ON THE PERMANENT OF A RANDOM BERNOULLI MATRIX

VAN VU AND TERENCE TAO

Abstract. We show that the permanent of an $n \times n$ matrix with iid Bernoulli entries ± 1 has probability $o(1)$ of vanishing (or of attaining any other fixed value) in the limit $n \to \infty$.

1. Introduction

Let n be a large integer (which we view as a parameter going to infinity), and let $M_n = (a_{ij})_{1 \leq i,j \leq n}$ be a random $n \times n$ Bernoulli matrix, thus the entries a_{ij} are iid random variables which take the values -1 and $+1$ with an equal probability of each. The *determinant*

$$\det(M_n) := \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{i\sigma(i)},$$

and in particular the *singularity probability*

$$\mathbf{P}(\det(M_n) = 0)$$

have been intensively studied [1, 3, 4, 5, 6, 7, 8, 9]. Several bounds for this probability are now known, starting with the initial bound[1]

$$\mathbf{P}(\det(M_n) = 0) = o(1) \tag{1}$$

of Komlós [4]; at this time of writing, the best known bounds are of the form

$$\left(\frac{1}{2} + o(1)\right)^n \leq \mathbf{P}(\det(M_n) = 0) \leq \left(\frac{1}{\sqrt{2}} + o(1)\right)^n$$

with the lower bound being trivial (arising from the probability that two rows or two columns are identical) and the upper bound being a recent result in [1]. The lower bound is conjectured to be sharp.

In this note we study the analogous question for the *permanent*

$$\operatorname{Per}(M_n) := \sum_{\sigma \in S_n} a_{i\sigma(i)} \tag{2}$$

of the random Bernoulli matrix M_n; this is an integer ranging between $-n!$ and $n!$. From the second moment method (as in [8]) one sees that the permanent has mean zero and variance $n!$. Higher moments of $\operatorname{Per}(M_n)$ can be computed (with some difficulty), but they do not appear to reveal any limiting distribution or significant concentration of measure effects[2].

Our main result is an analogue of (1):

V. Vu is supported by an NSF Career Grant.

[1]As usual, we use $o(1)$ to denote any quantity that goes to zero as $n \to \infty$.

[2]For the determinant $\det(M_n)$, one needs to take logarithms before one sees concentration; see [6]. We do not know if a similar result holds for the permanent.

Theorem 1.1. *We have*

$$\sup_x \mathbf{P}(\mathrm{Per}(M_n) = x) = o(1).$$

By optimising all the parameters in the argument, we can in fact replace the $o(1)$ bound by $O(e^{-c\sqrt{\log n}})$ for some absolute constant $c > 0$. This bound is very likely to be improvable, but we will not pursue this matter, other than by mentioning that by analogy with [5] (and in view of the Littlewood-Offord bound (Lemma 2.4) of Erdős[2]), it seems reasonable to hope for a bound of the form $O(1/\sqrt{n})$.

We also remark that our method here extends to the determinant $\det(M_n)$, providing a new (but inefficient) proof of (1).

The argument also extends to other random matrix ensembles with independent complex entries, as long as one has a uniform non-concentration bound $\mathbf{P}(a_{ij} = x) \leq p$ for all i, j and x, and some $p < 1$ independent of n. We leave the details of this generalisation to the interested reader.

2. Proof of theorem

Our argument is based on the cofactor expansion of the permanent in terms of permanents of minors arising from the first few rows of M_n. It is convenient to view this expansion as a "genetic" or "evolutionary" process[3], in which "parent" minors of order $k - 1$ pass on the trait of having non-zero permanent to their "children" minors of order k. To describe this more precisely, we need some notation.

Definition 2.1 (Generations)**.** We write $[n] := \{1, \ldots, n\}$, and for any integer k we write $\binom{[n]}{k} := \{A \subset [n] : |A| = k\}$, where $|A|$ denotes the cardinality of A; we refer to $\binom{[n]}{k}$ as the *population* at generation (or time) k, and the elements of $\binom{[n]}{k}$ as *individuals*. If $A \in \binom{[n]}{k}$, we define the *children* $\partial^+ A := \{B \in \binom{[n]}{k+1} : B \supset A\}$ and *parents* $\partial^- A := \{B \in \binom{[n]}{k-1} : B \subset A\}$ of an individual A, thus an individual at generation k for some $0 \leq k \leq n$ has $n - k$ children and k parents.

