

Book Review: An Introduction to Nonparametric Statistics

Kévin Allan Sales Rodrigues*^[0000–0003–4925–5883]

University of São Paulo, São Paulo, Brazil
kevin@usp.br

Book review of **An Introduction to Nonparametric Statistics** by John Edward Kolassa (2020). Chapman & Hall/CRC Press, Boca Raton. 224 pages. Price: USD 105.00 (print), 94.50 (e-book).

The book entitled *An Introduction to Nonparametric Statistics* (Kolassa, 2021) presents an exceptional introduction to nonparametric statistical techniques. To make it easier to read, I divided this book review into two sections: book description and comparisons with other books on nonparametric statistics.

1 Book description

According to the author, the book's target audience is graduate students in the applied statistical field. I find the book also suitable for graduate students researching applied mathematics and theoretical statistics. To make the best use of the book, the reader is expected to have a good knowledge of differential and multivariate integral calculus, matrix algebra, probability (mainly mean and variance of random variables), and classical inferential statistics.

A striking feature of the book is the logical and chronological chain of ideas. I classify the book as logical because it bothers to review the parametric statistics part corresponding to the nonparametric counterpart that will be covered immediately afterward. This logical approach favors learning because many undergraduate students study the parametric statistics before the nonparametric counterparts, and students can promptly relate what they previously learned in parametric statistics to what they study in nonparametric statistics. Thus, the book teaches nonparametric statistics more naturally and logically using the connection with the parametric methods. The first part of each chapter, except for two chapters: Chapter 1 (this is a leveling chapter of the book) and Chapter 8 (this is a chapter more focused on empirical probability density function graphing techniques), provides a brief review of the parametric technique equivalences to the nonparametric techniques that will be presented below.

I say that the book is chronological because it presents contents of nonparametric statistics in the same order as the articles that proposed such nonparametric techniques, favoring the understanding of the contents as the degree of complexity of the techniques increases gradually. For example, the author is constantly concerned with presenting the popular and traditional nonparametric tests: sign test, Mann-Whitney-Wilcoxon test, and Kruskal-Wallis test in chronological order, which are also included in other nonparametric statistics books.

The book has a brief introduction, 10 chapters and 2 appendices. The introduction leaves out something to be desired as to the fact that it does not list all the packages that will be used later in the text. It is good practice to list all packages used in the book, as it is helpful for the reader to prepare the computer environment in advance. It is also convenient to allow a university's computer lab with all the necessary packages already installed on the computers. For accessibility, I list here all the packages needed to run the R (R Core Team, 2022) code in the book: *MultNonParam*, *NonparametricHeuristic*, *devtools*, *BSDA*, *exactRankTests*, *DescTools*, *CvM2SL2Test*, *clinfun*, *muStat*, *crank*, *deming*, *Hotelling*, *ICSNP*, *KernSmooth*, *quantreg*, *MASS*, *boot*, *bootstrap* and *VGAM*¹.

Chapter 1 reviews the probability density functions and properties of random variables used throughout the book as well as the definitions and results of classical statistical inference (confidence intervals and hypothesis testing). Thus, this chapter is essential for the textbook to be self-sufficient, as far as possible, within the area of nonparametric statistics.

The content of nonparametric statistics starts from Chapter 2, which covers nonparametric inferential techniques for a single sample, such as the sign test.

Chapter 3 presents nonparametric inferential techniques for the case of two samples, and the theory behind the other tests and rank tests, is presented. Also, in Chapter 3, the widely used nonparametric tests such as Mann-Whitney and Mann-Whitney-Wilcoxon tests are presented; Mood's median test is presented for didactic purposes.

Chapter 4 extends the results presented to the case where we have three or more groups, explaining in detail the Kruskal-Wallis test. Nonparametric techniques analogous to the ANOVA (Analysis of Variance) of parametric inference are presented in Chapter 5.

Chapter 6 comments on the limitations of interpreting the Pearson correlation coefficient and aims to present alternative correlation measures to the Pearson correlation coefficient, which are the correlation coefficients proposed by Spearman and Kendall.

Chapter 7 presents techniques to perform multivariate nonparametric inference (context where observations contain more than two variables), such as population position parameter vector estimation and hypothesis testing about this same parameter vector.

