



Internship/Final project with Statistics Netherlands (CBS)

Subject: Statistical Disclosure Control when publishing on thematic maps

Introduction

Statistical Disclosure Control is an important aspect of the production of official statistics. Recently, the introduction of the GDPR (General Data Protection Regulation) and its national implementations such as the Dutch AVG (Algemene Verordening Gegevensbescherming) have drawn attention to the protection of personal data. Statistical publications, especially those from national institutes like CBS, should fulfill the requirement that no information can be deduced on individual persons, institutes or organisations from such a publication.

Traditionally, statistical institutes publish their main statistics as tabulated data. Nowadays, more and more publications make use of relatively new visualisation techniques. For some visualisation techniques, traditional disclosure control does not suffice. The goal of this internship is to study and improve disclosure control for a specific kind of visualisation: publishing distribution of variables on a thematic map, that is, as a function of location.

In a previous project, Hut proposed to protect thematic maps by adding Gaussian noise to the Nadaraya-Watson kernel estimator and extended the p-percent rule used to protect tables to maps. An alternative privacy definition was considered in a project by Ter Steege.

Possible research directions

Possible research questions that can be addressed in a project include the following:

- Use adaptive bandwidths to deal with spatial heterogeneities more effectively.
- Take care of anomalies (mass in unpopulated areas, non-sensitive areas that are seen as sensitive because of the smoothing process).
- Consider differential privacy, which has become popular in the statistics literature but is not used for thematic maps in official statistics because of a perceived problem with utility.
- Publications on thematic maps can be linked in the sense that a certain property is plotted on one map and its negation on another. Any statistical disclosure control map should be consistent over these maps.

Further information

This project is suitable for an internship and/or for a final project.

For additional information contact:

prof.dr. M.N.M. van Lieshout, or

dr. ir. J. Goseling.

Related literature

Abramson, I.A. (1982)

On bandwidth variation in kernel estimates. A square root law.
Annals of Statistics 10:1217-1223.

Wolf, P.-P. de, Jonge, E. de (2018)

Safely plotting continuous variables on a map.
In: Domingo-Ferrer, J., Montes, F. (eds)
Privacy in statistical databases. PSD 2018.

Lieshout, M.N.M. van (2019)
Theory of spatial statistics. A concise
introduction. Taylor and Francis/CRC Press.

Hut, D.A., Goseling, J., Lieshout, M.N.M. van, Wolf, P.-P. de,
Jonge, E. de (2020)
Statistical disclosure control when publishing on thematic maps.
In: Domingo-Ferrer, J., Muralidhar, K. (eds.)
Privacy in statistical databases. PSD 2020.
Based on, MSc Thesis, University of Twente. (<http://essay.utwente.nl/82311/>)

Steege, M. ter (2021)
Pufferfish privacy when publishing on thematic maps.
BSc Thesis, University of Twente. (<http://essay.utwente.nl/85628/>)

Lieshout, M.N.M. van
Infill asymptotics for adaptive kernel estimators
of spatial intensity functions.
Australian and New Zealand Journal of Statistics
63:159-181, 2021.

Lopuhaa-Zwakenberg, M., Goseling, J. (2021)
The privacy-utility tradeoff of robust local differential
privacy. ArXiv2101.09139.

Dong, J., Roth, A., Su, W.J. (2021+)
Gaussian differential privacy (with discussion).
Journal of the Royal Statistical Society, Series B, to appear.