Eigenvalues of Large Sample Covariance Matrices of Spiked Population Models

Jinho Baik[*] and **Jack W. Silverstein**[†]

July 15, 2005

Abstract

We consider a spiked population model, proposed by Johnstone, in which all the population eigenvalues are one except for a few fixed eigenvalues. The question is to determine how the sample eigenvalues depend on the non-unit population ones when both sample size and population size become large. This paper completely determines the almost sure limits of the sample eigenvalues in a spiked model for a general class of samples.

1 Introduction

The sample covariance matrix is fundamental to multivariate statistics. When the population size is fixed, as the number of samples tends to infinity, the sample covariance matrix is a good approximate of the population covariance matrix. However when the population size is large and comparable with the sample size, as is in many contemporary data, it is known that the sample covariance matrix is no longer a good approximation to the covariance matrix. A consequence of the main result in [17], along with refinements done in [30], show that with $n =$ the sample size, $p =$ the population size, as $n = n(p) \to \infty$ such that $\frac{p}{n} \to c$, the eigenvalues $s_j^{(p)}$, $j = 1, \dots, p$, of the sample covariance matrix of standardized i.i.d. Gaussian samples satisfy for any real x

$$\frac{1}{p}\#\{s_j^{(p)} : s_j^{(p)} < x\} \to F(x) \tag{1.1}$$

almost surely where

$$F'(x) = \frac{1}{2\pi xc}\sqrt{(b-x)(x-a)}, \qquad a < x < b, \tag{1.2}$$

and $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$ when $0 < c \le 1$. When $c > 1$, there is an additional Dirac measure at $x = 0$ of mass $1 - \frac{1}{c}$. Moreover, there are no stray eigenvalues in the sense that the largest and smallest eigenvalues converge to the edges of the support of F [10]:

$$s_1^{(p)} \to (1+\sqrt{c})^2 \tag{1.3}$$

[*]Department of Mathematics, University of Michigan, Ann Arbor, MI, 48109, USA, baik@umich.edu

[†]Department of Mathematics, North Carolina State University, Raleigh, NC, 27695, USA, jack@math.ncsu.edu

AMS 1991 *subject classifications.* Primary 15A52, 60F15; secondary 62H99.

Key words and phrases. Eigenvalues, sample covariance matrices, spiked population models, almost sure limits, non-null case.

almost surely and [22]

$$s^{(p)}_{\min\{p,n\}} \to (1-\sqrt{c})^2 \tag{1.4}$$

almost surely ($s^{(p)}_{n+1} = \dots s^{(p)}_p = 0$ when $n < p$). Such results apply to more general samples other than Gaussian (see e.g. [31], [32],[2]). One piece of information that can be extracted from this is that if there are non-zero eigenvalues of the sample covariance matrix well separated from the rest of the sample eigenvalues, one can infer that the samples are not i.i.d.

In many examples, indeed, a few eigenvalues of the sample covariance matrix are separated from the rest of the eigenvalues, the latter being packed together as in the support of the Marchenko-Pastur density (1.2). Examples include speech recognition [8, 14] , mathematical finance [20], [15], [16], wireless communications [27], physics of mixture [21], and data analysis and statistical learning [12]. The above results provide strong evidence that the samples have non-null covariance. Then it is a natural question to ask whether it is possible to determine which non-null population model can possibly result in the observed few sample eigenvalues separated from the Marchenko-Pastur density.

The simplest non-null case would be when the population covariance is a finite rank perturbation of a multiple of the identity matrix. In other words, all but finitely many eigenvalues of the population covariance matrix are the same, say equal to 1. Such a population model has been called the 'spiked population model': a null or purely noise model "spiked" with a few significant eigenvalues. The study of spiked models was proposed by Johnstone [14]. The question is how the eigenvalues of the sample covariance matrix would depend on the non-unit population eigenvalues as $p, n \to \infty$, as, for example, a few large population eigenvalues would possibly pull up a few sample eigenvalues. It is known [17, 24] that the Marchenko-Pastur result (1.1) still holds for the spiked model. But (1.3) and (1.4) are not guaranteed and some of the eigenvalues are not necessarily in the support of (1.2). In other words, there might be stray eigenvalues.

For example, suppose that the population covariance matrix has one non-unit eigenvalue, denoted by σ_1. If σ_1 is close to 1, one would expect that as the dimension p becomes large the population covariance matrix would be close to a large identity matrix, and hence σ_1 would have little effect on the eigenvalues of the sample covariance matrix. On the other hand, if σ_1 is much bigger than 1, then even if p becomes large, σ_1 might still pull up the eigenvalues of the sample covariance matrix. How big should σ_1 be in order to have any effect, how many eigenvalues of the sample covariance matrix would be pulled up and exactly where would the pulled-up eigenvalues be? We will see in the results below that the answers are $\sigma_1 > 1+\sqrt{c}$ (where $\frac{p}{n} \to c$), one eigenvalue at most, and $\sigma_1 + \frac{c\sigma_1}{\sigma_1 - 1}$, respectively.

The purpose of this paper is to provide a complete study of the almost sure limits of the sample eigenvalues in the spiked model *for a general class of samples* which are either real or complex and are not necessarily Gaussian, when both population size and sample size tend to infinity with finite ratio. Specializing to the Gaussian samples, our results obtain the almost sure limits of the eigenvalues of the Wishart matrix for the case when the covariance is a finite rank perturbation of the identity matrix. In particular, given non-unit population eigenvalues, we determine how many stray sample eigenvalues the spiked model has and where they are located. For *complex Gaussian* samples, such results were obtained in [19, 6] for the *largest* eigenvalue of the sample covariance matrix. While this paper was being prepared, the authors learned that Debashis Paul [18] obtained results for the spiked model for *real Gaussian* samples independently at the same time, which have some overlap with this paper. For real Gaussian samples when $c < 1$, the almost sure limits as in (1.10) and (1.11) below were obtained for large sample eigenvalues. On the other hand, this paper (i) is concerned with more general samples, not necessarily Gaussian, (ii) includes all choices of c and (iii) studies both large and small sample eigenvalues.

A very general study of the sample eigenvalues with non-null covariance matrix was conducted in [3, 4]. The spiked population model is a particular case of the non-null covariance model. However, in general, to apply the results of [3, 4], one needs to determine all the real roots of a polynomial of high degree (see (2.11) and Lemma 3.1). For the spiked model, we show how to use a perturbation argument in complex analysis to determine the roots of the polynomials with sufficient error bounds.

This paper concerns only the almost sure limits of the sample eigenvalues for the spiked model. It is also of great interest to study the limiting distributions of the sample eigenvalues. See Subsection 1.3 below for a discussion.

1.1 Model

Let T_p be a fixed $p \times p$ non-negative definite Hermitian matrix. Let Z_{ij}, $i, j = 1, 2, \ldots$, be independent and identically distributed complex valued random variables satisfying

$$\mathbb{E}(Z_{11}) = 0, \qquad \mathbb{E}(|Z_{11}|^2) = 1, \quad \text{and} \quad \mathbb{E}(|Z_{11}|^4) < \infty, \tag{1.5}$$

and set $Z_p = (Z_{ij})$, $1 \le i \le p$, $1 \le j \le n$. We take the sampled vectors to be the columns of $T_p^{1/2} Z_p$, where $T_p^{1/2}$ is a Hermitian square root of T_p. Hence T_p is the population covariance matrix. Of course, not all random vectors are realized as such, but this model is still very general. When Z_{ij} are i.i.d (real or complex) Gaussian, the model becomes the Gaussian sample with population covariance matrix T_p. Outside the Gaussian case we see that these vectors cover a broad range of random vectors, completely real or complex, with arbitrary population covariance matrix.

Let

$$B_p := \frac{1}{n} T_p^{1/2} Z_p Z_p' T_p^{1/2} \tag{1.6}$$

be the sample covariance matrix, where Z_p' denotes conjugate transpose. When Z_{11} is Gaussian, B_p is also known as a Wishart matrix. Denote the eigenvalues of B_p by $s_1^{(p)}, \ldots, s_p^{(p)}$: for some unitary matrix U_B,

$$U_B B_p U_B^{-1} = \begin{pmatrix} s_1^{(p)} & & & \\ & s_2^{(p)} & & \\ & & \ddots & \\ & & & s_p^{(p)} \end{pmatrix} = diag(s_1^{(p)}, s_2^{(p)}, \ldots, s_p^{(p)}). \tag{1.7}$$

For definiteness, we order the eigenvalues as $s_1^{(p)} \ge s_2^{(p)} \ge \cdots \ge s_p^{(p)} \ge 0$.

Let $\alpha_1 > \cdots > \alpha_M > 0$ be fixed real numbers for some fixed $M \ge 0$, which is independent of p and n. Let $k_1, \ldots, k_M$ be fixed non-negative integers and set $r = k_1 + \cdots + k_M$, which are also independent of p and n. We assume that all the eigenvalues of T_p are 1 except for, say, the first r eigenvalues. This is the 'spiked population model' proposed in [14]. Let the first r eigenvalues be equal to $\alpha_1, \ldots, \alpha_M$ with multiplicity

$k_1, \ldots, k_M$, respectively: for some unitary matrix U_T,

$$U_T T_p U_T^{-1} = diag(\underbrace{\alpha_1, \ldots, \alpha_1}_{k_1}, \underbrace{\alpha_2, \ldots, \alpha_2}_{k_2}, \ldots, \underbrace{\alpha_M, \ldots, \alpha_M}_{k_M}, \underbrace{1, \ldots, 1}_{p-r}). \tag{1.8}$$

We set $k_0 = 0$.

1.2 Results

Theorem 1.1 (case $c < 1$)**.** *Assume that $n = n(p)$ and $p \to \infty$ such that*

$$\frac{p}{n} \to c \tag{1.9}$$

for a constant $0 < c < 1$. Let M_0 be the number of j's such that $\alpha_j > 1 + \sqrt{c}$, and let $M - M_1$ be the number of j's such that $\alpha_j < 1 - \sqrt{c}$. Then the following holds.

- *For each $1 \le j \le M_0$,*

$$s^{(p)}_{k_1+\cdots+k_{j-1}+i} \to \alpha_j + \frac{c\alpha_j}{\alpha_j - 1}, \qquad 1 \le i \le k_j. \tag{1.10}$$

  almost surely.

