AN ANALOGUE OF THE ERDŐS–KAC THEOREM FOR THE SPECIAL LINEAR GROUP OVER THE INTEGERS

DANIEL EL-BAZ

ABSTRACT. We investigate the number of prime factors of individual entries for matrices in the special linear group over the integers. We show that, when properly normalised, it satisfies a central limit theorem of Erdős–Kac-type. To do so, we employ a sieve-theoretic set-up due to Granville and Soundararajan. We also make use of an estimate coming from homogeneous dynamics due to Gorodnik and Nevo.

1. INTRODUCTION

The celebrated Erdős–Kac theorem is a central limit theorem for the number of (distinct) prime factors $\omega$ of a “random” integer, in the following sense: for every $x \in \mathbb{R}$, we have

$$\lim_{n \to +\infty} \frac{1}{n} \# \{ 1 \leq m \leq n : \frac{\omega(m) - \log \log n}{\sqrt{\log \log n}} \leq x \} = \frac{1}{\sqrt{2\pi}} \int \limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

There is an abundant number of results of this type for the number of prime factors of various sequences of integers: shifted primes [8], values of integer polynomials [9] and friable numbers [10] [1] [11] to cite but a few examples.

In this note, we study the number of prime factors of the entries in integer matrices of unit determinant. More precisely, we define, for $n \geq 2$ and $T > 0$,

$$(1) \quad V_T(\mathbb{Z}) = \{ g \in \text{SL}_n(\mathbb{Z}) : \|g\| \leq T \},$$

where for $g \in \text{SL}_n(\mathbb{Z})$, $\|g\| = \sqrt{\text{Tr}(g^T g)}$ is the Frobenius norm.

For an integer $n \geq 1$, we let $\omega(n)$ be the number of distinct prime factors of $n$. We extend $\omega$ to $\mathbb{Z}$ by defining $\omega(0) = 0$ and for $n \geq 1$, $\omega(-n) = \omega(n)$. Our main theorem can now be stated.

**Theorem 1.1.** For every $n \geq 2$, every $x \in \mathbb{R}$ and for each pair $(i, j) \in \{1, \ldots, n\}^2$, we have

$$(2) \quad \lim_{T \to +\infty} \frac{1}{\# V_T(\mathbb{Z})} \# \left\{ g \in V_T(\mathbb{Z}) : \frac{\omega(g_{i,j}) - \log \log T}{\sqrt{\log \log T}} \leq x \right\} = \frac{1}{\sqrt{2\pi}} \int \limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Our proof uses the sieve-theoretic framework unravelled by Granville and Soundararajan in 2006 [7], which we recall in subsection 3.1 so as to be self-contained. To obtain the necessary estimates to feed into the sieve, we apply a deep result of Gorodnik and Nevo from 2010 [6].

We note that since this effective congruence counting is the key input, one can also apply such arguments to thin linear groups, thanks to the work of Bourgain, Gamburd and Sarnak [3]. Indeed, this was done in the case of Apollonian circle packings with integral curvatures by Djanković in [4].

Acknowledgements: I would like to thank Efthymios Sofos for asking the question answered in this paper. I am grateful to Zeev Rudnick for several helpful conversations and encouraging me to write up this short note. I am also grateful to Carlo Pagano for the comments he made on a previous version of this draft that led to an improved presentation. This research was supported by the European Research Council under...
the European Union’s Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 320755.

2. Counting with congruences

We need the following special case of \([\text{6, Corollary 5.2}].\)

For \(n \geq 2\) and a positive integer \(Q\), define the following — principal congruence — subgroup of \(\text{SL}_n(\mathbb{Z})\):

\[
\Gamma(Q) = \{ g \in \text{SL}_n(\mathbb{Z}) : g \equiv I_n \pmod{Q} \}.
\]

**Claim 2.1** (Gorodnik–Nevo). For every \(n \geq 2\), there exists \(\delta > 0\) such that for every \(M \in \text{SL}_n(\mathbb{Z})\) and every \(Q \geq 1\),

\[
\#\{A \in \Gamma(Q)M : \|A\| \leq T\} = \frac{\#V_T(\mathbb{Z})}{[\text{SL}_n(\mathbb{Z}) : \Gamma(Q)]} + O(\#V_T(\mathbb{Z})^{1-\delta}),
\]

where the implied constant is independent of \(Q\).

In fact, we shall require the following consequence.

