RANDOM MATRICES: LAW OF THE DETERMINANT

HOI H. NGUYEN AND VAN H. VU

Abstract. Let A_n be an n by n random matrix whose entries are independent real random variables with mean zero, variance one and with sub-exponential tail. We show that the logarithm of $|\det A_n|$ satisfies a central limit theorem. More precisely,

$$\sup_{x\in\mathbf{R}}\Big|\mathbf{P}\big(\frac{\log(|\det A_n|)-\frac{1}{2}\log(n-1)!}{\sqrt{\frac{1}{2}\log n}}\leq x\big)-\mathbf{P}(\mathbf{N}(0,1)\leq x)\Big|\leq \log^{-\frac{1}{3}+o(1)} n.$$

1. Introduction

Let A_n be an n by n random matrix whose entries $a_{ij}, 1\leq i,j\leq n$, are independent real random variables of zero mean and unit variance. We will refer to the entries a_{ij} as the *atom* variables.

As determinant is one of the most fundamental matrix functions, it is a basic problem in the theory of random matrices to study the distribution of $\det A_n$ and indeed this study has a long and rich history. The earliest paper we find on the subject is a paper of Szekeres and Turán [20] from 1937, in which they studied an extremal problem. In the 1950s, there is a series of papers [6, 16, 29, 17] devoted to the computation of moments of fixed orders of $\det A_n$ (see also [9]). The explicit formula for higher moments get very complicated and in general not available, except in the case when the atom variables have some special distribution (see, for instance [4]).

One can use the estimate for the moments and Markov inequality to obtain an upper bound on $|\det A_n|$. However, no lower bound was known for a long time. In particular, Erdős asked whether $\det A_n$ is non-zero with probability tending to one. In 1967, Komlós [13, 14] addressed this question, proving that almost surely $|\det A_n| > 0$ for random Bernoulli matrices (where the atom variables are iid Bernoulli, taking values ± 1 with probability 1/2). His method also works for much more general models. Following [13], the upper bound on the probability that $\det A_n = 0$ has been improved in [12, 22, 23, 3]. However, these results do not say much about the value of $|\det A_n|$ itself.

In a recent paper [22], Tao and the second author proved that for Bernoulli random matrices, with probability tending to one (as n tends to infinity)

2000 *Mathematics Subject Classification.* 60B20, 60F05.
Key words and phrases. random matrices, random determinant.

$$\sqrt{n!}\exp(-c\sqrt{n\log n}) \leq |\det A_n| \leq \sqrt{n!}\omega(n) \tag{1}$$

for any function $\omega(n)$ tending to infinity with n. This shows that almost surely, $\log|\det A_n|$ is $(\frac{1}{2}+o(1))n\log n$, but does not provide any distributional information. For related works concerning other models of random matrices, we refer to [19].

In [10], Goodman considered random gaussian matrices where the atom variables are iid standard gaussian variables. He noticed that in this case the determinant is a product of independent Chi-square variables. Therefore, its logarithm is the sum of independent variables and thus one expects a central limit theorem to hold. In fact, using properties of Chi-square distribution, it is not very hard to prove

$$\frac{\log(|\det A_n|) - \frac{1}{2}\log(n-1)!}{\sqrt{\frac{1}{2}\log n}} \to \mathbf{N}(0,1). \tag{2}$$

We refer the reader to [18, Section 4] for further discussion on this model.

In [7], Girko stated that (2) holds for general random matrices under the additional assumption that the fourth moment of the atom variables is 3. Twenty years later, he claimed a much stronger result which replaced the above assumption by the assumption that the atom variables have bounded $(4+\delta)$-th moment [8]. However, there are points which are not clear in these papers and we have not found any researcher who can explain the whole proof to us. In our own attempt, we could not pass the proof of Theorem 2 in [8]. In particular, definition (3.7) of this paper requires the matrix $\Xi\binom{1}{k}$ to be invertible, but this assumption can easily fail.

In this paper, we provide a transparent proof for the central limit theorem of the log-determinant. The next question to consider, naturally, is the rate of convergence. We are able to obtain a rate which we believe to be near optimal.

We say that a random variable ξ satisfies condition **C0** (with positive constants C_1, C_2) if

$$P(|\xi| \geq t) \leq C_1\exp(-t^{C_2}) \tag{3}$$

for all $t > 0$.

Theorem 1.1 (Main theorem)**.** *Assume that all atom variables a_{ij} satisfy condition* **C0** *with some positive constants C_1, C_2. Then*

$$\sup_{x\in\mathbf{R}}\Big|\mathbf{P}\big(\frac{\log(|\det A_n|) - \frac{1}{2}\log(n-1)!}{\sqrt{\frac{1}{2}\log n}} \leq x\big) - \Phi(x)\Big| \leq \log^{-1/3+o(1)} n. \tag{4}$$

Here and later, $\Phi(x) = \mathbf{P}(\mathbf{N}(0,1) < x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp(-t^2/2)dt$. In the remaining part of the paper, we will actually prove the following equivalent form,

$$\sup_{x\in\mathbf{R}}\Big|\mathbf{P}\big(\frac{\log(\det A_n^2)-\log(n-1)!}{\sqrt{2\log n}}\leq x\big)-\Phi(x)\Big|\leq \log^{-1/3+o(1)} n. \tag{5}$$

To give some feeling about (5), let us consider the case when a_{ij} are iid standard gaussian. For $0\leq i\leq n-1$, let V_i be the subspace generated by the first i rows of A_n. Let Δ_{i+1} denote the distance from $\mathbf{a}_{i+1}$ to V_i, where $\mathbf{a}_{i+1}=(a_{i+1,1},\dots,a_{i+1,n})$ is the $(i+1)$-th row vector of A_n. Then, by the "base times height" formula, we have

$$\det A_n^2=\prod_{i=0}^{n-1}\Delta_{i+1}^2. \tag{6}$$

Therefore,

$$\log\det A_n^2=\sum_{i=0}^{n-1}\log\Delta_{i+1}^2. \tag{7}$$

As the a_{ij} are iid standard gaussian, Δ_{i+1}^2 are independent Chi-square random variables of degree $n-i$. Thus, the right hand side of (7) is a sum of independent random variables. Notice that Δ_{i+1}^2 has mean $n-i$ and variance $O(n-i)$ and is very strongly concentrated. Thus, with high probability $\log\Delta_{i+1}^2$ is roughly $\log((n-i)+O(\sqrt{n-i}))$ and so it is easy to show that $\log\Delta_{i+1}^2$ has mean close to $\log(n-i)$ and variance $O(\frac{1}{n-i})$. So the variance of $\sum_{i=0}^{n-1}\log\Delta_{i+1}^2$ is $O(\log n)$. To get the precise value $\sqrt{2\log n}$ one needs to carry out some careful (but rather routine) calculation, which we leave as an exercise.

The reason for which we think that the rate $\log^{-1/3+o(1)} n$ might be near optimal is that (as the reader will see though the proofs) $2\log n$ is only an asymptotic value of the variance of $\log|\det A_n|$. This approximation has error term of order at least $\Omega(1)$ and since $\sqrt{2\log n+\Omega(1)}-\sqrt{2\log n}=\Omega(\log^{-1/2} n)$, it seems that one cannot have rate of convergence better than $\log^{-1/2+o(1)} n$. It is a quite interesting question whether one can obtain a polynomial rate by replacing $\log(n-1)!$ and $2\log n$ by other, relatively simple, functions of n.

Our arguments rely on recent developments in random matrix theory and look quite different from those in Girko's papers. In particular, we benefit from the arguments developed in [22, 24, 26]. We also use Talagrand's famous concentration inequality frequently to obtain most of the large deviation results needed in this paper.

