Glasgow
Mathematical
Journal

RESEARCH ARTICLE

Gaps between prime divisors and analogues in Diophantine geometry

Efthymios Sofos

Department of Mathematics, University of Glasgow, Glasgow G12 8QQ, UK
E-mail: efthymios.sofos@gmail.com

Abstract
Erdős considered the second moment of the gap-counting function of prime divisors in 1946 and proved an upper bound that is not of the right order of magnitude. We prove asymptotics for all moments. Furthermore, we prove a generalisation stating that the gaps between primes p for which there is no $\mathbb{Q}_p$-point on a random variety are Poisson distributed.

1. Introduction

What are the typical gaps between the prime divisors of a randomly selected integer? For $m \in \mathbb{N}$, we let $\omega(m)$ be the number of distinct prime divisors of m and $p_i(m)$ be the i-th smallest prime divisor of m, so that

$$\log \log p_1(m) < \ldots < \log \log p_{\omega(m)}(m)$$

is a finite sequence that depends on m. It is not difficult to show that for almost all m and almost all $1 \leqslant i \leqslant \omega(m)$, one has $\log \log p_i(m) \sim i$; hence, $\log \log p_{i+1}(m) - \log \log p_i(m)$ is typically bounded. A natural question is to count the number of gaps exceeding a fixed constant $z \geqslant 0$, i.e. estimate

$$\omega_z(m) := \sharp \left\{ 1 \leqslant i < \omega(m) \colon \log \log p_{i+1}(m) - \log \log p_i(m) > z \right\}.$$

Erdős [6, p. 534] was the first to study this question. He showed that for almost all m, the function $\omega_z(m)$ is well-approximated by $\mathrm{e}^{-z}\omega(m)$ by proving an upper bound for the second moment:

$$\frac{1}{n} \sum_{m \in \mathbb{N} \cap [1,n]} \left(\omega_z(m) - \mathrm{e}^{-z} \log \log n \right)^2 = o((\log \log n)^{3/2}), \text{ as } n \to +\infty.$$

However, it turns out that this is not of the right order of magnitude. Here, we prove asymptotics not just for the second moment, but for all moments:

Theorem 1.1. *Fix any $z \geqslant 0$ and $r \geqslant 0$. Then*

$$\frac{1}{n} \sum_{m \in \mathbb{N} \cap [1,n]} \left(\omega_z(m) - \frac{\log \log n}{\mathrm{e}^z} \right)^r = \mu_r ((1 - 2z\mathrm{e}^{-z})\mathrm{e}^{-z} \log \log n)^{r/2}(1 + o(1)), \text{ as } n \to +\infty,$$

where μ_r is the r-th moment of the standard normal distribution.

As a consequence, for all $\alpha < \beta \in \mathbb{R}$ one has

$$\lim_{n \to +\infty} \frac{1}{n} \sharp \left\{ m \in \mathbb{N} \cap [1,n] \colon \frac{\omega_z(m) - \mathrm{e}^{-z} \log \log m}{((1 - 2z\mathrm{e}^{-z})\mathrm{e}^{-z} \log \log m)^{1/2}} \in (\alpha, \beta] \right\} = \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta \mathrm{e}^{-t^2/2} \mathrm{d}t.$$

Setting $z = 0$, we recover the much celebrated Erdős–Kac theorem [7]. Our method is different from that of Erdős [6] in that it relies on Stein's method on normal approximations [18]. This allows us to deal with certain sums of dependent random variables that arise when modelling $\omega_z(m)$. Stein's method has been rarely used in number theory, for example, by Harper [11].

There are many generalisations of the Erdős–Kac theorem to functions of the form $\sum_{p\mid m} g(p)$ but they do not cover $\omega_z(m)$, as $g(p)$ would have to be a function of m as well. Galambos [9, Theorem 2] studied the values of a function that is somewhat related to our ω_z, namely the cardinality of $i < \omega(m)$ for which $\log\log p_{i+1}(m) - \log\log p_i(m) > z + \log\log\log m$. His results and method are rather different as they are suited to values of large gaps, while our result relates to small gaps. A function similar to Galambos' occurs in the recent work of Chan–Koymans–Milovic–Pagano [4, Section 4] on the negative Pell equation.

Remark 1.2. *At the cost of a non-self sufficient argument, the number theoretic part of the proof of Theorem* 1.1 *(namely, Lemma* 2.9*) can be alternatively verified via the Kubilius model* [5, *Section* 12]. *The approximation of* $\omega(m)$ *by* $e^{-z}\log\log m$ *means that the gaps in the sequence* $\{\log\log p_i(m)\}_{i\geqslant 1}$ *are Poissonian. It is worth mentioning that the occurrence of Poisson distribution in other areas of Probabilistic Number Theory is not uncommon, see the work of de Koninck–Galambos* [3], *Harper* [11], *Granville* [10] *and Kowalski–Nikeghbali* [12], *for example.*

Remark 1.3 (Further developments). *The interested reader may wonder whether one can use tools from analysis to make explicit the term* $o(1)$ *in Theorem* 1.1. *In the case of the Erdős–Kac theorem, this was done by Rényi and Turán* [15] *using complex analysis. After seeing the first version of this paper on arXiv, R. de la Bretèche and G. Tenenbaum proved an explicit error term using methods quite different from ours (namely, Fourier analysis); see their preprint* [4] *for details.*

1.1. Generalisations in Diophantine geometry

In Section 3, we provide a generalisation of Theorem 1.1, given by Theorem 3.2. In brief terms, it states that the gaps between primes p for which a typical variety over $\mathbb{Q}$ has no $\mathbb{Q}_p$-points obey the Poisson distribution. A statement analogous to the Erdős–Kac theorem was proved by Loughran–Sofos [14] by using geometric input from the work of Loughran–Smeets [13].

2. The proof of Theorem 1.1

2.1. Defining the model

The letter z will denote a fixed non-negative real number throughout Section 2. As usual, we denote $\exp(z) := e^z$. For a prime p and a positive integer m, we define

$$\delta_{p,z}(m) := \begin{cases} 1, & \text{if } p \mid m \text{ and } m \text{ is not divisible by any prime in } (p, p^{\exp(z)}], \\ 0, & \text{otherwise.} \end{cases}$$

In particular, $\omega_z(m) = \sum_p \delta_{p,z}(m)$, where the sum is over all primes. Our plan, initially, is to follow the Kubilius model idea (see Billingsley [1, equations (1.8),(1.9)]) to define Bernoulli random variables B_p that model the behaviour of $\delta_{p,z}$. For this, we use the random variables X_p as follows: for every prime p the random variable X_p is defined so that

$$P[X_p = 1] = \frac{1}{p}, \quad P[X_p = 0] = 1 - \frac{1}{p}$$

and such that X_p are independent. In particular, the mean $E[X_p]$ equals $1/p$, thus, $X_p = 1$ models the event that a random integer m is divisible by a fixed prime p. Let $[\cdot]$ denote the integer part. The independence of X_p is related to the Chinese Remainder Theorem.

To model $\delta_{p,z}$, we must also take into account the fact that each prime q in the range $(p, p^{\exp(z)}]$ must not divide m. Thus, we are naturally led to define

$$B_p := X_p \prod_{\substack{q \text{ prime} \\ p < q \leqslant p^{\exp(z)}}} (1 - X_q). \tag{2.1}$$

We will later prove that $\sum_p B_p$ is a good model for $\omega_z = \sum_p \delta_{p,z}$ in the sense that their moments agree asymptotically.

Remark 2.1 (Independence break-down). *Definition* (2.1) *leads to a major difference between this paper and the proofs of the Erdős–Kac theorem, namely, the variables B_p are dependent. Indeed, for all primes $p < q$ with $q \leqslant p^{\exp(z)}$, the quantity $E[B_p B_q]$ vanishes while none of $E[B_p]$, $E[B_q]$ does.*

2.2. Distribution and moments of the model via Stein's method

For any positive N, we define

$$S_N = \sum_{p \leqslant N} B_p$$

and denote its expectation and variance, respectively, by

$$c_N := E[S_N] \text{ and } s_N^2 := \text{Var}[S_N].$$

Our goal in this section is to prove that $(S_N - c_N)/s_N$ converges in law to the standard normal distribution as $N \to \infty$ and that its moments are asymptotically Gaussian. This will be done, respectively, in Propositions 2.5 and 2.7. We first need a few preparatory estimates.

