**Comparing the Predictiveness of Nested Statistical Models**

Shulamith T Gross, Baruch College of the City University of NY

This work was done in collaboration with Catherine Huber of University of Paris-5.

Our objective is to compare the predictive capacities of several nested models. The problem derives from the following problem in epidemiology: the occurrence of a certain disease is to be predicted to happen within a fixed period of time based on the values of a number of items measured on the observed patients. It may happen that one or several items, proved to be relevant for the best fitting model, have a non-significant contribution to the prediction of who is at risk of developing the disease. Two indices used to compare the respective predictive ability of the two models are the Integrated Discrimination Improvement (IDI) and the BRier's calibration Improvement (BRI). Estimation of the models and their relative IDI and BRI are conducted on the same sample, and their respective asymptotic properties are proved, permitting model selection by discrimination or calibration capacity. Our contribution may be compared to the use of the Akaike Information Criterion for model selection for fit (minimizing the Kullback-Leibler distance) which is a large sample approximation as well. Results of simulation studies for two different logistic models suggest the accuracy of the new standard error estimates and the normality of the sample indices in moderate and large samples. The special case of null true indices IDI and BRI is also treated by simulation. The unusual asymptotic behavior of the estimated IDI, but not of the estimated BRI, in the case of null true index, also observed by others (Kerr et al Amer J of Epidem 2011), is explained by the concept of linear equivalence introduced in Lai, Gross and Shen (Ann Statist 2011). Extended simulations also confirmed that the nonparametric Bootstrap confidence intervals retained their nominal coverage even in the null IDI case, and can therefore be used for testing the hypotheses of null IDI distance.

We apply model selection for prediction to cohort of Alzheimer patients. One of the genetic factors proved to be relevant for goodness of fit of the model is shown to be irrelevant for prediction purpose. Because this is a case of near null IDI and BRI we followed the direct analysis by a nonparametric Bootstrap which confirmed our finding. This result allows us to consider a larger sample of patients as we can include among the patients under study those for whom this genetic factor is not available. This feature can be extended to all studies where a factor that is missing for a subset of the patients under study is shown to be relevant for goodness of fit while it is irrelevant for prediction purpose. This interpretation remains controversial.