Single-sample predictive model stability assessment via variance components estimated through re-sampling and cross-validation

Michael P. Jones

Macquarie University, Sydney, Australia

Abstract

Background: Predictive models developed on a training sample are prone to over-optimism. In small samples, perhaps due to studying a rare disease or other outcome, an independent validation sample is not always feasible. Various approaches to single-sample model assessment have been proposed including shrinkage estimators for parameter bias and leave-k-out cross-validation.

Aim: To develop a measure of model predictive stability based on a training sample alone. Method: Ratio of within- and between-subject variance components (σ_w/σ_b) from leave-k-out cross-validated predicted probabilities across bootstrapped binary logistic models is derived as an index of model stability. In a simulation study σ_w/σ_b is compared with differences in model performance metrics between independent training and validation samples (Δ).

Results: The variance component index exhibits good stability across models with varying number of predictors and correlations between predictors. The index varies in both mean and variance across predictors simulated from Normal and Exponential distributions, as expected. Importantly, the variance components index correlates well with the difference in model parameters across independent samples. Conclusion: When independent sample validation is not feasible, useful insight into model stability can be gained from variance components derived from the training sample alone.

Keywords

Predictive model validation, Re-sampling, Variance components in predicted values.

References

Altman, D.G. and Andersen, P.K. (1989). Bootstrap investigation of the stability of a Cox regression model. *Stat. Med. 8*, 771–783.

Browne, M.J. (2000). Cross-validation methods. *J. Math. Psych. 44*, 108–132.

Harrell, F.E., Margolis, P.A., Gove, S., Mason, K.E., Mulholland, E.K., Lehmann, D., Muhe, L., Gatchalian, S., Eichenwald, H.F., and the who/ari young infant multicentre study group (1998). Development of a clinical prediction model for an ordinal outcome. *Stat. Med. 17*, 909–944.

Steyerberg, E.W., Bleeker, S.A., Moll, H.A., Grobbee, D.E., and Moons, K.G.M. (2000). Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J. Clin. Epidemiol. 56*, 441–447.

Steyerberg, E.W., Harrell, F.E., Borsboom, G., Eijkemans, M.J, Vergouwe, Y., and Habbema, J. (2001). Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J. Clin. Epidemiol. 54*, 774–781.