Anomaly Detection in Multivariate Profiles with Conformal Bayesian Inference

Nina Deliu NINA.DELIU@UNIROMA1.IT
MEMOTEF, Sapienza University of Rome, Rome, Italy
MRC – Biostatistics Unit, University of Cambridge, Cambridge, UK

Brunero Liseo BRUNERO.LISEO@UNIROMA1.IT
MEMOTEF, Sapienza University of Rome, Rome, Italy

Editor: Simone Vantini, Matteo Fontana, Aldo Solari, Henrik Boström and Lars Carlsson

Abstract

This work addresses the problem of detecting anomalies or abnormal values in multivariate longitudinal data. Specifically, motivated by the international mission of antidoping agencies, we are interested in identifying potential doping abuse in sports by analyzing athletes' individual profiles over time (WADA, 2021). In practice, the goal is to sequentially construct, at each time measurement t and for each athlete, individual reference regions, against which a new observed sample can be compared.

We propose a solution based on conformal predictive inference within a multivariate Bayesian hierarchical framework. To facilitate tractable and accurate estimation of multivariate profiles, we leverage the use of copula models, which allow for separate modeling of the marginal distributions and their dependency structure (Nelsen, 2006). Bayesian principles are employed to estimate the multivariate predictive distribution, and the correspondent multidimensional reference region is derived using a conformal theory approach (Vovk et al., 2005). We show that the resulting region is equivalent to the highest-predictive region when the predictive distribution is used as a conformal measure.

Extensive simulation studies are performed to showcase the advantages of the proposed copula-based solution over other existing univariate and multivariate frameworks, including scenarios of misspecification. Furthermore, this approach offers a flexible alternative with the possibility of integrating any parametric distribution for the marginals and potentially complex dependency structure through different existing copula models. The practical implementation and relevance of this framework are finally demonstrated in real profile data on a control (non-doped) population and on a doped athlete.

Keywords: Anomaly detection, Bayesian hierarchical modeling, Copula models, Highest-density regions, Posterior predictive density, Reference ranges

Acknowledgments

This project has been carried out with the support of the World Anti-Doping Agency and the Partnership for Clean Competition.

References

Roger B. Nelsen. *An introduction to copulas*. Springer Series in Statistics. Springer, New York, 2nd ed edition, 2006. ISBN 9780387286785.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.

WADA. The World Anti-Doping Code. Technical report, World Anti-Doping Agency, 2021. URL https://www.wada-ama.org/en/what-we-do/world-anti-doping-code.