Definition 2.2 (Dominance)**.** Fix a $n \times n$ matrix $M_n = (a_{ij})_{1 \leq i,j \leq n}$. For any time $0 \leq k \leq n$ and individual $A = \{j_1, \ldots, j_k\} \in \binom{[n]}{k}$, we define the *minor* $M_{k,A}$ to be the $k \times k$ submatrix $M_{k,A} := (a_{ij_s})_{1 \leq i,s \leq k}$. We say that the individual A has the *dominant trait* if $\mathrm{Per}(M_{k,A}) \neq 0$ and the *recessive trait* otherwise. We define the *dominant population* $\Omega_k \subset \binom{[n]}{k}$ at generation k to be the set of all individuals $A \in \binom{[n]}{k}$ with the dominant trait. We also define the $k \times n$ matrices $M_k := (a_{ij})_{1 \leq i \leq k; 1 \leq j \leq n}$, thus $M_{k,A}$ is a submatrix of M_k. One should view M_k as describing all the information available up to time k.

In our situation, M_n is a random Bernoulli matrix, and so Ω_k is a random subset of $\binom{[n]}{k}$ for each k. We observe the "initial condition" that the entire first generation has the dominant trait:

$$\Omega_1 = \binom{[n]}{1}. \tag{3}$$

The strategy of proof will be to exploit the fact that the trait of having non-zero permanent is indeed *dominant* in the sense that it is likely to pass to a child as long

[3]One could also interpret this process as a sort of percolation process on the Hamming cube, but the author found the evolutionary biology analogy was better at providing a guiding intuition.

as at least one of its parents carries the trait (and *very* likely to pass to a child if *many* parents carry the trait). Our basic tool for this will be[4]

Lemma 2.3 (Dominance)**.** *For any $1 \leq k \leq n$ and $A \in \binom{[n]}{k}$, we have the conditional probability bound*

$$\sup_x \mathbf{P}\left(\operatorname{Per}(M_{k,A}) = x | M_{k-1}\right) \ll |\Omega_{k-1} \cap \partial^- A|^{-1/2} \tag{4}$$

and in particular

$$\mathbf{P}\left(A \in \Omega_k | M_{k-1}\right) \geq 1 - O(|\Omega_{k-1} \cap \partial^- A|)^{-1/2}. \tag{5}$$

(Of course, these bounds are vacuously true when $\Omega_{k-1} \cap \partial^- A$ is empty.)

Proof. From (2) we have the *cofactor expansion*

$$\operatorname{Per}(M_{k,A}) = \sum_{j \in A} a_{kj} \operatorname{Per}(M_{k-1,A\backslash\{j\}}) \tag{6}$$

for all $1 \leq k \leq n$ and $A \in \binom{[n]}{k}$. Observe that the coefficients $\operatorname{Per}(M_{k-1,A\backslash\{j\}})$ depend only on M_{k-1}, while the signs a_{kj} are iid Bernoulli variables that are independent of M_{k-1}. The number of non-zero coefficients in (6) is of course $|\Omega_{k-1} \cap \partial^- A|$. The claim now follows from the following inequality of Erdős [2]:

Lemma 2.4 (Littlewood-Offord bound)**.** [2] *Let $a_1, \ldots, a_k \in \{-1, +1\}$ be iid Bernoulli random variables, let $v_1, \ldots, v_k \in \mathbf{R}$, and let $x \in \mathbf{R}$. If at least m of the $v_1, \ldots, v_k$ are non-zero for some $m \geq 0$, then*

$$\mathbf{P}(a_1 v_1 + \ldots + a_k v_k = x) \ll m^{-1/2}. \tag{7}$$

□

Remark 2.5. For the purposes of establishing the qualitative result in Theorem 1.1, the (sharp) right-hand side $m^{-1/2}$ in (7) could be replaced by any quantity that went to zero as $m \to \infty$.

Applying (4) with $k = n$ and $A = [n]$, we see that Theorem 1.1 will follow if we can show[5]

Proposition 2.6 (Persistence of dominant trait)**.** *The quantity $|\Omega_{n-1}|$ diverges to infinity in probability, or in other words*

$$\lim_{n\to\infty} \mathbf{P}(|\Omega_{n-1}| < N) = 0 \tag{8}$$

for every N.

We now begin the proof of Proposition 2.6. We need a slowly growing function $K = K(n)$ of n to be chosen later (the optimal choice of K turns out to be $K \sim \sqrt{\log n}$).