¹ Note: some packages are already installed along with *R* or need to be installed from GitHub or CRAN.

Chapter 8 is the shortest in the book and presents techniques for estimating a sample's probability density function and plotting empirical probability density functions. The histogram graph is also presented in detail in this chapter.

Alternative regression techniques such as Kernel regression, Local regression, Isotonic regression, Quantile regression, and “Resistant regression” are presented in Chapter 9. When commenting on “Resistant regression”, the author missed the opportunity to present basic results about the classes of M-estimators and even to comment on the existence of other classes of estimators (L-estimators, S-estimators). Knowing the specific terminology would be useful if the reader were interested to learn more about robust estimators. The topic of Splines is also covered in this chapter.

The final chapter introduces two resampling techniques: bootstrap and Jack-knife. Several types of bootstrap are presented, as well as the advantages and disadvantages of each type of bootstrap.

Appendix A presents the codes in *SAS* equivalent to the codes presented in *R* throughout the course. Appendix B contains the codes in *R* used to generate the various tables and graphs presented throughout the text to support the author's argument. Thus, Appendix B makes the material presented in the book reproducible by any reader with basic knowledge of *R*.

Two packages of *R* are used extensively throughout the book, to illustrate the results of different tests and nonparametric techniques or to do data analysis, and they are *MultNonParam* e *NonparametricHeuristic*. The *MultNonParam* package is available on CRAN, and the *NonparametricHeuristic* package is available on GitHub. The book's author created both packages. As the package names suggest, *NonparametricHeuristic* is useful for teaching nonparametric statistics while *MultNonParam* is actually used in data analysis.

2 Comparisons to *Nonparametric Statistical Methods Using R*

An important component of the book is illustrating applications of the techniques to real data through free software *R*. This greatly enriches the text, as the text presents both theory and applications as well as provides the code in *R* and *SAS* so that the reader can carry out their own statistical analyses. Other books also follow this same approach of presenting code for nonparametric data analysis using *R*, such as *Nonparametric Statistical Methods Using R* (Kloke & McKean, 2015). I now compare the book *Introduction to Nonparametric Statistics* (INS) to the book *Nonparametric Statistical Methods Using R* (NSMUR). Below, I list the point-by-point similarities and differences between them.

2.1 Similarities

1. Both books illustrate the theory through applications using *R*.
2. Both books provide the code for the examples.

3. Both books are concise, INS has about 200 pages of content while NSMUR has about 250 pages of content.
4. Both books cover resampling techniques such as bootstrapping.
5. Both books cover the most widespread nonparametric techniques while having additional topics of its own. For instance, INS provides a brief introduction to the concepts of local regression, isotonic regression, and quantile regression. NSMUR has a chapter dedicated to techniques for survival analysis such as Kaplan-Meier estimator and log rank test as well as a chapter dedicated to cluster correlated data analysis.

2.2 Differences

1. Only NSMUR includes a chapter reviewing basic R topics.
2. Only INS presents the codes in *SAS*.
3. Only INS makes a chapter reviewing basic topics of probability and classical statistical inference.
4. NSMUR has more exercises than INS, which is an advantage for the reader who likes to fix learning through exercise solving.
5. Only INS features Jackknife.

The INS book integrates the examples and R codes with the text, making the reading very fluid. Compared to more traditional books like Hettmansperger (1984) and Hettmansperger and McKean (2011), it is notable that INS is more focused on the applied part of nonparametric statistics and offers more comments on R code. In summary, the INS and NSMUR books fulfill the role of introductory books on nonparametric statistics very well.

It would be interesting to see in a possible next edition of the INS an extension of the chapter dealing with alternative regression methods, approaching isotonic regression and quantile regression with a little more theoretical depth.

Acknowledgements

The author gratefully acknowledges CNPq (Brazilian National Council for Scientific and Technological Development) for the financial support by grant #141836/2020-2.

References

- Hettmansperger, T. P. (1984). *Statistical inference based on ranks*. Krieger.
- Hettmansperger, T. P., & McKean, J. W. (2011). *Robust nonparametric statistical methods*. CRC Press.
- Kloke, J., & McKean, J. W. (2015). *Nonparametric statistical methods using r*. CRC Press Boca Raton.
- Kolassa, J. E. (2021). *An introduction to nonparametric statistics*. CRC Press.
- R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org>