- $$s^{(p)}_{k_1+\cdots+k_{M_0}+1} \to (1 + \sqrt{c})^2 \tag{1.11}$$

  almost surely.

- $$s^{(p)}_{p-r+k_1+\cdots+k_{M_1}} \to (1 - \sqrt{c})^2 \tag{1.12}$$

  almost surely (recall $r = k_1 + \cdots + k_M$).

- *For each $M_1 + 1 \le j \le M$,*

$$s^{(p)}_{p-r+k_1+\cdots+k_{j-1}+i} \to \alpha_j + \frac{c\alpha_j}{\alpha_j - 1}, \qquad 1 \le i \le k_j \tag{1.13}$$

  almost surely.

Therefore, when all non-unit population eigenvalues are 'close to 1' (i.e. when $M_0 = 0, M_1 = M$), Marchenko-Pastur density is not disturbed and no sample eigenvalues have almost sure limits outside the support $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$ of the Marchenko-Pastur density. The quantitative measure for the population eigenvalues to be 'close to 1' turns out to be that population eigenvalues are in the interval $[1-\sqrt{c}, 1+\sqrt{c}]$. When there are population eigenvalues outside $[1-\sqrt{c}, 1+\sqrt{c}]$, precisely the same number of sample eigenvalues are outside the support $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$ of the Marchenko-Pastur density. Each population eigenvalue α outside $[1-\sqrt{c}, 1+\sqrt{c}]$ pulls one sample eigenvalue from the support $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$ of the Marchenko-Pastur density and positions it at $\alpha + \frac{c\alpha}{\alpha-1}$ in the limit.

As an example, when $r = 1$, by denoting the only non-unit eigenvalue by σ_1, the largest sample eigenvalue $s_1^{(p)}$ satisfies

$$s_1^{(p)} \to \begin{cases} (1+\sqrt{c})^2, & \sigma_1 \le 1 + \sqrt{c} \\ \sigma_1 + \frac{c\sigma_1}{\sigma_1 - 1}, & \sigma_1 > 1 + \sqrt{c} \end{cases} \tag{1.14}$$

almost surely. When $r = 2$, by denoting the two non-unit eigenvalues by σ_1, σ_2, the largest sample eigenvalue $s_1^{(p)}$ satisfies

$$s_1^{(p)} \to \begin{cases} (1+\sqrt{c})^2, & \max\{\sigma_1, \sigma_2\} \le 1+\sqrt{c} \\ \max\{\sigma_1, \sigma_2\} + \frac{c\max\{\sigma_1,\sigma_2\}}{\max\{\sigma_1,\sigma_2\}-1}, & \max\{\sigma_1, \sigma_2\} > 1+\sqrt{c} \end{cases} \tag{1.15}$$

almost surely.

For complex Gaussian samples, Theorem 1.1 for the largest sample eigenvalue $s_1^{(p)}$ was first obtained in [19, 6]. When the samples are real Gaussian, the results (1.10) and (1.11) were independently obtained in [18].

Theorem 1.2 (case $c > 1$). *Assume that $n = n(p)$ and $p \to \infty$ such that*

$$\frac{p}{n} \to c \tag{1.16}$$

for a constant $c > 1$. Let M_0 be the number of j's such that $\alpha_j > 1 + \sqrt{c}$. Then the following holds.

- *For each $1 \le j \le M_0$,*
$$s^{(p)}_{k_1+\cdots+k_{j-1}+i} \to \alpha_j + \frac{c\alpha_j}{\alpha_j - 1}, \qquad 1 \le i \le k_j. \tag{1.17}$$
almost surely.

- $$s^{(p)}_{k_1+\cdots+k_{M_0}+1} \to (1+\sqrt{c})^2 \tag{1.18}$$
almost surely.

- $$s^{(p)}_n \to (1-\sqrt{c})^2 \tag{1.19}$$
almost surely.

- *For all p,*
$$s^{(p)}_{n+1} = \cdots = s^{(p)}_p = 0. \tag{1.20}$$

Thus, unlike the case of $c < 1$, small eigenvalues of T_p do not affect the eigenvalues of B_p when $c > 1$.

Theorem 1.3 (case $c = 1$). *Assume that $n = n(p)$ and $p \to \infty$ such that*

$$\frac{p}{n} \to 1. \tag{1.21}$$

Let M_0 be the number of j's such that $\alpha_j > 2$. Then the following holds.

- *For each $1 \le j \le M_0$,*
$$s^{(p)}_{k_1+\cdots+k_{j-1}+i} \to \alpha_j + \frac{\alpha_j}{\alpha_j - 1}, \qquad 1 \le i \le k_j. \tag{1.22}$$
almost surely.

- $$s^{(p)}_{k_1+\cdots+k_{M_0}+1} \to 4 \tag{1.23}$$

  almost surely.

- $$s^{(p)}_{\min\{n,p\}} \to 0 \tag{1.24}$$

  almost surely.

Theorem 1.3 for $s_1^{(p)}$ was first obtained in [19, 6] for complex Gaussian samples.

1.3 Discussion

Loosely speaking, the location of the eigenvalues in the spiked model are due to interactions between the non-unit population eigenvalues, which are finite in number, and the unit population eigenvalues whose size tends to infinity. It would be interesting to have a simple heuristic argument which shows how to obtain the critical values $1 \pm \sqrt{c}$ of the population eigenvalues and the location $\alpha_j + \frac{c\alpha_j}{\alpha_j - 1}$ of the pulled sample eigenvalues. For the complex Gaussian case, the paper [6] (see section 6) shows that the distribution of the largest sample eigenvalues is the same as the last passage time in a directed percolation model, and gives a heuristic argument that determines the critical value and the location of the pulled eigenvalues; the interaction is basically a competition between a 1-dimensional last passage time and a 2-dimensional last passage time.

It is also interesting to consider the limiting distributions of the eigenvalues. For the null case when T_p is the identity matrix, under the Gaussian assumption, the limiting distribution for the largest eigenvalue is obtained for the complex case in [9, 13] and for the real case in [14]. For the latter case [26] shows that the limiting distribution does not depend on the Gaussian assumption when $c = 1$. The limiting distributions are the Tracy-Widom distributions [28, 29], originating from the mathematical physics side of random matrix theory. For the spiked model with complex Gaussian samples, when $c \leq 1$, the limiting distributions of the largest eigenvalue are obtained in [19, 6]. The paper [6] determines the limiting distribution of $s_1^{(p)}$ for complete choices of the largest population eigenvalue α_1 and its multiplicity k_1: the distribution is (i) the Tracy-Widom distribution when $\alpha_1 < 1 + \sqrt{c}$, (ii) certain generalizations of the Tracy-Widom distribution (see also [5]) when $\alpha_1 = 1 + \sqrt{c}$, and (iii) the Gaussian distribution ($k_1 = 1$) and its generalization ($k_1 \geq 2$, the Gaussian unitary ensemble) when $\alpha_1 > 1 + \sqrt{c}$. For real Gaussian samples [18] showed that when $c < 1$, $M_0 \geq 1$ and $k_1 = \cdots = k_{M_0} = 1$ (i.e. all non-unit population eigenvalues larger than $1 + \sqrt{c}$ are simple), the limiting distribution of the pulled eigenvalues $s_j^{(p)}$, $1 \leq j \leq M_0$, is Gaussian. For the case of when the population eigenvalues are of higher multiplicity, the limiting distributions of all pulled eigenvalues are considered in [7] without the Gaussian assumption. However, it is an open question to determine the limiting distribution of the sample eigenvalue converging to the edge $(1 \pm \sqrt{c})^2$ of the support of the Marchenko-Pastur dentity for real samples. See section 1.3 of [6] for a conjecture on the scaling.

We include several plots for the case when $c = 0.5$ and there are three non-unit population eigenvalues given by 0.1, 3 and 4 (of multiplicity 1 each). In this case, the critical values of the eigenvalues are

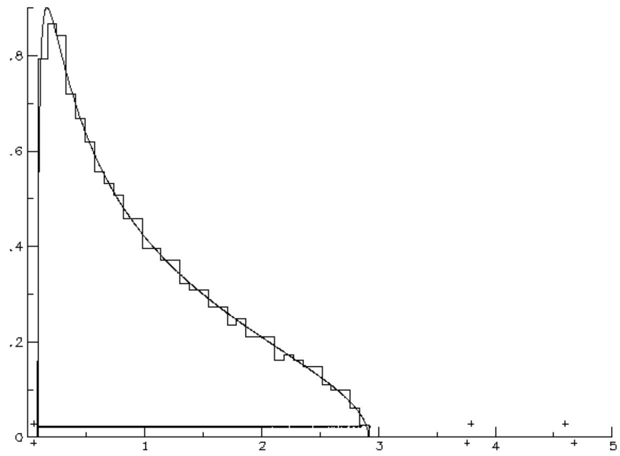

Figure 1: Gaussian samples when $p = 1000, n = 2000$

$1 + \sqrt{c} \simeq 1.70711$ and $1 - \sqrt{c} \simeq 0.29289$. Hence theoretically we expect that three sample covariance eigenvalues of values $\alpha_j + \frac{c\alpha_j}{\alpha_j - 1} \simeq 0.04444, 3.75$ and 4.66667 are away from the interval $[(1 - \sqrt{c})^2, (1 + \sqrt{c})^2] \simeq [0.08578, 2.91422]$. The histogram and the scatterplot of Figure 1 is from Gaussian samples when $p = 1000, n = 2000$. The smooth curve is the theoretical limiting density and the theoretical locations of the three separated eigenvalues are plotted with + signs below the horizontal axis. The smallest and largest two sample eigenvalues are plotted with + signs about the horizontal axis. Figure 2 is from Gaussian samples when $p = 100, n = 200$ while Figures 3 and 4 from samples of Bernoulli variables taking values -1 or 1 when $p = 1000, n = 2000$, and $p = 100, n = 200$ respectively. The observed values of the four separated eigenvalues in each case are as follows:

	smallest eigenvalue	2nd largest eigenvalue	largest eigenvalue
theoretical	0.04444	3.75	4.66667
Gaussian $p = 1000$	0.04369	3.78400	4.59127
Gaussian $p = 100$	0.03979	3.55388	4.66192
Bernoulli $p = 1000$	0.04555	3.75706	4.66594
Bernoulli $p = 100$	0.05015	3.62337	4.70786

