Journal of Multivariate Analysis 137 (2015) 161–172

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional Gaussian distributions

T. Tony Cai [a,*], Tengyuan Liang [a], Harrison H. Zhou [b]

[a] Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, United States
[b] Department of Statistics, Yale University, New Haven, CT 06511, United States

ARTICLE INFO

Article history:
Received 30 August 2013
Available online 20 February 2015

AMS 2000 subject classifications:
primary 62H12
62H10
secondary 62F12
94A17

Keywords:
Asymptotic optimality
Central limit theorem
Covariance matrix
Determinant
Differential entropy
Minimax lower bound
Sharp minimaxity

ABSTRACT

Differential entropy and log determinant of the covariance matrix of a multivariate Gaussian distribution have many applications in coding, communications, signal processing and statistical inference. In this paper we consider in the high-dimensional setting optimal estimation of the differential entropy and the log-determinant of the covariance matrix. We first establish a central limit theorem for the log determinant of the sample covariance matrix in the high-dimensional setting where the dimension $p(n)$ can grow with the sample size n. An estimator of the differential entropy and the log determinant is then considered. Optimal rate of convergence is obtained. It is shown that in the case $p(n)/n \to 0$ the estimator is asymptotically sharp minimax. The ultra-high-dimensional setting where $p(n) > n$ is also discussed.

1. Introduction

The determinant of a random matrix is an important functional that has been actively studied in random matrix theory under different settings. See, for example, [15,18,19,12–14,11,27,28,25,23]. In particular, central limit theorems for the log-determinant have been established for random Gaussian matrices in [15], for general real i.i.d. random matrices in [23] under an exponential tail condition on the entries, and for Wigner matrices in [28]. The determinant of random matrices has many applications. For example, the determinant is needed for computing the volume of random parallelotopes, which is of significant interest in random geometry (see [20,24]). More specifically, let $Z = (Z_1, \dots, Z_p)$ be linearly independent random vectors in $\mathbb{R}^n$ with $p \le n$. Then the convex hull of these p points in $\mathbb{R}^n$ almost surely determines a p-parallelotope and the volume of this random p-parallelotope is given by $\nabla_{n,p} = \det(Z^T Z)^{1/2}$, the squared root of the determinant of th random matrix $Z^T Z$.

The differential entropy and the determinant of the covariance matrix of a multivariate Gaussian distribution play a particularly important role in information theory and statistical inference. The differential entropy has a wide range of applications in many areas including coding, machine learning, signal processing, communications, biosciences and chemistry.

* Corresponding author.
E-mail address: tcai@wharton.upenn.edu (T.T. Cai).

http://dx.doi.org/10.1016/j.jmva.2015.02.003
0047-259X/

See [26,16,4,10,21]. For example, in molecular biosciences, the evaluation of entropy of a molecular system is important for understanding its thermodynamic properties. In practice, measurements on macromolecules are often modeled as Gaussian vectors. For a multivariate Gaussian distribution $\mathcal{N}_p(\mu, \Sigma)$, it is well-known that the differential entropy $\mathcal{H}(\cdot)$ is given by

$$\mathcal{H}(\Sigma) = \frac{p}{2} + \frac{p\log(2\pi)}{2} + \frac{\log\det\Sigma}{2}. \quad (1)$$

In this case, estimation of the differential entropy of the system is thus equivalent to estimation of the log determinant of the covariance matrix from the sample. For other applications, the relative entropy (a.k.a. the Kullback–Leibler divergence), which involves the difference of the log determinants of two covariance matrices in the Gaussian case, is important. The determinant of the covariance matrices is also needed for constructing hypothesis tests in multivariate statistics (see [2,22]). For example, the likelihood ratio test for testing linear hypotheses about regression coefficients in MANOVA is based on the ratio of the determinants of two sample covariance matrices [2]. In addition, quadratic discriminant analysis, which is an important technique for classification, requires the knowledge of the difference of the log determinants of the covariance matrices of Gaussian distributions. For these applications, it is important to understand the properties of the log determinant of the sample covariance matrix. The high-dimensional setting where the dimension $p(n)$ grows with the sample size n is of particular current interest.