Lemma 2.2. *We have*

$$E[B_p] = \frac{e^{-z}}{p} + O\left(\frac{1}{p \log p}\right), \tag{2.2}$$

$$c_N = e^{-z} \log \log N + O(1) \tag{2.3}$$

and

$$s_N^2 = \left(1 - \frac{2z}{e^z}\right) \frac{\log \log N}{e^z} + O(1). \tag{2.4}$$

Proof. Recall that Mertens' theorem states that $\sum_{p \leqslant T} 1/p = \log \log T + c + O(1/\log T)$ for some constant c. The independence of X_p yields

$$E[B_p] = \frac{1}{p} \prod_{p < q \leqslant p^{\exp(z)}} \left(1 - \frac{1}{q}\right),$$

which, by the approximation $1 - \varepsilon = \exp(-\varepsilon + O(\varepsilon^2))$ for $|\varepsilon| \leqslant 1$ and Mertens' theorem is

$$\frac{1}{p} \exp\left(-\sum_{p < q \leqslant p^{\exp(z)}} \frac{1}{p} + O\left(\sum_{p < q \leqslant p^{\exp(z)}} \frac{1}{p^2}\right)\right) = \frac{\exp(-\log \log p^{\exp(z)} + \log \log p + O(1/\log p))}{p}.$$

Since $\exp(O(1/\log p)) = 1 + O(1/\log p)$, this is sufficient for (2.2). The estimate (2.3) is directly deduced from it and the fact that $\sum_p (p\log p)^{-1}$ converges. Next, denoting $h_p = E[B_p]$ we have

$$s_N^2 = \sum_{p \leqslant N} E\big[(B_p - h_p)^2\big] + 2\sum_{p < q \leqslant N} E\big[(B_p - h_p)(B_q - h_q)\big].$$

First note that $E\big[(B_p - h_p)^2\big] = E[B_p] - h_p^2 = h_p(1 - h_p)$. Further, if $q > p^{\exp(z)}$ then B_p and B_q are independent, hence, $E\big[(B_p - h_p)(B_q - h_q)\big] = 0$. If $p < q \leqslant p^{\exp(z)}$ then $E[B_p B_q]$ vanishes, hence

$$E\big[(B_p - h_p)(B_q - h_q)\big] = -E[B_p]\,h_q - h_p E[B_q] + h_p h_q = -h_p h_q.$$

We obtain

$$s_N^2 = \sum_{p \leqslant N} h_p(1 - h_p) - 2\sum_{\substack{p < q \leqslant \min\{N, p^{\exp(z)}\}}} h_p h_q$$

$$= c_N - \sum_{p \leqslant N} h_p^2 - 2\sum_{\substack{p \leqslant N^{\exp(-z)} \\ p < q \leqslant p^{\exp(z)}}} h_p h_q - 2\sum_{\substack{N^{\exp(-z)} < p \leqslant N \\ p < q \leqslant N}} h_p h_q.$$

By (2.2) we have $h_p \ll 1/p$, hence, $\sum_p h_p^2 = O(1)$ and

$$\sum_{\substack{N^{\exp(-z)} < p \leqslant N \\ p < q \leqslant N}} h_p h_q \ll \left(\sum_{N^{\exp(-z)} < p \leqslant N} \frac{1}{p}\right)^2 = O(1).$$

Hence, (2.3) gives

$$s_N^2 = \mathrm{e}^{-z}\log\log N - 2\sum_{p \leqslant N^{\exp(-z)}} h_p \sum_{p < q \leqslant p^{\exp(z)}} h_q + O(1). \tag{2.5}$$

Using (2.2) we see that

$$\sum_{p \leqslant N^{\exp(-z)}} h_p \sum_{p < q \leqslant p^{\exp(z)}} h_q = \sum_{p \leqslant N^{\exp(-z)}} h_p \sum_{p < q \leqslant p^{\exp(z)}} \left(\frac{\mathrm{e}^{-z}}{q} + O\left(\frac{1}{q\log q}\right)\right),$$

which, by Mertens' theorem and $\sum_{q > t}(q\log q)^{-1} \ll (\log t)^{-1}$, equals

$$\sum_{p \leqslant N^{\exp(-z)}} h_p \left(\frac{z}{\mathrm{e}^z} + O\left(\frac{1}{\log p}\right)\right) = \sum_{p \leqslant N^{\exp(-z)}} \left(\frac{\mathrm{e}^{-z}}{p} + O\left(\frac{1}{p\log p}\right)\right)\left(\frac{z}{\mathrm{e}^z} + O\left(\frac{1}{\log p}\right)\right)$$

$$= \frac{z}{\mathrm{e}^{2z}}\left(\sum_{p \leqslant N^{\exp(-z)}} \frac{1}{p}\right) + O(1) = \frac{z}{\mathrm{e}^{2z}}(\log\log N) + O(1).$$

Injecting this into (2.5) concludes the proof. □

Lemma 2.3. *For all* $u \in \mathbb{N}, \mathbf{r} \in \mathbb{N}^u$ *and primes* $p_1, \ldots, p_u$, *we have*

$$E\left[\prod_{i=1}^u |B_{p_i} - E[B_{p_i}]|^{r_i}\right] = O_{\mathbf{r}}\left(\frac{1}{\mathrm{rad}(p_1 \cdots p_u)}\right),$$

where rad *denotes the radical.*

Proof. We write the factorisation into prime powers of $\prod_{i=1}^u p_i^{r_i}$ as $\prod_{j=1}^v q_j^{s_j}$, where q_j are v distinct primes. This implies that

$$E\left[\prod_{i=1}^u |B_{p_i} - E[B_{p_i}]|^{r_i}\right] = E\left[\prod_{j=1}^v |B_{q_j} - E[B_{q_j}]|^{s_j}\right].$$

Using $|B_{q_j} - E[B_{q_j}]| \leqslant B_{q_j} + E[B_{q_j}] \leqslant X_{q_j} + E[X_{q_j}] = X_{q_j} + 1/q_j$ and the binomial theorem yields

$$|B_{q_j} - E[B_{q_j}]|^{s_j} \leqslant (X_{q_j} + 1/q_j)^{s_j} = \sum_{t_j \in [0,s_j]} \binom{s_j}{t_j} \frac{X_{q_j}^{t_j}}{q_j^{s_j-t_j}} \ll_{\mathbf{s}} \max_{t_j \in [0,s_j]} \frac{X_{q_j}^{t_j}}{q_j^{s_j-t_j}},$$

hence,

$$E\left[\prod_{j=1}^{v} |B_{q_j} - E[B_{q_j}]|^{s_j} \right] \ll_{\mathbf{s}} \max_{\mathbf{t} \in [0,s_1] \times \cdots \times [0,s_v]} E\left[\prod_{j=1}^{v} \frac{X_{q_j}^{t_j}}{q_j^{s_j-t_j}} \right].$$

By the independence of the X_q, we infer that

$$E\left[\prod_{j=1}^{v} \frac{X_{q_j}^{t_j}}{q_j^{s_j-t_j}} \right] = \prod_{j=1}^{v} \frac{E[X_{q_j}^{t_j}]}{q_j^{s_j-t_j}} = \prod_{\substack{j=1 \\ t_j=0}}^{v} \frac{1}{q_j^{s_j}} \prod_{\substack{j=1 \\ t_j \geqslant 1}}^{v} \frac{E[X_{q_j}]}{q_j^{s_j-t_j}} \leqslant \prod_{\substack{j=1 \\ t_j=0}}^{v} \frac{1}{q_j} \prod_{\substack{j=1 \\ t_j \geqslant 1}}^{v} E[X_{q_j}] = \prod_{j=1}^{v} \frac{1}{q_j}.$$

The proof now concludes by noting that $\prod_{j=1}^{v} q_j$ is the radical of $\prod_{i=1}^{u} p_i^{r_i}$. $\qquad\square$

The following lemma is the main tool in the proof of Theorem 1.1. It is due to Stein [18, Corollary 2, p. 110].

Lemma 2.4 (Stein). *Let T be a finite set, and for each $t \in T$, let Z_t be a real random variable and T_t a subset of T such that $E[Z_t] = 0$, $E[Z_t^4] < \infty$ and $E[\sum_{t \in T} Z_t \sum_{s \in T_t} Z_s] = 1$. Then for all real b,*

$$\left| P\left[\sum_{t \in T} Z_t \leqslant b \right] - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} e^{-t^2/2} dt \right| \leqslant 4(\Psi_1 + \Psi_2 + \Psi_3), \tag{2.6}$$

where the terms Ψ_i are defined through

$$\Psi_1 = E\left[\sum_{t \in T} |E[Z_t | Z_s, s \notin T_t]| \right], \quad \Psi_2^2 = E\left[\sum_{t \in T} |Z_t| \left(\sum_{s \in T_t} Z_s \right)^2 \right]$$

and

$$\Psi_3^2 = E\left[\left\{ \sum_{t \in T} \sum_{s \in T_t} (Z_t Z_s - E[Z_s Z_t]) \right\}^2 \right].$$

Proposition 2.5. *Fix $z \geqslant 0$ and $b \in \mathbb{R}$. For any $N \in \mathbb{N}$, we have*

$$\left| P[S_N \leqslant c_N + b s_N] - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} e^{-t^2/2} dt \right| \ll_z (\log \log N)^{-1/4},$$

where the implied constant depends at most on z. In particular, $(S_N - c_N)/s_N$ converges in law to the standard normal distribution as $N \to \infty$.

