**densityratio: An R-package for density ratio estimation**

Thom Benjamin Volker, Carlos Gonzalez Poses, Erik-Jan van Kesteren

The density ratio (i.e., the ratio of the distribution of two datasets) is a workhorse in many computational social science tasks, such as sample selection bias adjustment, non-parametric two-sample testing, change-point detection and synthetic data utility evaluation. The key advantage of the density ratio in these applications lies in its ability to identify where and how two distributions differ. Over the past years, advanced methods have been developed to accurately estimate the density ratio from two samples. Despite these innovations, tools for density ratio estimation are rather inaccessible, because existing software only implements a narrow range of estimation techniques, is relatively slow, and/or lacks user-friendliness. To make the tools from the density ratio estimation literature available to computational social science researchers and beyond, we present the R-package *densityratio*.

The *densityratio*-package is designed to support novice and advanced users in a wide range of practical situations. It contains a comprehensive suite of methods for density ratio estimation, including novel extensions to deal with high-dimensional data. All methods efficiently estimate the density ratio between two input datasets using non-parametric kernel-based estimation techniques implemented in C++. Automatic hyperparameter tuning through fast, multi-core cross-validation minimizes the need for model specification on behalf of the end-user, allowing researchers to focus on their substantive questions. Densityratio makes it easy for users to not only estimate density ratios, but also to inspect, validate, and extend their functionality; two-sample testing, prediction, and plotting are built-in, allowing researchers to use the estimated density ratio in subsequent tasks and visualize the output of the model.

In the presentation, we demonstrate the *densityratio*-package in two empirical examples in the domain of sample selection bias and two-sample testing. As such, we illustrate how *densityratio* makes density ratio estimation a useful and accessible tool in the toolbox of computational social scientists.

**Literature**

1. Sugiyama, M., Suzuki, T. & Kanamori, T. *Density Ratio Estimation in Machine Learning*. (Cambridge University Press, 2012).

2. Liu, A.-C., Scholtus, S. & De Waal, T. Correcting Selection Bias in Big Data by Pseudo-Weighting. *J. Surv. Stat. Methodol.* **11**, 1181–1203 (2023).

3. Elliott, M. R. & Valliant, R. Inference for Nonprobability Samples. *Stat. Sci.* **32**, 249–264 (2017).

4. Sugiyama, M., Suzuki, T., Itoh, Y., Kanamori, T. & Kimura, M. Least-squares two-sample test. *Neural Netw.* **24**, 735–751 (2011).

5. Wornowizki, M. & Fried, R. Two-sample homogeneity tests based on divergence measures. *Comput. Stat.* **31**, 291–313 (2016).

6. Liu, S., Yamada, M., Collier, N. & Sugiyama, M. Change-point detection in time-series data by relative density-ratio estimation. *Neural Netw.* **43**, 72–83 (2013).

7. Volker, T. B., de Wolf, P.-P. & van Kesteren, E.-J. Assessing the utility of synthetic data: A density ratio perspective. In *UNECE Expert Meeting on Statistical Data Confidentiality* (Wiesbaden, 2023). doi:10.5281/zenodo.8315054.

8. Kanamori, T., Hido, S. & Sugiyama, M. A Least-squares Approach to Direct Importance Estimation. *J Mach Learn Res* **10**, 1391–1445 (2009).

9. Sugiyama, M. *et al.* Direct importance estimation for covariate shift adaptation. *Ann. Inst. Stat. Math.* **60**, 699–746 (2008).

10. Izbicki, R., Lee, A. & Schafer, C. High-Dimensional Density Ratio Estimation with Extensions to Approximate Likelihood Computation. in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics* 420–429 (PMLR, 2014).

11. Makiyama, K. densratio: An R Package for Density Ratio Estimation.

12. DensityRatioEstimation. JuliaML (2024).

13. Gretton, A. *et al.* Covariate Shift by Kernel Mean Matching. in *Dataset Shift in Machine Learning* (eds. Quiñonero-Candela, J., Sugiyama, M., Schwaighofer, A. & Lawrence, N. D.) 131–160 (The MIT Press, 2008). doi:10.7551/mitpress/7921.003.0013.

14. Sugiyama, M. *et al.* Direct density-ratio estimation with dimensionality reduction via least-squares hetero-distributional subspace search. *Neural Netw.* **24**, 183–198 (2011).