Thesis topic (B.Sc. or M.Sc.)

A count data regression model for arbitrary count data distributions with connected support

Regression models explicitly designed for count data have a long tradition including Poisson regression and other log-linear models like negative binomial regression. However, relatively simple parametric models for count data fail to capture the intricacies of empirically observed count distributions. While truncation or the inclusion of zero-inflation addresses some irregular count distribution, very flexible models like those presented by Huang (2013, 2014) are the exception.

In the context of item response theory current work at the group for Statistical Methods in the Social Sciences has led to the Sequential Item Responce Count Model (SIRCM), a flexible model for count data. This model generalizes the idea of Tutz' (1990, 1991) sequential model for ordinal data with the help of Eilers and Marx P-splines (e.g., Eilers & Marx, 2010).

The thesis will explore a regression model that is derived from the SIRCM. Depending on the interest and prior qualification of the candidate, the thesis could contribute to the exact model specification, the model's estimation, simulation work, applications or theory (especially the relationship to generalized linear models is interesting). R code for the SIRCM exists and the group has ample experience with count data modelling. Prior knowledge on generalized linear models is a plus but not a requirement.

This thesis will be supervised by Prof. Dr. Philipp Doebler.

Contact: doebler@statistik.tu-dortmund

Date: 26.5.2022

References

Eilers, P. H., & Marx, B. D. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(6), 637-653.

Huang, A. (2013). Density estimation and nonparametric inferences using maximum likelihood weighted kernels. *Journal of Nonparametric Statistics*, *25*(3), 561-571.

Huang, A. (2014). Joint estimation of the mean and error distribution in generalized linear models. *Journal of the American Statistical Association*, *109*(505), 186-196.

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*, *43*(1), 39-55.

Tutz, G. (1991). Sequential models in categorical regression. *Computational Statistics & Data Analysis*, *11*(3), 275-295.