II Joint Conference of the Belgian, Royal Spanish and Luxembourg Mathematical Societies
Logroño, June 6–8, 2016

Targeted penalized regression for cost effective causal effect estimation

Machteld Varewyck[1], **Stijn Vansteelandt**[1], **Marie Eriksson**[2] **Els Goetghebeur**[1]

In the absence of unmeasured confounders, standardised risks derived from confounder adjusted regression of outcome on exposure can form the basis of causal effect estimators. The method is in principle simple and beats the more sophisticated double robust estimators in specific settings. Both can be applied for the measurement of quality of care, e.g. to evaluate the added effect of hospital choice on patient survival. In practice, even the simple regression based measure meets substantial statistical and numerical challenges (cf. [2]) when a large number of centres is evaluated in the single model. Because of overfitting the resulting estimator can face convergence and stability problems, and it may suffer substantial finite sample bias and large variance. Normal random effects models have been applied to overcome this, but loose substantial power for the detection of poor performance in small centres. We show how Firth's correction (cf. [1]), originally developed as an asymptotic first order bias correction for maximum likelihood estimators, strikes a balance between fixed and normal random effects regression that gives enough prior weight to outlying risks to recover much needed power (cf. [3]). Moving on from this we develop an adapted lasso penalisation for model selection that accounts for the cost of measuring specific covariates. In our setting it involves a cross validation method that targets causal effects which are not directly observed. We develop and apply the approach to estimate quality of care based on the Swedish RiksStroke register.

References

[1] D. Firth, Bias reduction of maximum likelihood estimates. *Biometrika* **80** (1993) 27-38.

[2] D. Fouskakis and D. Draper, Comparing stochastic optimization methods for variable selection in binary outcome prediction, with application to health policy. *Journal of the American Statistical Association* **103** (2008) 1367–1381.

[3] M. Varewyck, E. Goetghebeur, M. Eriksson and S. Vansteelandt, On shrinkage and model extrapolation in the evaluation of clinical center performance. *Biostatistics* **15** (2014) 651–664.

[1]Department of Applied Mathematics, Computer Science and Statistics,
Ghent University, Krijgslaan 281-S9, 9000 Ghent,
Belgium.
els.goetghebeur@ugent.be

[2]Department of Statistics, Umea University,
Sweden
marie.eriksson@umu.se