Random matrices

Marianna Bolla, DSc. Prof. BME Math. Inst.

October 16, 2020

The first random matrix of the history is the Wishart matrix, defined in the 1930ies as the sample covariance matrix of a multivariate Gaussian sample.

Definition 1 *Let* $\mathbf{Z}_1, \dots, \mathbf{Z}_n \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{I}_p)$ *be i.i.d. p-dimensional normal sample. The distribution of the* $p \times p$ *matrix* $\widetilde{\boldsymbol{W}} = \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^T$ *is called standard Wishart with parameters* p, n *(n is also called degree of freedom as in the* $p = 1$ *case this is the* χ^2 *distribution of degree of freedom n). More generally, let* $\mathbf{X}_1, \dots, \mathbf{X}_n \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{C})$ *be i.i.d. p-dimensional normal sample. The distribution of the* $p \times p$ *matrix* $\boldsymbol{W} = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T$ *is called (central) Wishart with parameters* p, n*, and* $\boldsymbol{C}$*.*

In view of the relation between the $\mathbf{Z}_i$'s and the $\mathbf{X}_i$'s, the relation $\boldsymbol{W} = \boldsymbol{C}^{1/2} \widetilde{W} \boldsymbol{C}^{1/2}$ also holds, and it can be proved that n times the empirical covariance matrix of a p-dimensional normal sample, i.e. the matrix $\boldsymbol{S}$ in Definition **??**, is a (central) Wishart matrix with parameters p, $n-1$ and the common covariance matrix. The Wishart matrix is symmetric and positive semidefinite, and in the $p < n$ case it is positive definite, with probability 1. The Wishart entries are far not independent (in and above the main diagonal, otherwise the matrix is symmetric), the joint density of the entries (the density of the matrix) is revisited by means of easy transformations in [15]. The author of this paper also derived the distribution of eigenvalues and singular values of matrices arising from matrix factorizations, for fixed n and p, but he did not investigate asymptotics for the singular values of the $p \times n$ data matrix, whose entries are usually not independent.

The Wigner type matrices were introduced later, in the 1950ies, the eigenvalues of which modelled energy levels in slow nuclear reactions. These symmetric matrices are more easy to treat as their diagonal- and upper-diagonal entries are independent random variables, however, the distribution of them need not be defined uniquely. Therefore, there are many variants of theorems applying to the limiting behavior of the eigenvalues of Wigner type matrices depending on the assumptions for the entries and the kind of the convergence.

In accord with the description in [11, 16, **?**], first we formulate the famous *Wigner Semicircle Law* for the bulk spectrum of expanding symmetric matrices in a more general form than stated in the original papers [20, 21], see also [3].

Theorem 1 *Let* a_{ij} $(i \le j)$ *be independent real-valued random variables with the following properties.*

- *The distribution of* a_{ij}*'s is symmetric.*

- *All moments are finite. In view of symmetry, all odd moments vanish, especially,* $\mathbb{E}(a_{ij}) = 0$ $(i \leq j)$.
- $\mathbb{E}(a_{ij}^2) = \sigma^2$ $(i < j)$ *and* $\mathbb{E}(a_{ii}^2) < C$ $(i = 1, 2, \dots)$ *with* $0 < \sigma, C < \infty$ *constants.*
- $\mathbb{E}(a_{ij}^{2k}) \leq (Ck)^k$ $(k = 1, 2, \dots)$ *with* $0 < C < \infty$ *constant (called sub-Gaussian moments).*

Define a_{ij} *for* $i > j$ *by* $a_{ij} = a_{ji}$. *Let* $\lambda_1^{(n)} \geq \lambda_2^{(n)} \geq \dots \geq \lambda_n^{(n)}$ *be the spectrum of the random symmetric matrix* $\boldsymbol{A}^{(n)} = (a_{ij})_{i,j=1}^{n,n}$. *Denoting by*

$$\tilde{\lambda}_i^{(n)} = \frac{1}{2\sigma\sqrt{n}} \lambda_i^{(n)}, \quad i = 1, \dots, n,$$

the rescaled eigenvalues of of $\boldsymbol{A}^{(n)}$, *their empirical distribution function*

$$F_n(x) = \frac{1}{n} |\{i \, : \, \tilde{\lambda}_i^{(n)} \leq x, \quad i = 1, \dots, n\}$$

converges to a non-random limit $F(x) = \int_{-\infty}^{x} f(x)\, dx$, *which is the c.d.f. corresponding to the semicircle density*

$$f(x) = \begin{cases} \frac{2}{\pi}\sqrt{1 - x^2} & \text{if} \quad |x| \leq 1, \\ 0 & \text{if} \quad |x| > 1. \end{cases} \tag{1}$$