Our derivation of (8) shall be done in two stages:

- (Early period) We evolve the population for generations $1 \leq k \leq n - K$, ending up with $\gg n/K$ "disjoint lineages" of dominant individuals with high probability.

[4]In this paper we use the usual asymptotic notation $X = O(Y)$, $Y = \Omega(X)$, or $X \ll Y$ to denote the estimate $|X| \leq CY$ for some absolute constant C.

[5]In order to obtain the optimised bound $O(e^{-c\sqrt{\log n}})$, one would take $N = O(e^{\Theta(\sqrt{\log n})})$, and need an upper bound of $O(e^{-\Omega(\sqrt{\log n})})$ for the event $\mathbf{P}(|\Omega_{n-1}| < N)$.

- (Late period) We evolve the population for generations $n-K \leq k \leq n-1$. There will be enough "independent" lineages from the preceding stage that at least N of them will survive to generation $n-1$ with high probability.

Our main tools here shall be the domination inequality (5), together with Azuma's inequality to exploit the independence of the signs a_{kj} that appear in the cofactor expansion (6).

2.7. **The early period.** We begin by showing that most individuals at generation $n-2K$ will have at least one descendant with the dominant trait at generation $n-K$:

Proposition 2.8 (Dominant descendants)**.** *For every $A \in \binom{[n]}{n-2K}$, we have*

$$\mathbf{P}(B \supset A \textit{ for some } B \in \Omega_{n-K}) \geq 1 - O(e^{-\Omega(K)}).$$

Proof. By relabeling we may take $A = \{1, \ldots, n-2K\}$. For every generation $1 \leq k \leq n-K$, we define the random variable

$$r_k := \sup\{|B \cap A| : B \in \Omega_k\}$$

with the convention that $r_k = 0$ if Ω_k is empty; note that r_k ranges between 0 and $\min(k, n-2K)$ and depends only on M_k. Informally, this quantity measures the largest portion of A that the domination trait has been able to "conquer" at generation k. Our task is to show that

$$\mathbf{P}(r_{n-K} < n-2K) \ll e^{-\Omega(K)}. \tag{9}$$

On the other hand, from (3) we have $r_1 = 1$. To get from this to (9), we need

Lemma 2.9 (Spreading lemma)**.** *Let $2 \leq k \leq n-K$, and condition on M_{k-1} (so in particular, r_{k-1} is deterministic).*

(i) *If $r_{k-1} > 0$, then $\mathbf{P}(r_k \geq r_{k-1}) \geq 1 - 2^{-(n-k)}$.*
(ii) *If $0 < r_{k-1} < n-2K$, then $\mathbf{P}(r_k \geq r_{k-1}+1) = 1 - 2^{-(n-2K-r_{k-1})}$.*

Proof. We first establish (i). If $r_{k-1} > 0$, then we may find $B \in \Omega_{k-1}$ such that $|B \cap A| = r_{k-1}$. To establish the claim, it will suffice to show that with probability $1 - 2^{-(n-k)}$, that some child of B has the dominant trait, i.e. that $B \cup \{t\} \in \Omega_k$ for some $t \in [n] \backslash B$.

Suppose temporarily that some child $B \cup \{t\}$ of B had the recessive trait. Observe from (6) (and the dominant nature of B) that if one reversed the sign of a_{kt}, then $B \cup \{t\}$ would then have the dominant trait. Furthermore, this sign reversal does not affect whether any other child of B had the dominant or recessive trait. Since the a_{kt} are independent, and B has $n-k$ children, the claim follows.

The claim (ii) is proven similarly, except that we now restrict attention to the $n-2K-r_{k-1}$ children $B \cup \{t\}$ of B with $t \in [n-2K] \backslash B$. □

From Lemma 2.9(i) and the union bound, we have

$$\mathbf{P}(1 = r_1 \leq r_2 \leq r_3 \leq \ldots \leq r_{n-K} \leq n-2K) \geq 1 - O(e^{-\Omega(K)}). \tag{10}$$

From Lemma 2.9(ii), the union bound, (10), and induction, we have

$$\mathbf{P}(r_k \geq k) \geq 1 - O(e^{-\Omega(K)})$$

for all $1 \leq k \leq n - 2.1K$. We thus have

$$\mathbf{P}(n - 2.1K \leq r_{n-2.1K} \leq \ldots \leq r_{n-K} \leq n-2K) \geq 1 - O(e^{-\Omega(K)}). \tag{11}$$

