Investigating the Statistical Properties of the Double Kernel Density Estimator and its Applicability to a Multivariate Analysis of Real World Health Data

Harold Ship

Advisors: Prof. Boris Portnov, Dr. Itai Dattner, Prof. Em. Benjamin Reiser

In chronic disease epidemiology, health outcomes are often aggregated over geographic areas. However, the level of aggregation required can result in a loss of fine-grained information, including spatial information.

An alternative method, the Double Kernel Density (DKD) can help address some of the drawbacks of aggregation. This non-parametric approach makes use of smoothing of point data over an area of interest without the use of fixed boundaries.

The DKD method has an additional important advantage related to privacy concerns. For reasons of information privacy, information related to the health of individuals cannot be released by health authorities to researchers. The smoothing effect of the DKD hides the location the individual cases thus protecting the privacy of the patients thereby allowing the provision of incident data.

The goal of this research is to examine the double kernel density technique from a statistical perspective. The performance of kernel smoothing is sensitive to a tuning parameter, the bandwidth, which controls the amount of smoothing. We use Monte Carlo simulations to study the empirical properties of the DKD, emphasizing two areas: the overall accuracy of the DKD and the ability of the DKD to estimate peak values and locations while varying the bandwidth.

Preliminary simulation results concerning the accuracy of the DKD estimator appear to be in line with the known theoretical performance of kernel estimators. However, attempts at computing confidence bands have so far not been successful.