Journal of Statistical Planning and Inference 150 (2014) 66–83

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Inference on the eigenvalues of the covariance matrix of a multivariate normal distribution—Geometrical view

Yo Sheena

Faculty of Economics, Shinshu University, Japan

A R T I C L E I N F O

Article history:
Received 22 October 2013
Received in revised form
5 March 2014
Accepted 11 March 2014
Available online 20 March 2014

MSC:
Primary, 62H05;
Secondary, 62F12

Keywords:
Curved exponential family
Information loss
Fisher information metric
Embedding curvature
Affine connection
Positive definite matrix

A B S T R A C T

We consider inference on the eigenvalues of the covariance matrix of a multivariate normal distribution. The family of multivariate normal distributions with a fixed mean is seen as a Riemannian manifold with Fisher information metric. Two submanifolds naturally arises: one is the submanifold given by the fixed eigenvectors of the covariance matrix; the other is the one given by the fixed eigenvalues. We analyze the geometrical structures of these manifolds such as metric, embedding curvature under e-connection or m-connection. Based on these results, we study (1) the bias of the sample eigenvalues, (2) the asymptotic variance of estimators, (3) the asymptotic information loss caused by neglecting the sample eigenvectors, (4) the derivation of a new estimator that is natural from a geometrical point of view.

1. Introduction

Consider a normal distribution with zero mean and an unknown covariance matrix, $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Let denote the eigenvalues of $\boldsymbol{\Sigma}$ by

$$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p), \quad \lambda_1 > \cdots > \lambda_p$$

and eigenvectors matrix by $\boldsymbol{\Gamma}$, hence we have the spectral decomposition

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^t, \quad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda), \tag{1}$$

where $\mathrm{diag}(\lambda)$ means the diagonal matrix with the ith diagonal element λ_i. It is needless to say that the inference on $\boldsymbol{\Sigma}$ is an important task in many practical situations in such a diversity of fields as engineering, biology, chemistry, finance, psychology, etc. Especially we often encounter the cases where the property of interest depends on $\boldsymbol{\Sigma}$ only through its eigenvalues λ. We treat an inference problem on the eigenvalues λ from a geometrical point of view.

Treating the family of normal distributions $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ($\boldsymbol{\mu}$ is not necessarily zero) as a Riemannian manifold has been done by several authors. For example, see Fletcher and Joshi (2007), Lenglet et al. (2006), Skovgaard (1984), Smith (2005), and Yoshizawa and Tanabe (1999). When μ equals zero, the family of normal distributions $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ can be taken as a manifold (say $\mathcal{S}$) with a single coordinate system $\boldsymbol{\Sigma}$. Hence, $\mathcal{S}$ is identified with the space of symmetric positive definite matrices. Geometrically analyzing the space of symmetric positive definite matrices has been an interesting topic in a mathematical

http://dx.doi.org/10.1016/j.jspi.2014.03.004

or engineering point of view. Refer to Moakher and Zéraï (2011), Ohara et al. (1996) and Zhang et al. (2009) as well as the above literature.

In this paper, we analyze $\mathcal{S}$ from the standpoint of information geometry while focusing on the inference on the eigenvalues of $\boldsymbol{\Sigma}$. The paper is aimed to make a contribution in two regards: (1) the geometrical structure of $\mathcal{S}$ is analyzed in view of the eigenvalues and eigenvectors of $\boldsymbol{\Sigma}$; (2) some statistical problems on the inference for λ are explained in the geometrical terms.

We summarize the inference problem for λ. Based on independent n samples $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})', \ i = 1, \dots, n$ from $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$, we want to make inference on the unknown λ. We confine ourselves to the classical case where $n \geq p$. It is well-known that the product–sum matrix

$$\boldsymbol{S} = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^t$$

is sufficient statistic for both unknown λ and $\boldsymbol{\Gamma}$. The spectral decomposition of $\boldsymbol{S}$ is given by

$$\boldsymbol{S} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t, \quad \boldsymbol{L} = \mathrm{diag}(\boldsymbol{l}),$$

where

$$\boldsymbol{l} = (l_1, \dots, l_p), \quad l_1 > \cdots > l_p > 0 \quad \text{a.e.}$$