The convergence is understood almost surely (with probability 1) if entries of all matrices $\boldsymbol{A}^{(n)}$ $(n = 1, 2, \dots)$ *are defined on the same probability space.*

We remark that in the original theorem the convergence was understood in distribution; further, the entries were assumed to have an identical distribution in, and another identical distribution above the main diagonal. Note that in case of Gaussian entries, because of the zero expectation and equal variances, the upper diagonal entries are, indeed, identically distributed. The original proof of the Semicircle Law used the method of moments. For Gaussian distributed entries one can find the marginal distribution of the eigenvalues' joint distribution which is obtained from the joint distribution of the entries of $\boldsymbol{A}^{(n)}$ by means of the Jacobian transformation containing the derivatives of the matrix with respect to its eigenvalues and eigenvectors (with an appropriate parameterization). The first factor of this determinant is

$$\prod_{i<j} |\lambda_i^{(n)} - \lambda_j^{(n)}|, \tag{2}$$

supposing that there are no multiple eigenvalues (which is a zero probability event), while the second one depends on the eigenvectors (and follows Haar distribution), but in view of the independence of the two, this last one will be eliminated after integrating with respect to the eigenvectors to obtain the joint density of the eigenvalues (more precisely, it will only contribute to the constant term of the density). This derivation of the proof together with other generalizations can be found in [14].

In [11] it is concluded that on the conditions of the above theorem, all eigenvalues of $\boldsymbol{A}^{(n)}$ lie in the interval $2\sigma\sqrt{n}[1 - \varepsilon, 1 + \varepsilon]$ for any $\varepsilon > 0$ with overwhelming probability, whereas the spacings between subsequent eigenvalues are of order $n^{-1/2}$ in the bulk of the spectrum and much larger close to the edge. However, the Semicircle Law applies to the bulk spectrum, and it provides limited information about the asymptotic behavior of any particular eigenvalue. In the last decades several finer results have been obtained for the largest eigenvalues and the spectral radius of $\boldsymbol{A}^{(n)}$, sometimes under mildly modified conditions.

Authors in [18, 19] characterized the limiting distribution of the first k eigenvalues (for any fixed $k \geq 1$) of the so-called Gaussian Orthogonal Ensemble (GOE), corresponding to the case when the upper diagonal entries are i.i.d. normally distributed random variables. The limiting distributions are called Tracy–Widom type distributions. Later, it turned out that the same distributions are obtained under the conditions of the Semicircle Law. This so-called universality at the edge of the spectrum of Wigner random matrices was proved in [16] in the following sense. After proper rescaling, the first, second, third, etc. eigenvalues of a random symmetric matrix, meeting the conditions of Theorem 1, converge to the distributions established by Tracy and Widom for the GOE, hence giving a generalization of their results.

To investigate the spectral radius of a Wigner type matrix, other authors relaxed the condition that the entries have symmetric distribution and hence, zero expectation. In [10] the author considered the case when the entries in and above the main diagonal are independent, but the diagonal entries have an identical, and the off-diagonal ones another identical distribution with a positive expectation. The following more general statement of [7] contains the aforementioned result as a special case.

Theorem 2 *Let a_{ij} $(i \leq j)$ be independent (not necessarily identically distributed) random variables bounded with a common bound K, i.e. $|a_{ij}| \leq K$, $\forall i, j$. Assume that for $i < j$, $\mathbb{E}(a_{ij}) = \mu$ and $\operatorname{Var}(a_{ij}) = \sigma^2$, further that $\mathbb{E}(a_{ii}) = \nu$. Define a_{ij} for $i > j$ by $a_{ij} = a_{ji}$. Let $\lambda_1^{(n)} \geq \lambda_2^{(n)} \geq \cdots \geq \lambda_n^{(n)}$ be the eigenvalues of the random symmetric matrix $\boldsymbol{A}^{(n)} = (a_{ij})_{i,j=1}^{n,n}$. The numbers K, μ, σ^2, ν will be kept fixed as $n \to \infty$.*

- *If $\mu > 0$, then the distribution of $\lambda_1^{(n)}$ can be approximated in order $\frac{1}{\sqrt{n}}$ by a normal distribution of expectation*

$$(n-1)\mu + \nu + \frac{\sigma^2}{\mu}$$

and variance $2\sigma^2$. Further,

$$\max_{2 \leq i \leq n} |\lambda_i^{(n)}| < 2\sigma\sqrt{n} + \mathcal{O}(n^{1/3} \ln n)$$

in probability (with probability tending to 1 as $n \to \infty$).

- *If $\mu = 0$, then*

$$\max_{1 \leq i \leq n} |\lambda_i^{(n)}| = 2\sigma\sqrt{n} + \mathcal{O}(n^{1/3} \ln n)$$

in probability.