Notation. We say that an event E holds almost surely if $\mathbf{P}(E)$ tends to one as n tends to infinity. For an event A, we use the subscript $\mathbf{P}_{\mathbf{x}}(A)$ to emphasize that the probability under consideration is taking according to the random vector $\mathbf{x}$. For $1\leq s\leq n$, we denote by $\mathbf{e}_s$ the unit vector $(0,\dots,0,1,0,\dots,0)$, where all but the s-th component are zero. All standard asymptotic notation such as $O,\Omega,o,...$ etc are used under the assumption that $n\to\infty$.

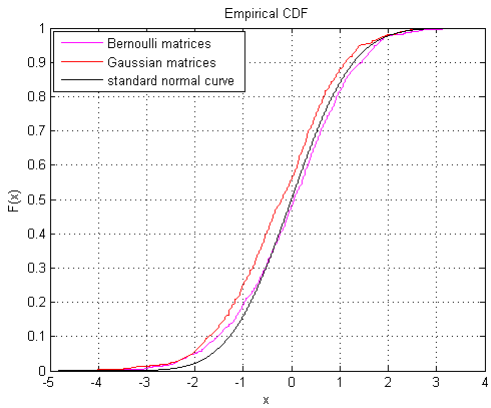

Figure 1: The plot compares the distributions of $(\log(\det A^2) - \log(n-1)!)/\sqrt{2\log n}$ for random Bernoulli matrices, random Gaussian matrices, and $\mathbf{N}(0,1)$. We sampled 1000 matrices of size 1000 by 1000 for each ensemble.

2. Our approach and main lemmas

We first make two extra assumptions about A_n. We assume that the entries a_{ij} are bounded in absolute value by $\log^\beta n$ for some constant $\beta > 0$ and A_n has full rank with probability one. We will prove Theorem 1.1 under these two extra assumptions. In Appendix A, we will explain why we can implement these assumptions without violating the generality of Theorem 1.1.

Theorem 2.1 (Main theorem with extra assumptions)**.** *Assume that all atom variables a_{ij} satisfy condition* **C0** *and are bounded in absolute value by $\log^\beta n$ for some constant β. Assume furthermore that A_n has full rank with probability one. Then*

$$\sup_{x\in\mathbf{R}} \Big|\mathbf{P}\big(\frac{\log(|\det A_n|) - \frac{1}{2}\log(n-1)!}{\sqrt{\frac{1}{2}\log n}} \le x\big) - \Phi(x)\Big| \le \log^{-1/3+o(1)} n. \tag{8}$$

In the first, and main, step of the proof, we prove the claim of Theorem 2.1 but with the last $\log^\alpha n$ rows being replaced by gaussian rows (for some properly chosen constant α). We remark that the replacement trick was also used in [8], but for an entire different reason. Our reason here is that for the last few rows, Lemma 2.4 is not very effective.

Theorem 2.2. *For any constant $\beta > 1$ the following holds for any sufficiently large constant $\alpha > 0$. Let A_n be an n by n matrix whose entries $a_{ij}, 1 \le i \le n_0, 1 \le j \le n$, are independent real random variables of zero mean, unit variance and absolute values at most $\log^\beta n$. Assume furthermore that A_n has full rank with probability one and the components of the last $\log^\alpha n$ rows of A are independent standard gaussian random variables. Then*

$$\sup_{x\in\mathbf{R}} \Big|\mathbf{P}\big(\frac{\log(\det A_n^2) - \log(n-1)!}{\sqrt{2\log n}} \le x\big) - \Phi(x)\Big| \le \log^{-1/3+o(1)} n. \tag{9}$$

In the second (and simpler) step of the proof, we carry out a replacement procedure, replacing the gaussian rows by the original rows once at a time and show that the replacement does not effect the central limit theorem. This step is motivated by the Lindeberg replacement method used in [26].

We present the verification of Theorem 2.1 using Theorem 2.2 in Section 8. In the rest of this section, we focus on the proof of Theorem 2.2.

Notice that in the setting of this theorem, the variables Δ_i are no longer independent. However, with some work, we can make the RHS of (7) into a sum of martingale differences plus a negligible error, which lays ground for an application of a central limit theorem of martingales. (In [8], Girko also used the CLT for martingales via the base times height formula, but his analysis looks very different from ours.) We are going to use the following theorem, due to Machkouri and Ouchti [15].

Theorem 2.3 (Central limit theorem for martingales)**.** [15, Theorem 1] *There exists an absolute constant L such that the following holds. Assume that $X_1, \dots, X_m$ are martingale differences with respect to the nested σ-algebras $\mathcal{E}_0, \mathcal{E}_1, \dots, \mathcal{E}_{m-1}$. Let $v_m^2 := \sum_{i=0}^{m-1} \mathbf{E}(X_{i+1}^2|\mathcal{E}_i)$, and $s_m^2 := \sum_{i=1}^m \mathbf{E}(X_i^2)$. Assume that $\mathbf{E}(|X_{i+1}^3||\mathcal{E}_i) \le \gamma_i \mathbf{E}(X_{i+1}^2|\mathcal{E}_i)$ with probability one for all i, where $(\gamma_i)_1^m$ is a sequence of positive real numbers. Then we have*

$$\sup_{x\in\mathbf{R}} \Big|\mathbf{P}(\frac{\sum_{0\le i<m} X_{i+1}}{s_m} < x) - \Phi(x)\Big| \le L\Big(\frac{\max\{\gamma_0, \dots, \gamma_{m-1}\}\log m}{\min\{s_m, 2^m\}} + \mathbf{E}^{1/3}(|\frac{v_m^2}{s_m^2} - 1|)\Big).$$

To make use of this theorem, we need some preparation. Condition on the first i rows $\mathbf{a}_1, \dots, \mathbf{a}_i$, we can view Δ_{i+1} as the distance from a random vector to $V_i := \mathrm{Span}\,(\mathbf{a}_1, \dots, \mathbf{a}_i)$. Since A_n has full rank with probability one, $\dim V_i = i$ with probability one for all i. The following is a direct corollary of [26, Lemma 43].

Lemma 2.4. *For any constant $\beta > 0$ there is a constant $C_3 > 0$ depending on β such that the following holds. Assume that $V \subset \mathbf{R}^n$ is a subspace of dimension $\dim(V) \le n-4$. Let $\mathbf{a}$ be a random vector whose components are independent variables of zero mean and unit variance and absolute values at most $\log^\beta n$. Denote by Δ the distance from $\mathbf{a}$ to V. Then we have*

$$\mathbf{E}(\Delta^2) = n - \dim(V) = n - i$$

and for any $t > 0$

$$\mathbf{P}(|\Delta - \sqrt{n - \dim(V)}| \ge t) = O\big(\exp(-\frac{t^2}{\log^{C_3} n})\big).$$

Set

$$n_0 := n - \log^\alpha n,$$

where α is a sufficiently large constant (which may depend on β). We will use short hand k_i to denote $n-i$, the co-dimension of V_i (and the expected value of Δ_i^2),

$$k_i := n - i.$$

We next consider each term of the right hand side of (7) where $0 \leq i < n_0$. Using Taylor expansion, we write

$$\begin{aligned}\log \frac{\Delta_{i+1}^2}{k_i} &= \log(1 + \frac{\Delta_{i+1}^2 - k_i}{k_i}) \\ &= \frac{\Delta_{i+1}^2 - k_i}{k_i} - \frac{1}{2}(\frac{\Delta_{i+1}^2 - k_i}{k_i})^2 + R_{i+1} \\ &:= X_{i+1} - \frac{X_{i+1}^2}{2} + R_{i+1},\end{aligned}$$

where

$$X_{i+1} := \frac{\Delta_{i+1}^2 - k_i}{k_i}, \text{ and } R_{i+1} := \log(1 + X_{i+1}) - (X_{i+1} - \frac{X_{i+1}^2}{2}).$$

By applying Lemma 2.4 with $t = k_i^{1/8} \geq \log^{\alpha/8} n$ and by choosing α sufficiently large, we have with probability at least $1 - O(\exp(-\log^2 n))$ (the probability here is with respect to the random $(i+1)$-th row, fixing the first i rows arbitrarily)

$$|X_{i+1}| = O(k_i^{-3/8}) = O((n-i)^{-3/8}) = o(1). \tag{10}$$

Thus, with probability at least $1 - O(\exp(-\log^2 n))$

$$|R_{i+1}| = O(|X_{i+1}|^3) = O((n-i)^{-9/8}).$$

Hence, by a uniform bound, the following holds with probability at least $1 - n \cdot O(\exp(-\log^2 n)) = 1 - O(\exp(-\log^2 n/2))$

$$\sum_{i<n_0} R_{i+1} = O(\sum_{i<n_0} (n-i)^{-9/8}) = o(\log^{-2} n),$$

again by having α sufficiently large.

