Journal of Multivariate Analysis 139 (2015) 207–218

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Sharp lower and upper bounds for the Gaussian rank of a graph

Emanuel Ben-David

Department of Statistics, Columbia University, United States

ARTICLE INFO

Article history:
Received 1 November 2014
Available online 24 March 2015

AMS subject classification:
62H12
97K30

Keywords:
Gaussian graphical models
Maximum likelihood estimation
Tree-width
Connectivity number
Degeneracy number
Random graphs
Symmetric graphs

ABSTRACT

An open problem in graphical Gaussian models is to determine the smallest number of observations needed to guarantee the existence of the maximum likelihood estimator of the covariance matrix with probability one. In this paper we formulate a closely related problem in which the existence of the maximum likelihood estimator is guaranteed for all generic observations. We call the number determined by this problem the Gaussian rank of the graph representing the model. We prove that the Gaussian rank is strictly between the subgraph connectivity number and the graph degeneracy number. These bounds are sharper than the bounds known in the literature and furthermore computable in polynomial time.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

An open problem in graphical Gaussian models is to determine the smallest number of observations needed to guarantee the existence of the maximum likelihood estimator (MLE) of the covariance matrix. This problem first arose in Dempster's paper [3] and has been frequently brought to attention by Steffen Lauritzen, as evident in [1] and Lauritzen's lectures in "Durham Symposium on Mathematical Aspects of Graphical Models 2008". Hence we refer to this problem as the D&L problem using the initials of Dempster and Lauritzen. Note that the D&L problem as stated above is not well-posed, because the existence of the MLE cannot be guaranteed for all observations, no matter how large the number of observations is. To make this well-posed, a necessary condition is to ignore a set of observations of probability zero and thus require the existence of the MLE only for observations in a set of probability one. It is now clear that the D&L problem in its most general form can be expressed as follows.

D&L problem (I) For a given graphical Gaussian model with respect to a graph $\mathcal{G}$, determine the smallest number of observations needed to guarantee the existence of the MLE with probability one.

A slightly different well-posed formulation of the D&L problem is to require the existence of the MLE (of the covariance matrix, henceforth) for all generic observations in the following sense. Let $n \leq p$. Then the observations[1] $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ are said to be generic if each $n \times n$ principal submatrix of the sample covariance matrix $\mathcal{S} = 1/n \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\top$ is non-singular. Note that with probability one every n observations are generic (thus the set of non-generic observations is of probability zero). This formulation can now be expressed as follows.

D&L problem (II) For a given graphical Gaussian model with respect to a graph $\mathcal{G}$, determine the smallest number of observations needed to guarantee the existence of the MLE for all generic observations.

E-mail address: bendavid@stat.columbia.edu.

[1] Throughout the paper, we always assume that the observations are independent and identically distributed.

http://dx.doi.org/10.1016/j.jmva.2015.03.004
0047-259X/© 2015 Elsevier Inc. All rights reserved.

The number determined by the D&L problem (II) is said to be the Gaussian rank of $\mathcal{G}$, denoted by $\mathrm{r}(\mathcal{G})$. The primary goal of this paper is to obtain sharp lower and upper bounds on $\mathrm{r}(\mathcal{G})$. In parallel to the D&L problem (II), the number determined by the D&L problem (I) is said to be the weak Gaussian rank of $\mathcal{G}$, denoted by $\mathrm{r}_w(\mathcal{G})$. Note again that being in general position is a generic property of any n observations, that is, with probability one for every n observations each $n \times n$ principal submatrix of the sample covariance matrix is non-singular. Thus obviously $\mathrm{r}(\mathcal{G}) \geq \mathrm{r}_w(\mathcal{G})$, for every graph $\mathcal{G}$.

As we mentioned earlier, the D&L problem first arose in [3], a celebrated work of Dempster in which, under the principle of parsimony in fitting parametric models, Dempster introduced graphical Gaussian models and gave a partial analytic expression for the MLE of the covariance matrix, assuming that the MLE exists. In practice, the MLE has to be computed by an iterative numerical procedure, such as the iterative proportional scaling procedure (IPS) [16]. However, if the number of observations is not sufficiently large there is no guarantee that the output matrix will indeed correspond to the MLE, as it might not be positive definite. It is clear that the number determined by the D&L problem under either formulations depends on the level of sparsity in graph, but the exact nature of this dependency is far from clear. Mainly because sparse graphs are not well understood, in particular it is not known what graph parameters best measure the graph sparsity. Hence the D&L problems (I) and (II) are important problems not only in relation to maximum likelihood estimation but also for gaining deeper insight into sparse graphs. In practice, the D&L problem (II) is equivalent to the D&L problem (I), because in general the data can be expected to be generic and therefore $n \geq \mathrm{r}(\mathcal{G})$ is a condition that minimally guarantees the existence of the MLE.

In a graphical Gaussian model the sparsity is given by a pattern of zeros in the inverse covariance matrix. The pattern of zeros is determined by the missing edges of a graph and each zero entry of a generic inverse covariance matrix indicates that the corresponding pair of variables are conditionally independent given the remaining variables. An attractive feature of graphical Gaussian models, or graphical models in general, is that if the graph representing the model is sparse, then the MLE of the covariance matrix can exist even if the number of observations is much smaller than the dimension of the multivariate Gaussian distribution. Intuitively, we expect that the Gaussian rank of the graph, the number determined by the D&L problem, to decrease as the graph becomes sparser. However, it is not clear what best measures the sparsity of a graph or how the Gaussian rank varies accordingly. The main theorem of this paper suggests two such measures of sparsity.

