

# The EBIC and a sequential procedure for feature selection in interactive linear models with high-dimensional data

Yawei He · Zehua Chen

Received: 14 December 2013 / Revised: 8 August 2014 / Published online: 3 December 2014  
© The Institute of Statistical Mathematics, Tokyo 2014

**Abstract** High-dimensional data arises in many important scientific fields. The analysis of high-dimensional data poses great challenges to statisticians. In high-dimensional data, the relationship among the variables is complex. It involves main effects as well as interaction effects of the covariates. The effect of some covariates is only realized through their interaction with the others. This makes the consideration of interactive models imperative in the analysis of high-dimensional data. Because of the existence of high spurious correlation among the covariates in high-dimensional data, conventional tools for dealing with interactive models become inappropriate. In this paper, we develop specific tools for feature selection in high-dimensional data with interactive models, including a version of the extended BIC (EBIC) for interactive models and a sequential feature selection procedure. Main-effect and interaction features are treated differently in the EBIC for interactive models and the sequential procedure due to their different natures. The selection consistency of the EBIC for interactive models and the sequential procedure is established. Simulation studies are carried out to vindicate the asymptotic property in finite samples as well as to compare with non-sequential procedures. The approach developed in this paper is also applied to a real data set.

**Keywords** High-dimensional data · EBIC · Feature selection · Interactive model · Sequential procedure · Selection consistency