We conclude that

Lemma 2.5. *With probability at least* $1 - O(\exp(-\log^2 n/2))$

$$\frac{\sum_{i<n_0} R_{i+1}}{\sqrt{2\log n}} = o(\frac{\log^{-2} n}{\sqrt{2\log n}}).$$

We will need three other lemmas.

Lemma 2.6 (Main contribution)**.**

$$\sup_{x\in\mathbf{R}} \Big|\mathbf{P}\big(\frac{\sum_{i<n_0} X_{i+1}}{\sqrt{2\log n}} \le x\big) - \Phi(x)\Big| \le \log^{-1/3+o(1)} n.$$

Lemma 2.7 (Quadratic terms)**.**

$$\mathbf{P}\Big(\Big|\frac{-\sum_{i<n_0} X_{i+1}^2/2 + \log n}{\sqrt{2\log n}}\Big| \ge \log^{-1/3+o(1)} n\Big) \le \log^{-1/3+o(1)} n.$$

Lemma 2.8 (Last few rows)**.** *For any constant* $0 < c < 1/100$

$$\mathbf{P}\Big(\Big|\frac{\sum_{n_0\le i} \log \frac{\Delta_{i+1}^2}{n-i}}{\sqrt{2\log n}}\Big| \ge \log^{-1/2+c} n\Big) = o(\exp(-\log^{c/2} n)).$$

Theorem 2.2 follows from the above four lemmas and the following trivial fact (used repeatedly and with proper scaling)

$$\mathbf{P}(A + B \le \sigma x) \le \mathbf{P}(A \le \sigma(x-\varepsilon)) + \mathbf{P}(B \le \sigma\varepsilon).$$

The reader is invited to fill in the simple details using the following observation

$$\begin{aligned}
\log(\det A_n^2) - \log(n-1)! &= \sum_{i=0}^{n-1} \log \Delta_{i+1}^2 - \log(n-1)! \\
&= \sum_{i=0}^{n-1} \log \frac{\Delta_{i+1}^2}{k_i} + \log n! - \log(n-1)! \\
&= \sum_{i<n_0} (X_{i+1} - \frac{X_{i+1}^2}{2} + R_{i+1}) + \sum_{n_0\le i} \log \frac{\Delta_{i+1}^2}{k_i} + \log n \\
&= \sum_{i<n_0} X_{i+1} - (\sum_{i<n_0} \frac{X_{i+1}^2}{2} - \log n) + \sum_{i<n_0} R_{i+1} + \sum_{n_0\le i} \log \frac{\Delta_{i+1}^2}{k_i}.
\end{aligned}$$

We will prove Lemma 2.6 using Theorem 2.3. Lemma 2.7 will be verified by the moment method and Lemma 2.8 by elementary properties of Chi-square variables. The key to the proof of Lemmas 2.6 and 2.7 is an estimate on the entries of the projection matrix onto the space $V_i^\perp$, presented in Section 4.

3. Proof of Lemma 2.6 and Lemma 2.7: Opening

We recall from the previous section that $X_{i+1} = \frac{\Delta_{i+1}^2 - k_i}{k_i}$. Denote by $P_i = (p_{st}(i))_{s,t}$ the projection matrix onto the orthogonal complement $V_i^{\perp}$. A standard fact in linear algebra is

$$\mathrm{tr}(P_i) = \sum_s p_{ss}(i) = k_i; \text{ and } \sum_{s,t} p_{st}(i)^2 = \sum_s p_{ss}(i) = k_i. \tag{11}$$

We now express X_{i+1} using P_i,

$$X_{i+1} = \frac{\|P_i \mathbf{a}_{i+1}\|^2 - k_i}{k_i} = \frac{\sum_{s,t} p_{st}(i) a_s a_t - k_i}{k_i} := \sum_{s,t} q_{st}(i) a_s a_t - 1, \tag{12}$$

where $a_1 = a_{i+1,1}, \dots, a_n = a_{i+1,n}$ are the coordinates of the vector $\mathbf{a}_{i+1}$ and

$$q_{st}(i) := \frac{p_{st}(i)}{k_i}.$$

By (11) we have $\sum_s q_{ss}(i) = 1$ and $\sum_{s,t} q_{st}(i)^2 = \frac{1}{k_i}$.

Because $\mathbf{E} a_s = 0$ and $\mathbf{E} a_s^2 = 1$, and the a_s are mutually independent, we can show by using a routine calculation that (see (21) from Section 6)

$$\mathbf{E}(X_{i+1}^2 | \mathcal{E}_i) = \frac{2}{k_i} - \sum_s q_{ss}(i)^2 (3 - \mathbf{E} {a_s}^4), \tag{13}$$

where $\mathcal{E}_i$ is the σ-algebra generated by the first i rows of A_n.

Define

$$Y_{i+1} := -\frac{X_{i+1}^2}{2} + \frac{1}{k_i} - \frac{1}{2} \sum_s q_{ss}(i)^2 (3 - \mathbf{E} {a_s}^4)$$

and

$$Z_{i+1} := \frac{1}{2} \sum_s q_{ss}(i)^2 (3 - \mathbf{E} {a_s}^4).$$

The reason we split $-\frac{X_{i+1}^2}{2} + \frac{1}{k_i}$ into the sum of Y_{i+1} and Z_{i+1} is that $\mathbf{E}(Y_{i+1} | \mathcal{E}_i) = 0$ and its variance can be easily computed.

Lemma 3.1.

$$\mathbf{P}\Big(|\frac{\sum_{i<n_0} Y_{i+1}}{\sqrt{2\log n}}| \geq \log^{-1/3+o(1)} n\Big) \leq \log^{-1/3+o(1)} n.$$

To complete the proof of Lemma 2.7 from Lemma 3.1, it suffices to show that the sum of the Z_i is negligible,

$$\mathbf{P}\Big(\frac{\sum_{i<n_0} Z_{i+1}}{\sqrt{2\log n}} = \Omega(\frac{\log\log n}{\sqrt{2\log n}})\Big) = O(n^{-100}). \tag{14}$$

Our main technical tool will be the following lemma.

Lemma 3.2. *With probability* $1 - O(n^{-100})$ *we have*

$$\sum_{i<n_0}\sum_s q_{ss}(i)^2 = O(\log\log n).$$

Notice that $\mathbf{E}a_s^4$ is uniformly bounded (by condition **C0**), it follows that with probability $1 - O(n^{-100})$,

$$\sum_{i<n_0}\sum_s q_{ss}(i)^2|3 - \mathbf{E}a_s^4| = O(\log\log n),$$

proving (14).

4. Proof of Lemma 2.6 and Lemma 2.7: Mid game

The key idea for proving Lemma 3.2 is to establish a good upper bound for $|q_{ss}(i)|$. For this, we need some new tools. Our main ingredient is the following delocalization result, which is a variant of a result from [24] (see also [5] and [27] for recent surveys), asserting that with high probability all unit vectors in the orthogonal complement of a random subspace with high dimension have small infinity norm.

