Harald Baayen

Using quantile regression and dynamic survival analysis to study the time course of the lexical processing of complex words

This talk draws attention to two statistical methods that make it possible to assess whether the effect of predictors on a response variable vary within the distribution of that response. Dynamic survival analysis is applicable to durational responses such as reaction times, fixation durations, and acoustic durations. Quantile regression can be applied to any kind of measurement, not only durations but also tongue positions or the amplitude of the brain's electrophysiological response to some stimulus.

Schmidtke et al. (2017) used nonparametric survival analysis (Reinhold & Sheridan, 2014) to show that whole-word frequency effects emerge earlier in the distribution of reaction times in visual lexical decision and eye movement fixation durations compared to constituent frequency effects. However, in the vast statistical literature on survival analysis, parametric methods are available (see Scheike and Martinussen, 2007) that enable the analyst to take into account the different causes of exit times: in lexical decision, an exit time can be due to a word or a non-word decision, and for fixation duration, a saccade can be initiated to either a position within the word or to a position elsewhere. Dynamic survival analysis applied to auditory lexical decisions to English compounds revealed early effects of compound frequency and late effects of modifier frequency, replicating Schmidtke et al. (2017). However, the competing risks setting of dynamic survival analysis enables a further analysis of the nonword responses to words. This analysis revealed that such error responses arise due to the intrusion of the modifier.

Quantile regression (Koenker, 2005) is a regression technique that allows the analyst to move beyond predicting the mean. How predictors work together can be scrutinized not only for the median, but also for deciles such as 0.1, or 0.9, or any other quantile of interest. The qgam package (Fasiolo et al. 2017) integrates quantile regression with the generalized additive model, and thus makes it possible to study how nonlinear trends change across the distribution. Quantile gams applied to experimental data on morphological processing lead to the same conclusions as dynamic survival analysis. Both argue against decompositional theories of morphology, and fit well with the discriminative perspective on lexical processing (Milin et al., 2017) as well as with Word and Paradigm morphology (Blevins, 2016).

References

Blevins, J. P. (2016). Word and paradigm morphology. Oxford: Oxford University Press.

Fasiolo, M., Goude, Y., Nedellec, R., and Wood, S. (2017). Fast calibrated additive quantile regression. Manuscript, University of Bristol.

Koenker, R. (2005). Quantile regression. Number 38. Cambridge university press.

Milin, Feldman, Ramscar, Hendrix, and Baayen (2017). Discrimination in lexical decision. PLOS-One, 12 (2), e0171935.

Reingold, E. M. and Sheridan, H. (2014). Estimating the divergence point: A novel distributional analysis procedure for determining the onset of the influence of experimental variables. Frontiers in Psychology, 5. http://dx.doi.org/10.3389/fpsyg.2014.01432.

Scheike, T. H. and Martinussen, T. (2007). Dynamic Regression models for survival data. Springer, New York.

Scheike, T. H. and Zhang, M.-J. (2008). Flexible competing risks regression modeling and goodness- of-fit. Lifetime Data Analysis, 14(4):464-483.

Scheike, T. H. and Zhang, M.-J. (2011). Analyzing competing risk data using the R timereg package. Journal of Statistical Software, 38(2):1-15.

Schmidtke, D., Matsuki, K., and Kuperman, V. (2017). Surviving blind decomposition: a distributional analysis of the time course of complex word recognition. Journal of Experimental Psychology: Learning, Memory and Cognition.

Wood, S. N. (2006). Generalized Additive Models. Chapman & Hall/CRC, New York.