

## Project Sheet

for a: ( ) Bachelor's thesis ( ) Semester project ( ) Master's thesis (X) CSE Seminar

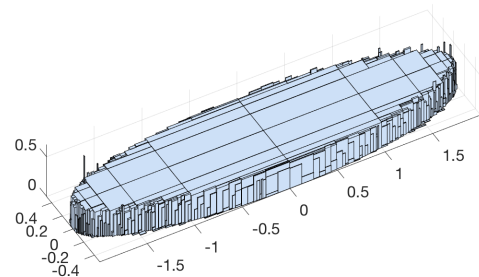
carried out in the semester: FS2019

by: Philip Müller

Subject Parallel Implementation of Density Estimator Using Distribution Element Trees

### Project Description

Recently, we have proposed a new adaptive estimator for probability densities in a potentially high-dimensional probability space [[doi:10.1007/s11222-017-9751-9](https://doi.org/10.1007/s11222-017-9751-9)]. This estimator is based on the decomposition of the probability space into so-called distribution elements (DEs) that are organized in a tree structure and accordingly our estimator is referred to as distribution element tree (DET) estimator. The DET estimator was found to be particularly accurate and computationally efficient for large samples containing many data points. Implementations were made available for the R language on [CRAN](#) and Matlab on the [MathWorks File Exchange](#).



*Figure: Distribution element tree estimate of uniformly distributed data in an ellipse.*

Both implementations are written in their respective scripting languages and thus offer limited computational performance. In this project, the existing implementations shall be complemented by a compiled implementation written in C/C++. To this end, the independent construction of tree branches in the DET shall be exploited by applying a shared-memory parallelism. Second, in order to make the advanced implementation available to a wider community, the C/C++ library shall be augmented by a Python interface and an R wrapper.

In addition to the existing implementations, the new DET library shall provide functionality for gradual data addition/estimator refinement as new data becomes available. This task is facilitated by the localized construction process of the DET. More specifically, a DE is comprised of a probability space cuboid and a locally-defined parametrized density model. During the tree construction process, a statistical test is applied to assess the consistency between the data in a DE and its local density model. Cuboids that fail the test are split into smaller cuboids, while others that pass the test are incorporated in the final estimator. In line with the outlined hierarchical tree construction process, the statistical tests have to be repeated only in cuboids that contain newly available data points. In a last step, the new library shall be documented to a similar extent as the existing R and Matlab codes.

Supervisors: Daniel W. Meyer

Professor: Patrick Jenny

This and accompanying documents handed out on:

01.04.2019

Report is due on:

28.06.2019

Signature Professor: \_\_\_\_\_