Global envelopes in R

Mari Myllymäki and Tomáš Mrkvička

October 11, 2022

1 Course objectives

Traditionally, one-dimensional or multivariate test statistics have been used to test different hypotheses. Today the handling of the data is more rich, and we can have data that can provide more insight on the studied phenomena if a functional test statistic is used. But this can be quite challenging: one has to take care of the multiple testing problem and changes in the distribution across the domain; often the test statistic is a specific one so its distribution is not known; and often the user requires a graphical interpretation of the testing results so that the interpretation of results is more understandable. Therefore, we have developed a non-parametric method that is based on Monte Carlo simulations or permutations, which handles all the above challenges. The methodology is implemented in the *freely available* R (R Core Team, 2021) package **GET** (Myllymäki and Mrkvička, 2020; Myllymäki *et al.*, 2017). The official release can be found in Comprehensive R Archive Network (CRAN) and its development version via a github repository:

- Official release: https://cran.r-project.org/package=GET

- Development version: https://github.com/myllym/GET

We introduced the methodology first for a general Monte Carlo test with any functional test statistic in the case where the functions are simulated under the null model (Myllymäki *et al.*, 2017).

We have further considered and developed the methodology for general multiple Monte Carlo testing (Mrkvička *et al.*, 2017), functional ANOVA (Mrkvička *et al.*, 2020) and general linear models (Mrkvička *et al.*, 2021; Mrkvička *et al.*, 2021), functional outlier detection and clustering (Dai *et al.*, 2020b,a). The methods have already shown their usefulness in many areas, e.g., spatial statistics (Myllymäki *et al.*, 2017; Mrkvička *et al.*, 2017; Mrkvička *et al.*, 2016; Myllymäki *et al.*, 2020), functional data analysis (Narisetty and Nair, 2016; Mrkvička *et al.*, 2020; Mrkvička *et al.*, 2021; Mrkvička *et al.*, 2021; Racine *et al.*, 2021) and image analysis (Mrkvička *et al.*, 2021) with applications to agriculture, architecture and art, astronomy and astrophysics, ecology and evolution, economics, eye movement research, fisheries, forestry, geography, material science, and medicine, health and neurosciences (see Myllymäki and Mrkvička, 2020, and references therein).

2 Learning outcomes

- The participants will have basic understanding of a Monte Carlo and permutation test based on a single, multivariate and functional test statistic. They know the basic principles of running a global envelope test in R, and have tested it in example situations.

- The participants are aware that the global envelope is a general tool that can be used also for other purposes than testing such as central regions of functional data, functional clustering and confidence and prediction bands.

3 Prerequisites

We expect the participants have some basic knowledge/understanding of statistics and the R software. The course is not intended to discuss very basics of the R code itself. Also the participants should have the newest version of R (https://cran.r-project.org/), and preferably also Rstu-

dio (https://www.rstudio.com/), installed on their computers, if they want to test things themselves. They could additionally also install the GET library in advance by starting R and typing `install.packages('GET')` there.

References

Dai W, Athanasiadis S, , Mrkvička T (2020a). "A New Functional Clustering Method with Combined Dissimilarity Sources and Graphical Interpretation." *IntechOpen*. doi:10.5772/intechopen.100124.

Dai W, Mrkvička T, Sun Y, Genton MG (2020b). "Functional Outlier Detection and Taxonomy by Sequential Transformations." *Computational Statistics & Data Analysis*, **149**, 106960. doi:https://doi.org/10.1016/j.csda.2020.106960.

Mrkvička T, Myllymäki M, Hahn U (2017). "Multiple Monte Carlo Testing, with Applications in Spatial Point Processes." *Statistics and Computing*, **27**(5), 1239–1255. doi:10.1007/s11222-016-9683-9.

Mrkvička T, Roskovec T, Rost M (2021). "A Nonparametric Graphical Tests of Significance in Functional GLM." *Methodology and Computing in Applied Probability*, **23**, 593–612. doi:10.1007/s11009-019-09756-y.

Mrkvička T, Myllymäki M, Jílek M, Hahn U (2020). "A One-Way ANOVA Test for Functional Data with Graphical Interpretation." *Kybernetika*, **56**(3), 432–458. doi:10.14736/kyb-2020-3-0432.

Mrkvička T, Myllymäki M, Kuronen M, Narisetty NN (2021). "New Methods for Multiple Testing in Permutation Inference for the General Linear Model." *Statistics in Medicine*. doi:10.1002/sim.9236.

Mrkvička T, Soubeyrand S, Myllymäki M, Grabarnik P, Hahn U (2016). "Monte Carlo Testing in Spatial Statistics, with Applications to Spatial Residuals." *Spatial Statistics*, **18**, 40–53. doi:10.1016/j.spasta.2016.04.005.

Myllymäki M, Kuronen M, Mrkvička T (2020). "Testing Global and Local Dependence of Point Patterns on Covariates in Parametric Models." *Spatial Statistics*, p. 100436. ISSN 2211-6753. doi:https://doi.org/10.1016/j.spasta.2020.100436.

Myllymäki M, Mrkvička T (2020). "GET: Global Envelopes in R." arXiv:1911.06583 [stat.ME]. URL https://arxiv.org/abs/1906.09004.

Myllymäki M, Mrkvička T, Grabarnik P, Seijo H, Hahn U (2017). "Global Envelope Tests for Spatial Processes." *Journal of the Royal Statistical Society B*, **79**, 381–404. doi:10.1111/rssb.12172.

Narisetty NN, Nair VJ (2016). "Extremal Depth for Functional Data and Applications." *Journal of the American Statistical Association*, **111**(516), 1705–1714. doi:10.1080/01621459.2015.1110033.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Racine E, Coops N, Bégin J, Myllymäki M (2021). "Tree Species, Crown Cover, and Age as Determinants of the Vertical Distribution of Airborne LiDAR Returns." *Trees*, **35**, 1845–1861. doi:https://doi.org/10.1007/s00468-021-02155-2.