For each $n - 2.1K < k \leq n - K$, let E_k be the event that $r_{k-1} = r_k < n - 2K$. This event depends on M_k, and by Lemma 2.9 we have the (crude) conditional probability bound $\mathbf{P}(E_k|M_{k-1}) \leq 1/2$ for any choice of M_{k-1}. On the other hand, we have $1.1K$ values of k. By Azuma's inequality, we conclude that with probability $1 - O(e^{-\Omega(K)})$, at most $0.6K$ (say) of the events E_k are true. From this and (10) we conclude (9). □

Corollary 2.10 (Existence of many disjoint lineages). *With probability* $1-O(e^{-\Omega(K)})$, *one can find* $A_1, \ldots, A_m \in \Omega_{n-K}$ *with* $m \gg n/K$, *such that the sets* $[n]\backslash A_1, \ldots, [n]\backslash A_m$ *are all disjoint.*

Proof. Clearly, we may find individuals $\tilde{A}_1, \ldots, \tilde{A}_{\tilde{m}} \in \binom{[n]}{n-2K}$ with $\tilde{m} \gg n/K$ such that $[n]\backslash\tilde{A}_1, \ldots [n]\backslash\tilde{A}_{\tilde{m}}$ are all disjoint. For each $1 \leq j \leq \tilde{m}$, let E_j be the event that $\tilde{A}_j$ has a descendant $\tilde{A}_j \subset A_j \in \Omega_{n-K}$ with the dominant trait. By Proposition 2.8, we know that $\mathbf{P}(E_j) \geq 1 - O(e^{-\Omega(K)})$ for each j. By the first moment method (which does not require any independence hypotheses on the E_j), we conclude that with probability $1 - O(e^{-\Omega(K)})$, that the events E_j are true for at least half of the j. This gives the claim with $m \geq \tilde{m}/2 \gg n/K$. □

2.11. **The late period.** The last step of the argument is to establish

Proposition 2.12 (Survival of many lineages). *Let* $n - K \leq k \leq n - 1$. *If* n *is sufficiently large depending on* K, *then with probability* $1 - O(e^{-\Omega(K)})$, *one can find* $A_1, \ldots, A_m \in \Omega_k$ *with* $m \gg e^{-\Omega(K^2)}n$, *such that the sets* $[n]\backslash A_1, \ldots, [n]\backslash A_m$ *are all disjoint.*

Applying this proposition with $k = n - 1$, we obtain Proposition 2.6 (and thus Theorem 1.1) as desired.

Remark 2.13. This will be the most inefficient part of the argument; in each of the last K generations, we need to sacrifice all but an exponentially small (in K) fraction of the lineages in order to create enough independence to ensure that a significant portion of the remaining lineages survive to the next generation. This is what causes the loss of $e^{-\Omega(K^2)}$ above, and is ultimately responsible for prevending K from exceeding $\sqrt{\log n}$. It is possible that a more sophisticated argument (perhaps using more advanced Littlewood-Offord theorems) could reduce this loss (thus allowing one to improve the upper bound of $O(e^{-c\sqrt{\log n}})$), but we have not pursued this matter.

Proof of Proposition 2.12. We induct on k. The case $n - K$ is just Corollary 2.10. Now suppose that $n - K < k \leq n - 1$ and that the claim has already been proven for $k - 1$. Thus, with probability $1 - O(e^{-\Omega(K)})$, we can find $B_1, \ldots, B_{m'} \in \Omega_{k-1}$ with $m' \gg e^{-\Omega(K^2)}n$ such that $[n]\backslash B_1, \ldots, [n]\backslash B_{m'}$ are all disjoint. These events depend only on M_{k-1}; henceforth we condition M_{k-1} so that these events hold, and then fix $B_1, \ldots, B_{m'}$.

Consider one of the lineages $B_j \in \Omega_{k-1}$. Let us say that B_j is *good* if there exists a child $A \in \partial^+ B_j$ of B_j such that $|\Omega_{k-1} \cap \partial^- A| \geq e^K$, and *bad* otherwise. We divide into two cases.

- **Case 1: At least half of the lineages are good.** For each good lineage B_j, we see from (5) that there will be a child of B_j in Ω_k with probability $1 - O(e^{-K/2})$. From the first moment method (which does not require any independence assumptions), we conclude that with probability $1 - O(e^{-K/2})$, over half of the good lineages will have a child in Ω_k. This gives the claim (with $m \geq m'/4$); note that the probability error of $O(e^{-K/2})$ incurred here (as well as the loss of 4 in the number of lineages) will be acceptable as we are only performing the induction K times.

