Journal of Multivariate Analysis 164 (2018) 1–21

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Local conditional and marginal approach to parameter estimation in discrete graphical models

Hélène Massam *, Nanwei Wang

Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada M3J 1P3

ARTICLE INFO

Article history:
Received 23 July 2016
Available online 27 October 2017

AMS 2000 subject classifications:
62H17
62F10

Keywords:
Discrete graphical models
Distributed estimation
Local conditional
Local marginal
Maximum composite likelihood estimate
"Large p, large N" asymptotics

ABSTRACT

Discrete graphical models are an essential tool in the identification of the relationship between variables in complex high-dimensional problems. When the number of variables p is large, computing the maximum likelihood estimate (henceforth abbreviated MLE) of the parameter is difficult. A popular approach is to estimate the composite MLE (abbreviated MCLE) rather than the MLE, i.e., the value of the parameter that maximizes the product of local conditional or local marginal likelihoods, centered around each vertex v of the graph underlying the model. The purpose of this paper is to first show that, when all the neighbors of v are linked to other nodes in the graph, the estimates obtained through local conditional and marginal likelihoods are identical. Thus the two MCLE are usually very close. Second, we study the asymptotic properties of the composite MLE obtained by averaging of the estimates from the local conditional likelihoods: this is done under the double asymptotic regime when both p and N go to infinity.

1. Introduction

Discrete graphical models are an essential tool in the analysis of complex high-dimensional categorical data. For these models, in this paper, we study properties of the maximum composite likelihood estimate of the parameter, which is commonly used when, due to a large number of variables, the maximum likelihood estimate is impossible to compute.

Let $V = \{1, \ldots, p\}$ be a finite index set. Let $G = (V, E)$ be an undirected graph, where E is the set of undirected edges in $V \times V$. Then the distribution of $X = (X_v, v \in V)$ is said to be Markov with respect to G if X_v is independent of X_u given $X_{V\setminus\{u,v\}}$ whenever the edge (v, u) is not in E. The set of distributions Markov with respect to a given graph G is called a graphical model. When the variables X_v take values in a finite set I_v, $v \in V$, the graphical model is said to be discrete. These models are used extensively to represent interactions between individuals in physical or human networks. Each data point is classified according to its values of $X_v = i_v$, $i_v \in I_v$, $v \in V$ and the data are thus gathered in a p-dimensional contingency table with cells $i = (i_v, v \in V)$ and cell counts $n(i)$, $i \in I = \prod_{v\in V} I_v$. As we shall recall in Section 2, the density of the cell counts can be written under a natural exponential family form as

$$f(t; \theta) = \exp\{\langle\theta, t\rangle - Nk(\theta)\}, \tag{1}$$

where t is a vector of marginal cell counts, $\langle\theta, t\rangle$ denotes the inner product of $t = t(x)$ and θ, the canonical loglinear parameter.

* Corresponding author.
E-mail addresses: massamh@mathstat.yorku.ca (H. Massam), wangnanw@yorku.ca (N. Wang).

https://doi.org/10.1016/j.jmva.2017.10.003
0047-259X/

For a given data set, the first task is to learn the underlying graph and once this is done, the second task is to estimate the parameter θ of the model. When p is large, to obtain the maximum likelihood estimate (abbreviated MLE) of θ through a simple maximization of the likelihood function is impossible because of the dimension of the parameter θ and the complexity of the cumulant generating function $k(\theta)$ in (1). Approximate techniques such as variational methods [7,16] or MCMC techniques [5] have been developed in recent years. More recently still, work has been done on a third type of approximate technique based on the maximization of composite likelihoods [1,10]. For a given data set $x^{(1)}, \ldots, x^{(N)}$, a composite likelihood is typically the product of local conditional likelihoods, coming from the local conditional probability of X_v given $X_{\mathcal{N}_v}$, $v \in V$, where $\mathcal{N}_v$ indicates the set of neighbors of v in G. The parameter of the local conditional distribution at v is a sub-vector of θ which, in the sequel, we will denote $\theta^{v,PS}$.