Figure 5 and Figure 6 are the cases when $c = 2$, $p = 2000, n = 1000$ with Gaussian and Bernoulli samples, respectively. Again three non-unit population eigenvalues are chosen: 0.1, 3 and 4. The critical value of the eigenvalues is $1 + \sqrt{c} \simeq 2.41421$ and the theory predicts that the two largest sample eigenvalues given by $\alpha_j + \frac{c\alpha_j}{\alpha_j - 1} \simeq 6$ and 6.66667 are separated from the interval $[(1 - \sqrt{c})^2, (1 + \sqrt{c})^2] \simeq [0.17157, 3.41209]$. Only non-zero eigenvalues are plotted in Figure 5 and Figure 6. The observed values of the separated eigenvalues in each case are as follows:

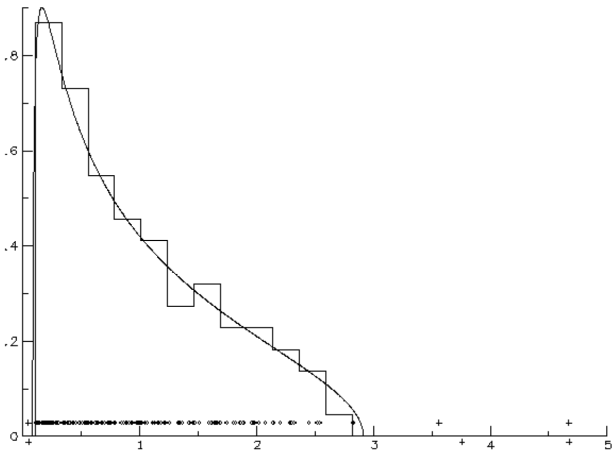

Figure 2: Gaussian samples when $p = 100, n = 200$

	2nd largest eigenvalue	largest eigenvalue
theoretical	6	6.66667
Gaussian $p = 2000$	5.8523	6.4013
Bernoulli $p = 2000$	6.01065	6.725

We mention here one possible application of the results obtained in this paper. Suppose some high dimensional data is believed to be due to a small number of independent (mean zero) factors corrupted by additive noise. However, the number of samples is not large enough to reliably estimate the population matrix. This matrix will have one eigenvalue, say σ^2, with high multiplicity. The remaining eigenvalues, each one corresponding to a factor, will be larger than σ^2. A scatterplot of the sample eigenvalues should reveal a separation with many grouped together in an interval to the left of the others. Using the fact that the mean of a Marchenko-Pastur density is one, the mean of the grouped eigenvalues can be taken as an estimate of σ^2. This estimate can then be scaled out of the sampled eigenvalues, and with c denoting the ratio of vector dimension to sample size, $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$ can be used as an estimate of the interval containing all the "noise" eigenvalues (which should be close to the interval observed to contain all the grouped eigenvalues), the ones outside this interval should correspond to factors. Taking into account the scaling and shifting caused by σ^2, estimates of the "uncorrupted" population eigenvalues (variances of the lengths of the factors) can be made. However, there is no guarantee all important eigenvalues will be detected. Indeed, any sample (scaled) eigenvalue corresponding to a population eigenvalue in $[1, 1+\sqrt{c}]$ will be close, with high probability, to $(1+\sqrt{c})^2$. It is then the decision of the user to either reconcile that population eigenvalues in this interval will never be detected, or, if $\sqrt{c}$ is suitably small, to dismiss those undetected sample eigenvalues as ones coming from insignificant factors (small length variances).

The paper is organized as follows. In section 2, we summarize the work of Z. D. Bai and J. W. Silverstein on which we heavily rely to prove our results. It turns out that the determination of the support of a Stieltjes

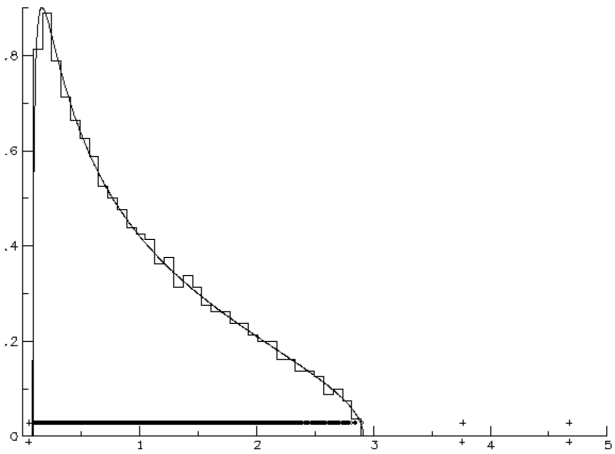

Figure 3: Bernoulli samples taking values -1 or 1 when $p = 1000, n = 2000$

transform plays the crucial role. This is obtained in section 3. The proofs of the main theorems are given in section 4.

Acknowledgments. Special thanks are due to Min Kang for kindly inviting J.B. to give a talk at North Carolina State University where the authors happened to have a chance to discuss about the problem, which eventually lead to this work. We would also like to thank Iain Johnstone for telling us the work of Debashis Paul [18] which was being done independently and at the same time. The work of J.B. was supported in part by NSF Grant #DMS-0350729.

2 Results of Z. D. Bai and J. W. Silverstein

Our analysis replies heavily on the work [3, 4] of Bai and Silverstein. In this section, we summarize the necessary results from [3, 4].

Notational Remark. We denote by p the population size and by n the sample size. The notations n and N are used in [4] for p and n, respectively.

The work [3, 4] is a refinement of the work of Marchenko and Pastur [17]. The key tool in those work is the so-called Stieltjes transform of a distribution. For a distribution function $G(\lambda)$, its Stieltjes transform $m_G(z)$ is defined by

$$m_G(z) = \int_{-\infty}^{\infty} \frac{1}{\lambda - z} dG(\lambda), \qquad z \in \mathbb{C}^+ := \{z \in \mathbb{C} : Im(z) > 0\}. \tag{2.1}$$

Recall the inversion formula

$$G([a, b]) = \frac{1}{\pi} \lim_{\eta \downarrow 0} \int_a^b Im\big(m_G(\xi + i\eta)\big) d\xi \tag{2.2}$$

for continuity points a, b of G.

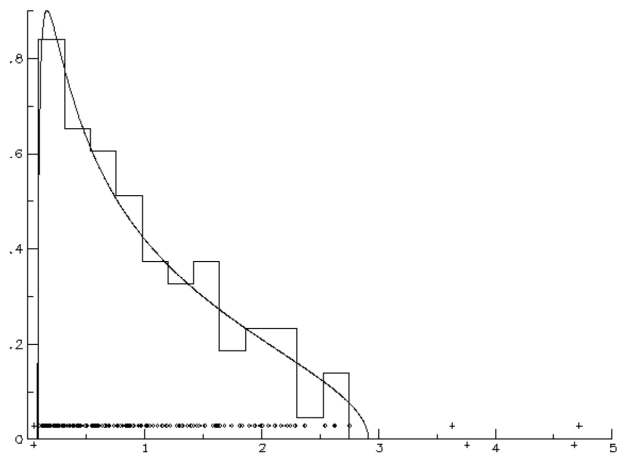

Figure 4: Bernoulli samples taking values -1 or 1 when $p = 100, n = 200$

We first sketch the idea of [17]. Note that the matrix $\underline{B}_p := \frac{1}{n} Z_p^* T_p Z_p$ has precisely the same set of eigenvalues as B_p except for $|p - n|$ zero eigenvalues. Sometimes it is easier to work with $\underline{B}_p$ than B_p. Let $F_{\underline{B}_p}$ denote the distribution function of the eigenvalues of $\underline{B}_p$. Set $\mathbf{m}_p(z)$ be the Stieltjes transform of $F_{\underline{B}_p}$, i.e.

$$\mathbf{m}_p(z) := m_{F_{\underline{B}_p}}(z) = \int_{-\infty}^{\infty} \frac{1}{\lambda - z} dF_{\underline{B}_p}(\lambda), \qquad z \in \mathbb{C}^+. \tag{2.3}$$

Suppose as before that $p, n \to \infty$ such that $\frac{p}{n} \to c$. Also denote by H_p the distribution of the population eigenvalues of T_p and suppose that H_p converges in distribution to a distribution H_∞. The result of [17] is that the random function $\mathbf{m}_p(z) \to m_\infty(z)$ for each $z \in \mathbb{C}^+$, for a non-random function $m_\infty(z)$ which satisfies the equation

$$z = -\frac{1}{m_\infty(z)} + c \int_{-\infty}^{\infty} \frac{t}{1 + t m_\infty(z)} dH_\infty(t), \qquad z \in \mathbb{C}^+. \tag{2.4}$$

It is shown that $m_\infty(z)$ is the Stieltjes transform of a distribution function, which we call F_∞:

$$m_\infty(z) = \int_{-\infty}^{\infty} \frac{1}{\lambda - z} dF_\infty(\lambda), \qquad z \in \mathbb{C}^+. \tag{2.5}$$

For the null-case, and also for the spiked model, H_∞ is the Dirac measure at $t = 1$, i.e. $dH_\infty(t) = \delta_1(t)dt$. In this case, the equation (2.4) becomes

$$z = -\frac{1}{m_\infty(z)} + \frac{c}{1 + m_\infty(z)}. \tag{2.6}$$

By solving the quadratic equation in $m_\infty(z)$, we find

$$m_\infty(z) = \frac{c - 1 - z + \sqrt{(z-a)(z-b)}}{2z} \tag{2.7}$$

with a suitable choice of the square-root, where $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$ as in (1.2). By using the inversion formula (2.2), it can be shown that $m_\infty(z)$ is the Stieltjes transform of the distribution function of

$$F_\infty(x) = (c - 1)1_{[0,\infty)} + F(x) \tag{2.8}$$

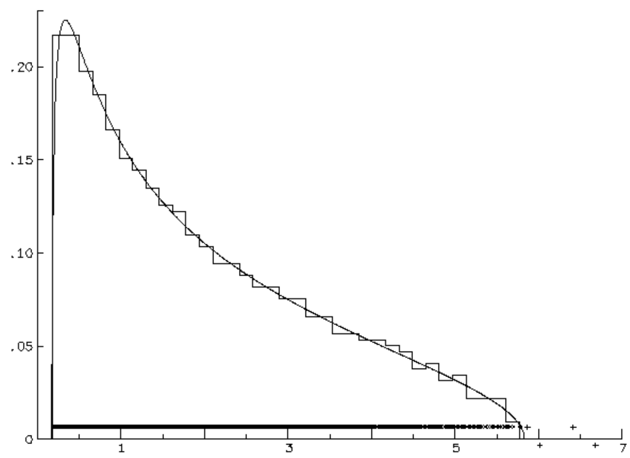

Figure 5: Gaussian samples when $p = 2000, n = 1000$

where $F(x)$ is in (1.2), and that $F_{\underline{B}_p} \to F_\infty$. By using the relation between $\underline{B}_p$ and B_p, (1.1) follows.