Lemma 4.1. *For any constant* $\beta > 0$ *the following holds for all sufficiently large constant* $\alpha > 0$. *Assume that the components of* $\mathbf{a}_1, \dots, \mathbf{a}_{n_1}$, *where* $n_1 := n - n\log^{-4\alpha} n$, *are independent random variables of mean zero, variance one and bounded in absolute value by* $\log^\beta n$. *Then with probability* $1 - O(n^{-100})$, *the following holds for all unit vectors* $\mathbf{v}$ *of the space* $V_{n_1}^\perp$,

$$\|\mathbf{v}\|_\infty = O(\log^{-2\alpha} n).$$

Proof. (of Lemma 3.2 assuming Lemma 4.1) Write

$$S = \sum_{i \le n_1} \sum_s q_{ss}(i)^2 + \sum_{n_1 < i < n_0} \sum_s q_{ss}(i)^2$$
$$:= S_1 + S_2.$$

Note that as $q_{st}(i) = p_{st}(i)/k_i$,

$$\sum_s q_{ss}(i)^2 \le \sum_{s,t} q_{st}(i)^2 = \sum_{s,t} \frac{p_{st}(i)^2}{k_i^2} = \frac{1}{k_i} = \frac{1}{(n-i)}.$$

Hence,

$$S_1 \le \sum_{i \le n_1} \sum_s q_{ss}(i)^2 \le \sum_{i \le n_1} \frac{1}{(n-i)} = O(\log \log n).$$

To bound S_2, note that

$$p_{ss}(i) = \mathbf{e}_s^T P_i \mathbf{e}_s = \|P_i \mathbf{e}_s\|^2 = |\langle \mathbf{e}_s, \mathbf{v} \rangle|^2$$

for some unit vector $\mathbf{v} \in V_i^\perp$.

Thus if $i > n_1$, then $V_i^\perp \subset V_{n_1}^\perp$, and hence by Lemma 4.1

$$p_{ss}(i) \le \|\mathbf{v}\|_\infty^2 = O(\log^{-4\alpha} n). \tag{15}$$

It follows that

$$\begin{aligned} S_2 &\le \sum_{n_1 < i < n_0} \max_s p_{ss}(i) \sum_s \frac{p_{ss}(i)}{(n-i)^2} \\ &= O(\log^{-4\alpha} n) \sum_{n_1 \le i < n_0} \frac{1}{(n-i)} = O(\log^{-4\alpha+1} n), \end{aligned}$$

completing the proof of Lemma 3.2. □

We now focus on the infinity norm of $\mathbf{v}$ and follow an argument from [24].

Proof. (of Lemma 4.1) By the union bound, it suffices to show that $|v_1| = O(\log^{-2\alpha} n)$ with probability at least $1 - O(n^{-101})$, where v_1 is the first coordinate of $\mathbf{v}$.

Let B be the matrix formed by the first n_1 rows $\mathbf{a}_1, \dots, \mathbf{a}_{n_1}$ of A. Assume that $\mathbf{v} \in V_{n_1}^{\perp}$ is a unit vector, then

$$B\mathbf{v} = 0.$$

Let $\mathbf{w}$ be the first column of B, and B' be the matrix obtained by deleting $\mathbf{w}$ from B. Clearly,

$$v_1 \mathbf{w} = -B'\mathbf{v}', \tag{16}$$

where $\mathbf{v}'$ is the vector obtained from $\mathbf{v}$ by deleting v_1.

We next invoke the following result, which is a variant of [24, Lemma 4.1]. This lemma was proved using a method of Guionet and Zeitouni [11], based on Talagrand's inequality.

Lemma 4.2 (Concentration of singular values)**.** *For any constant $\beta > 0$ the following holds for all sufficiently large constant $\alpha > 0$. Let A_n be a random matrix of size n by n, where the entries a_{ij} are independent random variables of mean zero, variance one and bounded in absolute value by $\log^\beta n$. Then for any $n/\log^\alpha n \le k \le n/2$, there exist $2k$ singular values of A_n in the interval $[0, ck/\sqrt{n}]$, for some absolute constant c, with probability at least $1 - O(n^{-101})$.*

We can prove Lemma 4.2 by following the arguments in [24, Lemma 4.1] almost word by word.

By the interlacing law and Lemma 4.2, we conclude that B' has $n - n_1$ singular values in the interval $[0, c(n-n_1)/\sqrt{n}]$ with probability $1 - O(n^{-101})$.

Let H be the space spanned by the left singular vectors of these singular values, and let π be the orthogonal projection onto H. By definition, the spectral norm of $\pi B'$ is bounded,

$$\|\pi B'\| \le c(n-n_1)/\sqrt{n}.$$

Thus (16) implies that

$$|v_1| \|\pi \mathbf{w}\| \le c(n-n_1)/\sqrt{n}, \tag{17}$$

here we used the fact that $\mathbf{w}$ is independent from B', and thus from π.

On the other hand, since the dimension of H is $n - n_1$, Lemma 2.4 implies that $\|\pi \mathbf{w}\| \ge \sqrt{n-n_1}/2$ with probability $1 - 4\exp(-(n-n_1)/16) = 1 - O(n^{-\omega(1)})$.

It thus follows from (17) that

$$|v_1| = O(\log^{-2\alpha} n).$$

□

5. Proof of Lemma 2.6: End game

Recall from (10) that conditioned on any first i rows, $|X_i| = O(k_i^{-3/8})$ with probability $1 - O(\exp(-\log^2 n/2))$. So, by paying an extra term of $O(\exp(-\log^2 n/2))$ in probability, it suffices to justify Lemma 2.6 for the sequence $X'_i := X_i \cdot \mathbf{I}_{|X_i|=O(k_i^{-3/8})}$.

On the other hand, the sequence X'_{i+1} is not a martingale difference sequence, so we slightly modify X'_{i+1} to $X''_{i+1} := X'_{i+1} - \mathbf{E}(X'_{i+1}|\mathcal{E}_i)$ and prove the claim for the sequence X'_{i+1}, here we recall that $\mathcal{E}_i$ is the σ-algebra generated by the first i rows of A_n. In order to show that this modification has no effect whatsoever, we first demonstrate that $\mathbf{E}(X'_{i+1}|\mathcal{E}_i)$ is extremely small.

Recall from (12) that $X_{i+1} = \sum_{s,t} q_{st}(i) a_s a_t - 1$. By Cauchy-Schwarz inequality and the assumption that a_s are bounded in absolute value by $\log^{O(1)} n$, we have with probability one

$$\begin{aligned} |X_{i+1}|^2 &\leq (1 + \sum_{s,t} q_{st}(i)^2)(1 + \sum_{s,t} a_s^2 a_t^2) = (1 + 1/k_i)(1 + \sum_{s,t} a_s^2 a_t^2) \\ &\leq 2(1 + \sum_{s,t} a_s^2 a_t^2) \leq n^2 \log^{O(1)} n. \end{aligned} \tag{18}$$

Thus with probability one

$$|\mathbf{E}(X'_{i+1}|\mathcal{E}_i)| = |\mathbf{E}(X'_{i+1}|\mathcal{E}_i) - \mathbf{E}(X_{i+1}|\mathcal{E}_i)| \leq \exp(-(\frac{1}{2} - o(1)) \log^2 n). \tag{19}$$

To justify Lemma 2.6 for the sequence X''_{i+1}, we apply Theorem 2.3.

The key point here is that thanks to the indicator function in the definition of X'_{i+1} and the fact that the difference between X''_{i+1} and X'_{i+1} is negligible, X''_{i+1} is bounded by $O(k_i^{-3/8})$ with probability one, so the conditions $\mathbf{E}(|X''_{i+1}|^3|\mathcal{E}_i) \leq \gamma_i \mathbf{E}({X''_{i+1}}^2|\mathcal{E}_i)$ in Theorem 2.3 are satisfied with

$$\gamma_i = O(k_i^{-3/8}) = O(\log^{-3\alpha/8} n).$$

We need to estimate s_{n_0}, v_{n_0} with respect to the sequence X''_{i+1}. However, thanks to the observations above X_{i+1} and X''_{i+1} are very close, and so it suffices to compute these values with respect to the sequence X_{i+1}.

