The Annals of Probability
1985, Vol. 13, No. 4, 1364–1368

THE SMALLEST EIGENVALUE OF A LARGE DIMENSIONAL WISHART MATRIX

BY JACK W. SILVERSTEIN[1]

Weizmann Institute of Science

For positive integers s, n let $M_s = (1/s)V_sV_s^T$, where V_s is an $n \times s$ matrix composed of i.i.d. $N(0,1)$ random variables. Assume $n = n(s)$ and $n/s \to y \in (0,1)$ as $s \to \infty$. Then it is shown that the smallest eigenvalue of M_s converges almost surely to $(1 - \sqrt{y})^2$ as $s \to \infty$.

For each $s = 1, 2\dots$ let $n = n(s)$ be a positive integer such that $n/s \to y > 0$ as $s \to \infty$. Let V_s be an $n \times s$ matrix whose entries are i.i.d. $N(0,1)$ random variables and let $M_s = (1/s)V_sV_s^T$. The random matrix $V_sV_s^T$ is commonly referred to as the Wishart matrix $W(I_n, s)$.

It is well known [Marĉenko and Pastur (1967), Wachter (1978)] that the empirical distribution function F_s of the eigenvalues of M_s $[F_s(x) \equiv (1/n) \times$ (number of eigenvalues of $M_s \le x$)] converges almost surely as $s \to \infty$ to a nonrandom probability distribution function F_y having a density with positive support on $[(1 - \sqrt{y})^2, (1 + \sqrt{y})^2]$, and when $y > 1$, F_y yields additional mass on $\{0\}$. It is also known [Geman (1980)] that the maximum eigenvalue $\lambda_{\max}^{(s)}$ of M_s converges a.s. to $(1 + \sqrt{y})^2$ as $s \to \infty$. [The statement of this result in Geman (1980) has all the M_s constructed from one doubly infinite array of i.i.d. random variables. However, it is obvious from the proof that no relation on the entries of V_s for different s is needed.] These results are established under assumptions more general on the entries of V_s than Gaussian distributed, involving conditions on the moments of these random variables.

The present paper will prove the following

THEOREM. *For $y < 1$ the smallest eigenvalue $\lambda_{\min}^{(s)}$ of M_s converges a.s. to $(1 - \sqrt{y})^2$ as $s \to \infty$.*

The proof relies on Gerŝgorin's theorem [Gerŝgorin (1931)] which states: Each eigenvalue of an $n \times n$ complex matrix $A = (a_{ij})$ lies in at least one of the disks

$$|z - a_{jj}| \le \sum_{i \ne j} |a_{ij}|, \qquad j = 1, 2, \dots, n,$$

in the complex plane.

Gerŝgorin's theorem will be applied to a tridiagonal matrix orthogonally similar to M_s. This result is relevant to areas in multivariate statistics, for example regression or tests using the central multivariate F matrix, where the

Received May 1984; revised November 1984.

[1]On leave from North Carolina State University. Supported under NSF Grant MCS-8101703A01.

AMS 1980 *subject classifications.* Primary 60F15; secondary 62H99.

Key words and phrases. Smallest eigenvalue of random matrix, Gerŝgorin's theorem, χ^2 distribution.

boundedness of the largest eigenvalue of M_s^{-1}, namely $[\lambda_{\min}(s)]^{-1}$, is needed. The truth of the theorem for non-Wishart matrices would also be important. However, as will be seen, the proof relies strongly on the variables being normal, so a different method appears to be necessary for more general sample covariance matrices.

PROOF OF THE THEOREM. Since F_y has positive support to the right of $(1-\sqrt{y})^2$ we immediately have

$$\limsup_{s\to\infty} \lambda_{\min}^{(s)} \le \left(1-\sqrt{y}\right)^2 \quad \text{a.s.} \tag{1}$$

Assume s is sufficiently large so that $n < s$. Let O_s^1 be $s \times s$ orthogonal, its first column being the normalization of the first row of V_s, the remaining columns independent of the rest of V_s. The columns of O_s^1 can be constructed, for example, by performing the Gram–Schmidt orthonormalization process to the first row of V_s, together with $s-1$ linearly independent nonrandom s-dimensional vectors. We have that $V_s^1 \equiv V_s O_s^1$ is such that its first row is $(X_s, 0, 0, \dots, 0)$, where X_s^2 is $\chi^2(s)$, $X_s \ge 0$, and the remaining rows are again made up of i.i.d. $N(0,1)$ random variables. (It will also follow that X_s is independent of the remaining elements of V_s^1 but this fact will not be needed.)