- **Case 2: At least half of the lineages are bad.** For each bad lineage B_j, we can find a child $A_j = B_j \cup \{t_j\} \in \partial^+ B_j$ with $|\Omega_{k-1} \cap \partial^- A_j| < e^K$. In other words, the set $R_j := \{t \in A_j : A_j \backslash \{t\} \in \Omega_{k-1}\}$ has cardinality at most e^K (and also contains t_j). Using the greedy algorithm, we can thus find a collection $B_{j_1}, \ldots, B_{j_{\tilde{m}}}$ of bad lineages with $\tilde{m} \gg e^{-K} m'$, such that

$$t_{j_s} \not\in R_{j_r} \text{ for any } 1 \leq r < s \leq \tilde{m}. \tag{12}$$

For each $1 \leq s \leq \tilde{m}$, let F_s be the event that A_{j_s} carries the dominant trait. As we have conditioned on M_{k-1}, this event depends only on the values of the random signs a_{kt} for $t \in R_{j_s}$. In particular, it depends on the sign a_{kt_s}. Furthermore, as B_{j_s} carries the dominant trait, the coefficient of a_{kt_s} in (6) is non-zero. This implies that if F_s fails (so A_{j_s} has the recessive trait), then reversing the sign of a_{kt_s} will make F_s true. Furthermore, by (12), this sign reversal does not affect the truth of $F_1, \ldots, F_{s-1}$. We thus conclude the conditional probability bound

$$\mathbf{P}(F_s | \mathcal{F}(F_1, \ldots, F_{s-1})) \geq 1/2,$$

where $\mathcal{F}(F_1, \ldots, F_{s-1})$ is the σ-algebra generated by the events $F_1, \ldots, F_{s-1}$. From Azuma's inequality, we conclude that with probability $1 - O(e^{-O(\tilde{m})}) \geq 1 - e^{-K}$ (if n is sufficiently large depending on K), the event F_s will hold for at least $m := \tilde{m}/4 \gg e^{-K} m'$ (say) values of s. The individuals A_{j_s} associated to these values of s then obey the required properties to close the induction. (Once again, the losses of e^{-K} are acceptable as we are only performing the induction K times; this is what causes the final loss of $e^{-\Omega(K^2)}$ in the m parameter.)

The proof of Proposition 2.12 (and thus Theorem 1.1) is complete. □

Remark 2.14. Note the dependence of m on n is linear for fixed K. Because of this, the above argument actually yields a slight strengthening of (8), namely that $\mathbf{P}(|\Omega_{n-1}| < o(n)) = o(1)$ for any choice of function $o(n)$ that grows slower than n. However, this strengthening does not appear to improve the quantitative bounds for Theorem 1.1, as the $o(1)$ exceptional probability overwhelms the $1/\sqrt{n}$-type terms arising from (4).

References

[1] J. Bourgain, V. Vu, P. Wood, *On the singularity probability of discretely random complex matrices*, preprint.

[2] P. Erdős, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. **51** (1945), 898–902.

[3] J. Kahn, J. Komlós, E. Szemerédi, *On the probability that a random* ± 1 *matrix is singular*, J. Amer. Math. Soc. **8** (1995), 223–240.

[4] J. Komlós, *On the determinant of* $(0, 1)$ *matrices*, Studia Sci. Math. Hungar. **2** (1967) 7-22.

[5] J. Komlós, *On the determinant of random matrices*, Studia Sci. Math. Hungar. **3** (1968) 387–399.

[6] T. Tao and V. Vu, *On random* ± 1 *matrices: Singularity and Determinant*, Random Structures Algorithms **28** (2006), no. 1, 1–23.

[7] T. Tao and V. Vu, *On the singularity probability of random Bernoulli matrices*, to appear in J. Amer. Math. Soc..

[8] P. Turán, On extremal problems concerning determinants, (Hungarian) *Math. Naturwiss. Anz. Ungar. Akad. Wiss.* 59, (1940). 95–105

[9] T. Voight, G. Ziegler, *Singular 0/1 matrices and the hyperplanes spanned by random 0/1 vectors*, preprint (arXiv:math.CO/0308050).

Department of Mathematics, Rutgers University, Piscataway NJ 08854-8019
E-mail address: `vanvu@@math.rutgers.edu`

Department of Mathematics, UCLA, Los Angeles CA 90095-1555
E-mail address: `tao@@math.ucla.edu`