SCALING OF EMPIRICAL COMPARTMENTAL DISTRIBUTIONS

Justas Kvedaravicius[1], dr. Aleksejus Kononovicius[2]

[1]Faculty of Physics, Vilnius university, Lithuania,
justas.kvedaravicius@ff.stud.vu.lt

[2]Institute of Theoretical Physics and Astronomy, Faculty of Physics, Vilnius university, Lithuania

Introduction

Opinion dynamics is a focal topic in sociophysics. Yet some authors note that there's a lack of comparison between voter model's theoretical results and empirical (experimental) data [1]. To overcome this issue some authors have suggested to include space (geography) in modelling opinion dynamics [1, 2]. To understand the implications of space to opinion dynamics, it's worth studying how empirical data can entail structures of opinion or such indicators. This research analyzes changes in empirical compartmental distributions by changing the scale of empirical data

United Kingdom 2011 census data

Empirical data from United Kingdom 2011 census have been chosen as it is available at various scales. Different demographic indicators of this data exhibit narrowing distributions as scale becomes less detailed (see Fig. 1). Yet it is unclear whether distributions stem from existing socio-demographic structures in United Kingdom or are of spatial nature.

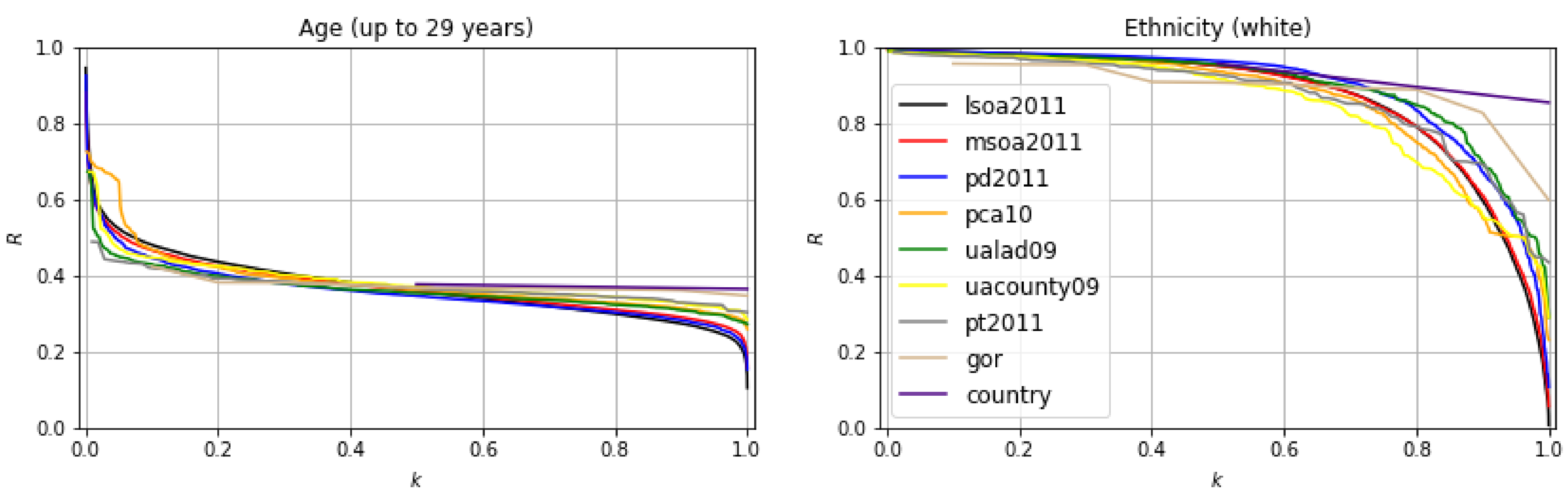

Fig. 1. Rank-size distributions of 2 of the selected demographic indices at various scales (scale abbreviations as per census notation). R marks the part of population sharing the demographic indicator, k is their relative rank.

Randomized spatial model

A randomized spatial model is proposed to discover the nature of observed distributions. It takes compartments (distinct territorial units) from the most detailed scale of empirical data and joins them to replicate the number of compartments of a less detailed scale. Thus the compartments of the new scale are not consistent with geography but rather randomized. The said procedure is repeated for all non-first data scales. Then two diversity quantities are chosen for which relative changes in their value are calculated. Changes of these quantities for randomized spatial model are compared to the ones produced by empirical data. Overall, the procedure used in this model resemble what is well-known in statistics as null-hypothesis testing.