Proof. We will apply Lemma 2.4 with

- T being the set of primes in $[2, N]$,
- T_p being the set of primes in $[p^{\exp(-z)}, p^{\exp(z)}] \cap [2, N]$,
- $Z_p = (B_p - E[B_p])/s_N$ for $p \in T$.

Let $Y_p := B_p - E[B_p]$. Note that if $q \notin T_p$ then Z_q and Z_p are independent, hence, $E[Y_p Y_q] = 0$. Therefore,

$$s_N^2 = \sum_{p,q \leqslant N} E[Y_p Y_q] = \sum_{\substack{p \leqslant N \\ q \in T_p}} E[Y_p Y_q],$$

which verifies $E\left[\sum_{p\in T} Z_p \sum_{q\in T_p} Z_q\right] = 1$. We next observe that since for every $q \notin T_p$, the random variables Z_q and Z_p are independent; one obtains $E[Z_p|Z_q, q \notin T_t] = E[Z_p] = 0$, therefore

$$\Psi_1 = 0. \tag{2.7}$$

Next, we use Lemma 2.3 to obtain

$$\Psi_2^2 s_N^3 = \sum_{p\leqslant N, q\in T_p} E\big[|Y_p|Y_q^2\big] + 2 \sum_{p\leqslant N, q_1 < q_2 \in T_p} E\big[|Y_p|Y_{q_1} Y_{q_2}\big]$$

$$\ll \sum_{p\leqslant N, q\in T_p} \frac{1}{pq} + \sum_{p\leqslant N, q_1, q_2 \in T_p} \frac{1}{pq_1 q_2}.$$

The sum $\sum_{q\in T_p} 1/q$ is bounded only in terms of z by Mertens' theorem. It shows that

$$\Psi_2^2 \ll s_N^{-3} \sum_{p\leqslant N} \frac{1}{p} \ll (\log\log N)^{-1/2}, \tag{2.8}$$

owing to (2.4).

To bound Ψ_3, we write $\mathscr{C}_p := \sum_{q\in T_p}\big(Y_p Y_q - E[Y_p Y_q]\big)$ to obtain

$$\Psi_3^2 s_N^4 = \sum_{p\leqslant N} E\big[\mathscr{C}_p^2\big] + 2 \sum_{p_1 < p_2 \leqslant N} E\big[\mathscr{C}_{p_1} \mathscr{C}_{p_2}\big]. \tag{2.9}$$

Furthermore, $E\big[\mathscr{C}_p^2\big]$ can be written as

$$\sum_{q\in T_p} E\big[\big(Y_p Y_q - E[Y_p Y_q]\big)^2\big] + 2 \sum_{q_1 < q_2 \in T_p} E\big[\big(Y_p Y_{q_1} - E[Y_p Y_{q_1}]\big)\big(Y_p Y_{q_2} - E[Y_p Y_{q_2}]\big)\big],$$

which can be seen to be

$$\ll \sum_{q\in T_p} \frac{1}{pq} + \sum_{q_1 < q_2 \in T_p} \frac{1}{pq_1 q_2}$$

by Lemma 2.3. Alluding to $\sum_{q\in T_p} 1/q \ll 1$ shows that

$$\sum_{p\leqslant N} E\big[\mathscr{C}_p^2\big] \ll \sum_{p\leqslant N} \frac{1}{p} \ll \log\log N. \tag{2.10}$$

Let us now observe that if $p_2 > p_1^{\exp(2z)}$ then $T_{p_1} \cap T_{p_2} = \emptyset$, therefore $\mathscr{C}_{p_1}$ and $\mathscr{C}_{p_2}$ are independent. Since for every p we have $E[\mathscr{C}_p] = 0$ by definition, we get $E[\mathscr{C}_{p_1} \mathscr{C}_{p_2}] = \prod_{i=1}^2 E[\mathscr{C}_{p_i}] = 0$. Thus,

$$\sum_{p_1 < p_2 \leqslant N} E\big[\mathscr{C}_{p_1} \mathscr{C}_{p_2}\big] = \sum_{\substack{p_1 < p_2 \leqslant N \\ p_2 \leqslant p_1^{\exp(2z)}}} \sum_{\substack{q_1 \in T_{p_1} \\ q_2 \in T_{p_2}}} E\big[\big(Y_{p_1} Y_{q_1} - E[Y_{p_1} Y_{q_1}]\big)\big(Y_{p_2} Y_{q_2} - E[Y_{p_2} Y_{q_2}]\big)\big]. \tag{2.11}$$

By Lemma 2.3, this is

$$\ll \sum_{p_1 \leqslant N} \sum_{p_1 < p_2 \leqslant p_1^{\exp(2z)}} \sum_{q_1 \in T_{p_1}} \sum_{q_2 \in T_{p_2}} \frac{1}{\operatorname{rad}(p_1 p_2 q_1 q_2)}.$$

For any positive integer c and prime q, we have $\operatorname{rad}(cq) = \operatorname{rad}(c)\frac{q}{\gcd(q,c)}$. Hence, the sum over q_2 is

$$\frac{1}{\operatorname{rad}(p_1 p_2 q_1)} \sum_{\substack{q_2 \in T_{p_2} \\ q_2 \in \{p_1, p_2, q_1\}}} 1 + \frac{1}{\operatorname{rad}(p_1 p_2 q_1)} \sum_{\substack{q_2 \in T_{p_2} \\ q_2 \notin \{p_1, p_2, q_1\}}} \frac{1}{q_2} \leqslant \frac{3 + \sum_{q\in T_{p_2}} 1/q}{\operatorname{rad}(p_1 p_2 q_1)} \ll_z \frac{1}{\operatorname{rad}(p_1 p_2 q_1)}$$

by Mertens' theorem. Hence, (2.11) is

$$\ll \sum_{\substack{p_1 \leqslant N \\ p_1 < p_2 \leqslant p_1^{\exp(2z)}}} \sum_{q_1 \in T_{p_1}} \frac{1}{\mathrm{rad}(p_1 p_2 q_1)} = \sum_{\substack{p_1 \leqslant N \\ p_1 < p_2 \leqslant p_1^{\exp(2z)}}} \frac{1}{\mathrm{rad}(p_1 p_2)} \left\{ \sum_{\substack{q_1 \in T_{p_1} \\ q_1 \in \{p_1, p_2\}}} 1 + \sum_{\substack{q_1 \in T_{p_1} \\ q_1 \notin \{p_1, p_2\}}} \frac{1}{q_1} \right\}.$$

The two sums over q_1 in the right-hand side are both bounded only in terms of z. This can be proved similarly as before with the sum over q_2. We obtain the bound

$$\ll \sum_{\substack{p_1 \leqslant N \\ p_1 < p_2 \leqslant p_1^{\exp(2z)}}} \frac{1}{\mathrm{rad}(p_1 p_2)} = \sum_{p_1 \leqslant N} \frac{1}{p_1} \sum_{p_1 < p_2 \leqslant p_1^{\exp(2z)}} \frac{1}{p_2} \ll \sum_{p_1 \leqslant N} \frac{1}{p_1} \ll \log \log N.$$

This shows that the quantity in (2.11) is $\ll \log \log N$, which, when combined with (2.10), can be fed into (2.9) to yield $\Psi_3^2 s_N^4 \ll \log \log N$. Invoking (2.4) provides us with $\Psi_3 \ll 1/\sqrt{\log \log N}$. Together with (2.7)–(2.8), it implies that

$$\left| P\left[S_N \leqslant c_N + b s_N \right] - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} e^{-t^2/2} \mathrm{d}t \right| \ll_z (\log \log N)^{-1/4}$$

owing to Stein's bound (2.6). Finally, letting $N \to \infty$ shows that $(S_N - c_N)/s_N$ converges in law to the standard normal distribution. $\qquad \square$

Remark 2.6. *We next prove asymptotics for the moments of $(S_N - c_N)/s_N$. This is possibly the central proof in the present paper. The argument is a modification of the one by Billingsley [1, Lemma 3.2], which relies on a version of the dominated convergence theorem. However, the underlying random variables are now dependent; thus, we need to introduce the notion of* linked indices.

