Contemporary Mathematics
Volume 50, 1986

EIGENVALUES AND EIGENVECTORS OF LARGE DIMENSIONAL SAMPLE COVARIANCE MATRICES

Jack W. Silverstein*

ABSTRACT. Limit theorems will be reviewed on the eigenvalues of a class of sample covariance matrices where the number of vector samples and the vector dimension are on the same order of magnitude. The main result states that the empirical distribution function of the eigenvalues converges almost surely to a nonrandom distribution function, as the dimension approaches infinity. The author will then present his results in describing the behavior of the eigenvectors of these matrices. The results suggest similarity between the measure on the appropriate orthogonal group induced by the matrix of eigenvectors and Haar measure.

1. INTRODUCTION

A sample covariance matrix is a random matrix whose entries consist of sample variances and covariances of the components of a random vector $\underline{X}$. When these components are known to have 0 mean the matrix takes on the form $(1/s)VV^T$, where V is $n \times s$ and its columns form an i.i.d. sample of the n-dimensional vector $\underline{X}$. These matrices are fundamental to multivariate statistics, the eigenvalues (all real, nonnegative) and eigenvectors (forming an orthonormal set) being used, for example, in principal component analysis, and hypothesis testing. Any knowledge of the spectral behavior is crucial to present applications, and at the same time would suggest new approaches to statistical problems.

Except for the Wishart case where $\underline{X}$ is multivariate normal, most of the analysis done on sample covariance matrices has been on large samples, keeping n fixed and letting $s \to \infty$. The following is a review of results when the components of $\underline{X}$ are themselves i.i.d.. They are limit theorems, providing information about the behavior of the eigenvalues and eigenvectors when n and s are both large and on the same order of magnitude. The theorems pertain to a sequence $\{M_n\}_{n=1}^{\infty}$ of matrices defined as follows:

*Supported under N.S.F. Grant MCS-8101703A01.

For each n let V_n be $n \times s$ consisting of i.i.d. mean 0 variance 1 random variables v_{ij} with distribution common for all n. Assume $s = s(n)$ with $s/n \to y > 0$ as $n \to \infty$. Let $M_n = (1/s)V_nV_n^T$.

The next section reviews the major results on the eigenvalues of M_n for n large. The third section outlines the work of the author in describing the behavior of the eigenvectors of M_n.

2. EIGENVALUES

The main result on the eigenvalues of M_n states that the empirical distribution function of the eigenvalues of F_n (that is, for every x, $F_n(x) = \frac{1}{n} \times$ (numbers of eigenvalues of $M_n \leq x$)) converges to a nonrandom limit. We have

THEOREM 1. ([5], [8], [16]) If there exists a $\delta > 0$ such that $E(|v_{11}|^{2+\delta}) < \infty$, then for every $x \in \mathbb{R}$, $F_n(x) \xrightarrow{as} F_y(x)$ as $n \to \infty$, where for $0 < y \leq 1$,

$$F_y'(x) = f_y(x)$$
$$= \begin{cases} \frac{1}{2\pi yx} \sqrt{(x - (1-\sqrt{y})^2)((1+\sqrt{y})^2 - x)} & \text{if } (1-\sqrt{y})^2 < x < (1+\sqrt{y})^2) \\ 0 & \text{otherwise} \end{cases}$$

and for $1 < y < \infty$,

$$F_y(x) = (1 - \tfrac{1}{y})I_{[0,\infty)}^{(x)} + \int_{(1-\sqrt{y})^2}^{x} f_y(t)dt.$$

This result forms a common intersection of the three references cited, each one considering a different class of matrices. Only [16] deals specifically with sample covariance matrices, the random vector $\underline{X}$ having independent but not necessarily identically distributed components. This paper establishes the a.s. convergence under the specified moment condition on v_{11}.

The other two papers, motivated from topics outside of multivariate statistics, allow the columns of V_n to vary in distribution. In [8] the matrices are formed as an analogue of the operator defined by the one-dimensional Schroedinger equation with random potential. In [5] they arise from a neural network model for the generation of neural connections of a hypothetical organism at birth. The techniques used in [5] and [16] (and in much of the work on M_n) involve the study of the limiting behavior of $\{\operatorname{tr} M^r\}_{r=1}^{\infty}$ which, divided by n, are the moments of F_n. Moments were not used in [8], but rather the Stieltjes transform of F_n, which is shown to converge in probability to the solution of a certain integral equation.