Selected diversity measures

A total of 4 statistical variables have been chosen to evaluate changes in distributions. These are namely range, standard deviation, Gini coefficient, Theil index. Two of those variables, Gini and Theil, are often used to measure inequality (i.e. in finances distribution). Relative changes in their values for coarsening scales are presented in Fig. 2. Two variables, standard deviation and Theil, are further used to compare results from proposed model and empirical data.

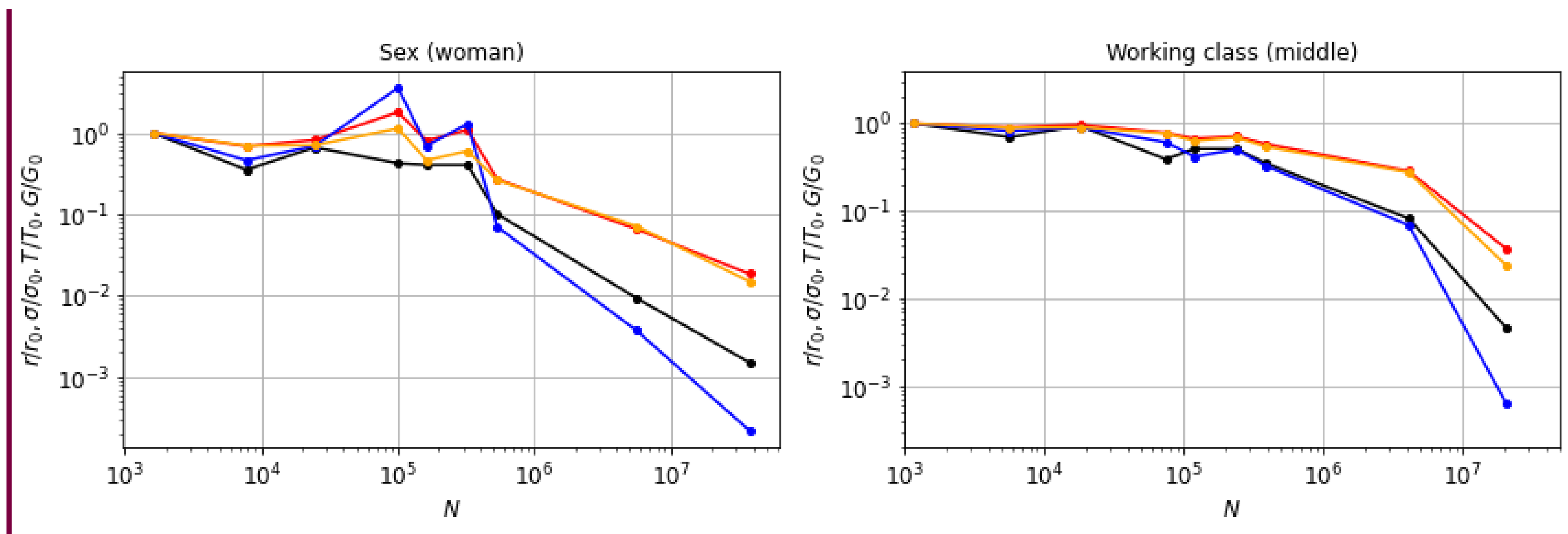

Fig. 2. Relative changes in diversity variables for changing scales. Data is shown for selected demographic indicators. Y axis represents changes in diversity variables relative values, N is mean compartment population (increases in size for more robust scales).

Results

Results of the randomized spatial model with the ones produced by empirical data are shown in Fig. 3. It is shown that relative changes of diversity indices in random model decline more rapidly than the ones of empirical data. Furthermore, we show by analytical derivation that changes of proposed model follow the $N^{-1/2}$ power-law for standard deviation whilst for Theil index these correspond to N^{-1} (see Fig. 3) These results also demonstrate that the changes in selected demographic indicators are not of stochastic (random) nature as they decay noticeably slower.

