Commun. Math. Res.
doi: 10.4208/cmr.2022-0030

Vol. **39**, No. 3, pp. 414-436
August 2023

Decomposition of Covariate-Dependent Graphical Models with Categorical Data

Binghui Liu and Jianhua Guo*

School of Mathematics and Statistics & KLAS, Northeast Normal University, Changchun, Jilin 130024, P.R. China.

Received 22 June 2022; Accepted 21 October 2022

Abstract. Graphical models are wildly used to describe conditional dependence relationships among interacting random variables. Among statistical inference problems of a graphical model, one particular interest is utilizing its interaction structure to reduce model complexity. As an important approach to utilizing structural information, decomposition allows a statistical inference problem to be divided into some sub-problems with lower complexities. In this paper, to investigate decomposition of covariate-dependent graphical models, we propose some useful definitions of decomposition of covariate-dependent graphical models with categorical data in the form of contingency tables. Based on such a decomposition, a covariate-dependent graphical model can be split into some sub-models, and the maximum likelihood estimation of this model can be factorized into the maximum likelihood estimations of the sub-models. Moreover, some sufficient and necessary conditions of the proposed definitions of decomposition are studied.

AMS subject classifications: 62-09

Key words: Collapsibility, contingency tables, covariate-dependent, decomposition, graphical models.

1 Introduction

In the last few decades, graphical models have been widely used as intuitive and efficient tools of data analysis in several application domains, including biology,

*Corresponding author. *Email address:* `jhguo@nenu.edu.cn` (J. Guo)

economics and sociology [8,13,16,17,20,21]. It is well known that graphical models have significant advantages in representation and computation. To start with, graphical models facilitate intuitive representation of conditional dependence relationships among interacting variables. Secondly, by making use of structural information, computational efficiency of a statistical inference problem can be greatly improved, especially in situation of high-dimensional cases with sparse structure.

To make full use of structural information of graphical models, many strategies were studied, including two complementary ones: collapsibility and decomposition. When considering a statistical inference problem, collapsibility means that all information of interest only depends on some of the variables rather than the remaining uninformative ones, whereas decomposition means that the problem can be divided into some separated sub-problems with lower complexities. In the literature of graphical models, collapsibility and decomposition have been defined from many different points of view, such as models, parameters, estimations and hypothesis tests [1,3–7,9,14,18,23,24], among which a particular interest is to intuitively characterize a definition of collapsibility or decomposition via the corresponding interaction graph.

In this paper, we mainly consider covariate-dependent graphical models, also called graphical regression models or conditional graphical models, which were proposed in situation where the interactions among the object variables may further depend on some explanatory variables or covariates. For instance, in cancer studies, genome-wide alterations, such as gene expressions, DNA copy numbers and SNPs, were often investigated conditionally depending on some clinical phenotypes, such as tumor category and mutation status, which may influence the interactions among these genome-wide alterations.

Researches on the covariate-dependent graphical models are mainly twofold: structural recovery and statistical inference, which can be implemented simultaneously or step by step. Most of the current methods that simultaneously achieve structural learning and statistical inference were proposed to seek sparse estimations of the conditional covariance matrices of a covariate-dependent Gaussian graphical model, based on regularized conditional likelihoods [2, 10, 15]. Alternatively, the two tasks can be completed step by step: making statistical inferences based on the resulting structures of a structural learning procedure in advance [4, 14].

In this paper, we center our attention on statistical inference problems of a covariate-dependent graphical model given its interaction graph. To improve computational efficiency of such problems via the interaction graph, existing studies focused on the strategy of collapsibility [4, 14], however ignoring decomposition