**Corollary 2.1.** For every \(n \geq 2\), there exists \(\delta > 0\) such that for every square-free integer \(q\) and every \((i,j) \in \{1, \ldots, n\}^2\),

\[
\# \{g \in V_T(\mathbb{Z}) : q \mid g_{i,j}\} = \prod_{p \in \mathbb{P}}\frac{p^{n-1}-1}{p^n-1}\#V_T(\mathbb{Z}) + O\left(q^{n^2-2}\#V_T(\mathbb{Z})^{1-\delta}\right)
\]

**Proof.** Fix \(n \geq 2\) and \((i,j) \in \{1, \ldots, n\}^2\).

The group \(\Gamma(q)\) acts by left multiplication on the set \(\{g \in \text{SL}_n(\mathbb{Z}) : q \mid g_{i,j}\}\), which we may therefore view as a disjoint union of finitely (a priori, at most \(q^{n^2-1}\)) \(\Gamma(q)\)-orbits. We restrict the orbits to elements having norm at most \(T\) and note that each orbit has the same number of points, given by the formula in **Claim 2.1**

\[
\frac{\#V_T(\mathbb{Z})}{[\text{SL}_n(\mathbb{Z}) : \Gamma(q)]} + O(\#V_T(\mathbb{Z})^{1-\delta}).
\]

First, \(\text{SL}_n(\mathbb{Z})/\Gamma(q) \cong \text{SL}_n(\mathbb{Z}/q\mathbb{Z})\) — a manifestation of “strong approximation” — so that

\[
[\text{SL}_n(\mathbb{Z}) : \Gamma(q)] = \#\text{SL}_n(\mathbb{Z}/q\mathbb{Z})
\]

\[
= \prod_{p \in \mathbb{P}, p \mid q} \#\text{SL}_n(\mathbb{Z}/p\mathbb{Z})
\]

\[
= \prod_{p \in \mathbb{P}, p \mid q} \frac{(p^n - 1)(p^n - p)(p^n - p^2)\cdots(p^n - p^{n-1})}{p - 1}
\]

where the second equality follows from the Chinese remainder theorem.

Next, the number of orbits is precisely

\[
\# \{g \in \text{SL}_n(\mathbb{Z}/q\mathbb{Z}) : g_{i,j} = 0\}.
\]

By the Chinese remainder theorem we simply need to compute, for \(p\) prime,

\[
\# \{g \in \text{SL}_n(\mathbb{F}_p) : g_{i,j} = 0\}.
\]

We estimate the number of such matrices in \(\text{GL}_n(\mathbb{F}_p)\) and then divide by \(\#\mathbb{F}_p^\times = p - 1\) to get the desired count. We have \(p^{n-1} - 1\) choices for the \(j\)th column, then \(p^n - p\) choices for another column to be linearly independent from that column, \(p^n - p^2\) choices for a third column to be linearly independent from the span of those two columns, etc.

Therefore

\[
\# \{g \in \text{SL}_n(\mathbb{F}_p) : g_{i,j} = 0\} = \frac{(p^{n-1} - 1)(p^n - p)(p^n - p^2)\cdots(p^n - p^{n-1})}{p - 1}
\]
and
\[(13) \quad \#\{g \in \text{SL}_n(\mathbb{Z}/q\mathbb{Z}) : g_{i,j} = 0\} = \prod_{p \in \mathcal{P}} \frac{(p^{n-1} - 1)(p^n - p)(p^n - p^2) \cdots (p^n - p^{n-1})}{p - 1}.
\]

It thus follows from (9) and (13) that
\[(14) \quad \#\{g \in V_T(\mathbb{Z}) : q \mid g_{i,j}\} = \frac{\#\{g \in \text{SL}_n(\mathbb{Z}/q\mathbb{Z}) : g_{i,j} = 0\}}{[\text{SL}_n(\mathbb{Z}) : \Gamma(q)]} \#V_T(\mathbb{Z}) + O\left(\prod_{p \in \mathcal{P}} p^{n^2 - 2} \#V_T(\mathbb{Z})^{1 - \delta}\right),
\]

as claimed. \(\square\)

