-M scaling, basic space scaling allows for missing data by

incorporating weighting parameters and allows analysis of multiple issue scales in multiple dimensions (whereas A-M scaling restricts each issue scale to a separate dimension). Basic space scaling is carried out with the `blackbox` (or `blackbox.transpose`) function in **basicspace**. Again, detailed examples are worked in R including a demonstration of bootstrapping to estimate uncertainty in the locations of raters (`blackbox`) and issues (`blackbox.transpose`).

What sets this chapter apart from the others, and from most R instructional works, is the depth and clarity of its mathematical presentation. The authors show how certain psychological aspects of issue scale data are reflected in the mathematics of these modeling methods. Since the source of this data is, typically, the voting public, as opposed to officeholders, it presents difficulties due to its increased subjectivity and to sometimes (or often) poorly informed voters. A-M scaling accounts for these difficulties with a very simple and clever mathematical approach. Related to this issue is a presentation on anchoring vignettes for standardizing scales across rater groups. A very interesting and instructive demonstration of anchoring using the R package **anchors** (Wand, King, and Lau 2011) is carried out on data involving political efficacy rankings from China and Mexico. It is accessible mathematically and gives perhaps the best illustration in the book of how psychological phenomena can be revealed and evaluated with basic modeling techniques.

It is in this chapter that we encounter the first of the exercise sets that the authors include in this work. These are certainly a valuable addition to this type of book. Most involve, at the outset, fairly straightforward adaptations of the code used in the corresponding chapter, but there are some instructive data formatting issues to be dealt with. Also, the plotting code they require raises their level of difficulty somewhat and calls for an increased focus in studying the code examples. Finally, there are some particularly challenging exercises involving bootstrapping and Bayesian methods. With just a small amount of mathematical supplementation, the exercises throughout the book provide ample content for a graduate level course in spatial modeling with R.

Chapters 4 and 5 deal with multidimensional scaling (MDS) and unfolding (MDU), respectively. Keeping in mind the authors' goals in writing this book, one should not expect to encounter here (and one does not) a primer on these techniques such as is found in, say, Borg and Groenen (2010), the unrivaled classic in this field to which readers are referred for background. After a 2-page section on classical MDS, the authors here turn immediately to MDS using simulated annealing and genetic algorithms. Their purpose is thus made clear, and the approach is, it should be said, justified by the aims of this work: to move the discussion quickly to more advanced computational and modeling issues that can be studied in particular by using R.

As such, the **smacof** package of de Leeuw and Mair (2009) plays a central role in both chapters. This implements de Leeuw's SMACOF (scaling by majorization of a complicated function) technique for MDS (`smacofSym` and `smacofIndDIff`) and MDU (`smacofRect`) and is the method of choice for carrying these out in R. The authors very capably demonstrate its use on several datasets and, in Chapter 5, compare `smacofRect` to Poole's MLSMU6 (Poole 1984) procedure (the mathematics for which are well presented) for MDU in a very interesting Monte Carlo study.

Most interesting, it is in these chapters that the authors most directly address the conflicting schools of thought, mentioned above, on stochastics in spatial modeling. In sections on Bayesian MDS and MDU, particularly the latter, they cogently make the case for stochastic

modeling of distance data. Central to the presentation are dealing with the flanking phenomenon which often occurs in the non-stochastic methods and the need for a suitable error theory. As in maximum likelihood MDS, the authors use the lognormal distribution for the distance data and demonstrate how its use in Bayesian MDS can produce not only interval estimations of locations but, more importantly, improved visualizations. This discussion makes these chapters, though not quite as elegant as Chapter 3, well worth very close reading.