For finite n, p, $\mathbf{m}_p(z)$ satisfies, with $c_p := \frac{p}{n}$,

$$z = -\frac{1}{\mathbf{m}_p(z)} + c_p \int_{-\infty}^{\infty} \frac{t}{1 + t\,\mathbf{m}_p(z)} dH_p(t) + \text{error}, \tag{2.9}$$

where 'error' $\to 0$ as $p \to \infty$, from which (2.4) follows. Now [3, 4] derived a sharp estimate on 'error', and then showed that by setting 'error'$= 0$ in (2.9) and solving the resulting equation for $\mathbf{m}_p(z)$, one obtains finer asymptotics of the eigenvalues of $\underline{B}_p$, and hence B_p. Namely, the precise intervals in $\mathbb{R}$ which contains no sample eigenvalues for all sufficiently large p are determined. They are in general subsets of $\text{supp}(F_\infty)^c$. Moreover, for the intervals in $\mathbb{R}$ which contains sample eigenvalues for all large p, the precise number of sample eigenvalues in each set is determined in terms of the solution to the truncated version of the equation (2.9). To state the result of [3, 4], we need some assumptions.

Assume the following:

(a) Z_{ij} are i.i.d. random variables in $\mathbb{C}$ with $\mathbb{E}(Z_{11}) = 0$, $\mathbb{E}|Z_{11}|^2 = 1$ and $\mathbb{E}|Z_{11}|^4 < \infty$.

(b) $n = n(p)$ with $c_p := \frac{p}{n} \to c > 0$ as $p \to \infty$.

(c) For each p, $U_T T_p U_T^{-1} = diag(\sigma_1^{(p)}, \ldots, \sigma_p^{(p)})$ for some unitary matrix U_T such that $H_p \to H_\infty$ in distribution for some distribution function H_∞ where H_p is the empirical distribution function of the eigenvalues of T_p defined by

$$dH_p(\lambda) = \frac{1}{p} \sum_{j=1}^{p} \delta_{\sigma_j^{(p)}}(\lambda). \tag{2.10}$$

(d) $\max\{\sigma_1^{(p)}, \ldots, \sigma_p^{(p)}\}$ is bounded in p.

(e) Set $Z_p = (Z_{ij})$, $1 \le i \le p$, $1 \le j \le n$ and $B_p = \frac{1}{n} T_p^{1/2} Z_p Z_p^* T_p^{1/2}$.

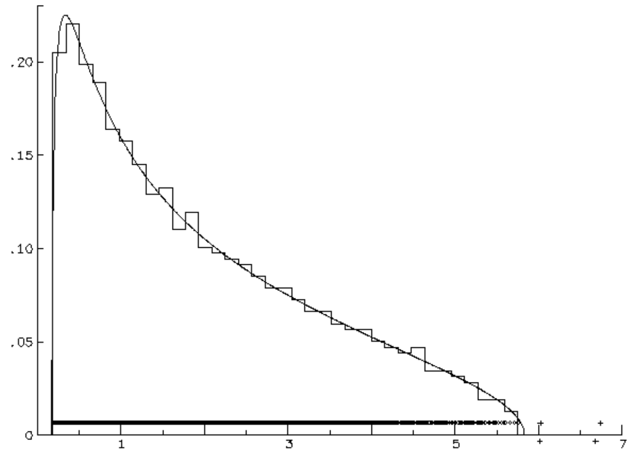

Figure 6: Bernoulli samples taking values -1 or 1 when $p = 2000, n = 1000$

(f) Set (cf. (2.9))

$$z_p(m) = -\frac{1}{m} + c_p \int \frac{t}{1+tm} dH_p(t). \tag{2.11}$$

¿From [24] and [23], it is known that there is a unique inverse function $m_p(z)$ such that $m_p(z) \in \mathbb{C}^+$ for $z \in \mathbb{C}^+$. It is also shown in [24, 23] that $m_p(z)$ is the Stieltjes transform of a distribution, which will be denoted by F_p:

$$m_p(z) = \int_{-\infty}^{\infty} \frac{1}{\lambda - z} dF_p(\lambda), \qquad z \in \mathbb{C}^+. \tag{2.12}$$

Suppose that the interval $[a, b]$ with $a > 0$ lies in an open interval outside $\mathrm{supp}(F_p)$ for all large p.

Remark. We emphasize that the function F_p is not the empirical distribution of $\underline{B}_p$. The distribution function F_p is defined only through (2.12) as the inverse transform of m_p, which solves the truncated version of the equation (2.9) for $\mathbf{m}_p$.

Remark. If $[a, b]$ satisfies condition (f) above, it is easy to check that $[a, b] \subset \mathrm{supp}(F_\infty)^c$, where F_∞ is the inverse Stieltjes transform of m_∞ given in (2.4).

Definition 1. Let $i_p \geq 0$ be the integer satisfying the conditions

$$\sigma_{i_p}^{(p)} > -\frac{1}{m_\infty(b)}, \qquad \sigma_{i_p+1}^{(p)} < -\frac{1}{m_\infty(a)}. \tag{2.13}$$

(Here $\sigma_0^{(p)} := \infty$.)

It is shown in [4] that, given an interval $[a, b]$ satisfying condition (f) above and $m_\infty(b) < 0$, such i_p exists for large p.

Proposition 2.1 (Theorem 1.2 [4])**.** *Assume (a)-(f) above. Let x_0 be the smallest value in the support of F_∞.*

(i) If $c(1 - H_\infty(0)) > 1$, then $x_0 > 0$, and $s_n^{(p)} \to x_0$ with probability 1. The value x_0 is the maximum of the function $z_\infty(m)$ for $m \in \mathbb{R}_+$.

(ii) If $c(1 - H_\infty(0)) \le 1$ or $c(1 - H_\infty(0)) > 1$ but $[a, b]$ is not contained in $[0, x_0]$, then $m_\infty(b) < 0$ and

$$\mathbb{P}\big(s_{i_p}^{(p)} > b \text{ and } s_{i_p+1}^{(p)} < a \text{ for all large } p\big) = 1 \tag{2.14}$$

with i_p defined in (2.13). (Here $s_0^{(p)} := \infty$.)

Therefore for the case of (ii), the i_pth sample eigenvalue and the i_{p+1}th sample eigenvalue are separated by an interval in $\text{supp}(F_p)^c$. Hence this result determines the precise number of the sample eigenvalues in (a small neighborhood of) each interval of $\text{supp}(F_p)^c$.

3 Determination of $\text{supp}(F_p)$

The key part in applying Proposition 2.1 turns out to be determining the support of F_p. This can be extracted from the following result due to Silverstein and Choi.

Lemma 3.1 ([25]; see also Lemma 1.3 [4]). *If $x \in \mathbb{R} \setminus \text{supp}(F_p)$, then $m := m_p(x) = \lim_{\epsilon \downarrow 0} m_p(x + i\epsilon)$ satisfies*

(i) $m \in \mathbb{R} \setminus \{0\}$

(ii) $-\frac{1}{m} \notin \text{supp}(H_p)$

(iii) $z_p'(m) = \lim_{\epsilon \downarrow 0} z'(m + i\epsilon) > 0$.

Conversely, if m satisfies (i)-(iii), then $x = z_p(m) = \lim_{\epsilon \downarrow 0} z(m + i\epsilon) \in \mathbb{R} \setminus \text{supp}(F_p)$.

Note that from (2.12), $m_p'(z) > 0$ for $z \in \mathbb{R} \setminus \text{supp}(F_p)$. Hence $z_p'(m) > 0$ at such points. Therefore $\text{supp}(F_p)^c$ and the points at which $z_p'(m) > 0$ are intimately related. The above Lemma gives the exact relationship.

Remark. Lemma 1.3 of [4] is stated for H_∞. But the proof of Lemma 1.3 in [25] applies also to the finite p case of H_p and c_p without any change. Indeed, the proposition applies to any distribution defined by its Stieltjes transform satisfying (2.12).

Remark. It is also shown in [25] that F_p has continuous density on $\mathbb{R}_+$.

When T_p is as in (1.8),

$$dH_p(x) = \frac{1}{p} \sum_{j=1}^{M} k_j \delta_{\alpha_j}(x) + \left(1 - \frac{r}{p}\right) \delta_1(x) \tag{3.1}$$

and

$$z_p(m) = -\frac{1}{m} + \frac{c_p}{1+m} + \frac{1}{n}\left[\sum_{j=1}^{M} \frac{k_j \alpha_j}{1 + \alpha_j m} - \frac{r}{1+m}\right], \tag{3.2}$$

where we recall that $r = k_1 + \cdots + k_M$. We first determine the set of real m such that $z_p'(m) > 0$.

Now

$$
\begin{aligned}
z_p'(m) &= \frac{1}{m^2} - \frac{c_p}{(1+m)^2} + \frac{1}{n}\left[\sum_{j=1}^{M} \frac{-k_j\alpha_j^2}{(1+\alpha_j m)^2} + \frac{r}{(1+m)^2}\right] \\
&= \frac{f(m) + \frac{1}{n}g(m)}{m^2(1+m)^2\prod_{\ell=1}^{M}(1+\alpha_\ell m)^2},
\end{aligned} \tag{3.3}
$$

where

$$
f(m) := \left((1+m)^2 - c_p m^2\right)\prod_{\ell=1}^{M}(1+\alpha_\ell m)^2 \tag{3.4}
$$

and

$$
g(m) := \left[\sum_{j=1}^{M} \frac{-k_j\alpha_j^2}{(1+\alpha_j m)^2} + \frac{r}{(1+m)^2}\right] m^2(1+m)^2 \prod_{\ell=1}^{M}(1+\alpha_\ell m)^2. \tag{3.5}
$$

We need the following basic lemmas of complex variables to determine the solution of $z_p'(m) = 0$.