Recall from (13) that

$$\mathbf{E}(X_{i+1}^2|\mathcal{E}_i) = \frac{2}{k_i} - \sum_s q_{ss}(i)^2(3 - \mathbf{E}{a_s}^4).$$

Also, recall from Section 4 that with probability $1 - O(n^{-100})$,

$$\sum_{i<n_0}\sum_s q_{ss}(i)^2(3 - \mathbf{E}a_s^4) = O(\log\log n).$$

This bound, together with (13) and (18), imply that with probability one

$$\mathbf{E}(\sum_{i<n_0} X_{i+1}^2|\mathcal{E}_i) = \sum_{i<n_0}\frac{2}{k_i} + O(\log\log n) = 2\log n + O(\log\log n),$$

which in turn implies that $v_{n_0}^2 = 2\log n + O(\log\log n)$ with probability $1 - O(n^{-100})$.

Using (18) again, because $n^{-100}n^2\log^{O(1)} n = o(1)$, we deduce that

$$s_{n_0}^2 = 2\log n + O(\log\log n). \tag{20}$$

With another application of (18), we obtain

$$\mathbf{E}|\frac{v_{n_0}^2}{s_{n_0}^2} - 1| \leq O(\frac{\log\log n}{\log n}) + n^{-100}n^2\log^{O(1)} n.$$

It follows that

$$\mathbf{E}^{1/3}|\frac{v_{n_0}^2}{s_{n_0}^2} - 1| \leq \log^{-1/3+o(1)} n.$$

By the conclusion of Theorem 2.3 and setting α sufficiently large, we conclude

$$\begin{aligned}\sup_{x\in\mathbf{R}}\Big|\mathbf{P}(\frac{\sum_{i<n_0} X''_{i+1}}{s_{n_0}} < x) - \Phi(x)\Big| &\leq L\Big(\frac{\log^{-3\alpha/8} n \times \log n_0}{s_{n_0}} + \mathbf{E}^{1/3}(|\frac{v_{n_0}^2}{s_{n_0}^2} - 1|)\Big)\\ &\leq \log^{-1/3+o(1)} n,\end{aligned}$$

completing the proof of Lemma 2.6.

6. Proof of Lemma 2.7: End Game

Our goal is to justify Lemma 3.1, which together with (14) verify Lemma 2.7.

We will show that the variance $\mathbf{Var}(\sum_{i<n_0} Y_{i+1})$ is small and then use Chebyshev's inequality. The proof is based on a series of routine, but somewhat tedious calculations. We first show that the expectations of the Y_{i+1}'s are zero, and so are the covariances $\mathbf{E}(Y_{i+1}Y_{j+1})$ by an elementary manipulation. The variances $\mathbf{Var}(Y_{i+1})$ will be bounded from above by Cauchy-Schwarz inequality.

We start with the formula $X_{i+1}^2 = (\sum_{s,t} q_{st}(i)a_s a_t)^2 - 2\sum_{s,t} q_{st}(i)a_s a_t + 1$. Observe that

$$\begin{aligned}(\sum_{s,t} q_{st}(i)a_s a_t)^2 &= (\sum_s q_{ss}(i)a_s^2 + \sum_{s\neq t} q_{st}(i)a_s a_t)^2 \\ &= (\sum_s q_{ss}(i){a_s}^2)^2 + (\sum_{s\neq t} q_{st}(i)a_s a_t)^2 + 2(\sum_s q_{ss}(i)a_s^2)(\sum_{s\neq t} q_{st}(i)a_s a_t).\end{aligned}$$

Expanding each term, using the fact that $\sum_s q_{ss}(i) = 1$ and $\sum_{s,t} q_{st}(i)^2 = \frac{1}{k_i}$, we have

$$\begin{aligned}(\sum_s q_{ss}(i){a_s}^2)^2 &= (\sum_s q_{ss}(i))^2 - \sum_s q_{ss}(i)^2(1-a_s^4) + 2\sum_{s\neq t} q_{ss}(i)q_{tt}(i)(a_s^2a_t^2 - 1) \\ &= 1 - \sum_s q_{ss}(i)^2(1-a_s^4) + 2\sum_{s\neq t} q_{ss}(i)q_{tt}(i)(a_s^2a_t^2 - 1),\end{aligned}$$

and

$$\begin{aligned}(\sum_{s\neq t} q_{st}(i)a_s a_t)^2 &= 2\sum_{s\neq t} q_{st}(i)^2 + 2\sum_{s\neq t} q_{st}(i)^2(a_s^2a_t^2 - 1) + 2\sum_{\substack{s_1\neq t_1, s_2\neq t_2\\ \{s_1,t_1\}\neq\{s_2,t_2\}}} q_{s_1t_1}(i)q_{s_2t_2}(i)a_{s_1}a_{t_1}a_{s_2}a_{t_2} \\ &= \frac{2}{k_i} - 2\sum_s q_{ss}(i)^2 + 2\sum_{s\neq t} q_{st}(i)^2(a_s^2a_t^2 - 1) + 2\sum_{\substack{s_1\neq t_1, s_2\neq t_2\\ \{s_1,t_1\}\neq\{s_2,t_2\}}} q_{s_1t_1}(i)q_{s_2t_2}(i)a_{s_1}a_{t_1}a_{s_2}a_{t_2}\big),\end{aligned}$$

as well as

$$2(\sum_s q_{ss}(i)a_s^2)(\sum_{s\neq t} q_{st}(i)a_s a_t) = 2(\sum_s q_{ss}(i)(a_s^2 - 1))(\sum_{s\neq t} q_{st}(i)a_s a_t) + 2\sum_{s\neq t} q_{st}(i)a_s a_t.$$

It follows that

$$
\begin{aligned}
&-2Y_{i+1} = X_{i+1}^2 - \frac{2}{k_i} + \sum_s q_{ss}(i)^2(3-\mathbf{E}{a_s}^4) = (\sum_{s,t} q_{st}(i)a_sa_t - 1)^2 - \frac{2}{k_i} + \sum_s q_{ss}(i)^2(3-\mathbf{E}{a_s}^4)\\
&= (\sum_{s,t} q_{st}(i)a_sa_t)^2 - 1 - 2\sum_s q_{ss}(i)(a_s^2-1) - 2\sum_{s\neq t} q_{st}(i)a_sa_t - \frac{2}{k_i} + \sum_s q_{ss}(i)^2(3-\mathbf{E}{a_s}^4)\\
&= -2\sum_s q_{ss}(i)(a_s^2-1) + \sum_s q_{ss}(i)^2(a_s^4-\mathbf{E}a_s^4) + 2\sum_{s\neq t} q_{ss}(i)q_{tt}(i)(a_s^2a_t^2-1) + 2\sum_{s\neq t} q_{st}(i)^2(a_s^2a_t^2-1)\\
&+ 2\sum_{\substack{s_1\neq t_1, s_2\neq t_2\\ \{s_1,t_1\}\neq\{s_2,t_2\}}} q_{s_1t_1}(i)q_{s_2t_2}(i)a_{s_1}a_{t_1}a_{s_2}a_{t_2} + 2(\sum_s q_{ss}(i)(a_s^2-1))(\sum_{s\neq t} q_{st}(i)a_sa_t). \qquad (21)
\end{aligned}
$$

As $\mathbf{E}a_s = 0, \mathbf{E}a_s^2 = 1$, and the a_s's are mutually independent with each other and with every row of index at most i (and in particular with $q_{st}(i)$'s), every term in the last formula is zero, and so we infer that $\mathbf{E}(Y_{i+1}) = 0$ and $\mathbf{E}(Y_{i+1}|\mathcal{E}_i) = 0$, confirming (13). With the same reasoning, we can also infer that the covariance $\mathbf{E}(Y_{i+1}Y_{j+1}) = 0$ for all $j < i$.