Motivated by the applications mentioned above, in the present paper we first study the limiting law of the log determinant of the sample covariance matrix for the high-dimensional Gaussian distributions. Let $X_1, \ldots, X_{n+1}$ be an independent random sample from the p-dimensional Gaussian distribution $\mathcal{N}_p(\mu, \Sigma)$. The sample covariance matrix is

$$\hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n+1}(X_k - \bar{X})(X_k - \bar{X})^T. \quad (2)$$

A central limit theorem is established for the log determinant of $\hat{\Sigma}$ in the high-dimensional setting where the dimension p grows with the sample size n with the only restriction that $p(n) \le n$. In the case when $\lim_{n\to\infty} \frac{p(n)}{n} = r$ for some $0 \le r < 1$, the central limit theorem shows

$$\frac{\log\det\hat{\Sigma} - \sum_{k=1}^{p}\log\left(1 - \frac{k}{n}\right) - \log\det\Sigma}{\sqrt{-2\log\left(1 - \frac{p}{n}\right)}} \xrightarrow{L} \mathcal{N}(0,1) \quad \text{as } n \to \infty. \quad (3)$$

The result for the boundary case $p = n$ yields

$$\frac{\log\det\hat{\Sigma} - \log(n-1)! + n\log n - \log\det\Sigma}{\sqrt{2\log n}} \xrightarrow{L} \mathcal{N}(0,1), \quad \text{as } n \to \infty. \quad (4)$$

In particular, this result recovers the central limit theorem for the log determinant of a random matrix with i.i.d. standard Gaussian entries. See [15,23].

We then consider optimal estimation of the differential entropy and the log-determinant of the covariance matrix in the high-dimensional setting. In the conventional fixed-dimensional case, estimation of the differential entropy has been considered by using both Bayesian and frequentist methods. See, for example, [21,26,1]. A Bayesian estimator was proposed in [26] using the inverse Wishart prior which works without the restriction that dimension is smaller than the sample size. However, how to choose good parameter values for the inverse Wishart prior remains an open question when the population covariance matrix is nondiagonal. A uniformly minimum variance unbiased estimator (UMVUE) was constructed in [1]. It was later proved in [21] that this UMVUE is in fact dominated by a Stein-type estimator and is thus inadmissible. The construction of an admissible estimator was left as an open problem in [21].

Based on the central limit theorem for the log determinant of the sample covariance matrix $\hat{\Sigma}$, we consider an estimator of the differential entropy and the log determinant of Σ and study its properties. A non-asymptotic upper bound for the mean squared error of the estimator is obtained. To show the optimality of the estimator, non-asymptotic minimax lower bounds are established using Cramer–Rao's information inequality. The lower bound results show that consistent estimation of $\log\det\Sigma$ is only possible when $\frac{p(n)}{n} \to 0$. Furthermore, it is shown that the estimator is asymptotically sharp minimax in the setting of $\frac{p(n)}{n} \to 0$.

The ultra-high-dimensional setting where $p(n) > n$ is important due to many contemporary applications. It is a common practice in high-dimensional statistical inference, including compressed sensing and covariance matrix estimation, to impose structural assumption such as sparsity on the target in order to effectively estimate the quantity of interest. It is of significant interest to consider estimation of the log determinant of the covariance matrix and the differential entropy in the case $p(n) > n$ under such structural assumptions. A minimax lower bound is given in Section 4 using Le Cam's method which shows that it is in fact not possible to estimate the log determinant consistently even when the covariance matrix is known to be diagonal with equal values. This negative result implies that consistent estimation of $\log\det\Sigma$ is not possible when $p(n) > n$ over all the collections of the commonly considered structured covariance matrices such as bandable, sparse, or Toeplitz covariance matrices.