3. Sieving

3.1. The Granville–Soundararajan framework. For a multiset \(A = \{a_1, \ldots, a_x\}\) and a positive integer \(d\), define
\[(16) \quad A_d = \#\{n \leq x : d \mid a_n\}.
\]
Suppose that, for square-free \(d\), \(A_d\) can be written in the form \(\frac{h(d)}{d} x + r_d\) for some non-negative multiplicative function \(h\) with \(0 \leq h(d) \leq d\) for all square-free \(d\). For a set of primes \(\mathcal{P}\) and \(a \in A\), define \(\omega_{\mathcal{P}}(a) = \#\{p \in \mathcal{P} : p \mid a\}\). Define \(\mu_{\mathcal{P}} = \sum_{p \in \mathcal{P}} \frac{\frac{h(p)}{p}}{\sigma_p}\) and \(\sigma_p^2 = \sum_{p \in \mathcal{P}} \frac{\frac{h(p)}{p}}{\sigma_p}\). Define \(D_k(\mathcal{P})\) to be the set of square-free integers which are the products of at most \(k\) elements of \(\mathcal{P}\). Finally, define \(C_k = \frac{\Gamma(k + 1)}{2^k \Gamma\left(\frac{k}{2} + 1\right)}\).

The following is [7] Proposition 3.

Claim 3.1 (Granville–Soundararajan). Uniformly for all natural numbers \(k \leq \sigma_p^{2/3}\), we have
\[(17) \quad \sum_{a \in A} (\omega_{\mathcal{P}}(a) - \mu_{\mathcal{P}})^k = C_k x \sigma_p^k \left(1 + O\left(\frac{k^3}{\sigma_p^2}\right)\right) + O\left(\mu_{\mathcal{P}}^k \sum_{d \in D_k(\mathcal{P})} |r_d|\right)
\]
if \(k\) is even and
\[(18) \quad \sum_{a \in A} (\omega_{\mathcal{P}}(a) - \mu_{\mathcal{P}})^k \ll C_k x \sigma_p^k \frac{k^{3/2}}{\sigma_p} + \mu_{\mathcal{P}}^k \sum_{d \in D_k(\mathcal{P})} |r_d|
\]
if \(k\) is odd.

3.2. Proof of the main theorem. In this section, we prove Theorem 1.1.

Fix \(n \geq 2\) and \((i,j) \in \{1, \ldots, n\}^2\).

Adopting the notation in [2], we define
\[(19) \quad A = \{g_{i,j} : g \in V_T(\mathbb{Z})\},
\]
a multiset of size \(x := \#V_T(\mathbb{Z})\).

Note that
\[(20) \quad x \sim c_n T^{n^2 - n}
\]
by a special case of equation (1.12) in [3] Example 1.6, with the explicit formula for \(c_n\) given by \(c_n = \frac{n^{n^2/2}}{\Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{n^2 - n + 2}{2}\right) \zeta(2) \cdots \zeta(n)}\).
For our application, we take
\begin{equation}
P = \{\text{all primes } \leq T^{\varepsilon(T)}\}
\end{equation}
where \(\varepsilon\) is a real function which tends to 0 at infinity that we shall choose later. More precisely, we shall choose \(\varepsilon(T)\) to be of the form
\[\varepsilon(T) := \frac{1}{(\log \log T)^\psi},\]
where \(\psi\) is a fixed positive constant to be chosen later.

Finally, we define the multiplicative function \(h: \mathbb{Z}_{\geq 1} \rightarrow \mathbb{R}_{\geq 0}\) by
\begin{equation}
h(q) = \mu^2(q)q \prod_{p | q} \frac{p^{n-1} - 1}{p^n - 1}
\end{equation}
In particular, for \(p\) prime,
\begin{equation}
h(p) = \frac{p^{n-1} - 1}{p^n - 1} = \frac{1}{p} + O\left(\frac{1}{p^n}\right).
\end{equation}

Therefore, using Mertens’ theorem we have
\begin{equation}
\mu_P = \log \log T + O(\log \log \log T)
\end{equation}
and likewise
\begin{equation}
\sigma^2_P = \log \log T + O(\log \log \log T).
\end{equation}

Now [Corollary 2.1] can be restated as follows: there exists a positive \(\delta\) such that for every square-free positive integer \(q\),
\begin{equation}
A_q = \frac{h(q)}{q}x + r_q
\end{equation}
where
\begin{equation}
r_q = O(q^{2-2x}1^{-\delta}).
\end{equation}

To help simplify the presentation, we introduce some probabilistic language.