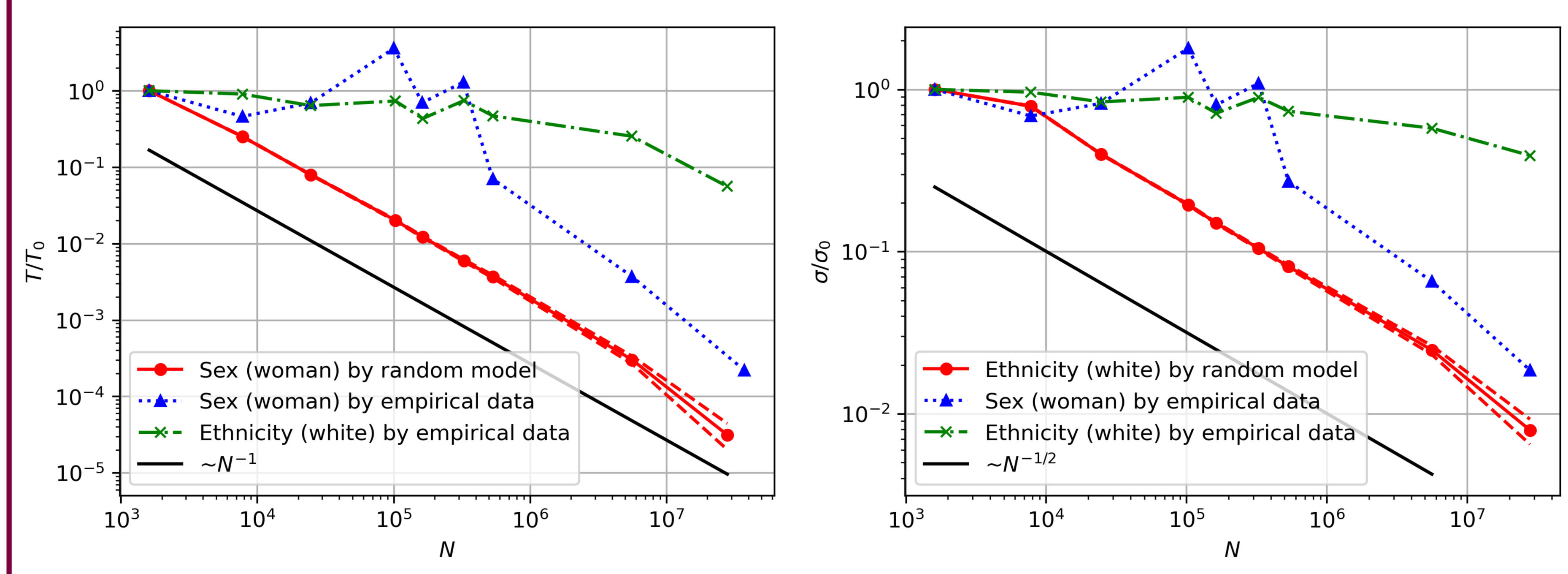

Fig. 3. Relative changes in Theil (left) and standard deviation (right) for empirical data and Randomized spatial model. Y axis represents their relative values, N is mean compartment population (see Fig. 2). See legend for corresponding power-laws.

Among other results, it has also been noted that chosen diversity indices describe changes in different quality (see Fig. 2). Other two quantities (range and Gini coefficient) have also been tested for results with empirical data. Their relative changes are somewhat similar to the previously mentioned ones: range to Theil index and Gini coefficient to standard deviation.

Implications

In conclusion, the proposed randomized spatial model provides researchers with easy yet delicate measure to assess whether or not empirical data (or data generated by spatial models) is of stochastic nature. Different diversity and inequality quantities may be chosen for data evaluation as they are bound only to calculability via compartments. Lastly, the model is rather flexible so further adaptations can prove beneficial in testing other hypothesis.

References

[1] J. Fernandez-Garcia, K. Suchecki, J. J. Ramasco, M. Miguel, V. Eguiluz, *Is the voter model a model for voters?*, Physical Review Letters 112, 158701 (2014)

[2] A. Kononovicius, *Compartmental voter model*, Journal of Statistical Mechanics: Theory and Experiment 2019, 103402 (2019).

Source of the top image: datashine.org.uk