It is thus enough to work with the diagonal terms $\mathbf{Var}(Y_{i+1})$. We have

$$
\begin{aligned}
\mathbf{Var}(Y_{i+1}) = \mathbf{E}\Big[&-\sum_s q_{ss}(i)(a_s^2-1) + \frac{1}{2}\sum_s q_{ss}(i)^2(a_s^4-\mathbf{E}a_s^4)\\
&+ \sum_{s\neq t} q_{ss}(i)q_{tt}(i)(a_s^2a_t^2-1) + \sum_{s\neq t} q_{st}(i)^2(a_s^2a_t^2-1)\\
&+ \sum_{\substack{s_1\neq t_1, s_2\neq t_2\\ \{s_1,t_1\}\neq\{s_2,t_2\}}} q_{s_1t_1}(i)q_{s_2t_2}(i)a_{s_1}a_{t_1}a_{s_2}a_{t_2} + (\sum_s q_{ss}(i)(a_s^2-1))(\sum_{s\neq t} q_{st}(i)a_sa_t)\Big]^2.
\end{aligned}
$$

After a series of cancellation, and because of Condition **C0**, we have

$$\begin{aligned}\mathbf{Var}(Y_{i+1}) \le O\Big(\mathbf{E}\Big[&\sum_s q_{ss}(i)^2 + \sum_s q_{ss}(i)^4 + \sum_{s\neq t_1, s\neq t_2} q_{ss}(i)^2 q_{t_1t_1}(i) q_{t_2t_2}(i)\\ &+ \sum_{s\neq t_1, s\neq t_2} q_{st_1}(i)^2 q_{st_2}(i)^2 + \sum_{s_1\neq t_1, s_2\neq t_2} |q_{s_1t_1}(i) q_{s_1t_2}(i) q_{s_2t_1}(i) q_{s_2t_2}(i)| + \sum_{s,t} q_{ss}(i) q_{tt}(i) q_{st}(i)^2\\ &+ \sum_s q_{ss}(i)^3 + \sum_{s,t} q_{ss}(i)^2 q_{tt}(i) + \sum_{s,t} q_{ss}(i) q_{st}(i)^2 + \sum_{s,t} |q_{ss}(i) q_{tt}(i) q_{st}(i)|\\ &+ \sum_{s,t} q_{ss}(i)^3 q_{tt}(i) + \sum_{s\neq t} q_{ss}(i)^2 q_{st}(i)^2 + \sum_{s,t} |q_{ss}(i)^2 q_{tt}(i) q_{st}(i)|\\ &+ \sum_{s,t} q_{ss}(i) q_{tt}(i) q_{st}(i)^2 + \sum_{s\neq t} |q_{ss}(i)^2 q_{tt}(i) q_{st}(i)|\\ &+ \sum_{s,t} |q_{st}(i)^3 q_{ss}(i)|\\ &+ \sum_{s\neq t_1, s\neq t_2, t_1\neq t_2} |q_{ss}(i) q_{st_1}(i) q_{st_2}(i) q_{t_1t_2}(i)|\Big]\Big),\end{aligned}$$

where the first two rows consist of the squares of the terms appearing in Y_{i+1} (after deleting several sums of zero expected value), and each of the following rows was obtained by expanding the product of each term with the rest in the order of their appearance.

Because $\sum_{s,t} q_{st}(i)^2 = \frac{1}{k_i}$, one has $\max_{s,t} |q_{st}(i)| \le \frac{1}{\sqrt{k_i}}$ for all s, t. Recall furthermore that $\sum_s q_{ss}(i) = 1$ and $0 \le q_{ss}(i)$ for all s. We next estimate the terms under consideration one by one as follows.

Firstly, the sums $\sum_s q_{ss}^3(i), \sum_s q_{ss}(i)^4, \sum_{s,t} q_{ss}(i) q_{st}(i)^2, \sum_{s,t} q_{ss}(i)^2 q_{st}(i)^2, \sum_{s,t} q_{ss}(i) q_{tt}(i) q_{st}(i)^2$, and $\sum_{s,t} |q_{st}(i)^3 q_{ss}(i)|$ can be bounded by $\max_{s,t} |q_{st}(i)| \sum_{s,t} q_{st}^2(i)$, and so by $k_i^{-3/2}$.

Secondly, by applying Cauchy-Schwarz inequality if needed, one can bound the sums $\sum_{s,t_1,t_2} q_{st_1}(i)^2 q_{st_2}(i)^2$, $\sum_{s_1,t_1,s_2,t_2} |q_{s_1t_1}(i) q_{s_1t_2}(i) q_{s_2t_1}(i) q_{s_2t_2}(i)|$, and $\sum_{s,t_1,t_2} |q_{ss}(i) q_{st_1}(i) q_{st_2}(i) q_{t_1t_2}(i)|$ by $2(\sum_{s,t} q_{st}^2(i))^2$, and so by $2k_i^{-2}$.

We bound the remaining terms as follows.

- $\sum_{s,t_1,t_2} q_{ss}(i)^2 q_{t_1t_1}(i) q_{t_2t_2}(i) = (\sum_s q_{ss}(i)^2)(\sum_t q_{tt}(i))^2 = \sum_s q_{ss}(i)^2$.

- $\sum_{s,t} q_{ss}(i)^2 q_{tt}(i) + \sum_{s,t} q_{ss}(i)^3 q_{tt}(i) \le 2(\sum_s q_{ss}(i)^2)(\sum_t q_{tt}(i)) = 2\sum_s q_{ss}(i)^2$.

- $\sum_{s,t} |q_{ss}(i) q_{tt}(i) q_{st}(i)| \le \sum_{s,t} q_{ss}(i)(q_{tt}(i)^2 + q_{st}(i)^2) \le \sum_t q_{tt}(i)^2 + \max_s q_{ss}(i) \sum_{s,t} q_{st}(i)^2$ $\le \sum_t q_{tt}(i)^2 + k_i^{-3/2}$.

- $\sum_{s,t} |q_{ss}(i)^2 q_{tt}(i) q_{st}(i)| \le \sup_{s,t} |q_{st}(i)| \sum_{s,t} q_{ss}(i)^2 q_{tt}(i) \le \sum_s q_{ss}(i)^2/\sqrt{k_i}$.

Putting all bounds together we have

$$\mathbf{Var}(\sum_{i<n_0} Y_{i+1}) = \sum_{i<n_0} \mathbf{Var}(Y_{i+1}) = O\Big(\mathbf{E}(\sum_{i<n_0}\sum_s q_{ss}(i)^2 + \sum_{i<n_0} k_i^{-3/2})\Big)$$
$$= O(\log\log n). \tag{22}$$

where we applied Lemma 3.2 in the last estimate.

To complete the proof, we note from the estimate of $s_{n_0}^2$ of Section 5 and from Lemma 3.2 that $|\sum_{i<n_0} \mathbf{E}Y_{i+1}| = O(\log\log n)$. Thus by Chebyshev's inequality

$$\mathbf{P}\Big(|\frac{\sum_{i<n_0} Y_{i+1}}{\sqrt{2\log n}}| \geq \log^{-1/3+o(1)} n\Big) = \log^{-1/3+o(1)} n.$$

7. Proof of Lemma 2.8

We recall that, with $i \geq n_0$, Δ_{i+1}^2 is a Chi-square random variable of degree $n-i$. Let us first consider the lower tail; it suffices to show

$$\mathbf{P}\big(\sum_{n_0\leq i} \frac{\log\frac{\Delta_{i+1}^2}{n-i}}{\sqrt{2\log n}} < -\log^{-1/2+c} n\big) = o(\exp(-\log^{c/2} n)) \tag{23}$$

for any constant $0 < c < 1/100$.

By properties of the normal distribution, it is easy to show that Δ_n^2 and Δ_{n-1}^2 are at least $\exp(-\frac{\sqrt{2}}{4}\log^c n)$ with probability $1-\exp(-\Omega(\log^c n))$, so we can omit these terms from the sum. It now suffices to show that

$$\mathbf{P}\big(\sum_{n_0\leq i\leq n-3} \frac{\log\frac{\Delta_{i+1}^2}{n-i}}{\sqrt{2\log n}} < -\frac{1}{2}\log^{-1/2+c} n\big) = o(\exp(-\log^{c/2} n)) \tag{24}$$

for any small constant $0 < c < 1/100$.