Other work on the eigenvalues of M_n will be reviewed briefly.

In [3], [15] it was shown that if there exists a $\delta > 0$ such that $E(|v_{11}^{6+\delta}|) < \infty$ (or $E(|v_{11}|^{4+\delta}) < \infty$) then the largest eigenvalue of M_n converges almost surely (a.s.) or in probability (i.p.) to $(1 + \sqrt{y})^2$. In [14] the a.s. convergence of the smallest eigenvalue to $(1 - \sqrt{y})^2$ for $y < 1$ is established in the Wishart case ($v_{11} = N(0,1)$).

Central limit theorems have been proven (under various additional assumptions) for the standardized sums of powers of the eigenvalues ([1], [6]) as well as for the standardized log of the determinant ([4], [6]).

The last result to be mentioned concerns the eigenvalues (all real, nonnegative) of the central multivariate F matrix, formed from two independent Wishart matrices. Let $\{M_n^a\}$ denote matrices with $v_{11} = N(0,1)$ and $\frac{s}{n} \to a$ as $n \to \infty$. In [12], [13], [17], [18] the following is shown.

For M_n^y, $M_n^{y'}$ independent $y > 0$, $0 < y' < 1$, the empirical distribution function of the eigenvalues of $M_n^y(M_n^{y'})^{-1}$ converges a.s. to $F_{y,y'}$, where for $0 < y \leq 1$,

$$F'_{y,y'}(x) = f^{(x)}_{y,y'}$$

$$= \begin{cases} \dfrac{(1 - y')}{2\pi x(xy'+y)} \sqrt{(x-b_1)(b_2-x)} & \text{if } b_1 < x < b_2, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$b_1 = \left(\frac{1 - \sqrt{1 - (1-y)(1-y')}}{1 - y'}\right)^2, \quad b_2 = \left(\frac{1 + \sqrt{1 - (1-y)(1-y')}}{1 - y'}\right)^2$$

and for $y > 1$,

$$F_{y,y'}(x) = (1 - \tfrac{1}{y})I_{[0,\infty)} + \int_{b_1}^{x} f_{y,y'}(t)dt.$$

3. EIGENVECTORS

Let $O_n\Lambda_nO_n^T$ be the spectral decomposition of M_n with the eigenvalues of M_n arranged in nondecreasing order along the diagonal of Λ_n. The eigenvectors are then the columns of O_n and $O_n \in \mathcal{O}_n$, the $n \times n$ orthogonal group (this decomposition can always be constructed on a probability space with O_n an $\mathcal{O}_n$-valued random matrix). There is only one case where the eigenvectors of M_n are completely understood, namely, the case when v_{11} is $N(0,1)$. Then it is well known that O_n is Haar distributed on $\mathcal{O}_n$. From this fact the conjecture was raised in [9], [10] that O_n for general v_{11} is "close" to being Haar distributed for n large, or equivalently the measure O_n induces on $\mathcal{O}_n$ approaches Haar measure in some sense as $n \to \infty$.

Since it is difficult to make statements about measures defined on different spaces, mappings were considered from the orthogonal groups $\{\mathcal{O}_n\}$ onto a common space, inducing a sequence of measures on the space. The limiting behavior of these measures should be known for $v_{11} = N(0,1)$. The conjecture could then be formulated in terms of whether the same limiting behavior holds for other distributions on v_{11}.

Consider the following mappings into $D[0,1]$, the space of right continuous functions with left-hand limits (e.g., [2], p. 109) on $[0,1]$:

Choose $x_n \in \mathbb{R}^n$, $||x_n|| = 1$ (nonrandom), Let $(y_1, y_2, \ldots, y_n)^T = O_n^T x_n$ and define $X_n \in D[0,1]$ as

$$X_n(t) = (n/2)^{1/2}\left(\sum_{i=1}^{[nt]} (y_i^2 - \tfrac{1}{n})\right) \quad ([\ ] \equiv \text{greatest integer function}).$$

These mappings are considered mainly because they carry over much of the uniformity of Haar measure, they are a natural extension of mappings studied earlier ([9], [10]), and because their limiting behavior is known, namely

$$X_n \xrightarrow{\mathcal{D}} W^0 \quad \text{as} \quad n \to \infty$$

where W^0 is Brownian bridge and $\mathcal{D}$ denotes weak convergence of random elements in $D[0,1]$ ([2]). This property follows from the fact that when O_n is Haar distributed, $O_n^T x_n$ is uniformly distributed on the unit sphere in $\mathbb{R}^n$, which in turn can be represented by normalizing a vector of i.i.d. $N(0,1)$ random variables. The basic theory of weak convergence in $D[0,1]$ can then be applied.