For each \(T > 0\), we equip the finite set \(V_T(\mathbb{Z})\) with the uniform probability measure. For the given pair \((i, j)\) \(\in \{1, \ldots, n\}^2\), we define the random variable \(\omega_{i, j}\) on \(V_T(\mathbb{Z})\) by \(\omega_{i, j}: g \mapsto \omega(g_{i, j})\). We can thus restate Theorem 1.1 as the convergence in distribution, as \(T \to +\infty\), of the random variables \(\omega_{i, j}(g) - \log \log T\sqrt{\log \log T}\) to the standard Gaussian \(N(0, 1)\).

The estimates for the moments in Claim 3.1 apply to \(\omega_P\) which is a truncated version of the random variable we are interested in; still, choosing \(\varepsilon\) carefully allows us to extract all the information we need, as we make clear in the following lemma.

Lemma 3.1. As \(T \to +\infty\), the random variables \(\frac{\omega_{i, j} - \mu_P}{\sigma_P}\) converge in distribution to \(N(0, 1)\) if and only if the random variables \(\frac{\omega_P - \mu_P}{\sigma_P}\) do.

Proof. We have
\[c \in \mathbb{Z} \setminus \{0\}, z > 1 \Rightarrow \#\{p \mid c : p > z\} \leq \frac{\log |c|}{\log z},\]
where the primes are counted with multiplicity. Therefore if \(0 < |c| \leq T\) and choosing \(z = T^{\varepsilon(T)}\), then
\[\omega(c) - \omega_P(c) \leq \frac{\log T}{\log T^{\varepsilon(T)}} \leq \frac{1}{\varepsilon(T)}.
\]
Observe that if we can choose the function \(\varepsilon\) such that
\begin{equation}
\frac{1}{\varepsilon(T)} = o\left(\sqrt{\log \log T}\right),
\end{equation}
then by the first sentence of [2] Remark 1, we are done. We can simply pick any \(\psi \in (0, \frac{1}{2})\) to have our truncation function \(\varepsilon\) tend to 0 at infinity and satisfy (28), which completes the proof. \(\square\)
Since we are interested in proving a central limit theorem, we shall not worry about the uniformity in $k$ — it is, in principle, possible to keep track of that and obtain good estimates for all moments.

In particular, Claim 3.1 reads, as $T$ (or equivalently $x$) goes to infinity: for every even $k$,

$$
\sum_{a \in A} \left( \frac{\omega_p(a) - \mu_p}{\sigma_p} \right)^k = C_k x \left( 1 + O \left( \frac{1}{\log \log x} \right) \right) + O \left( \frac{\log x}{\sigma_p^k} \right) \sum_{q \in \mathcal{D}_k(p)} |r_q| \quad \text{and for every odd } k,
$$

$$
\sum_{a \in A} \left( \frac{\omega_p(a) - \mu_p}{\sigma_p} \right)^k \ll \frac{x}{\sigma_p} + \frac{\mu_p}{\sigma_p^k} \sum_{q \in \mathcal{D}_k(p)} |r_q|.
$$

To conclude using that result, we first need to show that the error term

$$
\mathcal{R}_k(T) := \sum_{q \in \mathcal{D}_k(p)} |r_q|
$$

is sufficiently small. To do so, we first observe that every $q$ in $\mathcal{D}_k(p)$ satisfies

$$
q \leq (\max\{p : p \in P\})^k \leq T^{k\varepsilon(T)}.
$$

Using this with (29) we obtain

$$
\mathcal{R}_k(T) \ll x^{1-\delta} \sum_{q \leq T^{k\varepsilon(T)}} q^{\sigma_p^2 - 2} \leq T^{(n^2 - 1)k\varepsilon(T)} x^{1-\delta} \ll x^{1-\delta + \frac{\log x}{\log \log x}} = \frac{1}{\gamma_n x^{1/(n^2 - n)}},
$$

where in the last inequality we recalled (20) and let $\gamma_n = \left( \frac{1}{c_n} \right)^{n^2 - n}$.

Since $\varepsilon(x) = o(1)$ we obtain that

$$
\mathcal{R}_k(T) \ll x^{1-\frac{1}{2}}.
$$

Injecting the above estimate into (29) and (30) respectively, we obtain that we have — recalling the main terms of (24) and (25) and noting that as $T \to +\infty$, $\log \log T \sim \log \log x$ due to (20) — for every even $k$,

$$
\frac{1}{x} \sum_{a \in A} \left( \frac{\omega_p(a) - \mu_p}{\sigma_p} \right)^k = C_k \left( 1 + O \left( \frac{1}{\log \log x} \right) \right) + O \left( (\log \log x)^{k/2} x^{-\frac{1}{2}} \right)
$$