Lemma 3.2. *Let $h(z)$ be an analytic function in a closed disk $\overline{D(z_0, r)}$ of radius $r > 0$ centered at z_0. Then there is $\epsilon_0 > 0$ such that for $0 \le \epsilon < \epsilon_0$, the equation*

$$
z - z_0 = \epsilon h(z) \tag{3.6}
$$

has a unique solution in $D(z_0, r)$, which satisfies

$$
z = z_0 + \epsilon h(z_0) + O(\epsilon^2). \tag{3.7}
$$

Furthermore, if z_0 is real and $h(z)$ is real for real z, the solution (3.7) is real.

Proof. As h is continuous, there is a constant $C > 0$ such that $|h(z)| \le C$ for $|z - z_0| \le r$. When $|\epsilon| < \frac{r}{C}$, for $|z - z_0| = r$,

$$
|z - z_0| = r > |\epsilon| C \ge |\epsilon h(z)|. \tag{3.8}
$$

Hence from Rouche's theorem, the number of zeros of $z - z_0 - \epsilon h(z)$ inside $D(z_0, r)$ is equal to the number of zeros of $z - z_0$ inside $D(z_0, r)$, which is one. The zero z_ϵ satisfies $z_\epsilon - z_0 = \epsilon h(z_\epsilon) = O(\epsilon)$. Thus

$$
z_\epsilon - z_0 - \epsilon h(z_0) = \epsilon(h(z_\epsilon) - h(z_0)) = O(\epsilon^2). \tag{3.9}
$$

If z_0 is real and $h(z)$ is real for real z, then by taking the complex conjugate of (3.6), we find that $\overline{z_\epsilon}$ is also a solution. Since there is only one solution, we find that z_ϵ is real. ☐

Lemma 3.3. *Let $h(z)$ be an analytic function in a closed disk $\overline{D(z_0, r)}$ of radius $r > 0$ centered at z_0 such that $h(z_0) \neq 0$. Then there are $0 < r_0 \le r$ and $\epsilon_0 > 0$ such that for $0 \le \epsilon < \epsilon_0$, the equation*

$$
(z - z_0)^2 = \epsilon h(z) \tag{3.10}
$$

has precisely two distinct solutions in $D(z_0, r_0)$, which satisfy

$$
z = z_0 \pm \sqrt{\epsilon}\sqrt{h(z_0)} + O(\epsilon) \tag{3.11}
$$

where $\sqrt{h(z_0)}$ is an arbitrary branch. Furthermore, suppose that z_0 is real and $h(z)$ is real for real z. Then if $h(z_0) > 0$, both solutions (3.11) are real. On the other hand, if $h(z_0) < 0$, both solutions (3.11) are non-real.

Proof. The proof of (3.11) follows from Lemma 3.2 by taking the square root of (3.10). When z_0 is real and $h(z)$ is real for real z, the complex conjugate of a solution of (3.10) is also a solution. Thus the two solutions (3.11) of (3.10) are either complex conjugates of each other or both real since there are precisely two distinct solutions. Hence the Lemma follows. □

For the remainder of this section, we assume that $c \neq 1$ and none of the α_j's are equal to $1 \pm \sqrt{c}$. We further assume that p and n are sufficiently large so that $c_p \neq 1$ and none of the α_j's are equal to $1 \pm \sqrt{c_p}$. Then the numerator of (3.3) is a polynomial of degree exactly $2M+2$, and we now determine all the solutions of $z_p'(m) = 0$.

For f defined in (3.4), the equation $f(m) = 0$ has distinct solutions

$$m = \frac{-1}{1+\sqrt{c_p}} =: m_+, \qquad m = \frac{-1}{1-\sqrt{c_p}} =: m_- \tag{3.12}$$

of multiplicity 1 and

$$m = \frac{-1}{\alpha_j}, \qquad j = 1, 2, \ldots, M, \tag{3.13}$$

of multiplicity 2. The roots of $z_p'(m)$ are expected to be perturbations of the roots of $f(m)$, which we will find. First consider m_+. Dividing the equation $f(m) + \frac{1}{n}g(m) = 0$ by $\frac{f(m)}{m-m_+}$, we obtain the equation

$$m - m_+ + \frac{1}{n}\frac{m^2(1+m)^2}{(1-c_p)(m-m_-)}\left[\sum_{j=1}^{M} \frac{-k_j\alpha_j^2}{(1+\alpha_j m)^2} + \frac{r}{(1+m)^2}\right] = 0. \tag{3.14}$$

Lemma 3.2 implies that there is a solution of $z_p'(m) = 0$ of the form

$$m = m_+ + O\left(\frac{1}{n}\right), \tag{3.15}$$

which is real. Similarly, there is a real solution of $z_p'(m) = 0$ of the form

$$m = m_- + O\left(\frac{1}{n}\right). \tag{3.16}$$

Now consider the root $m = \frac{-1}{\alpha_j}$ of $f(m) = 0$. Dividing $f + \frac{1}{n}g = 0$ by $\frac{f(m)}{(m+\frac{1}{\alpha_j})^2}$, we obtain the equation

$$\left(m + \frac{1}{\alpha_j}\right)^2 = \frac{1}{n}G_j(m) \tag{3.17}$$

where

$$G_j(m) = \frac{-(1+\alpha_j m)^2 m^2 (1+m)^2}{\alpha_j^2(1-c_p)(m-m_+)(m-m_-)}\left[\sum_{\ell=1}^{M} \frac{-k_\ell\alpha_\ell^2}{(1+\alpha_\ell m)^2} + \frac{r}{(1+m)^2}\right]. \tag{3.18}$$

Note that

$$G_j\left(-\frac{1}{\alpha_j}\right) = \frac{k_j(\alpha_j - 1)^2}{\alpha_j^4(1-c_p)(\frac{-1}{\alpha_j} - m_+)(\frac{-1}{\alpha_j} - m_-)} \tag{3.19}$$

is not zero and also $G_j(m)$ is real for real m. Thus Lemma 3.3 implies that there are precisely two solutions of $z_p'(m) = 0$ of the form

$$m = -\frac{1}{\alpha_j} \pm \frac{1}{\sqrt{n}}\sqrt{G_j\left(-\frac{1}{\alpha_j}\right)} + O\left(\frac{1}{n}\right), \qquad j = 1, \ldots, M, \tag{3.20}$$

where the pair for each j are either both real or both non-real depending on the sign of $G_j\big(-\frac{1}{\alpha_j}\big)$.

Now when $c_p < 1$, the condition $G_j\big(-\frac{1}{\alpha_j}\big) > 0$ is equivalent to

$$\frac{-1}{\alpha_j} > m_+ \quad \text{or} \quad \frac{-1}{\alpha_j} < m_-, \tag{3.21}$$

which is the same as

$$\alpha_j > 1 + \sqrt{c_p} \quad \text{or} \quad \alpha_j < 1 - \sqrt{c_p}. \tag{3.22}$$

On the other hand, when $c_p > 1$, we note that $m_+ < 0 < m_-$. The condition $G_j\big(-\frac{1}{\alpha_j}\big) > 0$ is now equivalent to

$$m_+ < \frac{-1}{\alpha_j} < m_-, \tag{3.23}$$

which is the same as (since $\alpha_j > 0$)

$$\alpha_j > 1 + \sqrt{c_p}. \tag{3.24}$$

We summarize the above calculations.

Lemma 3.4. *The solutions of $z_p'(m) = 0$ are*

$$m = -\frac{1}{1+\sqrt{c_p}} + O\Big(\frac{1}{n}\Big) =: m_+^{(n)}, \qquad m = -\frac{1}{1-\sqrt{c_p}} + O\Big(\frac{1}{n}\Big) =: m_-^{(n)}. \tag{3.25}$$

and

$$m = -\frac{1}{\alpha_j} \pm \frac{1}{\sqrt{n}}\sqrt{G_j\big(-\frac{1}{\alpha_j}\big)} + O\Big(\frac{1}{n}\Big) =: m_{j,\pm}^{(n)}, \qquad j = 1, \dots, M, \tag{3.26}$$

all of multiplicity 1. *Furthermore, the following holds.*

- *When $c_p < 1$, $m_-^{(n)} < m_+^{(n)} < 0$, and $m_{j,\pm}^{(n)}$ are real if and only if $\alpha_j > 1+\sqrt{c_p}$ or $\alpha_j < 1 - \sqrt{c_p}$. If $1 - \sqrt{c_p} < \alpha_j < 1 + \sqrt{c_p}$, $m_{j,+}^{(n)}$ and $m_{j,-}^{(n)}$ are complex conjugates of each other.*
- *When $c_p > 1$, $m_+^{(n)} < 0 < m_-^{(n)}$, and $m_{j,\pm}^{(n)}$ are real if and only if $\alpha_j > 1 + \sqrt{c_p}$. If $\alpha_j < 1 + \sqrt{c_p}$, $m_{j,+}^{(n)}$ and $m_{j,-}^{(n)}$ are complex conjugates of each other.*

We now consider the cases when $c < 1$ and when $c > 1$ separately.

3.1 When $c < 1$

Let the indices $0 \le M_0, M_1 \le M$ be defined as in Theorem 1.1 (recall that we assume that none of the α_j's are equal to $1 \pm \sqrt{c}$), so that

$$\alpha_1 > \cdots > \alpha_{M_0} > 1 + \sqrt{c} > \alpha_{M_0+1} > \cdots > \alpha_{M-M_1} > 1 - \sqrt{c} > \alpha_{M-M_1+1} > \cdots > \alpha_M. \tag{3.27}$$

We now find the intervals in which $z_p'(m) > 0$.

