



Journal of Statistical Software

September 2011, Volume 43, Book Review 2.

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

Reviewer: Hakan Demirtas
University of Illinois at Chicago

Statistical Simulation: Power Method Polynomials and Other Transformations

Todd C. Headrick

Chapman & Hall/CRC, Boca Raton, FL, 2010.

ISBN 978-1-4200-6490-2. 174 pp. USD 77.60.

<http://www.crcpress.com/product/isbn/9781420064902>

Headrick's book is a valuable addition to the simulation world in the sense that it is the first systematic book on power method polynomials that have been gaining popularity during recent years in a broad range of research disciplines where multivariate non-normal continuous data generation given cumulants and correlation matrices is needed. Although the focus of the book is on power polynomials, two other general classes (g-h and generalized lambda families) that have been in use for multivariate data simulation are also discussed at relative depth. Ideas that utilize these three frameworks appeared in the literatures of economics, finance, psychometrics, education, among many others, and will undoubtedly continue to attract attention. For this reason, this book has potential to be one of the key sources for researchers who are involved in random number generation and Monte Carlo studies.

The intended audience of the book is statisticians, biostatisticians, and quantitatively sophisticated social scientists and methodologists. Since the power polynomial approach is a very specialized topic, it would not be appropriate to classify it as a textbook. However, it can serve and seems to be designed as a supplemental text in graduate level simulation- and computation-oriented courses. It is assumed that readers have intermediate (and occasionally advanced) knowledge of statistics, calculus, and linear algebra.

There has been a growing interest regarding generalized classes of distributions that span a wide spectrum in terms of symmetry and peakedness behavior. In this respect, power polynomials appear to have been gaining popularity in statistical theory and practice because of its flexibility and ease of execution. It all started with the work of [Fleishman \(1978\)](#) who argued that real-life distributions of variables are typically characterized by their first four moments. He presented a moment-matching procedure that simulates non-normal distributions often used in Monte Carlo studies. It is based on the polynomial transformation, $Y = a + bZ + cZ^2 + dZ^3$, where Z follows a standard normal distribution, and Y is standardized (zero mean and unit variance). The distribution of Y depends on the constants a , b , c , and d , whose values were tabulated for selected values of skewness ($r_1 = E[Y^3]$) and kurtosis ($r_2 = E[Y^4] - 3$) in the original paper ([Fleishman 1978](#)). This procedure of expressing any

given variable by the sum of linear combinations of powers of standard normal variates is capable of covering a wide area in the skewness-elongation plane whose bounds are given by the general expression $r_2 \geq r_1^2 - 2$.

Assuming that $E[Y] = 0$, and $E[Y^2] = 1$, by the first 12 moments of the standard normal distribution, the following set of equations can be derived after simple but tedious algebra:

$$\begin{aligned} a &= -c \\ b^2 + 6bd + 2c^2 + 15d^2 - 1 &= 0 \\ 2c(b^2 + 24bd + 105d^2 + 2) - r_1 &= 0 \\ 24[bd + c^2(1 + b^2 + 28bd) + d^2(12 + 48bd + 141c^2 + 225d^2)] - r_2 &= 0 \end{aligned}$$

Solving these equations can be accomplished by the Newton-Raphson method, or any other plausible root-finding or non-linear optimization routine. Note that this notation is a simpler version of what Headrick presented in his book, the purpose of this simplification is to make it more understandable for people who are less intricately involved with power polynomials. The multivariate extension that is predicated upon computing intermediate correlations among normal variates which are the components of individual variables, was formulated by [Vale and Maurelli \(1983\)](#). The generalizability to the multivariate case makes the polynomial method more compelling in that it presents an advantage over some other general distributions whose multivariate versions are either non-existent or very formidable to specify due to mathematical and/or computational difficulties. Fleishman's method has been extended in other ways in the literature. One extension utilizes the fifth-order polynomials in the spirit of controlling for higher-order moments ([Headrick 2002](#)). It should be noted that the power approach has been criticized on the grounds that the exact distribution was unknown and thus lacked probability density and cumulative distribution functions (PDF and CDF, respectively). However, [Headrick and Kowalchuk \(2007\)](#) derived the power method's PDF and CDF in general form. One particularly significant augmentation is that the polynomial terms do not have to come from the normal distribution, they could also be logistic- and uniform-based, as Headrick beautifully explains in the book.

The organization of the book is as follows. Chapter 1 is a general introduction and includes historical development on power polynomials as well as g-h and generalized lambda families, along with their applications appeared in the literature. In Chapter 2, the general form of the power method's PDF and CDF for normal-, logistic-, and uniform-based polynomials are derived; systems of equations that determine the coefficients for third- and fifth-order polynomials are developed; a procedure to demonstrate the multivariate methodology using a numerical example is presented, with the steps for computing the intermediate correlation matrix. In Chapter 3, a more elaborate and applied illustration of the power methodology is given based on numerous theoretical densities. Comparisons between third- and fifth-order systems and three different polynomials mentioned above are made in terms of goodness of fit on some key statistical quantities, from an applied perspective. In Chapter 4, a procedure for applying the power method in the context of simulating correlated non-normal systems of linear statistical models based on fifth-order power method polynomials is developed. Furthermore, procedures that use intraclass correlation coefficients rather than correlations and that simulate controlled correlation structures between non-normal variates, ranks, and variates within ranks are developed. Chapter 5 is concerned with univariate and multivariate data generation via g-h and generalized lambda densities. It must be emphasized that these trans-

formations can be employed in conjunction with power polynomials for simulating correlated non-normal variables.

Mathematica functions are provided throughout the book, which is a nice feature in the context of applicability of the described procedures and reproducibility of results.

On a slightly negative note, the book is a little too concise, readers who do not have adequate statistical training may face comprehension problems, and may need to resort to published papers on this topic. In addition, end of chapter questions would have been nice. Finally, if the author sets up a web page specifically designed for this book, it would be beneficial for disseminating the ideas articulated in this work.

All in all, given the utmost importance of multivariate continuous data generation across many fields, Headrick's book is a fairly major contribution to the literature in the sense that it is the first methodically written book on power polynomials, which I think was much needed. I would highly recommend this book to anyone who deals with random number generation and more generally with Monte Carlo simulation and statistical computing.

References

- Fleishman AI (1978). "A Method for Simulating Non-Normal Distributions." *Psychometrika*, **43**, 521–532.
- Headrick TC (2002). "Fast Fifth-Order Polynomial Transforms for Generating Univariate and Multivariate Non-Normal Distributions." *Computational Statistics & Data Analysis*, **40**, 685–711.
- Headrick TC, Kowalchuk RK (2007). "The Power Method Transformation: Its Probability Density Function, Distribution Function, and Its Further Use for Fitting Data." *Journal of Statistical Computation and Simulation*, **77**, 229–249.
- Vale CD, Maurelli VA (1983). "Simulating Multivariate Nonnormal Distributions." *Psychometrika*, **48**, 465–471.

Reviewer:

Hakan Demirtas

University of Illinois at Chicago

Division of Epidemiology and Biostatistics

Chicago, 60612, United States of America

E-mail: demirtas@uic.edu

URL: <http://www.healthstats.org/members/hdemirtas.html>

http://www.cade.uic.edu/sphapps/faculty_profile/facultyprofile.asp?i=demirtas

Journal of Statistical Software

published by the American Statistical Association

Volume 43, Book Review 2

September 2011

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

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

Published: 2011-09-11