Flipping the inequality inside the probability (by changing the sign of the RHS and swapping the denominators and numerators in the logarithms of the LHS) and using the Laplace transform trick (based on the fact that the Δ_i^2 are independent), we see that the probability in question is at most

$$\frac{\mathbf{E}\prod_{i=n_0}^{n-3}\frac{n-i}{\Delta_{i+1}^2}}{\exp(\frac{1}{\sqrt{2}}\log^c n)} = \frac{\prod_{i=n_0}^{n-3}\mathbf{E}\frac{n-i}{\Delta_{i+1}^2}}{\exp(\frac{1}{\sqrt{2}}\log^c n)}.$$

Recall that Δ_{i+1}^2 is a Chi-square random variable with degree of freedom $n-i$, so $\mathbf{E}\frac{1}{\Delta_{i+1}^2} = \frac{1}{n-i-2}$. Therefore, the numerator in the previous formula is $\frac{(n-n_0)(n-n_0-1)}{2} \leq \log^{2\alpha} n$.

Because

$$\frac{\log^{2\alpha} n}{\exp(\frac{1}{\sqrt{2}}\log^c n)} = o(\exp(-\log^{c/2} n)),$$

the desired bound follows.

The proof for the upper tail is similar (in fact simpler as we do not need to treat the first two terms separately) and we omit the details.

8. Deduction of Theorem 2.1 from Theorem 2.2

Our plan is to replace one by one the last $n-n_0$ gaussian rows of A_n by vectors of components having zero mean, unit variance, and satisfying Condition **C0**. Our key tool here is the classical Berry-Eseen inequality. In order to apply this lemma, we will make a crucial use of Lemma 4.1.

Lemma 8.1. [1, Berry-Esseen inequality] *Assume that* $\mathbf{v} = (v_1, \dots, v_n)$ *is a unit vector. Assume that* $b_1, \dots, b_n$ *are independent random variables of mean zero, variance one and satisfying condition* **C0**. *Then we have*

$$\sup_x |\mathbf{P}(v_1 b_1 + \dots + v_n b_n \leq x) - \Phi(x)| \leq c\|\mathbf{v}\|_\infty,$$

where c *is an absolute constant depending on the parameters appearing in* (3).

We remark that in the original setting of Berry and Esseen, it suffices to assume finite third moment.

In application, $\mathbf{v}$ plays the role of the normal vector of the hyperplane spanned by the remaining $n-1$ rows of A, and $\Delta_n = |v_1 b_1 + \dots + v_n b_n|$, where $(b_1, \dots, b_n) = \mathbf{b}$ is the vector to be replaced.

For the deduction, it is enough to show the following.

Lemma 8.2. *Let* A_n *be a random matrix with atom variables satisfying condition* **C0** *and non-singular with probability one. Assume furthermore that* A_n *has at least one and at most* $\log^\alpha n$ *gaussian rows. Let* B_n *be the random matrix obtained from* A_n *by replacing a gaussian row vector* $\mathbf{a}$ *of* A_n *by a random vector* $\mathbf{b} = (b_1, \dots, b_n)$ *whose coordinates are independent atom variables satisfying condition* **C0** *such that the resulting matrix is non-singular with probability one. Then*

$$\sup_x \Big|\mathbf{P}_{B_n}\Big(\frac{\log(\det B_n^2)-\log(n-1)!}{\sqrt{2\log n}} \le x\Big) - \mathbf{P}_{A_n}\Big(\frac{\log(\det {A_n}^2)-\log(n-1)!}{\sqrt{2\log n}} \le x\Big)\Big| \le O(\log^{-2\alpha} n). \tag{25}$$

Clearly, Theorem 1.1 follows from Theorem 2.2 by applying Lemma 8.2 $\log^{\alpha} n$ times.

Proof. (Proof of Lemma 8.2) Without loss of generality, we can assume that B_n is obtained from A_n by replacing the last row $\mathbf{a}_n$. As A_n is non-singular, $\dim(V_{n-1}) = n-1$.

By Lemma 4.1, by paying an extra term of $O(n^{-100})$ in probability (which will be absorbed by the eventual bound $\log^{-2\alpha} n$), we may also assume that the normal vector $\mathbf{v}$ of V_{n-1} satisfies

$$\|\mathbf{v}\|_\infty = O(\log^{-2\alpha} n).$$

Next, observe that

$$\frac{\log(\det A^2)-\log(n-1)!}{\sqrt{2\log n}} = \frac{\sum_{i=0}^{n-2}\log(\Delta_{i+1}^2/(n-i)) + \log n}{\sqrt{2\log n}} + \frac{\log \Delta_n^2}{\sqrt{2\log n}}$$

and

$$\frac{\log(\det B^2)-\log(n-1)!}{\sqrt{2\log n}} = \frac{\sum_{i=0}^{n-2}\log(\Delta_{i+1}^2/(n-i)) + \log n}{\sqrt{2\log n}} + \frac{\log {\Delta'_n}^2}{\sqrt{2\log n}}$$

where Δ_n and Δ'_n are the distance from $\mathbf{a}_n$ and $\mathbf{b}_n$ to V_{n-1} respectively.

By Lemma 8.1, it is yielded that

$$\sup_x |\mathbf{P}_{\mathbf{a}_n}(\Delta_n^2 \le x) - \mathbf{P}_{\mathbf{b}_n}({\Delta'_n}^2 \le x)| \le c\|\mathbf{v}\|_\infty = O(\log^{-2\alpha} n).$$

Hence

$$\sup_x |\mathbf{P}_{\mathbf{a}_n}\Big(\frac{\log(\det A^2)-\log(n-1)!}{\sqrt{2\log n}} \le x\Big) - \mathbf{P}_{\mathbf{b}_n}\Big(\frac{\log(\det B^2)-\log(n-1)!}{\sqrt{2\log n}} \le x\Big) = O(\log^{-2\alpha} n),$$

completing the proof of Lemma 8.2.

□

Appendix A. Simplifying the model: Deducing Theorem 1.1 from Theorem 2.1

In this section, we show that the two extra assumptions that $|a_{ij}| \leq \log^\beta n$ and A_n has full rank with probability one do not violate the generality of Theorem 1.1.

To start with, we need a very weak lower bound on $|\det A_n|$.

Lemma A.1. *There is a constant C such that*

$$\mathbf{P}(|\det A_n| \leq n^{-Cn}) \leq n^{-1}.$$

Proof. (Proof of Lemma A.1). It follows from [25, Theorem 2.1] that there is a constant C such that $\mathbf{P}(\sigma_n(A_n) \leq n^{-C}) \leq n^{-1}$. Since $|\det A_n|$ is the product of its singular values, the bound follows. □

Remark A.2. The above bound is extremally weak. By modifying the proof in [22], one can actually prove Tao-Vu lower bound (1) for random matrices satisfying **C0**. Also, sharper bounds on the least singular value are obtained in [28, 21]. However, for the arguments in this section, we only need the bound on Lemma A.1.

Let us start with the assumption $|a_{ij}| \leq \log^\beta n$. We can achieve this assumption using the standard truncation method (see [2] or [26]). In what follows, we sketch the idea.

Notice that by condition **C0**, we have, with probability at least $1 - \exp(-\log^{10} n)$, that all entries of A_n have absolute value at most $\log^\beta n$, for some constant $\beta > 0$ which may depend on the constants in **C0**.

We replace the variable a_{ij} by the variable $a'_{ij} := a_{ij}\mathbf{I}_{|a_{ij}|\leq \log^\beta n}$, for all $1 \leq i, j \leq n$ and let A'_n be the random matrix formed by a'_{ij}. Since with probability at least $1 - \exp(-\log^{10} n)$, $A_n = A'_n$, it is easy to show that if A'_n satisfies the claim of Theorem 1.1 then so does A_n.