The denominator of (3.3) is non-negative. ¿From Lemma 3.4, the numerator of (3.3) is factored as

$$const \cdot (m - m_-^{(n)})(m - m_+^{(n)}) \prod_{j=1}^{M} (m - m_{j,-}^{(n)})(m - m_{j,+}^{(n)}) \tag{3.28}$$

The constant prefactor is, from (3.4) and (3.5),

$$(1-c_p)\prod_{j=1}^{M}\alpha_j^2+O\Big(\frac{1}{n}\Big), \tag{3.29}$$

which is positive when n is large enough. On the other hand, among the terms in the product of (3.28), $m_{j,\pm}^{(n)}$ corresponding to the indices $M_0+1\le j\le M_1$ are complex conjugates of each other. Thus

$$\prod_{j=M_0+1}^{M_1+1}(m-m_{j,-}^{(n)})(m-m_{j,+}^{(n)})\ge 0. \tag{3.30}$$

Hence using the fact that

$$0>m_{1,+}^{(n)}>m_{1,-}^{(n)}>\cdots>m_{M_0,+}^{(n)}>m_{M_0,-}^{(n)}>m_{+}^{(n)} \tag{3.31}$$

and

$$m_{-}^{(n)}>m_{M_1+1,+}^{(n)}>m_{M_1+1,-}^{(n)}>\cdots>m_{M,+}^{(n)}>m_{M,-}^{(n)}, \tag{3.32}$$

we find that the numerator of (3.3) is positve in the intervals

$$(-\infty,m_{M,-}^{(n)})\cup(m_{M,+}^{(n)},m_{M-1,-}^{(n)})\cup\cdots\cup(m_{M_1+2,+}^{(n)},m_{M_1+1,-}^{(n)})\cup(m_{M_1+1,+}^{(n)},m_{-}^{(n)}) \tag{3.33}$$

union

$$(m_{+}^{(n)},m_{M_0,-}^{(n)})\cup(m_{M_0,+}^{(n)},m_{M_0-1,-}^{(n)})\cup\cdots\cup(m_{2,+}^{(n)},m_{1,-}^{(n)})\cup(m_{1,+}^{(n)},\infty). \tag{3.34}$$

The singular points of (3.3) are not contained in any of the above intervals except for the singular point $m=0$. Hence the set of m such that $z_p'(m)>0$ is equal to (3.33) union

$$(m_{+}^{(n)},m_{M_0,-}^{(n)})\cup(m_{M_0,+}^{(n)},m_{M_0-1,-}^{(n)})\cup\cdots\cup(m_{2,+}^{(n)},m_{1,-}^{(n)})\cup(m_{1,+}^{(n)},0)\cup(0,\infty). \tag{3.35}$$

Now Lemma 3.1 determines $\mathrm{supp}(F_p)$.

Proposition 3.5. *Suppose that $c<1$ and none of α_j is equal to $1\pm\sqrt{c}$. With the indices M_0 and M_1 defined in Theorem 1.1, for n sufficiently large,*

$$\begin{aligned}\mathrm{supp}(F_p)^c=&(-\infty,0)\cup(0,z_{M,-}^{(n)})\cup(z_{M,+}^{(n)},z_{M-1,-}^{(n)})\cup\cdots\cup(z_{M_1+1,+}^{(n)},z_{-}^{(n)})\\&\cup(z_{+}^{(n)},z_{M_0,-}^{(n)})\cup(z_{M_0,+}^{(n)},z_{M_0-1,-}^{(n)})\cup\cdots\cup(z_{2,+}^{(n)},z_{1,-}^{(n)})\cup(z_{1,+}^{(n)},\infty)\end{aligned} \tag{3.36}$$

where

$$z_{\pm}^{(n)}=(1\pm\sqrt{c_p})^2+O\Big(\frac{1}{n}\Big) \tag{3.37}$$

and

$$z_{j,\pm}^{(n)}=\alpha_j+\frac{c_p\alpha_j}{\alpha_j-1}\pm\frac{A_j}{\sqrt{n}}+O\Big(\frac{1}{n}\Big),\qquad j=1,\ldots,M_0,\quad j=M_1+1,\ldots,M, \tag{3.38}$$

for some constant $A_j>0$. The intervals in (3.36) are disjoint.

Proof. We will first see that the intervals (3.33) union (3.35) satisfy conditions (i)-(iii) of Lemma 3.1. Condition (iii) is clearly satisfied. Also 0 is not contained in (3.33) and (3.35), and so condition (i) is fullfilled. Finally, as $\operatorname{supp}(H_p) = \{\alpha_1, \dots, \alpha_M, 1\}$ and

$$m_-^{(n)} < -1 < m_+^{(n)}, \qquad m_{j,-}^{(n)} < -\frac{1}{\alpha_j} < m_{j,+}^{(n)}, \tag{3.39}$$

condition (ii) is satisfied for m in (3.33) union (3.35).

We now need to find the image of the above intervals under z_p. Clearly, $z_p(-\infty) = 0$, $z_p(0-) = +\infty$, $z_p(0+) = -\infty$ and $z_p(+\infty) = 0$. A direct computation yields

$$z_p(m_\pm^{(n)}) = (1 \pm \sqrt{c_p})^2 + O\Big(\frac{1}{n}\Big). \tag{3.40}$$

and

$$z_p(m_{j,\pm}^{(n)}) = \alpha_j + \frac{c_p \alpha_j}{\alpha_j - 1} \pm \frac{A_j}{\sqrt{n}} + O\Big(\frac{1}{n}\Big) \tag{3.41}$$

where

$$A_j = \frac{1}{C_j}\Big\{ C_j^2 \alpha_j^2 \Big(1 - \frac{c_p}{(\alpha_j - 1)^2}\Big) + k_j \Big\}, \qquad C_j := \sqrt{G(-1/\alpha_j)}. \tag{3.42}$$

Note that $A_j > 0$ for $1 \le j \le M_0$ and $M_1 + 1 \le j \le M$ since $\alpha_j > 1 + \sqrt{c_p}$ or $\alpha_j < 1 - \sqrt{c_p}$. Also it is straightforward to check from the graph of the function

$$x + \frac{c_p x}{x - 1} \tag{3.43}$$

that

$$\begin{aligned} 0 < \alpha_M + \frac{c_p \alpha_M}{\alpha_M - 1} < \cdots < \alpha_{M_1+1} + \frac{c_p \alpha_{M_1+1}}{\alpha_{M_1+1} - 1} < (1 - \sqrt{c_p})^2 \\ < (1 + \sqrt{c_p})^2 < \alpha_{M_0} + \frac{c_p \alpha_{M_0}}{\alpha_{M_0} - 1} < \cdots < \alpha_1 + \frac{c_p \alpha_1}{\alpha_1 - 1}. \end{aligned} \tag{3.44}$$

This implies the Proposition. □

3.2 When $c > 1$

This case is similar to the previous case when $c < 1$. We indicate only the difference.

We again assume that p and n are large enough so that the set of j's satisfying $\alpha_j > 1 + \sqrt{c_p}$ is the same as the set of j's satisfying $\alpha_j > 1 + \sqrt{c}$. Let the index $0 \le M_0 \le M$ be defined, as in Theorem 1.1. We further assume that none of α_j is equal to $1 + \sqrt{c}$ so that

$$\alpha_{M_0} > 1 + \sqrt{c} > \alpha_{M_0+1}. \tag{3.45}$$

The denominator of (3.3) is non-negative and as before, the numerator of (3.3) is equal to (3.28). But this time, the constant prefactor (3.29) is negative when n is large enough. Also, as in (3.30),

$$\prod_{j=M_0+1}^{M} (m - m_{j,-}^{(n)})(m - m_{j,+}^{(n)}) \ge 0. \tag{3.46}$$

Now using the fact that

$$m_-^{(n)} > 0 > m_{1,+}^{(n)} > m_{1,-}^{(n)} > \cdots > m_{M_0,+}^{(n)} > m_{M_0,-}^{(n)} > m_+^{(n)}, \tag{3.47}$$

we find that the numerator of (3.3) is positive in the intervals

$$(m_+^{(n)}, m_{M_0,-}^{(n)}) \cup (m_{M_0,+}^{(n)}, m_{M_0-1,-}^{(n)}) \cup \cdots \cup (m_{2,+}^{(n)}, m_{1,-}^{(n)}) \cup (m_{1,+}^{(n)}, m_-^{(n)}). \tag{3.48}$$

Hence, taking into account the singular point $m = 0$ of $z_p'(m)$, the intervals where $z_p'(m) > 0$ are

$$(m_+^{(n)}, m_{M_0,-}^{(n)}) \cup (m_{M_0,+}^{(n)}, m_{M_0-1,-}^{(n)}) \cup \cdots \cup (m_{2,+}^{(n)}, m_{1,-}^{(n)}) \cup (m_{1,+}^{(n)}, 0) \cup (0, m_-^{(n)}). \tag{3.49}$$

The proof of the following proposition is parallel to Proposition 3.5.

Proposition 3.6. *Suppose that $c > 1$ and none of α_j is equal to $1 + \sqrt{c}$. With the index M_0 defined in Theorem 1.2, for n sufficiently large,*

$$\operatorname{supp}(F_p)^c = (-\infty, z_-^{(n)}) \cup (z_+^{(n)}, z_{M_0,-}^{(n)}) \cup (z_{M_0,+}^{(n)}, z_{M_0-1,-}^{(n)}) \cup \cdots \cup (z_{2,+}^{(n)}, z_{1,-}^{(n)}) \cup (z_{1,+}^{(n)}, \infty) \tag{3.50}$$

where

$$z_\pm^{(n)} = (1 \pm \sqrt{c_p})^2 + O\left(\frac{1}{n}\right) \tag{3.51}$$

and

$$z_{j,\pm}^{(n)} = \alpha_j + \frac{c_p \alpha_j}{\alpha_j - 1} \pm \frac{A_j}{\sqrt{n}} + O\left(\frac{1}{n}\right), \qquad j = 1, \dots, M_0, \tag{3.52}$$

for some constant $A_j > 0$. The intervals in (3.50) are disjoint.

4 Proof of Theorems 1.1, 1.2 and 1.3

When T_p is (1.8), H_p is equal to (3.1), and hence $dH_\infty(x) = \delta_1(x)dx$. In this case, (see (2.6) and (2.7))

$$dF_\infty(\lambda) = \begin{cases} \frac{1}{2\pi\lambda}\sqrt{((1+\sqrt{c})^2 - \lambda)(\lambda - (1-\sqrt{c})^2)}1_{[(1-\sqrt{c})^2,(1+\sqrt{c})^2]}(\lambda), & c > 1 \\ \frac{1}{2\pi\lambda}\sqrt{((1+\sqrt{c})^2 - \lambda)(\lambda - (1-\sqrt{c})^2)}1_{[(1-\sqrt{c})^2,(1+\sqrt{c})^2]}(\lambda) + (1-c)\delta_0, & 0 < c \leq 1. \end{cases} \tag{4.1}$$

4.1 When $c < 1$

We first assume that none of α_j is equal to $1 \pm \sqrt{c}$ so that Proposition 3.5 is applicable. The case when some of α_j are equal to $1 \pm \sqrt{c}$ will be discussed at the end of this subsection.