and for every odd $k$,

$$
\frac{1}{x} \sum_{a \in A} \left( \frac{\omega_p(a) - \mu_p}{\sigma_p} \right)^k \ll \frac{1}{\sqrt{\log \log x}} + (\log \log x)^{k/2} x^{-\frac{1}{2}}
$$

where the implied constants depend on $k$ and we recall that $C_k$ is the $k$th moment of the standard normal distribution. Equivalently, as $T \to +\infty$, we have that for every even $k$, (31)

$$
\frac{1}{\#V(T)} \sum_{a \in A} \left( \frac{\omega_p(a) - \mu_p}{\sigma_p} \right)^k \to C_k
$$

and for every odd $k$, (32)

$$
\frac{1}{\#V(T)} \sum_{a \in A} \left( \frac{\omega_p(a) - \mu_p}{\sigma_p} \right)^k \to 0,
$$

which shows (since the normal distribution is characterised by its moments) that as $T \to +\infty$, the random variables $\frac{\omega_p(a) - \mu_p}{\sigma_p}$ converge in distribution to $\mathcal{N}(0, 1)$.

Thanks to Lemma 3.1 this means that the random variables $\frac{\omega_p - \mu_p}{\sigma_p}$ converge in distribution to $\mathcal{N}(0, 1)$. This is almost what we want, with the following lemma making the final connection and concluding the proof of Theorem 1.1.
Lemma 3.2. As \( T \to +\infty \), the random variables \( \frac{\omega_{i,j} - \mu_p}{\sigma_p} \) converge in distribution to \( N(0,1) \) if and only if the random variables \( \frac{\omega_{i,j} - \log \log T}{\sigma_p \sqrt{\log \log T}} \) do.

Proof. We write

\[
\frac{\omega_{i,j} - \log \log T}{\sqrt{\log \log T}} = \frac{\sigma_p}{\sqrt{\log \log T}} \left( \frac{\omega_{i,j} - \mu_p}{\sigma_p} + \frac{\mu_p - \log \log T}{\sigma_p} \right)
\]

and note that, as \( T \to +\infty \), \( \frac{\sigma_p}{\sqrt{\log \log T}} \to 1 \), while \( \frac{\mu_p - \log \log T}{\sigma_p} \to 0 \), thanks to (25) and (24). The claim now follows from the third sentence in [2, Remark 1]. \( \square \)

References

[1] K. Alladi. An Erdős–Kac theorem for integers without large prime factors. *Acta Arith.*, 49(1):81–105, 1987.
[2] P. Billingsley. On the central limit theorem for the prime divisor functions. *Amer. Math. Monthly*, 76:132–139, 1969.
[3] J. Bourgain, A. Gamburd, and P. Sarnak. Affine linear sieve, expanders, and sum-product. *Invent. Math.*, 179(3):559–644, 2010.
[4] G. Djanković. The Erdős-Kac theorem for curvatures in integral Apollonian circle packings. *Publ. Inst. Math. (Beograd) (N.S.*), 89(103):11–17, 2011.
[5] W. Duke, Z. Rudnick, and P. Sarnak. Density of integer points on affine homogeneous varieties. *Duke Math. J.*, 71(1):143–179, 1993.
[6] A. Gorodnik and A. Nevo. Quantitative ergodic theorems and their number-theoretic applications. *Bull. Amer. Math. Soc. (N.S.*), 52(1):65–113, 2015.
[7] A. Granville and K. Soundararajan. Sieving and the Erdős-Kac theorem. In *Equidistribution in number theory, an introduction*, volume 237 of *NATO Sci. Ser. II Math. Phys. Chem.*, pages 15–27. Springer, Dordrecht, 2007.
[8] H. Halberstam. On the distribution of additive number-theoretic functions. *Journal of the London Mathematical Society*, s1-30(1):43–53, 1955.
[9] H. Halberstam. On the distribution of additive number-theoretic functions (ii). *Journal of the London Mathematical Society*, s1-31(1):1–14, 1956.
[10] D. Hensley. The number of positive integers \( \leq x \) and free of prime factors \( > y \). *Journal of Number Theory*, 21(3):286 – 298, 1985.
[11] A. Hildebrand. On the number of prime factors of integers without large prime divisors. *Journal of Number Theory*, 25(1):81 – 106, 1987.

School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel

Max Planck Institute for Mathematics, Vivatsgasse 7, 53111 Bonn, Germany

E-mail address: danielelbaz88@gmail.com