Chapter 6 presents the NOMINATE method of Poole and Rosenthal (1985), implemented in R through the **wnominate** (Poole, Lewis, Lo, and Carroll 2011) and **anominate** (Carroll, Lewis, Lo, Poole, and Rosenthal 2013) packages in R. This is the center-piece of the book, as it should be given the prominence of the method for analyzing voting data. The chapter is excellent from start to finish in its examples and R coding, but mostly so in its discussion of **anominate**. This package, not yet available (at the time of writing this review) on the Comprehensive R Archive Network (CRAN) but downloadable from the book website, implements the α-nominate method for determining whether the optimum voter utility function for a dataset is gaussian, quadratic, or some mixture of the two. The presentation is mathematically thorough and gives the reader useful insight into the crucial concept of the utility function and its role in the NOMINATE models.

Professor Poole's optimal classification (OC) method, implemented with the **oc** package (Poole, Lewis, Lo, and Carroll 2012), is also presented here. OC is a non-parametric unfolding technique in which a cut-planes algorithm is applied to a classic MDS configuration of the agreement score matrix to create maximum classification regions for voters. The demonstrations here are especially creative and substantive and feature some truly excellent R coding for analyses and plotting of the results. Two small points here: OC can be compared instructively to de Leeuw (2006) later work on principal component analysis of binary choice data and it would be valuable to have the authors' thoughts on this. Also, OC raises some interesting questions in computational geometry and discussion of these in subsequent editions might enhance the presentation for some readers. These comments do not, however, alter our view that this section on this important method is very engagingly written.

Chapter 7 briefly introduces the reader to some advanced topics in the field including a discussion of the interplay between spatial modeling and item response theory, which also plays a roll in Chapter 6. It is intended to point highly interested readers in promising research directions and there is just enough here to accomplish that.

In conclusion, *Analyzing Spatial Models of Choice and Judgment with R* is the rare R-instructional book that succeeds on three levels. It clearly sets forth the psychological theory underlying its modeling method. It demonstrates how the mathematics used for the modeling provide principles of construction and interpretation consistent with that theory. And, it features very well-presented and sophisticated R code – sophisticated enough to bring novice users of R very far along the path of proficiency and even enough, in some sections, to educate and challenge more advanced users. Students and practitioners interested in this field, or in latent space modeling in general, should consider it essential reading.

References

Borg I, Groenen PJF (2010). *Multidimensional Scaling: Theory and Applications*. 2nd edition. Springer-Verlag, New York.

Carroll R, Lewis J, Lo J, Poole K, Rosenthal H (2013). "The Structure of Utility in Spatial Models of Voting." *American Journal of Political Science*, **57**(4), 1008–1028.

de Leeuw J (2006). "Principal Component Analysis of Binary Data by Iterated Singular Value Decomposition." *Computational Statistics & Data Analysis*, **50**(1), 21–39.

de Leeuw J, Mair P (2009). "Multidimensional Scaling Using Majorization: SMACOF in R." *Journal of Statistical Software*, **31**(3), 1–30. URL http://www.jstatsoft.org/v31/i03/.

Poole K (1984). "Least Squares Metric, Unidimensional Unfolding." *Psychometrika*, **49**(3), 311–323.

Poole K, Lewis J, Lo J, Carroll R (2011). "Scaling Roll Call Votes with **wnominate** in R." *Journal of Statistical Software*, **42**(14), 1–21. URL http://www.jstatsoft.org/v42/i14/.

Poole K, Lewis J, Lo J, Carroll R (2012). ***oc**: OC Roll Call Analysis Software*. R package version 0.93, URL http://CRAN.R-project.org/package=oc.

Poole K, Rosenthal H (1985). "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science*, **29**(2), 357–384.

Poole K, Rosenthal H, Lewis J, Lo J, Carroll R (2013). ***basicspace**: A Package to Recover a Basic Space from Issue Scales*. R package version 0.07, URL http://CRAN.R-project.org/package=basicspace.

Wand J, King G, Lau O (2011). "**anchors**: Software for Anchoring Vignette Data." *Journal of Statistical Software*, **42**(3), 1–25. URL http://www.jstatsoft.org/v42/i03/.

Reviewer:

Gary Evans
Purdue University
Krannert School of Management
West Lafayette, IN 47907
E-mail: evans204@purdue.edu

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 58, Book Review 1 *Published:* 2014-06-10
June 2014