When T_p is (1.8), all the conditions (a)-(e) of Proposition 2.1 are satisfied or are defined accordingly.

Now suppose $[a, b]$ is an interval satisfying condition (f). Since

$$z_+^{(n)} \to (1 + \sqrt{c})^2, \qquad z_-^{(n)} \to (1 - \sqrt{c})^2, \tag{4.2}$$

and for any i,

$$z_{i,+}^{(n)}, \quad z_{i,-}^{(n)} \to \alpha_i + \frac{c\alpha_i}{\alpha_i - 1}, \tag{4.3}$$

we see that

$$
\begin{aligned}
[a,b] \subset &(-\infty,0) \cup \left(0, \alpha_M + \frac{c\alpha_M}{\alpha_M - 1}\right) \cup (\alpha_M + \frac{c\alpha_M}{\alpha_M-1}, \alpha_{M-1} + \frac{c\alpha_{M-1}}{\alpha_{M-1}-1}\Big) \\
&\cup \cdots \cup \left(\alpha_{M_1+1} + \frac{c\alpha_{M_1+1}}{\alpha_{M_1+1}-1}, (1-\sqrt{c})^2\right) \\
&\cup \left((1+\sqrt{c})^2, \alpha_{M_0} + \frac{c\alpha_{M_0}}{\alpha_{M_0}-1}\right) \cup \cdots \cup \left(\alpha_2 + \frac{c\alpha_2}{\alpha_2 - 1}, \alpha_1 + \frac{c\alpha_1}{\alpha_1-1}\right) \cup \left(\alpha_1 + \frac{c\alpha_1}{\alpha_1 - 1}, \infty\right).
\end{aligned} \tag{4.4}
$$

On the other hand,

$$\mathrm{supp}(F_\infty)^c = (-\infty, 0) \cup \big(0, (1-\sqrt{c})^2\big) \cup \big((1+\sqrt{c})^2, \infty\big). \tag{4.5}$$

Hence $[a,b] \subset \mathrm{supp}(F_\infty)^c$. Also from (2.5), it is easy to see that $m'_\infty(z) > 0$ for $z \in \mathrm{supp}(F_\infty)^c$. The first consequence of (ii) of Proposition 2.1 (note that $H_\infty(0) = 0$) is that $m_\infty(b) < 0$. Thus $m_\infty(a) < m_\infty(b) < 0$. Therefore, condition (2.13) is equivalent to the condition

$$[a,b] \subset \big[z_\infty\big(-1/\sigma^{(p)}_{i_p+1}\big), z_\infty\big(-1/\sigma^{(p)}_{i_p}\big)\big]. \tag{4.6}$$

We will consider four different choices of $[a,b]$. First fix $1 \le j \le M_0$. Take

$$[a,b] = \big[\alpha_j + \frac{c\alpha_j}{\alpha_j - 1} + \epsilon, \alpha_{j-1} + \frac{c\alpha_{j-1}}{\alpha_{j-1}-1} - \epsilon\big] \tag{4.7}$$

for an arbitrary fixed $\epsilon > 0$. (Here $\alpha_0 := +\infty$.) From (4.3), we see that

$$[a,b] \subset (z^{(n)}_{j,+}, z^{(n)}_{j-1,-}) \tag{4.8}$$

for all large p, and hence condition (f) is satisfied using Proposition 3.5. Set

$$i_p := k_1 + \cdots + k_{j-1}. \tag{4.9}$$

(When $j = 1$, $i_p := 0$.) For T_p given by (1.8),

$$\sigma^{(p)}_{i_p} = \alpha_{j-1}, \qquad \sigma^{(p)}_{i_p+1} = \alpha_j. \tag{4.10}$$

But

$$z_\infty(-1/\alpha_j) = \alpha_j + \frac{c\alpha_j}{\alpha_j - 1} \tag{4.11}$$

and hence condition (4.6) is satisfied. Therefore i_p is defined to satisfy condition (2.13). Proposition 3.5 now implies that

$$\mathbb{P}\left(s^{(p)}_{k_1+\cdots+k_{j-1}} > \alpha_{j-1} + \frac{c\alpha_{j-1}}{\alpha_{j-1}-1} - \epsilon \text{ and } s^{(p)}_{k_1+\cdots+k_{j-1}+1} < \alpha_j + \frac{c\alpha_j}{\alpha_j - 1} + \epsilon \text{ for all large } p\right) = 1. \tag{4.12}$$

This yields that, $1 \le j \le M_0 - 1$,

$$\mathbb{P}\left(\alpha_j + \frac{c\alpha_j}{\alpha_j - 1} - \epsilon < s^{(p)}_{k_1+\cdots+k_{j-1}+k_j} \le \cdots \le s^{(p)}_{k_1+\ldots k_{j-1}+1} < \alpha_j + \frac{c\alpha_j}{\alpha_j-1} + \epsilon \text{ for all large } p\right) = 1, \tag{4.13}$$

which implies (1.10) for $1 \le j \le M_0 - 1$, and

$$\mathbb{P}\left(s^{(p)}_{k_1+\cdots+k_{M_0-1}+1} < \alpha_{M_0} + \frac{c\alpha_{M_0}}{\alpha_{M_0} - 1} + \epsilon \text{ for all large } p\right) = 1. \tag{4.14}$$

For the second choice of $[a,b]$, set

$$[a,b] = \big[(1+\sqrt{c})^2 + \epsilon, \alpha_{M_0} + \frac{c\alpha_{M_0}}{\alpha_{M_0}-1} - \epsilon\big] \tag{4.15}$$

for an arbitrary fixed $\epsilon > 0$. Noting that

$$z_+^{(n)} \to (1+\sqrt{c})^2 \tag{4.16}$$

and setting $i_p := k_1 + \cdots + k_{M_0}$, a calculation similar to the above yields that

$$\mathbb{P}\Big(s^{(p)}_{k_1+\cdots+k_{M_0}} > \alpha_{M_0} + \frac{c\alpha_{M_0}}{\alpha_{M_0}-1} - \epsilon \text{ and } s^{(p)}_{k_1+\cdots+k_{M_0}+1} < (1+\sqrt{c})^2 + \epsilon \text{ for all large } p\Big) = 1. \tag{4.17}$$

Thus, together with (4.14), we obtain (1.10) for $j = M_0$. Also as (2.4) and discussions around (2.8) imply that the support of the limiting spectral distribution of B_p is $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$, we obtain (1.11).

As the third and fourth choices of $[a,b]$, we set

$$[a,b] = \big[\alpha_{M_1+1} + \frac{c\alpha_{M_1+1}}{\alpha_{M_1+1}-1} + \epsilon, (1-\sqrt{c})^2 - \epsilon\big] \tag{4.18}$$

and

$$[a,b] = \big[\alpha_{j+1} + \frac{c\alpha_{j+1}}{\alpha_{j+1}-1} + \epsilon, \alpha_j + \frac{c\alpha_j}{\alpha_j - 1} - \epsilon\big] \tag{4.19}$$

for some $M_1 + 1 \le j \le M$ ($\alpha_{M+1} := 0$), respectively. Arguments as above imply the remaining part of Theorem 1.1.

We now consider the case when an α_j is equal to $1 \pm \sqrt{c}$. We first observe certain monotonicity of the eigenvalues $s_j^{(p)}$ on α_js. Note that the matrix $\underline{B}_p := \frac{1}{n} Z_p' T_p Z_p$ has the same set of eigenvalues as B_p except for $|p-n|$ zero eigenvalues. Consider a set of parameters β_j, $1 \le j \le M$, such that $\alpha_j \ge \beta_j$. Let $\hat{T}_p$ be the matrix T_p with α_j's replaced by β_j's, and set $\hat{B}_p = \frac{1}{n}\hat{T}_p^{1/2} Z_p Z_p' \hat{T}_p^{1/2}$ and $\underline{\hat{B}}_p = \frac{1}{n} Z_p' \hat{T}_p Z_p$. Then clearly, $\underline{B}_p$ and $\underline{\hat{B}}_p$ are Hermitian, and $\underline{B}_p \ge \underline{\hat{B}}_p$. Hence from the min-max principle (see e.g. [11]), we find that

$$s_j^{(p)} \ge \hat{s}_j^{(p)} \tag{4.20}$$

for all non-zero eigenvalues, where $\hat{s}_j^{(p)}$ denotes the eigenvalues of $\hat{B}_p$.

Suppose that

$$\alpha_1 > \cdots > \alpha_{M_0} > 1 + \sqrt{c} = \alpha_{M_0+1} > \cdots > \alpha_{M-M_1} > 1 - \sqrt{c} > \alpha_{M-M_1+1} > \cdots > \alpha_M. \tag{4.21}$$

Replacing in (4.21) α_{M_0+1} by $(1\pm\epsilon)\alpha_{M_0+1} = (1\pm\epsilon)(1+\sqrt{c})$ for sufficiently small $\epsilon > 0$, the above monotonicity argument implies the following:

(i) For each $1 \le j \le M_0$,

$$\alpha_j + \frac{c\alpha_j}{\alpha_j - 1} \le \liminf s^{(p)}_{k_1+\cdots+k_{j-1}+i} \le \limsup s^{(p)}_{k_1+\cdots+k_{j-1}+i} \le \alpha_j + \frac{c\alpha_j}{\alpha_j-1}, \qquad 1 \le i \le k_j. \tag{4.22}$$

almost surely.

(ii)

$$(1+\sqrt{c})^2 \le \liminf s^{(p)}_{k_1+\cdots+k_{M_0}+1} \le \limsup s^{(p)}_{k_1+\cdots+k_{M_0}+1} \le (1+\epsilon)\alpha_{M_0+1} + \frac{c(1+\epsilon)\alpha_{M_0+1}}{(1+\epsilon)\alpha_{M_0+1}-1} \quad (4.23)$$

almost surely.

(iii)

$$(1-\sqrt{c})^2 \le \liminf s^{(p)}_{p-r+k_1+\cdots+k_{M_1}} \le \limsup s^{(p)}_{p-r+k_1+\cdots+k_{M_1}} \le (1-\sqrt{c})^2 \quad (4.24)$$

almost surely.