Under the assumption $E(|v_{11}|^{4+\delta}) < \infty$ (ensuring the convergence i.p. of the largest eigenvalue) it follows that $X_n \xrightarrow{\mathcal{D}} W^0$ is equivalent to

$$X_n(F_n(x)) \xrightarrow{\mathcal{D}} W_x^y \equiv W^0_{F_y(x)} \quad \text{as} \quad n \to \infty$$

on $D[0,\infty)$ ([7]) where $W^0_{F_y(x)}$ is Brownian bridge composed with $F_y(x)$ (we remark that weak convergence on $D[0,\infty)$ is equivalent to weak convergence on $D[0,b]$ (under the natural projection) for every $b > 0$).

It has been my goal to see whether $X_n(F_n(x)) \xrightarrow{\mathcal{D}} W_x^y$ holds more generally. Results consistent with this property holding, an important necessary condition, and a partial answer have been determined, and will be given below.

Previous work before considering X_n is summarized in

THEOREM 2. ([9], [10], [11]) If all moments of v_{11} exist, then for every

$$t \in [0,1], \quad \frac{X_n(t)}{\sqrt{n}} \xrightarrow{\text{a.s.}} 0 \quad \text{as} \quad n \to \infty \quad \text{for any } \{x_n\} \text{ with } ||x_n|| = 1.$$

This result is like a Law of Large Numbers. It indicates to some degree the uniformity of $O_n^T x_n$ over the unit sphere in $\mathbb{R}^n$.

From the next result we see that at least one additional condition must be satisfied on the distribution of v_{11}.

THEOREM 3 ([10]) If the maximum eigenvalue of M_n converges i.p. to $(1+\sqrt{y})^2$ and if $X_n \xrightarrow{\mathcal{D}} W^o$ for $x_n = (1,0,0,\ldots,0)^T$, then $E(v_{11}^4) = 3$.

Therefore some further similarity of the distribution of v_{11} to $N(0,1)$ is required.

The following theorem is like a Central Limit Theorem. It demonstrates again that there is some invariant behavior of the eigenvectors of M_n beyond Theorem 2, provided $E(v_{11}^4) = 3$.

THEOREM 4 ([11]) If $E(v_{11}^4) = 3$ and all moments of v_{11} exist, then for any $\{x_n\}$ with $||x_n|| = 1$,

$$(*) \qquad \left\{\int_0^\infty f_i(x)dX_n(F_n(x))\right\}_{i=1}^\infty \xrightarrow{\mathcal{D}} \left\{\int_{(1-\sqrt{y})^2}^{(1+\sqrt{y})^2} f_i(x)dW_x^y\right\}_{i=1}^\infty \quad \text{as } n \to \infty$$

for $f_i(x) = x^i$ (here $\mathcal{D}$ denotes convergence in distribution on $\mathbb{R}^\infty$). Moreover if $E(v_{11}^4) \neq 3$, then sequences $\{x_n\}$, $||x_n|| = 1$, exist for which $\left\{\int_0^\infty x^i dX_n(F_n(x))\right\}_{i=1}^\infty$ fails to converge in distribution to anything.

If $E(v_{11}^4) = 3$ and $E(|v_{11}|^m) \leq m^{\alpha m}$, $m = 1,2,\ldots$ and some α, then for any $\{x_n\}$, $||x_n|| = 1$, (*) holds for f_i analytic at 0, $f_i(0) = 0$, with radius of convergence greater than $(1+\sqrt{y})^2$.

The conclusions of Theorems 2 and 4 follow from the truth of $X_n \xrightarrow{\mathcal{D}} W^o$, so these theorems support this property. The theorems also demonstrate invariant behavior of the mappings $O_n \xrightarrow{\mathcal{D}} X_n/\sqrt{n}$ ($\in D[0,1]$) and the left side of (*) ($\in \mathbb{R}^\infty$) which lend support in itself to the similarity of O_n to Haar measure for n large.

