Journal of Multivariate Analysis 157 (2017) 45–52

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

A calibration method for non-positive definite covariance matrix in multivariate data analysis

Chao Huang [a], Daniel Farewell [b], Jianxin Pan [c,*]

[a] *South East Wales Trials Unit, Cardiff University, Heath Park, Cardiff, CF14 4YS, UK*
[b] *Division of Population Medicine, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4YS, UK*
[c] *School of Mathematics, University of Manchester, Manchester M13 9PL, UK*

ARTICLE INFO

Article history:
Received 6 May 2016
Available online 10 March 2017

AMS subject classification:
62H99

Keywords:
Covariance matrix calibration
Nearness problem
Non-positive definiteness
Spectral decomposition

ABSTRACT

Covariance matrices that fail to be positive definite arise often in covariance estimation. Approaches addressing this problem exist, but are not well supported theoretically. In this paper, we propose a unified statistical and numerical matrix calibration, finding the optimal positive definite surrogate in the sense of Frobenius norm. The proposed algorithm can be directly applied to any estimated covariance matrix. Numerical results show that the calibrated matrix is typically closer to the true covariance, while making only limited changes to the original covariance structure.

1. Introduction

The estimation of covariance matrices plays an essential role in multivariable data analysis. Covariances are required by many statistical modeling approaches, including multivariate regression and the analysis of spatial data. Often, well-estimated covariance matrices improve efficiency in estimating parameters in a mean function [22]. In some circumstances, the covariance matrix may itself be of direct scientific interest: for instance, in spatial variation analysis for geographical data, and in volatility analysis for financial data.

However, it is not uncommon that estimators of covariance matrices fail to be positive definite. A typical example is the sample covariance matrix, which is often singular when the sample size is close to, or less than, the dimension of the random samples [3]. If singularity is caused by collinearity, conventional ridge regression [18] or modern variable selection [6,21] approaches may solve the problem by excluding redundant variables. Dimension reduction approaches such as Principle Component Analysis [19] can also help to exclude eigenvalues with ignorable contributions.

However, these resolutions only apply in cases where such redundancy truly exists. More often, non-positive definiteness may be put down to the generic difficulty of maintaining positive definiteness in covariance estimation; resulting estimators may not even be positive semidefinite. Even for elaborately designed statistical approaches, the estimators of covariance matrices can be ill-conditioned [5,14]. A number of approaches have been proposed to resolve this issue. However, these are either limited to special circumstances or lack theoretical support. For instance, one alternative is to use the Moore–Penrose inverse of a non-positive definite matrix to replace the regular inverse typically used in statistical inferences [20]. However, this does not directly resolve the non-positive definiteness, and is lack of statistical interpretation. Alternatively, a smoothing

* Corresponding author.
E-mail addresses: HuangC12@cardiff.ac.uk (C. Huang), farewelld@cf.ac.uk (D. Farewell), Jianxin.Pan@manchester.ac.uk (J. Pan).

http://dx.doi.org/10.1016/j.jmva.2017.03.001
0047-259X/

approach exists [23] in which non-positive eigenvalues of the covariance matrix estimator are replaced by certain positive values. However, justification for the selection of these positive values was scant.

Based on the fundamental work of Halmos [7], Higham [9] proposed a solution for finding the nearest (in the sense of Frobenius norm) positive semidefinite matrix to an arbitrary input matrix. However, this surrogate positive semidefinite matrix is still singular [9,10], so difficulty persists in using the surrogate matrix in statistical practice. Rebonato and Jäckel [17] considered a correlation matrix calibration using the hyperspherical decomposition and eigenvalue correction, which again leads to positive semidefinite correlation matrices. Hendrikse et al. [8] proposed an eigenvalue correction method using bootstrap resampling in order to reduce the bias arising in sample eigenvalues. Their work focused on the correction of the sample covariance, where the performance of the correction method relies on the assumed distribution of the covariance matrix eigenvalues in the population.