Proposition 2.7. *Fix $z \geqslant 0$ and a positive integer r. Then we have*

$$\lim_{N \to \infty} E\left[\left(\frac{S_N - c_N}{s_N} \right)^r \right] = \mu_r,$$

where μ_r is the r-th moment of the standard normal distribution.

Proof. Take $2k$ to be the least strictly positive integer with $r < 2k$, so that Proposition 2.5 [19, Example 2.21] implies that it suffices to prove that

$$\sup_{N \geqslant 1} \left| E\left[\left(\frac{S_N - c_N}{s_N} \right)^{2k} \right] \right|$$

is bounded only in terms of k and z. Equivalently, by (2.4) it suffices to show

$$E\left[(S_N - c_N)^{2k} \right] = E\left[\left(\sum_{p \leqslant N} (B_p - E[B_p]) \right)^{2k} \right] \ll_{k,z} (\log \log N)^k.$$

The left side equals

$$\sum_{u=1}^{2k} \sum_{\substack{\mathbf{r} \in \mathbb{N}^u \\ 2k = \sum_{i=1}^{u} r_i}} \frac{(2k)!}{r_1! \cdots r_u!} \sum_{p_1 < \ldots < p_u \leqslant N} E\left[\prod_{i=1}^{u} (B_p - E[B_p])^{r_i} \right].$$

Using Lemma 2.3, we see that the contribution of the terms with $u \leqslant k$ is

$$\ll_k \max_{1 \leqslant u \leqslant k} \left(\sum_{p \leqslant N} \frac{1}{p} \right)^u \ll_k (\log \log N)^k.$$

Therefore,

$$E\big[(S_N - c_N)^{2k}\big] \ll \max_{\substack{u \in [k+1, 2k] \\ \mathbf{r} \in \mathbb{N}^u := \sum_i r_i = 2k}} \sum_{p_1 < \ldots < p_u \leqslant N} E\left[\prod_{i=1}^{u} (B_p - E[B_p])^{r_i}\right] + (\log \log N)^k, \qquad (2.12)$$

with an implied constant that is independent of N.

For given $u \in \mathbb{N}$, $z \geqslant 0$ and primes $p_1 < \ldots < p_u$, we say that two consecutive integers $i, i+1$ in $[1, u]$ are *linked* if and only if $p_{i+1} \leqslant p_i^{\exp(z)}$. In particular, p_{i+1} lies in a relatively small interval; hence, its contribution will be small. Denote the number of linked pairs $(i, i+1)$ by $\ell(\mathbf{p})$. By Lemma 2.3, we obtain

$$\sum_{p_1 < \ldots < p_u \leqslant N} E\left[\prod_{i=1}^{u} (B_p - E[B_p])^{r_i}\right] \ll_z \left(\sum_{p \leqslant N} \frac{1}{p}\right)^{u - \ell(\mathbf{p})} \ll (\log \log N)^{u - \ell(\mathbf{p})},$$

where we used the estimate $\sum_{p_i < p_{i+1} < p_i^{\exp(z)}} 1/p_i \ll_z 1$ whenever i and $i+1$ are linked. Hence, the contribution of all prime vectors $(p_1, \ldots, p_u)$ with at least $\ell(\mathbf{p}) \geqslant u - k$ linked pairs is at most

$$\ll (\log \log N)^{u - \ell(\mathbf{p})} \ll (\log \log N)^k,$$

which is acceptable. By (2.12), we obtain

$$E\big[(S_N - c_N)^{2k}\big] \ll \max_{\substack{u \in [k+1, 2k] \\ \mathbf{r} \in \mathbb{N}^u := \sum_i r_i = 2k}} \sum_{\substack{p_1 < \ldots < p_u \leqslant N \\ \ell(\mathbf{p}) < u - k}} E\left[\prod_{i=1}^{u} (B_p - E[B_p])^{r_i}\right] + (\log \log N)^k, \qquad (2.13)$$

We will now show that every sum over p_i in (2.13) vanishes. Denoting the cardinality of $1 \leqslant i \leqslant u$ with $r_i = 1$ by a, we see that the number of i with $r_i \geqslant 2$ is $u - a$. Since $2k = \sum_{i=1}^{u} r_i$, we get $2k \geqslant a + 2(u - a)$. Equivalently, $2(u - k) \leqslant a$, hence, by $\ell(\mathbf{p}) < u - k$ one gets

$$2\ell(\mathbf{p}) < \sharp\{i \in [1, u] : r_i = 1\}. \qquad (2.14)$$

We now partition the integers in $[1, u]$ into disjoint subsets $\mathscr{A}_1, \ldots, \mathscr{A}_r$ using the following rules:

- if i and $i+1$ are in S_j then they are linked,
- if $i \in S_a$ and $i+1 \in S_b$ for some $a \neq b$ then i and $i+1$ are not linked.

The inequality $s \leqslant 2(-1 + s)$ for $s \geqslant 2$ gives

$$\sharp\{i \in [1, u] : i \text{ linked to some index}\} = \sum_{\substack{1 \leqslant j \leqslant r \\ 2 \leqslant \sharp \mathscr{A}_j}} \sharp \mathscr{A}_j \leqslant \sum_{\substack{1 \leqslant j \leqslant r \\ 2 \leqslant \sharp \mathscr{A}_j}} 2(-1 + \sharp \mathscr{A}_j).$$

This equals $2\ell(\mathbf{p})$ since each $\mathscr{A}_j$ has $-1 + \sharp \mathscr{A}_j$ linked pairs and the total number of links is $\ell(\mathbf{p})$. By (2.14), we infer that there exists an index j for which $r_j = 1$ and that is not linked to any other index. This implies that the following random variables are independent:

$$\prod_{\substack{1 \leqslant i \leqslant u \\ i \neq j}} (B_{p_i} - E[B_{p_i}])^{r_i} \quad \text{and} \quad (B_{p_j} - E[B_{p_j}])^{r_j} = B_{p_j} - E[B_{p_j}].$$

Since $E\big[B_{p_j} - E[B_{p_j}]\big] = 0$, we infer that every expectation in the right-hand side of (2.13) vanishes. This concludes the proof. □

2.3. *Justifying the model*

Let n be a positive integer and denote by Ω_n the uniform probability space $\mathbb{N} \cap [1, n]$. Our goal now becomes to show that, as $n \to \infty$, the moments of $\omega_z(m)$ for m in Ω_n are asymptotically the same as the moments of S_N for some parameter $N = N(n) \to \infty$. Recall (2.1). For technical reasons, we will first

work with a truncated version of ω_z, namely,

$$\omega_{z,N}(m) = \sum_{p \leqslant N} \delta_{p,z}(m), \tag{2.15}$$

where $N = N(n)$. The function $\delta_{p,z}$ imposes simultaneous coprimality conditions of m with several primes in large intervals, and to deal with this, we shall need the Fundamental Lemma of Sieve Theory [8, Corollary 6.10].

Lemma 2.8 (Fundamental Lemma of Sieve Theory). *Let $\mathscr{P}$ be a set of primes. Given any sequence $a_m \geqslant 0$ for $m \in \mathbb{N}$ and any square-free $d \leqslant x$ that is only divisible by primes in $\mathscr{P}$, we assume that*

$$\sum_{\substack{m \leqslant x \\ m \equiv 0 (\mathrm{mod}\ d)}} a_m = Xg(d) + r_d$$

for some real numbers X, r_d and a multiplicative function g. Assume that $0 \leqslant g(p) < 1$ and that there exist constants $K > 1, \kappa > 0$ such that

$$\prod_{\substack{w \leqslant p < y \\ p \in \mathscr{P}}} (1 - g(p))^{-1} \leqslant K \left(\frac{\log y}{\log w} \right)^{\kappa}$$

holds for all $2 \leqslant w < y$. Then for all $D \geqslant y \geqslant 2$, we have

$$\sum_{\substack{m \leqslant x \\ p \in \mathscr{P}, p < y \Rightarrow p \nmid m}} a_m = X \left(\prod_{\substack{p < y \\ p \in \mathscr{P}}} (1 - g(p)) \right) \{1 + O(\mathrm{e}^{-s})\} + O \left(\sum_{\substack{d < D \\ p | d \Rightarrow p \in \mathscr{P}}} \mu^2(d) |r_d| \right), \tag{2.16}$$

where $s = \log D / \log y$ and the implied constants depend at most on κ and K.