The first part of Theorem 4 yields uniqueness of any weak limit X of a subsequence on $[0,b]$ provided $P(X \in C[0,b]) = 1$. Then $X_n(F_n(x)) \xrightarrow{\mathcal{D}} W_x^y$ would follow from tightness and the weak limit of any convergent subsequence being continuous.

We conclude with the partial answer promised earlier. Essentially, $X_n(F_n(x)) \xrightarrow{\mathcal{D}} W_x^y$ is now known to hold on a class of nongaussian v_{11}'s, but only for certain unit vectors $\{x_n\}$.

Recent work has shown the first part of Theorem 4 to be true under the condition $E(|v_{11}|^8) < \infty$. Also, by extending the results in Section 12 of [2], a certain tightness criterion has been established. With it I have proved:

If v_{11} is symmetrically distributed about 0, $E(v_{11}^4) = 3$ and $E(|v_{11}|^8) < \infty$, then $X_n(F_n(x)) \xrightarrow{\mathcal{D}} W_x^y$ follows for

$$x_n = \left(\pm \frac{1}{\sqrt{n}}, \pm \frac{1}{\sqrt{n}}, \ldots, \pm \frac{1}{\sqrt{n}}\right)^T.$$

I hope that subsequent work will show when $X_n(F_n(x)) \xrightarrow{\mathcal{D}} W_x^y$ more generally.

REFERENCES

[1] Arharov, L. V., "Limit theorems for the characteristic roots of a sample covariance matrix." Soviet Math. Dokl. 12(1971), pp. 1206-1209.

[2] Billingsley, P., Convergence of Probability Measures, Wiley, New York, 1968.

[3] Geman, S., "A limit theorem for the norm of random matrices." Ann. Probab. 8(1980), pp. 252-261.

[4] Girko, V. L., "The central limit theorem for random determinants." Theory Prob. Applic. 24(1979), pp. 729-740.

[5] Grenander, U. and Silverstein, J. W., "Spectral analysis of networks with random topologies." SIAM J. Appl. Math. 32(1977), pp. 499-519.

[6] Jonsson, D., "Some limit theorems for the eigenvalues of a sample covariance matrix." J. Multivariate Anal., 12(1982), pp. 1-38.

[7] Lindvall, T., "Weak convergence of probability measures and random functions in the function space D[0,∞)." J. Appl. Probab. 10(1973), pp. 109-121.

[8] Marcenko, V. A. and Pastur, L. A., "Distribution of eigenvalues for some sets of random matrices." Math. USSR-Sb. 1(1967), pp. 457-483.

[9] Silverstein, J. W., "On the randomness of eigenvectors generated from networks with random topologies." SIAM J. Appl. Math., 37(1979), pp. 235-245.

[10] ______________, "Describing the behavior of eigenvectors of random matrices using sequences of measures on orthogonal groups." SIAM J. Math. Anal., 12(1981), pp. 274-281.

[11] ______________, "Some limit theorems on the eigenvectors of large dimensional sample covariance matrices." J. Multivariate Anal., 15(1984), pp. 295-324.

[12] ______________, "Comments on a result of Yin, Bai, and Krishnaiah for large dimensional multivariate F matrices." J. Multivariate Anal., 15(1984), pp. 408-409.

[13] ______________, "The limiting eigenvalue distribution of a multivariate F matrix." SIAM J. Math. Anal., 16(1985), pp. 641-646.

[14] ______________, "The smallest eigenvalue of a large dimensional Wishart matrix." Ann. Probab. (to appear).

[15] ______________, "On the largest eigenvalue of a large dimensional sample covariance matrix." (submitted).

[16] Wachter, K. W., "The strong limits of random matrix spectra for sample matrices of independent elements." Ann. Probab. 6(1978), pp. 1-18.

[17] Yin, Y. Q., Bai, Z. D., and Krishnaiah, P. R., "Limiting behavior of the eigenvalues of a multivariate F matrix." J. Multivariate Anal., 13(1983), pp. 508-516.

[18] Yin. Y. Q. and Krishnaiah, P. R., "A limit theorem for the eigenvalues of product of two random matrices." J. Multivariate Anal. 13(1983), pp. 489-507.

DEPARTMENT OF MATHEMATICS
NORTH CAROLINA STATE UNIVERSITY
BOX 8205
RALEIGH, N. C. 27695-8205