DOI: 10.1111/rssb.12460

J. Goseling and M.N.M. van Lieshout's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al.*

J. Goseling[1] | **M.N.M. van Lieshout**[1,2]

[1]University of Twente, Enschede, Netherlands

[2]CWI, Amsterdam, Netherlands

Correspondence: M.N.M. van Lieshout, CWI, Amsterdam, Netherlands. Email: marie-colette.van.lieshout@cwi.nl

We congratulate Professors Dong, Roth and Su on their compelling work on *Gaussian Differential Privacy*.

In official statistics, the p%-rule (Hundepoel et al., 2012) is widely used to protect tabular data. In recent work (Hut et al., 2020) we adapted this concept to thematic maps, for example, of energy consumption per company. Usually such maps are drawn directly from an underlying table that is protected from disclosure. The resulting colour-coded map, however, is, by construction, discretised in regions defined by the cells in the table. These geographic regions are usually large, corresponding, for instance, to municipalities. The resulting protection is very conservative, leading to a map with reduced utility. Therefore, there is a need for smooth thematic maps.

One might use the Nadaraya–Watson kernel weighted average. This procedure, however, is not necessarily safe. Indeed, suppose that an attacker is able to read off the plotted, smoothed, values of the variables of interest at all measurement locations. Then their original values satisfy a linear system which in many cases (including that of a Gaussian kernel) can be solved exactly if the measurement locations are distinct.

To protect sensitive information we propose to add correlated Gaussian noise E with variance τ and map

$$\frac{\sum_{i=1,\cdots,N} g_i \kappa((r-r_i)/h) + E(r)}{\sum_{i=1,\cdots,N} \kappa((r-r_i)/h)}, \quad r \in D.$$

Here the $g_i > 0$ are the values of the variable at distinct locations r_i in a planar region D, κ is the Gaussian kernel and $h > 0$ the bandwidth that determines the amount of smoothing.

The counterpart of the p%-rule is as follows. Let $0 \leq \alpha < 1$. Then a map is *unsafe* if

$$\max_{1=1,\cdots,N} P\left(\left|\frac{\widehat{g}_i - g_i}{g_i}\right| < \frac{p}{100}\right) > \alpha.$$

In words, a map is safe when small relative errors happen with small probability. We proved that if

$$\sqrt{\tau} \geq \frac{p}{100\Phi^{-1}((1+\alpha)/2)} \max_{i=1,\cdots,N} \left\{ \frac{g_i}{\sqrt{(K_h^{-1})_{ii}}} \right\},$$

where $K_h = (\kappa((r_i - r_j)/h))_{i,j=1,\cdots N}$, the resulting thematic map is safe.

REFERENCES

Hundepoel, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Schulte Nordholt, E., Spicer, K. et al. (2012) *Statistical disclosure control*. Hoboken: Wiley.

Hut, D., Goseling, J., van Lieshout, M.C., de Wolf, P.P. & de Jonge, E. (2020) Statistical disclosure control when publishing on thematic maps. *LNCS*, 12276, 195–205.

How to cite this article: Goseling J, van Lieshout M.N.M. J. Goseling and M.N.M. van Lieshout's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al.* *J R Stat Soc Series B*. 2022;84:3–54. https://doi.org/10.1111/rssb.12460

DOI: 10.1111/rssb.12461

Jorge Mateu's contribution to the Discussion of 'Gaussian Differential Privacy' by Dong *et al.*

Jorge Mateu

Department of Mathematics, University Jaume I, Castellón, Spain

Correspondence
Jorge Mateu, Department of Mathematics, University Jaume I, E-12071, Castellón, Spain.
Email: mateu@uji.es

The authors are to be congratulated on a valuable and thought-provoking contribution motivating this new framework for private data analysis, the f-differential privacy. A key aspect is the use of trade-off functions of hypothesis testing as a measure of indistinguishability of two or group neighbouring data sets.