Let O_n^1 be $n \times n$ orthogonal of the form

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & O_{n-1}^1 & \\ 0 & & & \end{pmatrix},$$

where O_{n-1}^1 is orthogonal, its first row being the normalization of $\{(V_s^1)_{j1}\}_{j=2}^n$ (as a vector in $\mathbb{R}^{n-1}$), the rest independent of V_s^1. Then $V_s^2 \equiv O_n^1 V_s^1$ is of the form

$$\begin{pmatrix} X_s & 0 & \cdots & 0 \\ Y_{n-1} & & & \\ 0 & & & \\ \vdots & & W_{n-1,s-1} & \\ 0 & & & \end{pmatrix},$$

where Y_{n-1}^2 is $\chi^2(n-1)$, $Y_{n-1} \ge 0$ and $W_{n-1,s-1}$ is $(n-1)\times(s-1)$, made up of i.i.d. $N(0,1)$ random variables.

We then multiply V_s^2 on the right by an $s \times s$ orthogonal matrix O_s^2 of the form

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & & & & \\ 0 & & & & \\ \vdots & & O_{s-1}^2 & & \\ 0 & & & & \end{pmatrix},$$

where the first column of O_{s-1}^2 is the normalization of the first row of $W_{n-1,s-1}$, and then multiply $V_s^2 O_s^2$ on the left by an appropriate $n \times n$ orthogonal matrix, and so on. In the end we will have the existence of two orthogonal matrices O_n and O_s such that

$$O_n V_s O_s = \begin{pmatrix} X_s & 0 & 0 & 0 & & & & & \cdots & 0 \\ Y_{n-1} & X_{s-1} & 0 & 0 & & & & & \cdots & 0 \\ 0 & Y_{n-2} & X_{s-2} & 0 & & & & & \cdots & 0 \\ 0 & 0 & \vdots & \vdots & & & & & \cdots & 0 \\ \vdots & \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & Y_1 & X_{s-(n-1)} & 0 & \cdots & 0 \end{pmatrix},$$

where X_i^2 is $\chi^2(i)$, $X_i \ge 0$, and Y_j^2 is $\chi^2(j)$, $Y_j \ge 0$. The fact that these random variables are independent will not be needed.

It follows that M_s is orthogonally similar to a tridiagonal matrix, the first and last rows being, respectively,

$$(1/s)\left(X_s^2, X_s Y_{n-1}, 0, \dots, 0\right),$$
$$(1/s)\left(0, 0, \dots, 0, X_{s-n+2}Y_1, Y_1^2 + X_{s-n+1}^2\right),$$

while the three nonzero elements in the $j+1$st row ($j = 1, 2, \dots, n-2$) are

$$(1/s)\left(X_{s-j+1}Y_{n-j}, Y_{n-j}^2 + X_{s-j}^2, X_{s-j}Y_{n-j-1}\right).$$

By Gerŝgorin's theorem we have that

$$\lambda_{\min}^{(s)} \ge \min\Big[(1/s)\left(X_s^2 - X_s Y_{n-1}\right), (1/s)\left(Y_1^2 + X_{s-n+1}^2 - X_{s-n+2}Y_1\right),$$
$$\min_{j \le n-2} (1/s)\left(Y_{n-j}^2 + X_{s-j}^2 - \left(X_{s-j+1}Y_{n-j} + X_{s-j}Y_{n-j-1}\right)\right)\Big]. \tag{2}$$

We have $\chi^2(1)/m \to_{\text{a.s.}} 0$ and $\chi^2(m)/m \to_{\text{a.s.}} 1$ as $m \to \infty$. Since n/s $\cancel{s/n} \to y \in (0,1)$ as $s \to \infty$ we have

$$(1/s)\left(X_s^2 - X_s Y_{n-1}\right) \to_{\text{a.s.}} 1 - \sqrt{y},$$
$$(1/s)\left(Y_1^2 + X_{s-n+1}^2 - X_{s-n+2}Y_1\right) \to_{\text{a.s.}} 1 - y \quad \text{as } s \to \infty.$$

Notice $1 - y > 1 - \sqrt{y} > (1 - \sqrt{y})^2$.

Applying Markov's inequality to $P(\exp(t\chi^2(m) - tm) > \exp(ts\varepsilon))$ and $P(\exp(-t\chi^2(m) + tm) > \exp(ts\varepsilon))$ for sufficiently small $t > 0$, it is straightforward to show for any $\varepsilon > 0$ the existence of an $a \in (0,1)$ depending only on ε such that

$$P\left(\left|\left(\chi^2(m)/s\right) - (m/s)\right| > \varepsilon\right) \le 2a^s$$

for all $s > 0$ and all positive integers $m \le s$.

