A generalized information criterion for high-dimensional PCA rank selection

<u>洪弘</u>¹, 陳素雲², 銀慶剛³

1台灣大學,2中央研究院,3清華大學

Abstract

Principal component analysis (PCA) is a commonly used statistical procedure for dimension reduction. An important issue for PCA is to determine the rank, which is the number of dominant eigenvalues of the covariance matrix. Among informationbased criteria, Akaike information criterion (AIC) and Bayesian information criterion (BIC) are two most common ones. Both use the number of free parameters for assessing model complexity, which may suffer the problem of model misspecification. To alleviate this difficulty, we propose using the generalized information criterion (GIC) for PCA rank selection. The resulting GIC model complexity takes into account the sizes of eigenvalues and, hence, is more robust to model misspecification. The asymptotic properties and selection consistency of GIC are derived under the high-dimensional setting. Compared to AIC and BIC, the proposed GIC is better capable than AIC in excluding noise eigenvalues, and is more sensitive than BIC in detecting signal eigenvalues. Moreover, we discuss an application of GIC to selecting the number of factors for factor analysis. Our numerical study reveals that GIC compares favorably to the methods based on (deterministic) parallel analysis.

Keyword: GIC, high-dimensionality, model selection under misspecification, parallel analysis, PCA