Journal of Multivariate Analysis 147 (2016) 1–19

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Improved second order estimation in the singular multivariate normal model

Didier Chételat [a], Martin T. Wells [b,*]

[a] Department of Decision Sciences, HEC Montréal, Canada
[b] Department of Statistical Science, Cornell University, United States

ARTICLE INFO

Article history:
Received 30 June 2015
Available online 14 January 2016

AMS 2000 subject classifications:
primary 62C15
secondary 62F10
62H12

Keywords:
Covariance matrix
Precision matrix
Discriminant function
LDA
Unbiased risk estimator
Moore–Penrose inverse
Singular normal
Singular Wishart

ABSTRACT

We consider the problem of estimating covariance and precision matrices, and their associated discriminant coefficients, from normal data when the rank of the covariance matrix is strictly smaller than its dimension and the available sample size. Using unbiased risk estimation, we construct novel estimators by minimizing upper bounds on the difference in risk over several classes. Our proposal estimates are empirically demonstrated to offer substantial improvement over classical approaches.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

With the recent explosion of high throughput data, much interest has arisen in applications where the number of feature parameters is greater than the sample size. In this situation, it is typically assumed that, despite their number, the underlying components are linearly independent, or in other words that their covariance matrix has full rank. However, little attention has been given to the situation where there is dependence between the components, that is, where the covariance matrix would be singular.

Recently, Tsukuma and Kubokawa [25] investigated the problem of estimating the mean vector of a multivariate normal distribution when the unknown covariance matrix is singular. By deriving an unbiased risk estimator for the quadratic loss, they were able to give sufficient conditions for an estimator to dominate the maximum likelihood estimator.

This article is concerned with the same model as Tsukuma and Kubokawa [25], but we consider three different estimation problems. Unlike the mean estimation problem, all three estimation scenarios depend the second order moment of the distribution. In each case we provide decision-theoretic results that lead to improved inference. The first task is the estimation of the singular covariance matrix itself, under an invariant squared loss. This problem was first considered in the full rank case by Haff [13], and in the high-dimensional setting by Konno [17]. The second concern is the estimation of the Moore–Penrose

* Corresponding author.
E-mail addresses: didier.chetelat@hec.ca (D. Chételat), mtw1@cornell.edu (M.T. Wells).

http://dx.doi.org/10.1016/j.jmva.2016.01.001
0047-259X/© 2016 Elsevier Inc. All rights reserved.

pseudo-inverse of the covariance matrix, also known as the precision matrix, under the Frobenius loss. This problem was first considered in the full rank case by Haff [10,12] and in the high-dimensional setting by Kubokawa and Srivastava [18].

Finally, we consider the problem of estimating the discriminant coefficient that arise in Linear Discriminant Analysis (LDA) under the squared loss, a problem first considered in the full rank case by Haff [14] and Dey and Srinivasan [6]. LDA is a standard method for classification when the number of observations n is much larger than the number of features p. If data follows p-variate normal distribution with the same covariance structure across the groups, it provides an asymptotically optimal classification rule, meaning that its misclassification error converges to Bayes risk. However, it was noted by Dudoit et al. [7] that a naive implementation of LDA for high-dimensional data provides poor classification results in comparison to alternative methods. A rigorous proof of this phenomenon in the case $p \gg n$ is given by Bickel and Levina [1]. There are two main reasons for this. First, standard LDA uses the sample covariance matrix to estimate the covariance structure, and in high dimensional settings this results in a singular estimator. Secondly, by using all p features in classification, interpretation of the results becomes challenging. One of the popular approaches to deal with the singularity is to use the independence rule which overcomes the singularity problem of the sample covariance but ignores the dependency structure. This approach is very appealing because of its simplicity and was encouraged by the work of Bickel and Levina [1], who showed it performs better than the standard LDA in a $p \gg n$ setting when the population matrix is full rank. Unfortunately, independence is only an approximation and it is unrealistic in most applications: for instance, in a genomic context, gene interactions and low dimensional network structure are crucial for the understanding of biological processes. In this situation, one should aim for better estimators of the covariance matrix rather than relying on an independence structure that assumes a full rank population covariance matrix. Indeed, we will see in Section 3 that using the diagonal of the sample covariance matrix is a poor strategy if the true covariance matrix is rank deficient, that is, has a low dimensional structure.

