University of Cologne

University of Cologne • Albertus-Magnus-Platz • D- 50923 Köln

Faculty of Mathematics and Physics
Prof.RN Dr. Zdenek Dolezal
Charles University
Ke Karlovu 2027/3
12116 Praha 2, Czech Republic

Faculty of Management, Economics and Social Sciences

Institute for Econometrics and Statistics

Prof. Dr. Karl Mosler

Phone +49 221 470-2982
Fax +49 221 470-5074.
mosler@statistik.uni-koeln.de
http://www.uni-koeln.de/wiso-fak/wisostatsem/

Cologne,
April 3, 2023

Habilitation of Dr. Stanislav Nagy

Spectabilis, dear Professor Dolezal,

thank you for your letter of March 21. Since some nine years I know Dr. Nagy well from conferences and particularly from his papers. I'm glad to inform you about the scholarly and scientific achievements of his habilitation thesis. Please find here my remarks on it.

Multivariate data arise in virtually every field of statistical application. To analyse them without presupposing undue parametric models, depth statistics have been constructed and successfully applied in recent years. A depth statistic measures the centrality of a given point in *d*-space with respect to a given probability distribution. It gives rise to central regions (the upper sets oft the depth) that describe the distribution regarding its location, spread and shape. The oldest concept, and still most discussed one, is the halfspace depth (HD), due to Tukey (1975) and Donoho/Gasko (1992), which being based on probabilities of halfspaces builds on the linear structure of *d*-space in a natural way. But still, nearly fifty decades after, fundamental theoretical and practical problems of the halfspace depth are unsettled.

Visiting Address:
Universitätsstr. 22a, Room 4.324
D- 50937 Köln

With his habilitation thesis, Stanislav Nagy provides a seminal input to their solution. With seven already published articles and an elaborate summary of them he particularly contributes to five issues: When does a halfspace depth function completely characterize the underlying distribution (Articles (A), (B), (C))? If yes, how can the distribution be calculated from its halfspace depth (Article (D))? How can a depth function be extended to a function that vanishes nowhere even on an empirical distribution (Article (E))? How fast converges the random halfspace depth, which is a standard approximation using randomly chosen directions, to the true value (Article (F))? How is he relative performance of several other approximative procedures (Article (G)).

As the thesis includes short as well as long abstracts of all seven articles, I will only touch on some of the contents.

As a visual concept of data analysis, halfspace depth naturally asks for the use of geometric methods. Many authors have already advanced depth statistics with sophisticated instruments of convex analysis. However, Stanislav is the first who draws on results from General Geometry by diving deeply into classical concepts, specifically the long standing theory of floating bodies. By this, he succeeds in finding a new sufficient condition for unique characterization generalizing the so far strongest result of Kong/Zuo (2010), which says that the smoothness of boundaries of all HD central regions is sufficient for characterization. The new more general condition comprises the existence of all Dupin floating bodies. In addition, by exploiting the relationship between Dupin floating bodies. and HD central regions, Stanislav achieves a bulk of other interesting results, which are new either in Statistics or in Geometry.

If a probability distribution consists of finitely many atoms, its HD function obviously attains only a finite number of values. The reverse is also true, as has been proved in another paper by Stanislav. Moreover, he copes with the problem of reconstructing a probability from a depth function having finitely many values. Doing this, he simplifies the classic

Struyf/Rousseeuw (1998) approach and provides a much more efficient algorithm. This again is an example of a clever view on the geometry of the HD central regions; here in the finitely-valued case they are polytopes. Reconstruction in this seemingly simple case is not trivial at all as non-general positioned data and ties have to be taken into account. However, beyond finitely discrete probabilities and parametric classes of distributions (e.g. elliptical), the question of characterization is essentially answered in the negative, as the above sufficient conditions appear to be of limited use.

Many depth notions, including the halfspace depth, vanish outside the support of the distribution. Especially when employing depth in supervised classification, this yields tot he so called outsider problem, meaning that data outside the convex hull of an empirical distribution cannot be classified. Stanislav constructs a simple extension of depth notions, illumination depth, which vanishs nowhere in *d*-space and has good theoretical and computational properties.

To calculate the HD with respect to a given empirical distribution is an NP-hard computational problem. Many authors have proposed algorithms for this, some of which cleverly exploit the geometrical structure of the data. The simplest approximate approach, the so called random halfspace depth by Cuesta-Albertos/Nieto-Reyes (2008), consists in randomly choosing a finite number of halfspace directions and taking the minimum of univate depths in these directions. Here the uniform convergence random halfspace depth and its speed are investigated. It comes out that convergence of the random halfspace depth is too slow to be useful as an approximation. Finally, a number of more sophisticated approximations of HD is compared in a careful simulation study. Eight approaches are adapted to the case of directions (in the unit sphere) and programmed (to be implemented in an R-package), and their relative performance evaluated.

What's the bottom line of this thesis? After all these deep and imaginative investigations into its theory and computability, beyond empirical or finitely discrete distributions, the picture of halfspace depth remains mixed. The thesis greatly contributes to clarifying the potentials and, even more, the limitations of using the halfspace depth in statistical analysis. Other depth notions, some also having good theoretical properties, may prove more useful in practical applications.

The thesis contains some results that carry over to other depths (e.g. having the projection property). Also, the idea of illumination depth can be combined with many depth notions (e.g. the zonoid depth).

With this thesis Stanislav Nagy shows an impressive scholarship. He is a great reader, not only of literature in Statistics and Probability, but as well in more remote mathematical disciplines like General Geometry. He meticulously traces results back to their original publication and detects flaws and gaps in published proofs. He identifies problems and develops ideas to solve them, in theory as well as in computation and the implementation of algorithms. Beyond this thesis, as a PostDoc Stanislav has published many more results on different themes. In fact he has provided important contributions to statistical theory and methodology.

Needless to say, the results of this thesis are own results obtained by Stanislav with the specified coautors, which is also confirmed by the output of the Turnitin program to detect plagiarism.

Dear Colleague, I hope that these remarks will be of some value for the habilitation procedure. If there remain any questions, please don't hesitate to ask me.

With kind regards,