Lemma 2.9. *Assume that there exists a function $N : [1, \infty) \to [1, \infty)$ satisfying*

$$\lim_{n \to \infty} N(n) = +\infty, \tag{2.17}$$

$$\limsup_{n \to \infty} \frac{(\log N(n))(\log \log \log N(n))}{\log n} \neq +\infty. \tag{2.18}$$

Fix $z \geqslant 0$ and $k \in \mathbb{N}$. Then we have

$$\lim_{n \to \infty} E_{m \in \Omega_n} \left[\left(\frac{\omega_{z,N} - c_N}{s_N} \right)^k \right] = \mu_k,$$

where μ_k is the k-th moment of the standard normal distribution.

Proof. By Proposition 2.7 and (2.17), it is sufficient to prove

$$\lim_{n \to \infty} \left(E_{m \in \Omega_n} \left[\left(\frac{\omega_{z,N}(m) - c_N}{s_N} \right)^k \right] - E \left[\left(\frac{S_N - c_N}{s_N} \right)^k \right] \right) = 0. \tag{2.19}$$

Let $r \in \mathbb{N}$. By (2.15), the fact that $\delta_{p,z} \in \{0, 1\}$ and the binomial theorem, we obtain

$$E_{m \in \Omega_n} \left[\omega_{z,N}(m)^r \right] = \sum_{u=1}^{r} \sum_{\substack{r_1, \ldots, r_u \in \mathbb{N} \\ r_1 + \ldots + r_u = r}} \frac{r!}{r_1! \cdots r_u!} \sum_{p_1 < \cdots < p_u \leqslant N} E_{m \in \Omega_n} \left[\delta_{p_1,z}(m) \cdots \delta_{p_u,z}(m) \right]. \tag{2.20}$$

Let $\mathscr{P}$ be the set of all primes in $\bigcup_{i=1}^{u}(p_i, p_i^{\exp(z)}]$ and let a_m be the indicator function of integers divisible by $p_1 \cdots p_u$. In particular,

$$E_{m \in \Omega_n}\left[\delta_{p_1, z}(m) \cdots \delta_{p_u, z}(m)\right] = \frac{1}{n} \sum_{\substack{1 \leqslant m \leqslant n \\ p \in \mathscr{P} \Rightarrow p \nmid m}} a_m.$$

We assume that $p_{i+1} > p_i^{\exp(z)}$ for all $i = 1, 2, \ldots, u-1$ since otherwise the sum clearly vanishes. We will now use Lemma 2.8 with $X = n/(p_1 \cdots p_u)$, $g(d) = 1/d$, $D = \sqrt{n}$, $y = N^{2 \exp(z)}$. If d is divisible only by primes in $\mathscr{P}$, then it is coprime to $p_1 \cdots p_u$, hence,

$$\sum_{\substack{m \leqslant n \\ m \equiv 0 (\mathrm{mod}\ d)}} a_m = \left[\frac{n}{p_1 \cdots p_u d}\right],$$

thus, $|r_d| \leqslant 1$ because r_d is the fractional part of X/d. Furthermore, we can take K to be any large fixed positive constant and $\kappa = 1$, owing to

$$\prod_{\substack{w \leqslant p < y \\ p \in \mathscr{P}}} (1 - g(p))^{-1} = \prod_{\substack{w \leqslant p < y \\ p \in \mathscr{P}}} (1 - 1/p)^{-1} \leqslant \prod_{w \leqslant p < y} (1 - 1/p)^{-1} \ll \frac{\log y}{\log w}.$$

The bound $|r_d| \leqslant 1$, means that $\sum_{d \leqslant D} \mu^2(d)|r_d| \leqslant D = \sqrt{n}$. Since $p_u \leqslant N$, every prime in $\mathscr{P}$ is strictly smaller than y, hence, (2.16) gives

$$E_{m \in \Omega_n}\left[\delta_{p_1} \cdots \delta_{p_u}\right] = \left\{\prod_{i=1}^{u} \frac{1}{p_i} \prod_{p_i < q \leqslant p_i^{\exp(z)}} (1 - 1/p)\right\}\left\{1 + O\left(e^{-\frac{\log n}{4 \exp(z) \log N}}\right)\right\} + O(n^{-1/2}), \quad (2.21)$$

where the implied constant depends at most on r and z.

By the binomial theorem, we get

$$E[S_N^r] = E\left[\left(\sum_{p \leqslant N} B_p\right)^r\right] = \sum_{u=1}^{r} \sum_{\substack{r_1, \ldots, r_u \in \mathbb{N} \\ r_1 + \ldots + r_u = r}} \frac{r!}{r_1! \cdots r_u!} \sum_{p_1 < \cdots < p_u \leqslant N} E\left[B_{p_1} \cdots B_{p_u}\right]$$

and we note that we can restrict the sum over p_i to the terms with $p_{i+1} > p_i^{\exp(z)}$ for all i, since otherwise $E\left[B_{p_1} \cdots B_{p_u}\right] = 0$. Under this restriction, the random variables B_{p_i} are independent, hence,

$$\prod_{i=1}^{u} \frac{1}{p_i} \prod_{p_i < q \leqslant p_i^{\exp(z)}} (1 - 1/p) = E\left[B_{p_1} \cdots B_{p_u}\right].$$

We infer from (2.20) and (2.21) that

$$\left|E_{m \in \Omega_n}\left[\omega_{z, N}(m)^r\right] - E[S_N^r]\right| \ll_r E[S_N^r] e^{-\frac{\log n}{4 \exp(z) \log N}} + n^{-1/2} \sum_{u=1}^{r} \sum_{p_1 < \ldots < p_u \leqslant N} 1.$$

By (2.3), this is $\ll (\log \log N)^r \exp(-\frac{\log n}{4 \exp(z) \log N}) + n^{-1/2} N^r$. Thus, the difference in (2.19) is

$$\ll s_N^{-k} \sum_{r=0}^{k} \binom{k}{r} (-c_N)^{k-r} \left(E_{\Omega_n}\left[\omega_{z, N}^r\right] - E[S_N^r]\right)$$

$$\ll s_N^{-k} \left\{e^{-\frac{\log n}{4 \exp(z) \log N}} (c_N + \log \log N)^k + n^{-1/2}(N + c_N)^k\right\}.$$

We need to show that this vanishes asymptotically, and by (2.4) and (2.17), it suffices to show

$$(2 \log \log N)^k \leqslant \exp\left(\frac{\log n}{4 \exp(z) \log N}\right) \quad \text{and} \quad (2N)^k \leqslant n^{1/2}.$$

Both of these inequalities can be directly inferred from (2.17) to (2.18). $\square$

Lemma 2.10. *Assume that there exists a function* $N : [1, \infty) \to [1, \infty)$ *satisfying*

$$\lim_{n\to\infty} \frac{\log\log N(n)}{\log\log n} = 1, \tag{2.22}$$

$$\limsup_{n\to\infty} \frac{(\log N(n))\sqrt{\log\log n}}{\log n} = +\infty. \tag{2.23}$$

Fix $z \geqslant 0$. *Then* $s_N((1 - 2ze^{-z})e^{-z}\log\log n)^{-1/2} \to 1$ *as* $n \to \infty$ *and*

$$\lim_{n\to\infty} \frac{\max\left\{\left|(\omega_z - e^{-z}\log\log n) - (\omega_{z,N} - c_N)\right| : m \in \mathbb{N} \cap [1, n]\right\}}{\sqrt{\log\log n}} = 0.$$

Proof. Combining (2.4) and (2.22) one immediately gets

$$\lim_{n\to\infty} \frac{s_N}{((1 - 2ze^{-z})e^{-z}\log\log n)^{1/2}} = 1.$$

For any $m \in [1, n]$, we have

$$\left|(\omega_z - e^{-z}\log\log n) - (\omega_{z,N} - c_N)\right| \leqslant \sum_{p>N} \delta_{p,z}(m) + |e^{-z}(\log\log n) - c_N|.$$

Since $\delta_{p,z}$ takes only values in $\{0, 1\}$ and $\delta_{p,z}(m) = 1$ implies that p divides m, we see that

$$\sum_{p>N} \delta_{p,z}(m) \leqslant \sharp\{p \mid m : p > N\} \leqslant \frac{\log m}{\log N} \leqslant \frac{\log n}{\log N}.$$

Furthermore, (2.3) gives

$$e^{-z}(\log\log n) - c_N \ll 1 + \log\frac{\log n}{\log N} \ll \frac{\log n}{\log N}.$$

The proof now concludes by using (2.23). ☐

2.4. Proof of Theorem 1.1

The function

$$N(n) := n^{1/\log\log n}$$

fulfills (2.17)–(2.18)–(2.22)–(2.23). Hence, we can apply Lemmas 2.9–2.10.