Note that $\mathcal{O}(n^{1/3} \ln n)$ is also $o(\sqrt{n})$, therefore the second term in the above formulas is negligible compared to the first one. Theorem 2 also implies that the largest absolute value eigenvalues of a Wigner type matrix $\boldsymbol{A}^{(n)}$ (all entries have zero expectation) is exactly of order $2\sigma\sqrt{n}$. However, if the common non-zero expectation of the off-diagonal entries is $\mu > 0$, then there is a large eigenvalue having an asymptotic normal distribution with bounded variance around its expectation of order n. The result can easily be extended to the $\mu < 0$ case, therefore the spectral radius is of order n too. In fact, it is the shift in the expectation what puts off the edge of the spectrum.

The following sharp concentration result of [?] applies to general random symmetric matrices with uniformly bounded entries.

Theorem 3 *Let a_{ij} $(1 \le i \le j \le n)$ be independent random variables with absolute value at most 1. Define a_{ij} for $1 \le j < i \le n$ by $a_{ij} = a_{ji}$. Let $\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_n$ be the eigenvalues of the random symmetric matrix $\boldsymbol{A} = (a_{ij})_{i,j=1}^{n,n}$. Then for every positive integer $1 \le i \le \frac{n}{2}$, the probability that λ_i deviates from its median by more than t is at most $e^{-\frac{t^2}{32i^2}}$. The same estimate holds for the probability that λ_{n-i+1} deviates from its median by more than t.*

Observe that the estimate given in the above theorem does not depend on n (of course, the median does depend on it). The statement – apart from a constant factor in the exponent – remains valid if a_{ij}'s are uniformly bounded with some constant K. The authors of Theorem 3 prove that the eigenvalues are also highly concentrated on their own expectations, since λ_i and its median are $\mathcal{O}(i)$ apart. Indeed, let m_i denote the median of λ_i. Then

$$|\mathbb{E}(\lambda_i) - m_i| \le \mathbb{E}|\lambda_i - m_i| = \int_0^\infty \mathbb{P}(|\lambda_i - m_i| > t)\, dt$$
$$\le \int_0^\infty e^{-\frac{t^2}{32i^2}}\, dt = 8\sqrt{2\pi} i.$$

Therefore, for all $t \gg i$ we have

$$\mathbb{P}\left(|\lambda_i - \mathbb{E}(\lambda_i)| > t\right) \le e^{-\frac{(1-o(1))t^2}{32i^2}} \quad \text{when} \quad 1 \le i \le \frac{n}{2}, \tag{3}$$

and the same estimate holds for the probability $\mathbb{P}\left(|\lambda_{n-i+1} - \mathbb{E}(\lambda_{n-i+1})| > t\right)$.

The authors also show that their estimate is sharp for the deviation of λ_1. For this purpose, they consider the following $n \times n$ adjacency matrix of an Erdős–Rényi random graph (see [6]): the diagonal entries are zeros, while the upper diagonal ones are independent Bernoulli distributed random variables with parameter $\frac{1}{2}$ (vertex pairs are connected with probability $\frac{1}{2}$, independently of each other), in which case, by Theorem 2, $\lambda_1 = \frac{n}{2} + o(1)$. The concentration provided by Theorem 3 for larger values of i is weaker than that provided for $i = 1$. The authors also note that for the adjacency matrix of a random graph, when the entries are in the [0,1] interval, the estimate of their Theorem 3 can be improved to $4e^{-\frac{t^2}{8i^2}}$.

The proof of Theorem 3 is based on the Talagrand inequality (see [17]) which is an efficient large deviation tool for product spaces. In fact, this technique is only applicable when the entries are uniformly bounded. Possibly, with other techniques, similar sharp concentration results can be obtained in case of Gaussian entries or just under the conditions of the Semicircle Law. For example, under the conditions of Theorem 1, the individual eigenvalues are highly concentrated on the expected values of the corresponding order statistics based on the limiting distribution as follows.

Proposition 1 *With the notation of Theorem 1,*

$$\lim_{n\to\infty} \max_{1\le i\le n} |\tilde{\lambda}_i^{(n)} - \bar{\lambda}_i^{(n)}| = 0$$

almost surely, where $\bar{\lambda}_i^{(n)} = F^{-1}(\frac{2i-1}{2n})$, $i = 1, \dots, n$, *and* F *is the c.d.f. of the semicircle density (1).*

This issue together with other generalizations for the singular values of square Gaussian matrices are discussed in the survey of [5].

Analogous estimates for the concentration of different norms of random rectangular matrices of complex entries were obtained in [13] also using the Talagrand inequality. [1] generalized Theorem 2 to rectangular matrices in the following way.