In this paper, we propose a unified approach to calibrate a non-positive definite covariance matrix to ensure positive definiteness. The calibrated covariance matrix is usually closer to the true covariance matrix than the original covariance matrix estimator. Our proposed approach is implemented through a straightforward screening algorithm. In Section 2, we briefly review the matrix nearness problem, before proposing our novel calibration method together with its integrated criterion and algorithm. In Section 3 we conduct two simulation studies, and in Section 4 we discuss two case studies, including a calibration of the non-positive definite covariance matrix obtained by nonparametric regression in Diggle and Verbyla [5]. Conclusions are presented in Section 5.

2. Calibration method

2.1. The matrix nearness problem

In numerical analysis, a nearness problem involves finding, for a given matrix and a particular matrix norm, the nearest matrix that has certain important properties. Examples include finding the nearest covariance matrix [9] or correlation matrix [2,16] in the sense of the Frobenius norm (or 2-norm).

Given an arbitrary square matrix X of order n, we denote its Frobenius norm by $\|X\| = \mathrm{trace}(X^\top X)^{1/2}$. The nearness problem involves finding the nearest symmetric positive semidefinite matrix $P_0(X)$:

$$P_0(X) = \arg\min_{A \geq 0} \|X - A\|. \tag{1}$$

Throughout, we shall assume that $A \geq 0$ denotes both non-negative definiteness and symmetry $A = A^\top$. Higham [9] used a polar decomposition to show that the solution to (1) has the explicit form $P_0(X) = (B+H)/2$, where $B = S(X) = (X+X^\top)/2$ is the symmetric matrix version of X, and H is the symmetric polar factor of B, satisfying $B = UH$ with U a unitary matrix and $H \geq 0$. This solution has been compiled in a MATLAB file named `poldex.m`, which can be found in the Matrix Computation Toolbox [11]. Clearly, if X is symmetric then the solution becomes $P_0(X) = (X + H)/2$. If, further, we are given the spectral decomposition of a symmetric $X = X^\top$ (that is, $X = Q\Lambda Q^\top$ for $Q^\top Q = I$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$), we have $P_0(X) = Q\,\mathrm{diag}\{\max(\lambda_1, 0), \ldots, \max(\lambda_n, 0)\}Q^\top$. In other words, the nearest positive semidefinite matrix $P_0(X)$ can be obtained by replacing by zero any negative eigenvalues of a symmetric X [10], eliminating the corresponding columns of Q (and causing some information loss). A immediate alternative is to instead replace negative eigenvalues by positive values, so that a positive definite correction of X is formed without this loss of information about Q. However, the theory of this idea need to be justified, particularly on how to choose appropriate replacement positive values, for which we will address in this paper.

2.2. A new calibration approach

We now aim to find a positive definite matrix surrogate for a generic X. First, we formulate this question as a nearness problem. For $c \geq 0$, let $\mathcal{D}_c = \{A : A - cI \geq 0\}$ be the set of positive definite matrices with no eigenvalue smaller than c. Given X, finding the nearest matrix $P_c(X) \in \mathcal{D}_c$ to X in terms of the Frobenius norm amounts to defining

$$P_c(X) = \arg\min_{A \in \mathcal{D}_c} \|X - A\|. \tag{2}$$

An explicit expression for $P_c(X)$ is given in Theorem 1.

Theorem 1. *Given X and a constant $c \geq 0$, the nearest (in the sense of Frobenius norm) matrix $P_c(X) \in \mathcal{D}_c$ to X is of the form*

$$P_c(X) = P_0(X - cI) + cI \tag{3}$$

where (as before) $P_0(X - cI) = (B + H)/2$ for $B = S(X - cI)$ and H the polar factor of B. Furthermore, if X is symmetric with spectral decomposition $X = Q\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)Q^\top$ then $P_c(X)$ has the simplified form

$$P_c(X) = Q\,\mathrm{diag}\{\max(\lambda_1, c), \ldots, \max(\lambda_n, c)\}Q^\top. \tag{4}$$