For any $r \in \mathbb{N}$, $c \in \mathbb{C}$ any probability space Ω_n and any two sequences of random variables X_n, Y_n satisfying $\lim_{n\to\infty} \sup_{m\in\Omega_n} |X_n(m) - Y_n(m)| = 0$ and $\lim_{n\to\infty} E_{m\in\Omega_n}[X_n(m)^r] = c$ it is easy to see by the binomial theorem that $\lim_{n\to\infty} E_{m\in\Omega_n}[Y_n(m)^r] = c$. Using this with $\Omega_n = \mathbb{N} \cap [1, n]$,

$$X_n(m) = \frac{\omega_{z,N}(m) - c_N}{s_N} \quad \text{and} \quad Y_n(m) = \frac{\omega_z(m) - e^{-z}\log\log n}{s_N},$$

in combination with Lemmas 2.9–2.10, shows that for every $k \in \mathbb{N}$ one has

$$\lim_{n\to\infty} E_{m\in\Omega_n}\left[\left(\frac{\omega_z(m) - e^{-z}\log\log n}{s_N}\right)^k\right] = \mu_k.$$

Given any sequence $a_n \in \mathbb{R}$ having limit 1 and any sequence of random variables X_n with $E[X_n]$ having limit c, it is clear that $a_n E[X_n]$ has limit c. Using this with

$$a_n = \frac{s_N}{((1 - 2ze^{-z})e^{-z}\log\log n)^{1/2}} \quad \text{and} \quad X_n(m) = \frac{\omega_z(m) - e^{-z}\log\log n}{s_N}$$

and invoking Lemma 2.10 shows that for every $k \in \mathbb{N}$ one has

$$\lim_{n \to \infty} E_{m \in \Omega_n} \left[\left(\frac{\omega_z(m) - e^{-z} \log \log n}{((1 - 2ze^{-z})e^{-z} \log \log n)^{1/2}} \right)^k \right] = \mu_k. \tag{2.24}$$

This proves Theorem 1.1 whenever r is a positive integer and this is sufficient. To see that, take any $r \in [0, \infty)$ and note that (2.24) implies that

$$T_n = \frac{\omega_z(m) - e^{-z} \log \log n}{((1 - 2ze^{-z})e^{-z} \log \log n)^{1/2}}$$

converges in law to the standard normal distribution. Taking p to be the least even integer strictly exceeding r in [19, Example 2.21] shows that the r-th moment of T_n converges to the r-th moment of the standard normal distribution. $\qquad\square$

3. Poissonian gaps for local solubility in families of varieties

Serre's problem [16] on the probability that a random variety over $\mathbb{Q}$ has a $\mathbb{Q}$-rational point has recently received a lot of attention due to its extension by Loughran–Smeets [13] to a very general setting, namely, for any dominant morphism $f : V \to \mathbb{P}^n$, where, V is a smooth projective variety over $\mathbb{Q}$ and f has a geometrically integral generic fibre. The fibres of f form an infinite family of varieties and typically one is interested in how often they have a $\mathbb{Q}$-rational point. Imposing the harmless condition that the generic fibre of f is geometrically integral, it is easy to see that for every x outside of some proper Zariski closed set the function

$$\omega_f(x) := \sharp \left\{ p \text{ prime} : (f^{-1}(x))(\mathbb{Q}_p) = \emptyset \right\},$$

is bounded due to the Lang–Weil estimates and Hensel's lemma. This function helps us in understanding the density of fibres with a $\mathbb{Q}$-rational point. Ordering $\mathbb{P}^n(\mathbb{Q})$ by the standard Weil height H on $\mathbb{P}^n(\mathbb{Q})$ and assuming that a certain invariant $\Delta(\pi)$ is non-vanishing, Loughran and Sofos [14] recently proved the analogue of Erdős–Kac's theorem for $\omega_f(x)$, namely that

$$\frac{\omega_f(x) - \Delta(\pi) \log \log H(x)}{(\Delta(\pi) \log \log H(x))^{1/2}}$$

converges in law to the standard normal distribution. This was the first instance of an Erdős–Kac law in Diophantine geometry.

Our goal in this section is to go further and study the gaps between the primes p counted by $\omega_f(x)$. For $x \in \mathbb{P}^n(\mathbb{Q})$ with $f^{-1}(x)$ smooth, we let $p_i(x)$ be the i-th smallest prime number for which $f^{-1}(x)$ has no $\mathbb{Q}_p$-point. We then define for all $z \geqslant 0$,

$$\omega_{f,z}(x) := \sharp\{i \geqslant 1 : \log \log p_{i+1}(x) - \log \log p_i(x) > z\}.$$

Before stating our theorem, we must recall the definition of the invariant $\Delta(f)$ that is due to Loughran and Smeets [13].

Definition 3.1. *Let $f : V \to X$ be a dominant proper morphism of smooth irreducible varieties over a field k of characteristic 0. For each point $x \in X$ with residue field $\kappa(x)$, the absolute Galois group $\mathrm{Gal}(\overline{\kappa(x)}/\kappa(x))$ of the residue field acts on the irreducible components of*

$$f^{-1}(x)_{\overline{\kappa(x)}} := f^{-1}(x) \times_{\kappa(x)} \overline{\kappa(x)}$$

of multiplicity 1. Choose some finite group Γ_x through which this action factors and define

$$\delta_x(f) = \frac{\sharp \left\{ \gamma \in \Gamma_x : \begin{array}{c} \gamma \text{ fixes an irreducible component} \\ \text{of } f^{-1}(x)_{\overline{\kappa(x)}} \text{ of multiplicity } 1 \end{array} \right\}}{\sharp \Gamma_x}$$

and

$$\Delta(f) = \sum_{D \in X^{(1)}} (1 - \delta_D(f)),$$

where $X^{(1)}$ *denotes the set of codimension* 1 *points of X.*

Theorem 3.2. *Let V be a smooth projective variety over $\mathbb{Q}$ equipped with a dominant morphism $f : V \to \mathbb{P}^n$ with geometrically integral generic fibre and $\Delta(f) \neq 0$. Let H be the usual Weil height on $\mathbb{P}^n$. Fix any $z \geqslant 0$ and $r \geqslant 0$. Then*

$$\sum_{\substack{x \in \mathbb{P}^n(\mathbb{Q}), H(x) \leqslant B \\ f^{-1}(x) \text{ smooth}}} \left(\frac{\omega_{f,z}(x) - \Delta(f) \exp(-z\Delta(f)) \log \log B}{\sqrt{\Delta(f) \exp(-z\Delta(f)) \log \log B}} \right)^r$$

$$= \mu_r \left(1 - \frac{2\Delta(f)z}{\exp(\Delta(f)z)} \right)^{r/2} \sharp \{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leqslant B\} (1 + o(1)),$$

as $B \to \infty$, where μ_r is the r-th moment of the standard normal distribution.

The case $z = 0$ recovers Theorems 1.2–1.3 of Loughran–Sofos [14].

Taking $r = 2$ in Theorem 3.2 and [14, Theorem 1.2] shows the following after a use of Chebychev's inequality:

Corollary 3.3. *Let $f : V \to \mathbb{P}^n$ be a morphism as in Theorem 3.2. Fix any $z \geqslant 0$. Ordering $\mathbb{P}^n(\mathbb{Q})$ by the usual Weil height,* 100% *of fibres $f^{-1}(x)$ satisfy*

$$\left| \frac{\omega_{f,z}(x)}{\omega_f(x)} - \frac{\Delta(f)}{e^{z\Delta(f)}} \right| \leqslant (\log \log H(x))^{-1/4}.$$

Remark 3.4. *As the right-hand side vanishes asymptotically, the corollary means that for almost all fibres $f^{-1}(x)$, the proportion of gaps in the sequence $\{\log \log p_i(x)\}_{i \geqslant 1}$ exceeding z is roughly constant, independently of the fibre!*

In our proof, we use the arguments from Section 2, where the uniform probability space $\mathbb{N} \cap [1, n]$ is replaced by $\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leqslant B\}$. The main number–theoretic we use is Proposition 3.6. In sieve theory language, this is a level of distribution result for the fibres of f. The level of distribution it provides is less than B^ε for any constant $\varepsilon > 0$, which is well-known to be a problematic regime for any sieve theory problem; we overcome this by extirpating small primes $p \leqslant t_0(B)$ from $\widetilde{\omega}_{f,z}$, see (3.5).

