A/Prof. Olivier Thas
Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University,
Belgium, and
NIASRA, University of Wollongong, Australia
http://biomath.ugent.be/~othas/homepage_of_Olivier_Thas.html

Title:

Estimation in the Probabilistic Index Model

Abstract:
The Probabilistic Index Model (PIM) was introduced by Thas et al., 2012. It is a flexible class of semiparametric models that can be used to generate many classical rank tests, such as the Wilcoxon-Mann-Whitney, Kruskal-Wallis and Friedman tests, among others; see DeNeve and Thas (2015).

The PIM models the conditional probability P(Y, Y' ;X,X*), where Y and Y' are independent random outcomes associated with covariates X and X*, respectively. With g a link function, a PIM is often of the form g(P(Y,Y' ;X,X*)) = β_0 + β^t(X*- X). Thas et al. (2012) proposed an estimator for β which is consistent and asymptotically normal under mild conditions. However, no semiparametric efficiency was proven and their simulation results indicate that the convergence to the asymptotic normal distributions is too slow for the method to be recommended for use in small samples.

In this presentation we will present some recent results on improved estimation methods. First, we propose semiparametric efficient estimators, and next we show how empirical likelihood methods can be employed to result in well-behaved inference in small samples.

We derive the class of all consistent and asymptotically normal estimators for β in the semiparametric PIM by appealing to the theory of semiparametrics (Tsiatis, 2007) and identify the efficient influence function for β. Next, we propose estimating equations to solve the efficient influence function relying on the theory of semiparametric two-step estimators. The efficient estimator is computationally more demanding than the original estimator. In addition, we work out estimators that statistically improve the latter while retaining their computationally attractive properties. These improved estimators are chosen in such a way that they decrease the second-order finite-sample bias as compared to the original estimator of Thas et al. (2012).

We conclude that the semiparametric efficient estimator, which is computationally more intensive than the original estimator, does not result in much larger efficiencies in small to moderately large datasets. The biased reduced estimator seems to be a good compromise.

We explore methods that are designed to give better small sample results. Resampling techniques, such as the bootstrap and jackknife, are often used as alternative approaches to increase accuracy in many statistical applications.

However, they sometimes require strong computational power. We solve this issue by applying the bootstrapping U-statistics method of Jiang and Kalbeisch (2012). In addition to bootstrap, we also use methods based on empirical likelihood to improve small sample inference for probabilistic index models. In particular, we adapt the empirical likelihood methods of Jing et al. (2009) and Chen et al. (2008).

Our simulation results demonstrate that our empirical likelihood methods work well for samples with sizes as small as 20.