Statistica Sinica **37** (2027), 1-19
doi:https://doi.org/10.5705/ss.202023.0385

UNIVERSALLY CONSISTENT TESTS FOR THE GRAPH OF A GAUSSIAN GRAPHICAL MODEL

Thien-Minh Le, Ping-Shou Zhong* and Chenlei Leng

University of Tennessee at Chattanooga, University of Illinois Chicago and University of Warwick

Abstract: The Gaussian graphical model is routinely employed to model the joint distribution of multiple random variables. The graph it induces is not only useful for describing the relationship between these variables but also critical for improving statistical estimation precision. In high-dimensional data analysis, despite abundant literature on estimating this graph structure, tests for the adequacy of its specification at a global level are severely underdeveloped. To make progress, this paper proposes novel goodness-of-fit tests that are computationally easy and theoretically tractable. The first contribution of this paper is the development of a new direct plug-in test statistic. We show that its asymptotic distribution under the null follows a Gumbel distribution with a location parameter depending on the underlying true graph structure. The direct test, however, has no power for detecting structures including the truth but not equal. Our second contribution is the development of a novel consistency-empowered test statistic that gains power by, interestingly, amplifying the noise incurred in estimation. The improved test is shown to be universally consistent for all fixed alternatives. Extensive simulation illustrates that the proposed test procedures have the right size under the null, and is powerful under alternatives. As an application, we apply the tests to the analysis of a COVID-19 data set, demonstrating that our test can serve as a valuable tool in choosing a graph structure to improve estimation efficiency.

Key words and phrases: Dependence, Gaussian graphical model, goodness-of-fit test, Gumbel distribution, high-dimensional data.

1. Introduction

The Gaussian graphical model is commonly used for describing the joint distribution of multiple random variables (Lauritzen, 1996). The graph structure induced by this model not only delineates the conditional dependence between these variables, but also is critical for improving estimation precision. In estimating regression parameters in generalized estimating equations (GEE) for example, Zhou and Song (2016) found that incorporating a suitable dependence structure of covariates can improve estimation efficiency, sometimes substantially. In another example, Li and Li (2008) showed that a correctly specified dependence structure is useful to improve estimation efficiency in regularized estimation and variable

*Corresponding author. E-mail: pszhong@uic.edu