Further in the direction of composite likelihood, recent research has focused on studying each local conditional model and combining all the local results to yield a global estimate of either the underlying structure G or the parameter θ. For model selection when p is large, Ravikumar et al. [14], for example, introduced a local approach to discrete graphical model selection by looking at the regularized local conditional likelihood of X_v given $X_{V\setminus\{v\}}$, i.e., due to the Markov properties of the model, given $X_{\mathcal{N}_v}$. For parameter estimation, with p large and G given, the local approach has been used for Gaussian graphical models by Wiesel and Hero [18] and for discrete models by Liu and Ihler [11]. These two papers used, respectively, the ADDM optimization technique and the so-called linear consensus or maximum consensus to obtain a global estimate of the parameter from the estimates of the parameters of the local conditional distributions. Moreover, Liu and Ihler [11] study the asymptotic properties, for p fixed and N going to infinity, of the maximum composite likelihood estimate thus obtained by combining estimates of $\theta^{v,PS}$, $v \in V$.

For the estimation of the precision matrix in graphical Gaussian models, Meng et al. [12,13] depart from the ideas of the two papers just mentioned, in two ways. First, they do not consider local conditional models but rather local marginal models. Second they do not look only at "one-hop" marginal models, i.e., models built on v and its neighbors $\mathcal{N}_v$ but they consider "two-hop" local marginal models, i.e., marginal models with vertex set comprising v, its neighbors and the neighbors of the neighbors. With the two-hop local marginal likelihoods, they show with examples that they achieve such accuracy that, to obtain the overall estimate of the parameter, they need not use a method more sophisticated than simple averaging of the various local marginal likelihood estimates.

In this paper, we are concerned with the maximum composite likelihood estimation of θ and our purpose is twofold. First, we extend the methodology of [13] to discrete graphical models: as we will show in Lemmas 1 and 2, the marginal models are not convex in the parameter θ of the global model. So, in a way parallel to [13], in Section 3.3, we consider relaxed local marginal models (rather than regular marginal ones). We obtain estimates of θ by averaging estimates from local marginal models and compare them to the estimate obtained from local conditional models.

In Theorem 1, for the one-hop neighborhood, we show that, under the condition that all the neighbors of v are linked to vertices not in $\mathcal{N}_v$, the v-local estimate of the parameters of interest obtained from local conditional or marginal likelihoods is identical. We express this condition by saying that $\mathcal{N}_v = \mathcal{B}_v$ where $\mathcal{B}_v$, called the buffer set, is the set of vertices in $\mathcal{N}_v$ which are linked to vertices outside the neighborhood. A similar statement holds for the two-hop neighborhoods. In Section 3, we compare the MCLE obtained through local conditional and marginal likelihoods for three graphical models with 100 variables: one graph is random, the other is a hub graph and the third is a 10×10 grid. Whether the condition $\mathcal{N}_v = \mathcal{B}_v$ is satisfied for $v \in V$ depends on the type of graph. Computing the mean square error of the MCLEs, we see that, for one-hop neighborhoods, the slightly improved accuracy obtained through local marginal models is counterbalanced by a significant increase in computational complexity. And this improvement is only observed for the random graph. For two-hop neighborhoods, the mean square errors for the MCLE obtained from local conditional and marginal likelihoods are indistinguishable. Moreover, due to computational complexity, in the case of the hub graph, it was not possible to compute the MCLE based on marginal likelihoods.

We note that the subset B_v plays also an important role in Lemmas 1 and 2 which give us the relationship between the parameters of the local marginal models and the parameters of the global model: marginal parameters indexed by subsets not contained in $\mathcal{B}_v$ are equal to the global parameters indexed by the same subsets.

In Section 4, we look at the asymptotic properties of the MCLE of θ. We do so under the classical and double asymptotic regimes. Our main result, Theorem 3, states that, when both p and N go to infinity, under certain conditions, Conditions A and B, on the Fisher information matrices of the local conditional models, for $N/\ln p$ large enough, our estimate is close to the true value of the parameter with high probability. Our result under the classical regime, Theorem 2 where p is fixed, coincides with Theorem 4.1 of [11] and is given here for the sake of completeness.

To end this section, we comment on the existence of the maximum composite likelihood estimate. First when computing the estimates from the local conditional likelihoods, we need to make sure that they exist, i.e., that there exist finite estimates of $\theta^{v,PS}$ that maximize the local conditional likelihood. If they do not exist, our maximization software may return values that are erroneous. It may happen also that the global maximum likelihood estimate of θ does not exist and yet the local estimates of $\theta^{v,PS}$ exist: we can then obtain a maximum composite likelihood estimate of the parameter. We expand on these points in Lemma 3. Techniques to identify the existence of the global maximum likelihood estimate of θ have been developed in [3] and [17]. In the present paper, we will assume that all local estimates exist.