Therefore we can apply Boole's inequality on $2n - 2$ ($\le$ constant $\cdot\, s$) events to conclude that for any $\varepsilon > 0$

$$P\left(\max_{s-(n-2) \le m \le s} \left|\left(X_m^2/s\right) - m/s\right| > \varepsilon \text{ or } \max_{m \le n-1} \left|\left(Y_m^2/s\right) - m/s\right| > \varepsilon\right)$$

is summable. Therefore

$$\max\left[\max_{s-(n-2)\le m\le s}\left|\left(X_m^2/s\right)-m/s\right|,\ \max_{m\le n-1}\left|\left(Y_m^2/s\right)-m/s\right|\right]\to_{\text{a.s.}}0\quad\text{as } s\to\infty.$$

We have

$$\begin{aligned}A_j^s\equiv\Big|&(1/s)\left(Y_{n-j}^2+X_{s-j}^2-\left(X_{s-j+1}Y_{n-j}+X_{s-j}Y_{n-j-1}\right)\right)\\&-\Big((n-j)/s+(s-j)/s-\Big(\sqrt{(s-j+1)/s}\sqrt{(n-j)/s}\\&\qquad+\sqrt{(s-j)/s}\sqrt{(n-j-1)/s}\Big)\Big)\Big|\\\le{}&\left|\left(Y_{n-j}^2/s\right)-(n-j)/s\right|+\left|\left(X_{s-j}^2/s\right)-(s-j)/s\right|\\&+\left|\left(X_{s-j+1}/\sqrt{s}\right)\left(Y_{n-j}/\sqrt{s}\right)-\sqrt{(s-j+1)/s}\sqrt{(n-j)/s}\right|\\&+\left|\left(X_{s-j}/\sqrt{s}\right)\left(Y_{n-j-1}/\sqrt{s}\right)-\sqrt{(s-j)/s}\sqrt{(n-j-1)/s}\right|.\end{aligned}$$

Using the inequality $|\underline{a}\underline{b}-ab|\le|\underline{a}^2-a^2|^{1/2}|\underline{b}^2-b^2|^{1/2}+|a|\,|\underline{b}^2-b^2|^{1/2}+|b|\,|\underline{a}^2-a^2|^{1/2}$ for a, b, $\underline{a}$, $\underline{b}$ nonnegative, together with the fact that the nonrandom fractions making up A_j^s are bounded by 1, we conclude that

$$\max_{j\le n-2}A_j^s\to_{\text{a.s.}}0\quad\text{as } s\to\infty.$$

The expression

$$\begin{aligned}(n-j)/s+(s-j)/s-\Big(&\sqrt{(s-j+1)/s}\sqrt{(n-j)/s}\\&+\sqrt{(s-j)/s}\sqrt{(n-j-1)/s}\Big)\end{aligned}$$

achieves its smallest value when $j=1$, for which we get

$$\begin{aligned}&(n-1)/s+(s-1)/s-\left(\sqrt{(n-1)/s}+\sqrt{(s-1)/s}\sqrt{(n-2)/s}\right)\\&\quad\to y+1-2\sqrt{y}=\left(1-\sqrt{y}\right)^2\quad\text{as } s\to\infty.\end{aligned}$$

Therefore, from (2) we have

$$\liminf_{s\to\infty}\lambda_{\min}^{(s)}\ge\left(1-\sqrt{y}\right)^2\quad\text{a.s.}$$

which, together with (1) gives us

$$\lim_{s\to\infty}\lambda_{\min}^{(s)}=\left(1-\sqrt{y}\right)^2\quad\text{a.s.}\qquad\square$$

We note that the above proof can easily be modified to show $\lambda_{\max}^{(s)}\to(1+\sqrt{y})^2$ for all $y>0$.

REFERENCES

[1] GEMAN, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.* **8** 252–261.

[2] GERŠGORIN, S. A. (1931). Über die Abgrenzung der Eigenwerte einer Matrix. *Izv. Akad. Nauk SSSR Ser. Fiz.-Mat.* **6** 749–754.

[3] MARĈENKO, V. A. and PASTUR, L. A. (1967). Distributions of eigenvalues of some sets of random matrices. *Math. USSR-Sb.* **1** 507–536.

[4] WACHTER, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* **6** 1–18.

JACK W. SILVERSTEIN
DEPARTMENT OF MATHEMATICS
BOX 8205
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27695-8205