

TUTORIAL: STATISTICAL ANALYSIS OF ALGORITHMS WITH R

THOMAS BARTZ-BEIELSTEIN, COLOGNE UNIVERSITY OF APPLIED SCIENCE
AND
MIKE PREUSS, TECHNICAL UNIVERSITY DORTMUND

1. GOAL

We propose a hands-on tutorial, which is based on and will be held in close cooperation with Mark Wineberg's introductory tutorial "Statistical Analysis for Evolutionary Computation", to be abbreviated as "Stats I" in the following. The proposed tutorial "Statistical Analysis of Algorithms with R"

- (1) exemplifies and extends ideas presented during Stats I
- (2) demonstrates how to analyze results from complex experimental studies, e.g., experimental studies in EC
- (3) gives a comprehensive introduction in the R language
- (4) introduces the powerful GUI "rstudio" (<http://rstudio.org>, see Fig. 1)
- (5) discusses R alternatives such as Excel, matlab and JMP/SAS
- (6) will be held in an interactive manner, i.e., the analyses will be performed in real time.

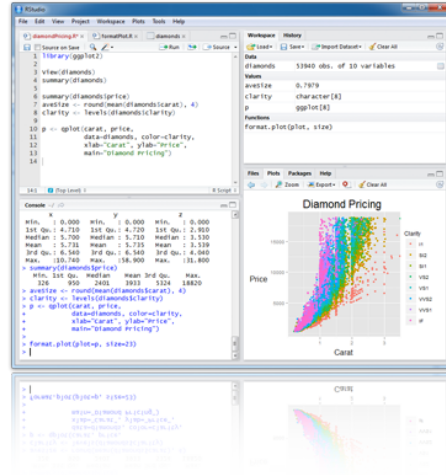


FIGURE 1. Graphical user interface to rstudio

2. CONTENTS

- (1) Interfacing with R
- (2) R basics and technical details
- (3) Case study 1: Analyzing algorithms
- (4) Design of Experiments: How to plan and set up experiments in an efficient manner
- (5) Case study 2: Tuning algorithms interactively
- (6) Some pitfalls in experimentation
- (7) Comparison of R and related approaches, e.g., Excel- or JMP- based analyses
- (8) R based packages for an automated analysis and tuning, e.g., sequential parameter optimization
- (9) Reporting results. Automated report generation using Sweave
- (10) Discussion

3. WHY R?

R is the most attractive and fastest growing open source computer language for statistical computing and graphics in the world. It provides a wide variety of statistical and graphical techniques: linear and nonlinear modeling, statistical tests, time series analysis, classification, clustering, etc. R is distributed over CRAN, which is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. The number of packages on CRAN (<http://cran.r-project.org>) has grown roughly exponentially over the last years [3].

4. PREREQUISITES

The tutorial will be held at an introductory level. However, Stats I attendance is mandatory. R source code examples and further materials will be provided in advance. Participants can perform experiments on their own computer during the workshop.

5. PRESENTERS

Thomas Bartz-Beielstein. Dr. Thomas Bartz-Beielstein is a professor for Applied Mathematics at the Faculty of Computer Science and Engineering Science at CUAS (Cologne University of Applied Sciences—Fachhochschule Köln). He is head of the research center CIOP (Computational Intelligence, Optimization & Data Mining) at CUAS. Thomas has presented several tutorials and workshops devoted to the experimental analysis of algorithms at GECCO, CEC, PPSN, etc. His research interests include optimization, simulation, and statistical analysis of complex real-world problems. Thomas is an R user with more than ten years of experience. He is the driving force behind the "sequential parameter optimization" (SPO) [2], which presents a framework for the statistical analysis of algorithms [1]. The related toolbox is freely available via CRAN, see <http://cran.r-project.org/web/packages/SPOT>. SPOT is applied as a tuner for numerous optimization algorithms such as evolution strategies, genetic programming, differential evolution, or particle swarm optimization. Thomas worked as a research assistant

in the Collaborative Research Center Design and Management of Complex Technical Processes and Systems by Means of Computational Intelligence Methods at the Technical University Dortmund, Chair Algorithm Engineering. He wrote his doctoral thesis "New Experimentalism Applied to Evolutionary Computation" under the supervision of Prof. Hans-Paul Schwefel. His recent book "Experimental methods for the analysis of optimization algorithms" (edited together with Marco Chiarandini, Luis Paquete, and Mike Preuss) is published by Springer (<http://www.springer.com/978-3-642-02537-2>).

REFERENCES

- [1] Thomas Bartz-Beielstein. SPOT: An R package for automatic and interactive tuning of optimization algorithms by sequential parameter optimization. CIOP Technical Report 05/10, Research Center CIOP (Computational Intelligence, Optimization and Data Mining), Cologne University of Applied Science, Faculty of Computer Science and Engineering Science, June 2010. Comments: Related software can be downloaded from <http://cran.r-project.org/web/packages/SPOT/index.html>.
- [2] Thomas Bartz-Beielstein, Christian Lasarczyk, and Mike Preuß. Sequential parameter optimization. In B. McKay et al., editors, *Proceedings 2005 Congress on Evolutionary Computation (CEC'05), Edinburgh, Scotland*, volume 1, pages 773–780, Piscataway NJ, 2005. IEEE Press.
- [3] J. Fox. Aspects of the social organization and trajectory of the R project. *The R Journal*, 1(2):5–13, 2009.