Theorem 4 *Let* $\boldsymbol{A}$ *be a random* $m \times n$ *matrix with independent, uniformly bounded entries* a_{ij}*'s such that* $|a_{ij}| \le K$, $\mathbb{E}(a_{ij}) = 0$, *and* $\mathrm{Var}(a_{ij}) \le \sigma^2$ *with some constants* $0 < K, \sigma^2 < \infty$ *for* $i = 1, \dots, m$; $j = 1, \dots, n$. *Under these conditions, for any* $\alpha > \frac{1}{2}$, *if*

$$K < \sigma\sqrt{m+n}(7\alpha \ln(m+n))^{-3},$$

then

$$\mathbb{P}\left(\|\boldsymbol{A}\| > \frac{7}{3}\sigma\sqrt{m+n}\right) < (m+n)^{\frac{1}{2}-\alpha}.$$

The theorem implies that the spectral norm (largest singular value) of the above type $m \times n$ random rectangular matrix is of order $\sqrt{m+n}$ in probability, i.e. with probability tending to 1 as $m, n \to \infty$.

Going back to the Wishart matrices, in the density of which the factor (2) also appears, the first asymptotic result for the bulk spectrum with appropriate normalization, analogous to the Wigner's theorem, is due to [12], where the limiting distribution depends on the ratio $\gamma = p/n$ when $p, n \to \infty$. [22] gave the exact almost sure limit (depending on γ) for the largest eigenvalue of Gram sequences under special conditions. The limiting distribution of the largest eigenvalue of the sample covariance matrix based on a multivariate normal distribution was also characterized in [9]. The author proved that the distribution of the appropriately centered and scaled largest eigenvalue of the standard Wishart matrix with parameters p and n approaches the Tracy–Widom law when $p, n \to \infty$ in such a way that $\frac{n}{p} = \gamma > 1$ is a fixed constant.

References

[1] Achlioptas D and McSherry F 2007 Fast computation of low-rank matrix approximations. *J. ACM* **54** (2), Article 9.

[2] 2002Alon1 Alon N, Krivelevich M and Vu VH 2002 On the concentration of eigenvalues of random symmetric matrices. *Isr. J. Math.* **131**, 259–267.

[3] Arnold, L 1967 On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.* **20**, 262–268.

[4] Bhatia R 1997 *Matrix Analysis*. Springer.

[5] Davidson KR and Szarek SJ 2001 Local operator theory, random matrices and Banach spaces. In *Handbook of the Geometry of Banach spaces* (Johnson WB and Lindenstrauss J eds), Vol. 1, Chapter 8, pp. 317–366. Elsevier Science, Amsterdam.

[6] Erdős P and Rényi A 1960 On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–61.

[7] Füredi Z and Komlós J 1981 The eigenvalues of random symmetric matrices. *Combinatorica* **1**, 233–241.

[8] Horn RA and Johnson CR 1985 *Matrix Analysis*. Cambridge Univ. Press, Cambridge.

[9] Johnstone IM 2001 On the distribution of the largest eigenvalue in principal components analysis. *Ann. Stat.* **29**, 295–327.

[10] Juhász F 1981 On the spectrum of random graphs. In *Algebraic methods in graph theory, Coll. Math. Soc. J. Bolyai* (ed. Lovász L), Vol. 25, pp. 313–316. North–Holland, Amsterdam.

[11] König W 2005 Orthogonal polynomial ensembles in probability theory. *Probab. Surv.* **2**, 385–447.

[12] Marchenko VA and Pastur LA 1967 Distribution of eigenvalues for some set of random matrices. *Mat. USSR-Sbornik* **1**, 457–486.

[13] Meckes M 2004 Concentration of norms and eigenvalues of random matrices. *J. Funct. Anal.* **211** (2), 508–524.

[14] Mehta LM 1990 *Random Matrices*, 2nd edn. Academic Press.

[15] Olkin I 2002 The 70th anniversary of the distribution of random matrices: A survey. *Linear Algebra Appl.* **354**, 231–243.

[16] Soshnikov A 1999 Universality at the end of the spectrum in Wigner random matrices. *Commun. Math. Phys.* **207**, 697–733.

[17] Talagrand M 1996 Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. I.H.E.S.* **81**, 73–205.

[18] Tracy CA and Widom H 1994 Level-spacing distributions and the Airy kernel. *Commun. Math. Phys.* **159**, 151–174.

[19] Tracy CA and Widom H 1996 On orthogonal and symplectic matrix ensembles. *Commun. Math. Phys.* **177**, 727–754.

[20] Wigner EP 1955 Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math.* **62**, 548–564.

[21] Wigner EP 1958 On the distribution of the roots of certain symmetric matrices. *Ann. Math.* **67**, 325–328.

[22] Yin YQ, Bai ZD and Krishnaiah PR 1988 On the limit of the large-dimensional sample covariance matrix. *Probab. Theory Relat. Fields* **78**, 509–521.