

# Preface

Numerical and statistical algorithms are typically confined within a specific programming language. For example, the R open-source data-analysis software uses a specialized scripting language, which is an implementation of the “S” programming language. Many commercial mathematical programs follow this trend. This book is about a platform for statistical calculations using algorithms that are not confined by a chosen language. For example, this platform allows mixing Python and Java numerical libraries, or using them on their own. Or, one can use this book to program statistical code using other languages, such as Groovy, Ruby, and BeanShell. This book is about an approach to scientific programming and visualization that does not set strict requirements on specific programming languages, nor on operating systems where such calculations are performed.

There are many books written about Java—one of the most popular programming languages. There are many books written about Python, which is another very popular programming language. This book explains how to mix them, bringing incredible algorithmic power and cutting-edge numeric libraries to scientific computations and data visualization.

In this book I did not go deep inside particular scientific research area, since the aim was to give concrete examples which illustrate which Java libraries should be used to perform computations. In the cases when I could not cover the subject in detail, a sufficient number of relevant references was given, so the reader can easily find necessary information for each chapter using external sources.

Thus this book presents practical approaches to numerical computations, data analysis, and knowledge discovery, focusing on programming techniques. Each chapter describes the conceptual underpinning for numerical and statistical calculations using Java libraries, covering many aspects from simple multidimensional arrays and histograms to clustering analysis, curve fitting, neural networks, and symbolic calculations. To make the examples as simple as possible from the computational point of view, I fully embrace the scripting approach in the course of this book. This leads to short and clear analysis codes, so you could concentrate on the logic of analysis flow rather than on language-specific details.

This book uses Python as the main programming language, since it is elegant and easy to learn. It is a great language for teaching scientific computation. For developers, this is an ideal language for fast prototyping and debugging. The book discusses how to design code snippets for numeric computation and statistics on the Java platform. To be more exact, we will use Jython (Python implemented in Java), a language that uses not only native Python modules, but can also access very comprehensive Java classes. The reader will learn how to write analysis codes, while numerous code snippets will give you some ideas on numeric algorithms which can easily be incorporated into realistic research application. The book includes more than 300 code snippets to produce data-visualization plots in 2D and 3D.

I am almost convinced myself that this book is self-contained and does not depend on detailed knowledge of computing language, although knowledge of Python and Java is desirable. However, the reader may still need some programming background in order to use this book with other languages, such as Groovy, BeanShell, and Ruby, since I did not give very detailed coverage of these languages.

## Who Is This Book for

This book is intended for general audiences, for those who use computing to make sense of data surrounding us. It can be used as a source of knowledge on data analysis and statistical calculations for students and professionals of all disciplines. This book was written for undergraduate and graduate students, academics, professors, and professionals of any field and any age. The book could be used as a textbook for students.

We also hope that this book will be useful for those who study financial markets, since the numeric algorithms discussed in this book are undoubtedly common to any knowledge discovery research. This book equips readers with the description of a computational platform for statistical calculations which can be viewed as an inexpensive alternative to costly commercial products used by financial-market analysts.

I assume the readers are not familiar with Python/Jython, the main programming language used for code snippets in this book. But some basic understanding of statistics and mathematics would be very helpful to understand the material of this book.

All example codes of this book can easily be transformed to Java, Groovy, Ruby/JRuby, or BeanShell codes. You are presumed to have knowledge of programming in Java, if you will choose the path of moving the examples to Java, or if you will decide to create Java libraries to be deployed as jar files for a new project. The book will discuss how to do this, and a few Java examples will be provided.

Transformations of the example snippets to scripting languages, such as Groovy, Ruby/JRuby, or BeanShell, may require some knowledge of these scripting languages. The good thing is that the analysis algorithms and numerical libraries will be exactly the same, so a little effort is required to move to other languages. Again, we will show you how to convert Jython codes to these languages. In most

cases, our examples should be sufficient to get started with a new language. The more knowledge about Groovy and Ruby/JRuby you can bring, the more you will get out of this book.

## Books You May Read Before

The material of this book is self-contained. However, to understand the material deeper, you may need to look at other sources. First of all, there are plenty of good books [1–5] on Python and Jython, which are more complete for language-specific topics than the information given in this book. If you program in Java that forms the backbone of numerical and graphical libraries discussed in this book, a great deal of supplementary information can be found in Java books [6–10].

Secondly, there are several books on Groovy, a popular scripting language that can be used to work with the Java numerical libraries discussed in this book [11, 12]. If your choice is JRuby, the Ruby programming language on the Java platform, look at these books [13, 14] to get started.

Thirdly, as you read, you may need to look at external sources to understand the material better, especially when we come to statistical interpretations of data. We will supply the reader with the necessary references, so he or she can choose the most appropriate (and affordable) books to discover the world of data analysis and data mining.

## References

1. Pilgrim M (2004), Dive into Python, Apress
2. Guzdial M (2005) Introduction to computing and programming in Python, a multimedia approach, Prentice Hall
3. Martelli A (2006) Python in a nutshell (in a nutshell (O'Reilly)), O'Reilly Media, Inc.
4. Lutz M (2007) Learning Python, 3rd edn. O'Reilly Media, Inc.
5. Langtangen H (2008) Python scripting for computational science. Springer-Verlag, Berlin, Heidelberg
6. Richardson C, Avondolio D, Vitale J et al (2005) Professional Java, JDK 5 Edition, Wrox
7. Arnold K, Gosling J, Holmes D (2005) Java(TM) programming language. In: The Java Series, 4th edn. Addison-Wesley Professional
8. Flanagan D (2005) Java in a nutshell, 5th edn. O'Reilly Media, Inc.
9. Eckel B (2006) Thinking in Java, 4th edn. Prentice Hall PTR
10. Bloch J (2008) Effective Java. In: The Java Series, 2nd edn. Prentice Hall PTR
11. Subramaniam V (2013) Programming Groovy 2: dynamic productivity for the Java Developer, Pragmatic Programmers, LLC
12. King D, Glover A (2013) Groovy in action, Manning Publications, 2007, <https://www.manning.com/books/groovy-in-action>
13. Edelson J, Liu H (2008) JRuby cookbook, O'Reilly Media
14. Bini O (2007) Practical JRuby on Rails Web 2.0 Projects: Bringing Ruby on Rails to Java, Apress

Numeric Computation and Statistical Data Analysis on  
the Java Platform

Chekanov, S.V.

2016, XXVI, 620 p. 92 illus., Hardcover

ISBN: 978-3-319-28529-0