While the entries of A'_n are bounded by $\log^\beta n$, there is still one problem we need to address, namely that the new variables a'_{ij} do not have mean 0 and variance one. We can achieve this by a simple normalization trick. First observe that by property **C0**, taking β sufficiently large, it is easy to show that $\mu_{ij} = \mathbf{E}a'_{ij}$ has absolute value at most $n^{-\omega(1)}$ and $|1 - \sigma_{ij}| \leq n^{-\omega(1)}$, where σ_{ij} is the standard deviation of a'_{ij}. Now define

$$a''_{ij} := a'_{ij} - \mu_{ij}$$

and

$$a'''_{ij} = \frac{a''_{ij}}{\sigma_{ij}}.$$

Note that a'''_{ij} now does have mean zero and variance one. Let A''_n and A'''_n be the corresponding matrices of a''_{ij} and a'''_{ij} respectively.

By Brun-Minkowski inequality we have,

$$|\det(A'_n)| \leq (|\det A''_n|^{1/n} + |\det N_n|^{1/n})^n,$$

where N_n is the matrix formed by μ_{ij}.

Since $|\mu_{ij}| = n^{-\omega(1)}$, by Hadamard's bound $|\det N_n|^{1/n} \leq n^{-\omega(1)}$. On the other hand, we have by Lemma A.1 that $\mathbf{P}(|\det A''_n|^{1/n} \geq n^{-C}) \geq 1 - n^{-1}$. It thus follows that

$$\mathbf{P}(|\det A'_n| \leq (1 + o(1))|\det A''_n|) \geq 1 - n^{-1}.$$

We can prove a matching lower bound by the same argument. From here, we conclude that if $|\det A''_n|$ satisfies the conclusion of Theorem 1.1, then so does $|\det A'_n|$.

To pass from $\det(A''_n)$ to $\det(A'''_n)$, we apply Brunn-Minkowski inequality again,

$$|\det(A'''_n)| \leq (|\det A''_n|^{1/n} + |\det N'_n|^{1/n})^n,$$

where N'_n is the matrix form by $a''_{ij}(1 - \sigma_{ij}^{-1})$. Note that $|1 - \sigma_{ij}^{-1}| \leq n^{-\omega(1)}$ and $|a''_{ij}| = \log^{O(1)} n$, we infer that $|\det(A''_n)|$ and $|\det(A'''_n)|$ are comparable with high probability

$$\mathbf{P}(|\det A''_n| = (1 + o(1))|\det A'''_n|) \geq 1 - n^{-1}.$$

Now we address the assumption that A_n has full rank with probability one. Notice that this is usually not true when the a_{ij} have discrete distribution (such as Bernoulli). However, we find the following simple trick that makes the assumption valid for our study.

Instead of the entry a_{ij}, consider $a'_{ij} := (1 - \varepsilon^2)^{1/2} a_{ij} + \varepsilon \xi_0$ where ξ_0 is uniform on the interval $[-1, 1]$ and ε is very small, say n^{-1000n}. It is clear that the matrix A'_n formed by the a'_{ij} has full rank with probability one. On the other hand, it is easy to show that by Brunn-Minkowski inequality and Hadamard's bound

$$|\det A_n| = (|\det A'_n|^{1/n} \pm O(n^{-500}))^n.$$

Furthermore, by Lemma A.1, $|\det A_n| \geq n^{-Cn}$ with probability $1 - n^{-1}$, and so we can conclude as in the previous argument.

References

[1] A. C. Berry, *The Accuracy of the Gaussian Approximation to the Sum of Independent Variates*, Transactions of the American Mathematical Society 49 (1941) (1), 122-136.

[2] Z. Bai and J. Silverstein, *Spectral analysis of large dimensional random matrices*, Science press, 2006.

[3] J. Bourgain, V. Vu and P. M. Woods, *On the singularity probability of discrete random matrices*, Journal of Functional Analysis 258 (2010), no.2, 559-603.
[4] A. Dembo, *On random determinants*, Quart. Appl. Math. 47 (1989), no. 2, 185-195.
[5] L. Erdős, *Universality of Wigner random matrices: a survey of recent results*, `arxiv.org/abs/1004.0861`.
[6] G. E. Forsythe and J. W. Tukey, *The extent of n random unit vectors*, Bull. Amer. Math. Sot. 58 (1952), 502.
[7] V. L. Girko, *The central limit theorem for random determinants* (Russian), translation in Theory Probab. Appl. 24 (1979), no. 4, 729-740.
[8] V. L. Girko, *A Refinement of the central limit theorem for random determinants* (Russian), translation in Theory Probab. Appl. 42 (1998), no. 1, 121-129.
[9] V. L. Girko, *Theory of random determinants*, Kluwer Academic Publishers, 1990.
[10] N. R. Goodman, *Distribution of the determinant of a complex Wishart distributed matrix*, Annal of Statistics, 34 (1963), 178-180.
[11] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Electronic Communication in Probability 5 (2000), 119-136.
[12] J. Kahn, J. Komlós and E. Szemerédi, *On the probability that a random* ± 1 *matrix is singular*, J. Amer. Math. Soc. 8 (1995), 223-240.
[13] J. Komlós, *On the determinant of* $(0,1)$ *matrices*, Studia Sci. Math. Hungar. 2 (1967), 7-21.
[14] J. Komlós, *On the determinant of random matrices*, Studia Sci. Math. Hungar. 3 (1968), 387-399.
[15] M. El Machkouri and L. Ouchti, *Exact convergence rate in the central limit theorem for a class of martingale*, Bernoulli 13 (4), 2007, 981-999.
[16] H. Nyquist, S. O. Rice and J. Riordan, *The distribution of random determinants*, Quart. Appl. Math. 12 (1954), 97-104.
[17] A. Prékopa, *On random determinants I*, Studia Sci. Math. Hungar. 2 (1967), 125-132.
[18] G. Rempała and J. Wesołowski, *Asymptotics for products of independent sums with an application to Wishart determinants*, Statist. Probab. Lett. 74 (2005), no. 2, 129-138.
[19] A. Rouault, *Assymptotioc beehavior of random determinants in the Laguerre, Gram and Jacobi ensembles*, Latin American Journal of Probability and Mathematical Statistics (ALEA), 3 (2007) 181-230.
[20] G. Szekered and P. Turán, *On an extremal problem in the theory of determinants* (Hungarian), Math. Naturwiss. Am. Ungar. Akad. Wiss. 56 (1937), 796-806.
[21] M. Rudelson and R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, Advances in Mathematics 218 (2008), 600-633.
[22] T. Tao and V. Vu, *On random* ± 1 *matrices: singularity and determinant*, Random Structures Algorithms 28 (2006), 1-23.
[23] T. Tao and V. Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. 20 (2007), 603-628.
[24] T. Tao and V. Vu, *Random matrices: the distribution of the smallest singualar values*, Geometric and Functional Analysis 20 (2010), no. 1, 260-297.
[25] T. Tao and V. Vu, *Random matrices: The Circular Law,* Communication in Contemporary Mathematics, 10 (2008) 261-307.
[26] T. Tao and V. Vu, *Random matrices: Universality of the local eigenvalue statistics*, Acta Math. 206 (2011), no. 1, 127-204.
[27] T. Tao and V. Vu, *Random matrices: the universality phenomenon for Wigner ensembles*, `arxiv.org/abs/1202.0068`.
[28] T. Tao and V. Vu, Smooth analysis of the condition number and the least singular value. Math. Comput. 79(272): 2333-2352 (2010).
[29] P. Turán, *On a problem in the theory of determinants* (Chinese), Acta Math. Sinica 5 (1955), 41 l-423.

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, USA

E-mail address: hoing@math.upenn.edu

Department of Mathematics, Yale University, New Haven , CT 06520, USA

E-mail address: van.vu@yale.edu