3.1. Proof of Theorem 3.2

For a prime p, we define

$$\sigma_p := \frac{\sharp\{x \in \mathbb{P}^n(\mathbb{F}_p) : f^{-1}(x) \text{ is non-split}\}}{\sharp \mathbb{P}^n(\mathbb{F}_p)},$$

where we use the term "non-split" in the sense of Skorobogatov [17, Def. 0.1]. We then introduce the random variable $\widetilde{X}_p$ so that

$$P[\widetilde{X}_p = 1] = \sigma_p, P[\widetilde{X}_p = 0] = 1 - \sigma_p$$

and such that $\widetilde{X}_p$ are independent. We then define

$$\widetilde{B}_p := \widetilde{X}_p \prod_{\substack{q \text{ prime} \\ p < q \leqslant p^{\exp(z)}}} (1 - \widetilde{X}_q).$$

Furthermore, for any positive N, we define

$$\widetilde{S}_N = \sum_{p \leqslant N} \widetilde{B}_p, \quad \widetilde{c}_N := E[\widetilde{S}_N] \quad \text{and} \quad \widetilde{s}_N^2 := \text{Var}[\widetilde{S}_N].$$

Using [14, Proposition 3.6] instead of Mertens' theorem and the estimate $\sigma_p \ll 1/p$ from [14, Lemma 3.3], the arguments in Lemma 2.2 can be modified to yield

$$E[\widetilde{B}_p] = \exp(-z\Delta(f))\sigma_p + O\left(\frac{1}{p \log p}\right), \tag{3.1}$$

$$\widetilde{c}_N = \Delta(f) \exp(-z\Delta(f)) \log \log N + O(1) \tag{3.2}$$

and

$$\widetilde{s}_N^2 = \left(1 - \frac{2\Delta(f)z}{\exp(\Delta(f)z)}\right) \frac{\Delta(f) \log \log N}{\exp(\Delta(f)z)} + O(1). \tag{3.3}$$

Next, the proof of Lemma 2.3 goes through easily upon replacing B_p by $\widetilde{B}_p$ owing to the inequality $E[\widetilde{B}_p] \leqslant E[\widetilde{X}_p] = \sigma_p \ll 1/p$. Replacing S_N by $\widetilde{S}_N$ in the statement of Proposition 2.5, we see that the proof goes through by replacing Z_p by $\widetilde{Z}_p := (\widetilde{B}_p - E[\widetilde{B}_p])/\widetilde{s}_N$. Finally, using all the analogues of results in Section 2 that we mentioned so far allows one to modify the arguments of the proof of Proposition 2.7 to obtain the following result:

Proposition 3.5. *Fix $z \geqslant 0$ and a positive integer r. Then we have*

$$\lim_{N \to \infty} E\left[\left(\frac{\widetilde{S}_N - \widetilde{c}_N}{\widetilde{s}_N}\right)^r\right] = \mu_r,$$

where μ_r is the r-th moment of the standard normal distribution.

This concludes the probabilistic part of the proof of Theorem 3.2. The number–theoretic part requires the Fundamental lemma of sieve theory and the following:

Proposition 3.6. *Keep the setting of Theorem 3.2. Then there exist constants $\delta > 1, A > 0$ that depends on V and f with the following property. Let $Q \in \mathbb{N}$ with $p \nmid Q$ for all $p \leqslant A$. Then for all $\varepsilon > 0$ and $Q \leqslant B^{1/6}$, we have*

$$\sharp \left\{ x \in \mathbb{P}^n(\mathbb{Q}) : \begin{array}{l} H(x) \leqslant B, f^{-1}(x) \text{ smooth} \\ f^{-1}(x)(\mathbb{Q}_p) = \emptyset \; \forall \, p \mid Q \end{array} \right\} = c_n B^{n+1} \prod_{p \mid Q} \sigma_p + O\left(\frac{\delta^{\omega(Q)} B^{n+1}}{Q \min\{p \mid Q\}}\right),$$

where the implied constant is independent of B and Q.

Proof. By [14, Proposition 3.4], there exist $\alpha > 0, d \in \mathbb{N}$ such that the left-hand side is at most

$$c_n B^{n+1}\left(\prod_{p \mid Q} (\sigma_p + \alpha/p^2)\right) + O\left((4d)^{\omega(Q)}(Q^{2n+1}B + QB^n(\log B)^{[1/n]})\right).$$

while, it exceeds a similar quantity with α replaced by $-\alpha$. As shown in [14, Lemma 3.7], we have

$$\prod_{p \mid Q} (\sigma_p + \alpha/p^2) = \prod_{p \mid Q} \sigma_p + O\left(\frac{(2\alpha d)^{\omega(Q)}}{Q \min\{p : p \mid Q\}}\right).$$

This is satisfactory by defining $\delta = 2 + \max\{4d, 2\alpha d\}$. Finally,

$$Q^{2n+1}B + QB^n \log B \ll \frac{B^{n+1}}{Q^2} \leqslant \frac{B^{n+1}}{Q \min\{p \mid Q\}}$$

owing to $Q \leqslant B^{1/6}$. □

Our next task is to show that the moments of a truncated version of $\omega_{f,z}$ are asymptotically Gaussian. For this we shall follow the arguments in Section 2.3, where $\Omega_n = \mathbb{N} \cap [1, n]$ is replaced by the uniform discrete probability space

$$\widetilde{\Omega}_B = \{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leqslant B, f^{-1}(x) \text{ smooth}\}$$

for $B > 0$. The condition that $f^{-1}(x)$ smooth is included in the definition of $\widetilde{\Omega}_B$ to make $\omega_{f,z}(x)$ well-defined for each $x \in \widetilde{\Omega}_B$. Choosing a polynomial which vanishes on the singular locus of f, we see that

$$\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leqslant B, f^{-1}(x) \text{ not smooth}\} = O(B^n).$$

Then the standard result

$$\sharp\{x \in \mathbb{P}^n(\mathbb{Q}) : H(x) \leqslant B\} = c_n B^{n+1} + O(B^n (\log B)^{[1/n]}),$$

where $c_n = 2^n / \zeta(n + 1)$, shows that

$$\sharp\widetilde{\Omega}_B = c_n B^{n+1} + O\left(B^n (\log B)^{[1/n]}\right).$$

We furthermore let for $x \in \mathbb{P}^n(\mathbb{Q})$,

$$\widetilde{\delta}_{p,z}(x) := \begin{cases} 1, & \text{if } f^{-1}(x)(\mathbb{Q}_p) = \varnothing \text{ and } f^{-1}(x)(\mathbb{Q}_q) \neq \varnothing \text{ for every prime } q \in (p, p^{\exp(z)}], \\ 0, & \text{otherwise.} \end{cases}$$

We shall choose any two functions $t_0, t_1 : (0, \infty) \to (0, \infty)$ satisfying

$$1 < t_0(B) < t_1(B) < B, \quad \lim_{B \to \infty} t_0(B) = \lim_{B \to \infty} t_1(B) = \infty. \tag{3.4}$$

They will be chosen optimally later. The analogue of (2.15) in our setting is defined as

$$\widetilde{\omega}_{z,B}(x) = \sum_{t_0(B) < p \leqslant t_1(B)} \widetilde{\delta}_{p,z}(x). \tag{3.5}$$

We obtain for $r \in \mathbb{N}$,

$$E_{x \in \widetilde{\Omega}_B}\left[\widetilde{\omega}_{z,B}(x)^r\right] = \sum_{u=1}^{r} \sum_{\substack{r_1,\ldots,r_u \in \mathbb{N} \\ r_1 + \ldots + r_u = r}} \frac{r!}{r_1! \cdots r_u!} \sum_{\substack{t_0(B) < p_1 < \cdots < p_u \leqslant t_1(B) \\ p_{i+1} > p_i^{\exp(z)} \; \forall i}} E_{x \in \widetilde{\Omega}_B}\left[\widetilde{\delta}_{p_1,z}(x) \cdots \widetilde{\delta}_{p_u,z}(x)\right], \tag{3.6}$$

where we added the assumption $p_{i+1} > p_i^{\exp(z)} \; \forall i$ since otherwise the expectation in the right-hand side vanishes.