are the eigenvalues of $\boldsymbol{S}$, and $\boldsymbol{H}$ is the corresponding eigenvectors matrix. This decomposition gives us two statistics available, i.e. the sample eigenvalues $\boldsymbol{l}$ and the sample eigenvectors $\boldsymbol{H}$. However it is almost customary that we only use the sample eigenvalues, discarding the information contained in $\boldsymbol{H}$. In the past literature on the inference for the population eigenvalues, every notable estimator is based simply on the sample eigenvalues. See Takemura (1984), Dey and Srinivasan (1985), Haff (1991), Yang and Berger (1994) for orthogonally invariant estimators of $\boldsymbol{\Sigma}$; Dey (1988), Hydorn and Muirhead (1999), Jin (1993), Sheena and Takemura (2011) for direct estimators of λ. Since we do not have enough space to state the concrete form of each estimator, we just mention Stein's estimator as a pioneering work for "shrinkage" estimator of $\boldsymbol{\Sigma}$. In general, an orthogonally invariant estimator of $\boldsymbol{\Sigma}$ is given by

$$\hat{\boldsymbol{\Sigma}} = \boldsymbol{H}\boldsymbol{\Phi}\boldsymbol{H}^t, \quad \boldsymbol{\Phi} = \mathrm{diag}(\phi_1(\boldsymbol{l}), \dots, \phi_p(\boldsymbol{l})). \tag{2}$$

The estimator of λ is given by the eigenvalues of $\hat{\boldsymbol{\Sigma}}$, that is, $(\phi_1(\boldsymbol{l}), \dots, \phi_p(\boldsymbol{l}))$. The sample covariance matrix (M.L.E. estimator) $\overline{\boldsymbol{S}} \triangleq n^{-1}\boldsymbol{S}$ gives the estimator of λ as $\phi_i(\boldsymbol{l}) = n^{-1} l_i, \ i = 1, \dots, p$, while Stein's "shrinkage" estimator gives birth to

$$\phi_i(\boldsymbol{l}) = l_i/(n+p+1-2i), \quad i = 1, \dots, p. \tag{3}$$

Stein's estimator assigns the lighter (heavier) weight to the larger (smaller) sample eigenvalues, hence the diversity of $\boldsymbol{l}$ is shrunk. This estimator is quite simple and performs much better than M.L.E. (see Dey and Srinivasan, 1985). Unlike Stein's estimator, many estimators in the above literature are not explicitly given or too complicated for immediate use. Nonetheless they all have one common feature. The derived estimators of λ only depend on $\boldsymbol{l}$.

In a sense it is natural to implicitly associate the sample eigenvalues to the population eigenvalues, and the sample eigenvectors to the population counterpart. However the sample eigenvalues are not sufficient for the unknown population eigenvalues. Therefore it is important to evaluate how much information is lost by neglecting the sample eigenvectors. Following Amari (1982), we gain an understanding of the asymptotic information loss with geometric terms such as Fisher information metric and embedding curvatures.

Another statistically interesting topic is the bias of $n^{-1}\boldsymbol{l}$. It is well known that $n^{-1}\boldsymbol{l}$ is largely biased and the estimators mentioned above are all modification of $n^{-1}\boldsymbol{l}$ to correct the bias, that is, "shrinkage estimators." We show that the bias is closely related to the embedding curvatures. Moreover the geometric structure of $\mathcal{S}$ naturally leads us to a new estimator, which is also a shrinkage estimator.

The organization of this paper is as follows: In the former part (Sections 2 and 3), we describe the geometrical structure of $\mathcal{S}$ in view of the spectral decomposition (1). In Section 2, we observe $\mathcal{S}$ as a Riemannian manifold endowed with Fisher information metrics. In Section 3, we treat two submanifolds of $\mathcal{S}$, a submanifold given by the fixed eigenvectors and the one given by the fixed eigenvalues. The embedding curvatures of these submanifolds are explicitly given. We will show that the bias of $\boldsymbol{l}$ is closely related to the curvatures. In the latter part (Sections 4 and 5), we consider the estimation problem of λ. In Section 4, we describe the asymptotic variance of estimators when $\boldsymbol{\Gamma}$ is known (Section 4.1) and the asymptotic information loss caused by discarding the sample eigenvectors $\boldsymbol{H}$ (Section 4.2). The asymptotic information loss could be measured by the difference in the asymptotic variance between two certain estimators. In Section 5 for the case when $\boldsymbol{\Gamma}$ is unknown, we propose a new estimator of λ, which is naturally derived from a geometric point of view. In the last section, some comments are made for further research. All the proofs are collected in Appendix.

Unfortunately we do not have enough space to explain the geometrical concepts used in this paper. Refer to Boothby (2002), Amari (1985), and Amari and Nagaoka (2000).