(iv) For each $M_1+1 \le j \le M$,

$$\begin{aligned} \alpha_j + \frac{c\alpha_j}{\alpha_j-1} &\le \liminf s^{(p)}_{p-r+k_1+\cdots+k_{j-1}+i} \to \alpha_j + \frac{c\alpha_j}{\alpha_j-1} \\ &\le \limsup s^{(p)}_{p-r+k_1+\cdots+k_{j-1}+i} \to \alpha_j + \frac{c\alpha_j}{\alpha_j-1} \le \alpha_j + \frac{c\alpha_j}{\alpha_j-1} \qquad 1 \le i \le k_j \end{aligned} \quad (4.25)$$

almost surely.

Since

$$\lim_{\epsilon\downarrow 0}(1+\epsilon)\alpha_{M_0+1} + \frac{c(1+\epsilon)\alpha_{M_0+1}}{(1+\epsilon)\alpha_{M_0+1}-1} = (1+\sqrt{c})^2 \quad (4.26)$$

and the above result is true for arbitrary sufficiently small $\epsilon > 0$, Theorem 1.1 follows for the case when the parameters are given by (4.21). For the case when $\alpha_{M-M_1} = 1-\sqrt{c}$, the argument is almost identical, and we skip the details.

4.2 When $c > 1$

¿From (4.1), when $c > 1$, the smallest value in the support of F_∞ is

$$x_0 = (1-\sqrt{c})^2 > 0. \quad (4.27)$$

Hence Proposition 2.1 (i) implies that

$$s^{(p)}_n \to (1-\sqrt{c})^2. \quad (4.28)$$

Since when $p > n$, at least $p-n$ eigevalues $s^{(p)}_j$ are equal to 0, we conclude that

$$s^{(p)}_{n+1} = \cdots = s^{(p)}_p = 0. \quad (4.29)$$

Therefore, (1.19) and (1.20) are obtained.

The proof of (1.17) and (1.18) is similar to the case when $c < 1$ by using Proposition 3.6 and noting that an interval $[a,b]$ satisfying condition (f) of Proposition 2.1 is contained in

$$\big(-\infty,(1-\sqrt{c})^2\big) \cup \left((1+\sqrt{c})^2, \alpha_{M_0} + \frac{c\alpha_{M_0}}{\alpha_{M_0}-1}\right) \cup \cdots \cup \left(\alpha_2 + \frac{c\alpha_2}{\alpha_2-1}, \alpha_1 + \frac{c\alpha_1}{\alpha_1-1}\right) \cup \left(\alpha_1 + \frac{c\alpha_1}{\alpha_1-1}, \infty\right), \quad (4.30)$$

which is a subset of

$$\mathrm{supp}(F_\infty)^c = \big(-\infty,(1-\sqrt{c})^2\big) \cup \big((1+\sqrt{c})^2,\infty\big). \quad (4.31)$$

4.3 When $c = 1$

Since the limiting distribution (1.2) for $c = 1$ has a continuous density on the interval $(0, 4)$, it is easy to see (1.24).

We first observe a monotonicity of $s_j^{(p)}$ in n. Let $\hat{Z}_p = (Z_{ij})$, $1 \le i \le p, 1 \le j \le \hat{n}$ and let $\hat{B}_p := \frac{1}{\hat{n}} T_p^{1/2} \hat{Z}_p \hat{Z}_p' T_p^{1/2}$. When $\hat{n} > n$, it is clear that

$$\hat{n}\hat{B}_p \ge nB_p. \tag{4.32}$$

Therefore, if the ordered eigenvalues of $\hat{B}_p$ are denoted by $\hat{s}_p^{(p)}$, the min-max principle implies that

$$\hat{n}\hat{s}_j^{(p)} \ge ns_j^{(p)} \tag{4.33}$$

for all $1 \le j \le p$.

Take $\hat{n} = \left[\frac{n}{1+\epsilon}\right]$ for $\epsilon > 0$ where $[x]$ denotes the largest integer $\le x$. Then for sufficiently small $\epsilon > 0$,

$$\alpha_1 > \cdots > \alpha_{M_0} > 1 + \sqrt{1+\epsilon} > \alpha_{M_0+1} > \cdots > \alpha_M. \tag{4.34}$$

By applying Theorem 1.2 and using (4.33), we obtain the following:

- For each $1 \le j \le M_0$,

$$\liminf s_{k_1+\cdots+k_{j-1}+i}^{(p)} \ge \frac{1}{1+\epsilon}\left(\alpha_j + \frac{(1+\epsilon)\alpha_j}{\alpha_j - 1}\right), \qquad 1 \le i \le k_j. \tag{4.35}$$

  almost surely.

- $$\liminf s_{k_1+\cdots+k_{M_0}+1}^{(p)} \ge \frac{1}{1+\epsilon}(1+\sqrt{1+\epsilon})^2 \tag{4.36}$$

  almost surely.

On the other hand, take $\hat{n} = \left[\frac{n}{1-\epsilon}\right]$ for $\epsilon > 0$. We first assume $2 > \alpha_{M_0+1}$. Then as $\alpha_M > 0$, for sufficiently small $\epsilon > 0$,

$$\alpha_1 > \cdots > \alpha_{M_0} > 1 + \sqrt{1-\epsilon} > \alpha_{M_0+1} > \cdots > \alpha_M > 1 - \sqrt{1-\epsilon}, \tag{4.37}$$

and hence $M_1 = M$. By applying Theorem 1.1 and using (4.33), we obtain the following:

- For each $1 \le j \le M_0$,

$$\limsup s_{k_1+\cdots+k_{j-1}+i}^{(p)} \le \frac{1}{1-\epsilon}\left(\alpha_j + \frac{(1-\epsilon)\alpha_j}{\alpha_j - 1}\right), \qquad 1 \le i \le k_j. \tag{4.38}$$

  almost surely.

- $$\limsup s_{k_1+\cdots+k_{M_0}+1}^{(p)} \le \frac{1}{1-\epsilon}(1+\sqrt{1-\epsilon})^2 \tag{4.39}$$

  almost surely.

If $\alpha_{M_0+1} = 2$, then for sufficiently small $\epsilon > 0$,

$$\alpha_1 > \cdots > \alpha_{M_0} > \alpha_{M_0+1} > 1 + \sqrt{1-\epsilon} > \alpha_{M_0+2} > \cdots > \alpha_M > 1 - \sqrt{1-\epsilon}. \tag{4.40}$$

Hence Theorem 1.1 implies (4.38) but (4.39) becomes

$$\limsup s^{(p)}_{k_1+\cdots+k_{M_0}+1} \le \frac{1}{1-\epsilon}\left(\alpha_{M_0+1} + \frac{(1-\epsilon)\alpha_{M_0+1}}{\alpha_{M_0+1}-1}\right) \tag{4.41}$$

almost surely.

Therefore (4.35) and (4.38) yield (1.22), and (4.36), (4.39) and (4.41) yield (1.23).

References

[1] Z. D. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica*, 9:611–677, 1999.

[2] Z. D. Bai and Y. Q. Yin. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.* 21:1275–1294, 1993.

[3] Z. D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1):316–345, 1998.

[4] Z. D. Bai and J. W. Silverstein. Exact separation of eigenvalues of large dimensional sample covariance matrices. *Ann. Probab.*, 27(3):1536–1555, 1999.

[5] J. Baik. Painlevé formulas of the limiting distributions for non-null complex sample covariance matrices. `http://xxx.lanl.gov/abs/math.PR/0504606`.

[6] J. Baik, G. Ben Arous and S. Péché. Phase transition of the largest eigenvalue for non-null complex sample covariance matrices. to appear in *Ann. Probab.*, 2005.

[7] J. Baik and J. W. Silverstein. Distributional Behavior of the Stray Eigenvalues of Large Sample Matrices of Spiked Population Models, in preparation.

[8] A. Buja, T. Hastie, and R. Tibshirani. Penalized discriminant analysis. *Ann. Statist.*, 23:73–102, 1995.

[9] P.J. Forrester. The spectrum edge of random matrix ensembles. *Nuclear Physics B*, 402:709–728, 1993.

[10] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.

[11] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge Press, New York, 1985.

[12] D. Hoyle and M. Rattray. Limiting form of the sample covariance eigenspectrum in PCA and kernel PCA. to appear in proceedings of Neural Information Processing Systems 2003.

[13] K. Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. of Math. (2)*, 153:259–296, 2001.

[14] I. Johnstone. On the distribution of the largest Principal Component. *Ann. Statist.*, 29:295–327, 2001.

[15] L. Laloux, P. Cizeau, M. Potters, and J. Bouchaud. Random matrix theory and financial correlations. *Intern. J. Theor. Appl. Finance*, 3(3):391–397, 2000.

[16] Y. Malevergne and D. Sornette. Collective origin of the coexistance of apparent RMT noise and factors in large sample correlation matrices. `http://xxx.lanl.gov/abs/cond-mat/0210115`.

[17] V.A. Marcenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik*, 1:457–486, 1967.

[18] D. Paul. Asymptotic behaviour of the largest eigenvalues of sample covariance matrix when true covariance is a finite perturbation of identity. preprint, 2004.

[19] S. Péché. Universality of local eigenvalue statistics for random sample covariance matrices. *Ph.D. Thesis*, Ecole Polytechnique Fédérale de Lausanne, 2003.

[20] V. Plerous, P. Gopikrishnan, B. Rosenow, L. Amaral, T. Guhr, and H. Stanley. Random matrix approach to cross correlations in financial data. *Phys. Rev. E*, 65(6):066126, 2002.

[21] R. Sear and J. Cuesta. Instabilities in complex mixtures with a large number of components. *Phys. Rev. Lett.*, 91(24):245701, 2003.

[22] J. W. Silverstein. The smallest eigenvalue of a large dimensional wishart matrix. *Ann. Probab.*, 13(4):1364–1368, 1985.

[23] J. W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.*, 55(2):331–339, 1995.

[24] J. W. Silverstein and Z. D. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.*, 54(2):175–192, 1995.

[25] J. W. Silverstein and S. I. Choi. Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.*, 54(2):295–309, 1995.

[26] A. Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys*, 108(5-6):1033–1056, 2002.

[27] E. Telatar. Capacity of multi-antenna Gaussian channels. *European transactions on Telecommunications*, 10(6):585–595, 1999.

[28] C. Tracy and H. Widom. Level spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159:33–72, 1994.

[29] C. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177:727–754, 1996.

[30] K. W. Wachter. The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* 6:1–18, 1978.

[31] Y. Q. Yin. Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* 20:50–68, 1986.

[32] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah On limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probab. Theory Related Fields* 78: 509–521, 1988.