The benchden Package: Benchmark Densities for Nonparametric Density Estimation

Henrike Weinert[1,⋆], **Thoralf Mildenberger**[2]

1. Faculty of Statistics, TU Dortmund University, 44221 Dortmund, Germany
2. Department of Mathematics, University of Bayreuth, 95440 Bayreuth, Germany

⋆Contact author: weinert@statistik.tu-dortmund.de

Keywords: nonparametic statistics, density estimation, benchmark densities

This talk gives an introduction to the *R*-package **benchden** which implements a set of 28 example densities for nonparametric density estimation (Mildenberger and Weinert, 2011). In the last decades, nonparametric curve estimation has become an important field of research. To assess the performance of nonparametric methods aside from theoretical analysis often simulation studies are used. Indeed, most published articles suggesting a new method contain a simulation study in which the proposed method is compared to at least a few competitors. The comparison of methods on real-life data sets has the disadvantage that the correct solution is unknown. By using artificial data sets generated under a completely known mechanism one can overcome this disadvantage. Examples of such data sets are the Donoho-Johnstone functions *Blocks, Bumps, Doppler* and *HeaviSine* originally introduced in Donoho and Johnstone (1994). These four functions have well-known features like discontinuities or certain textures that resemble the difficulties encountered in specific applications.

In the case of density estimation there seems to be no generally used set of test densities. Our package **benchden** (Mildenberger et al., 2011) aims at closing this gap. The set of 28 test bed densities first introduced by Berlinet and Devroye (1994) is sufficiently large to cover a wide variety of situations that are of interest for the comparison of different methods. These densities also differ widely in their mathematical properties such as smoothness or tail behaviour and include some densities with infinite peaks that are not square-integrable. They include both densities from standard families of distributions as well as some examples specifically constructed to pose special challenges to estimation procedures.

The package contains the usual *R*-functions for densities that evaluate the density, distribution and quantile functions or generate random variates. In addition, a function which gives some information on special properties of the densities is included, which should be useful in large simulation studies.

References

Berlinet, A. and L. Devroye (1994). A comparison of kernel density estimates. *Publications de l'Institute de Statistique de L'Universite de Paris 38*, 3–59.

Donoho, D. L. and I. Johnstone (1994). Ideal spatial adaption by wavelet shrinkage. *Biometrika 81*, 425–455.

Mildenberger, T. and H. Weinert (2011). The **benchden** package: Benchmark densities for nonparametric density estimation. Discussion Paper 14/2011, SFB 823, Technische Universitt Dortmund.

Mildenberger, T., H. Weinert, and S. Tiemeyer (2011). *benchden: 28 benchmark densities from Berlinet/Devroye (1994)*. R package version 1.0.4.