Despite some efforts since 90's not much progress has been made to resolve the D&L problem. The existing results are limited to a handful of publications as follows.

(R1) [5] For a decomposable graph $\mathcal{G}$, a graph that has no induced cycle of length larger than or equal to four, the Gaussian rank is equal to $\omega(\mathcal{G})$, the size of the largest complete subgraph of $\mathcal{G}$.
(R2) [5] For every graph $\mathcal{G}$, $\omega(\mathcal{G}) \leq \mathrm{r}_w(\mathcal{G}) \leq \mathrm{r}(\mathcal{G}) \leq \mathrm{tw}(\mathcal{G}) + 1$, where $\mathrm{tw}(\mathcal{G})$ denotes the treewidth of $\mathcal{G}$ (See [7] for the definition).
(R3) [1] For C_p, a cycle of length $p \geq 3$, $2 = \omega(C_p) < \mathrm{r}_w(C_p) = \mathrm{r}(C_p) = 3 = \mathrm{tw}(C_p) + 1$.
(R4) [17] For $G_{3,3}$, the 3×3 grid, $2 = \omega(G_{3,3}) < \mathrm{r}_w(G_{3,3}) = \mathrm{r}(G_{3,3}) = 3 < \mathrm{tw}(G_{3,3}) + 1 = 4$.

In the literature the bounds given by (R2) are currently the best known bounds for the Gaussian rank. Restricted to the class of decomposable graphs these bounds are tight, since $\mathrm{tw}(\mathcal{G}) + 1 = \omega(\mathcal{G})$ for a decomposable graph $\mathcal{G}$, but we may note that (R4) in [17] shows that for non-decomposable graphs the bounds in (R2) are not necessarily tight. Intuitively, it is apparent that $\omega(\mathcal{G})$ overestimates and $\mathrm{tw}(\mathcal{G})$ underestimates the sparsity of $\mathcal{G}$ and therefore sharper bounds may exist. In this paper we give sharper bounds on the Gaussian rank. The lower and upper bounds we give are the subgraph connectivity number, denoted by $\kappa^*(\mathcal{G})$, and the graph degeneracy number, denoted by $\delta^*(\mathcal{G})$. Both $\kappa^*(\mathcal{G})$ and $\delta^*(\mathcal{G})$ will be defined later in Section 2.2. Formally we prove the following theorem.

Theorem 1.1. *Let $\mathcal{G} = (V, E)$ be a graph. Then*

$$\kappa^*(\mathcal{G}) + 1 \leq \mathrm{r}(\mathcal{G}) \leq \delta^*(\mathcal{G}) + 1 \tag{1.1}$$

The proof of this theorem is given in Section 4. The upper bound is proved by mathematical induction using two key observations: suppose a graph $\mathcal{H}$ is obtained from $\mathcal{G}$ by removing a vertex v and its adjacent edges. Then (1) $\mathrm{r}(\mathcal{H}) \geq \mathrm{r}(\mathcal{G}) - 1$ and (2) if $\mathrm{r}(\mathcal{G})$ is larger than the number of the vertices adjacent to v, then $\mathrm{r}(\mathcal{H}) = \mathrm{r}(\mathcal{G})$. The proof of the lower bound relies on Lovász–Saks–Schrijver's Theorem in [11].

All the results stated in (R1) through (R4) now immediately follow from Theorem 1.1, since by some simple calculations we can show that

(a) for a decomposable graph $\mathcal{G}$, $\omega(\mathcal{G}) = \kappa^*(\mathcal{G}) + 1 = \delta^*(\mathcal{G}) + 1 = \mathrm{tw}(\mathcal{G}) + 1$;
(b) for any (arbitrary) graph $\mathcal{G}$, $\omega(\mathcal{G}) \leq \kappa^*(\mathcal{G}) + 1 \leq \mathrm{r}(\mathcal{G}) \leq \delta^*(\mathcal{G}) + 1 \leq \mathrm{tw}(\mathcal{G}) + 1$;
(c) for a cycle (of any length) $\mathcal{G}$, $\kappa^*(\mathcal{G}) = \delta^*(\mathcal{G}) = 2$;
(d) for a $k \times m$ grid (with k and $m \geq 2$), $\kappa^*(\mathcal{G}) = \delta^*(\mathcal{G}) = 2$.

Note that by Part (d) the Gaussian rank of every grid is 3 which is substantially less than the upper bound given by (R2) for grids of large dimensions. The reason is that the treewidth of a $k \times m$ grid is $\min\{k, m\}$ which tends to $+\infty$ as k and $m \to +\infty$ [14]. Here we mention that by relating $\mathrm{r}_w(\mathcal{G})$ to the rigidity of the graph $\mathcal{G}$, Gross and Sullivant in a recent manuscript [6] also obtain the upper bound in Theorem 1.1 for $\mathrm{r}_w(\mathcal{G})$.

Download English Version:

https://daneshyari.com/en/article/1145503

Download Persian Version:

https://daneshyari.com/article/1145503

Daneshyari.com