The presentation of our approach to these three estimation problems is divided as follows. The decision-theoretic results are described in Section 2. For each of the three problems, we construct an appropriate unbiased estimator of the risk (URE) using Stein's and Haff's lemmas [24,11,25]. We then consider the class of estimator given by constant multiples of a naive estimator, and minimize an upper bound on the difference in risk to obtain estimators that dominate the naive estimator. Finally, we consider a larger class given by the sum of this estimator and an appropriate trace, and obtain an unbiased risk estimate of its risk. By a heuristic argument, we obtain an estimator that appears to dominate the previous estimator in simulations.

The remainder of this paper is organized as follows. The primary decision-theoretic results are presented in Section 2. In Section 3 we investigate the amount of improvement provided by the proposed estimators through numerical study. A discussion of our results as well as pointers to future research directions are given in Section 4. Finally, the proofs of the propositions and theorems stated in Section 2 are provided in Section 5.

2. Estimation results

2.1. Model

Our setting is similar to the one used in [25]. We observe an n-sample $X_1, \ldots, X_n$ identically and independently distributed from a p-dimensional multivariate normal distribution $\mathcal{N}_p(\mu, \Sigma)$, where μ and Σ are unknown. However, the p-dimensional covariance matrix Σ is rank-deficient with respect to the dimension and the sample size, in the sense that

$$r = \mathrm{rk}\Sigma < \min(n, p). \tag{2.1}$$

The resulting singular multivariate normal distribution does not have a density with respect to the Lebesgue measure on $\mathbb{R}^p$, but lives in the r-dimensional linear subspace spanned by the columns of Σ. More details can be found, for example, in [23, Section 2.1].

Define the $n \times p$ data matrix $X = (X_1, \ldots, X_p)^\top$. The sample covariance matrix $S = (X - 1_n\bar{X}^\top)^\top(X - 1_n\bar{X}^\top)/n$ then follows a Wishart distribution $\mathcal{W}_p(n-1, \Sigma/n)$ with $n-1$ degrees of freedom. Since Σ is rank-deficient, it is singular in the terminology of Srivastava and Khatri [23, Section 3.1]. We warn the reader that the expression "singular Wishart" has also been used in the literature to describe the different situation where the covariance is positive-definite and the dimension exceeds the degrees of freedom, as in [22]. Let $S = O_1 L O_1^\top$ denote the reduced spectral decomposition of S, where $L = \mathrm{diag}(l_1, \ldots, l_r)$ denote the r non-zero eigenvalues and O_1 is $p \times r$ semi-orthogonal.

In this situation, neither S nor Σ are invertible. Since inverses of covariance matrix are of considerable interest in multivariate statistical analysis, some generalized inverse of these quantities is desirable. In this article, we will focus on the Moore–Penrose pseudoinverse, which will be denoted A^+ for a matrix A. Definitions and theoretical properties can be found in [15, Chapter 20].

The singular multivariate normal model is amenable to decision-theoretic analysis through a key insight of Tsukuma and Kubokawa [25, Section 2.2]. The authors proved that when (2.1) holds, the subspace spanned by the sample covariance matrix is almost surely constant and matches the subspace spanned the true covariance matrix, in the sense that the remarkable identity holds

$$SS^+ = \Sigma\Sigma^+. \tag{2.2}$$

This fact will be repeatedly used in Section 5 proofs and is essential to our derivations.

Download English Version:

https://daneshyari.com/en/article/1145240

Download Persian Version:

https://daneshyari.com/article/1145240

Daneshyari.com