Let us now define the function $m_B : \mathbb{P}^n(\mathbb{Q}) \to \mathbb{N}$ given by

$$m_B(x) := \prod_{\substack{t_0(B) < p \leqslant t_1(B) \\ f^{-1}(x)(\mathbb{Q}_p) = \varnothing}} p.$$

Letting $\widetilde{x}$ be the product of all primes $p \leqslant t_1(B)$ we note that $m_B(x) \leqslant \widetilde{x}$. Now let $\mathscr{P}$ be the set of all primes in $\bigcup_{i=1}^{u} (p_i, p_i^{\exp(z)}]$ and for $m \in \mathbb{N}$ let

$$\widetilde{a}_m := \begin{cases} \sharp\{x \in \widetilde{\Omega}_B : m_B(x) = m\} / \sharp\widetilde{\Omega}_B, & \text{if } m \equiv 0 \pmod{p_1 \cdots p_u}, \\ 0, & \text{otherwise.} \end{cases}$$

This gives

$$E_{x \in \widetilde{\Omega}_B}\left[\widetilde{\delta}_{p_1,z}(x) \cdots \widetilde{\delta}_{p_u,z}(x)\right] = \sum_{\substack{m \leqslant \widetilde{x} \\ p \in \mathscr{P} \Rightarrow p \nmid m}} \widetilde{a}_m.$$

We shall use Lemma 2.8 with $a_m := \widetilde{a}_m$, $\kappa = \Delta(f)$,

$$X = \prod_{i=1}^{u} \sigma_{p_i}, \quad g(d) = \prod_{p|d} \sigma_p, \quad D = B^{1/10}, \quad y = t_1(B)^{2\exp(z)}.$$

The assumption $0 \leqslant g(p) < 1$ is satisfied here due to $\sigma_p \ll 1/p$ and $p > t_0(B) \to \infty$. Note that for square-free d that is only divisible by primes in $\mathscr{P}$, we have

$$r_d = -g(d)X + \sum_{\substack{m \leqslant \widetilde{x} \\ m \equiv 0 (\mathrm{mod}\ d)}} \widetilde{a}_m.$$

Assuming that

$$\log t_1(B) = o(\log B) \tag{3.7}$$

we see that when $d \leqslant D$, one has

$$dp_1 \cdots p_u \leqslant dt_1(B)^u \leqslant Dt_1(B)^u = B^{\frac{1}{10} + \frac{\log t_1(B)}{\log B}} \leqslant B^{1/6} \tag{3.8}$$

for all large B. This allows us to employ Proposition 3.6 with $Q = dp_1 \cdots p_u$ to obtain

$$\sum_{\substack{m \leqslant \widetilde{x} \\ m \equiv 0 (\mathrm{mod}\ d)}} \widetilde{a}_m = \frac{c_n B^{n+1} X g(d)}{\sharp \widetilde{\Omega}_B} + O\left(\frac{\delta^{u+\omega(d)} B^{n+1}}{\sharp \widetilde{\Omega}_B dp_1^2 p_2 \cdots p_u}\right) = X g(d) + O_u\left(\frac{(\log B)^{[1/n]}}{B} + \frac{\delta^{\omega(d)}}{dp_1^2 \prod_{i=2}^{u} p_i}\right),$$

where we used $\sharp \widetilde{\Omega}_B = c_n B^{n+1} + O(B^n (\log B)^{[1/n]})$ and $X g(d) \ll 1$. The inequality (3.8) shows that

$$\frac{dp_1^2 \prod_{i=2}^{u} p_i}{\delta^{\omega(d)}} \leqslant \left(d \prod_{i=1}^{u} p_i\right)^2 \leqslant B^{1/3} \leqslant \frac{B}{\log B},$$

hence,

$$r_d = -X g(d) + \sum_{m \leqslant \widetilde{x}, m \equiv 0 (\mathrm{mod}\ d)} \widetilde{a}_m \ll \frac{\delta^{\omega(d)}}{dp_1^2 \prod_{i=2}^{u} p_i}.$$

This shows that the error term occurring in (2.16) is

$$\ll \frac{X}{\mathrm{e}^s} \prod_{\substack{p < y \\ p \in \mathscr{P}}} (1 - \sigma_p) + \sum_{\substack{d \leqslant B \\ p|d \Rightarrow p \in \mathscr{P}}} \frac{|\mu(d)| \delta^{\omega(d)}}{dp_1^2 \prod_{i=2}^{u} p_i} \leqslant \frac{X}{\mathrm{e}^s} + \frac{1}{p_1^2 \prod_{i=2}^{u} p_i} \prod_{p \in \mathscr{P}} \left(1 + \frac{1}{p}\right)^{\delta}.$$

The product over $p \in \mathscr{P}$ equals

$$\prod_{i=1}^{u} \prod_{p_i < p \leqslant p_i^{\exp(z)}} \left(1 + \frac{1}{p}\right)^{\delta} \ll \prod_{i=1}^{u} \left(\frac{\log(p_i^{\exp(z)})}{\log p_i}\right)^{\delta} \ll 1.$$

Furthermore, the estimates $p_1 > t_0(B)$ and $X \ll 1/(p_1 \cdots p_u)$ show that

$$\frac{X}{\mathrm{e}^s} + \frac{1}{p_1^2 \prod_{i=2}^{u} p_i} \prod_{p \in \mathscr{P}} \left(1 + \frac{1}{p}\right)^{\delta} \ll \frac{1}{p_1 \cdots p_u} \frac{1}{\min\{\mathrm{e}^s, t_0(B)\}}.$$

The main term occurring in Lemma 2.8 is

$$X \prod_{p \in \mathscr{P}} (1 - \sigma_p) = \prod_{i=1}^{u} \left(\sigma_{p_i} \prod_{p_i < p \leqslant p_i^{\exp(z)}} (1 - \sigma_p)\right),$$

hence, the expectation $E_{x \in \widetilde{\Omega}_B}\left[\widetilde{\delta}_{p_1, z}(x) \cdots \widetilde{\delta}_{p_u, z}(x)\right]$ in the right-hand side of (3.6) equals

$$\prod_{i=1}^{u} \left(\sigma_{p_i} \prod_{p_i < p \leqslant p_i^{\exp(z)}} (1 - \sigma_p)\right) + O\left(\frac{1}{p_1 \cdots p_u} \frac{1}{\min\{\mathrm{e}^s, t_0(B)\}}\right).$$

Injecting this into (3.6) produces the error term

$$\ll_r \frac{1}{\min\{e^s, t_0(B)\}} \sum_{u=1}^{r} \sum_{p_1 < \cdots < p_u \leqslant t_1(B)} \frac{1}{p_1 \cdots p_u} \ll_r \frac{(\log \log t_1(B))^r}{\min\{e^s, t_0(B)\}}.$$

Following arguments similar to the ones in the proof of Lemma 2.8, the main term is

$$\sum_{u=1}^{r} \sum_{\substack{r_1,\ldots,r_u \in \mathbb{N} \\ r_1+\ldots+r_u=r}} \frac{r!}{r_1! \cdots r_u!} \sum_{\substack{t_0(B) < p_1 < \cdots < p_u \leqslant t_1(B) \\ p_{i+1} > p_i^{\exp(z)} \ \forall i}} \prod_{i=1}^{u} \sigma_{p_i} \prod_{p_i < p \leqslant p_i^{\exp(z)}} (1 - \sigma_p) = E[T_B^r],$$

where

$$T_B := \sum_{t_0(B) < p \leqslant t_1(B)} \widetilde{B}_p.$$

We have shown that for all $r \in \mathbb{N}$ one has

$$\left| E_{x \in \widetilde{\Omega}_B}[\widetilde{\omega}_{z,B}(x)^r] - E[T_B^r] \right| \ll_r \frac{(\log \log t_1(B))^r}{\min\{e^s, t_0(B)\}}.$$

Noting that $T_B = \widetilde{S}_{t_1(B)} - \widetilde{S}_{t_0(B)}$ gives

$$E[T_B^r] = E[\widetilde{S}_{t_1(B)}^r] + O_r\left(\max_{0 \leqslant k \leqslant r-1} E[\widetilde{S}_{t_0(B)}^{r-k} \widetilde{S}_{t_1(B)}^k] \right).$$

and the Cauchy–Schwarz inequality shows that

$$E[T_B^r] = E[\widetilde{S}_{t_1(B)}^r] + O_r\left(\max_{0 \leqslant k \leqslant r-1} E[\widetilde{S}_{t_0(B)}^{2(r-k)}]^{1/2} \, E[\widetilde{S}_{t_1(B)}^{2k}]^{1/2} \right).$$

Since $0 \leqslant \widetilde{B}_p \leqslant \widetilde{X}_p$, we infer that $0 \leqslant \widetilde{S}_N \leqslant \sum_{p \leqslant N} \widetilde{X}_p$, hence,

$$E[\widetilde{S}_N^r] \leqslant E\left[\left(\sum_{p \leqslant N} \widetilde{X}_p \right)^r \right] = \sum_{u=1}^{r} \sum_{\substack{r_1,\ldots,r_u \in \mathbb{N} \\ r_1+\ldots+r_u=r}} \frac{r!}{r_1! \cdots r_u!} \sum_{p_1 < \cdots < p_u \leqslant N} \prod_{i=1}^{u} E[\widetilde{X}_p^{r_i}].$$

But $E[\widetilde{X}_p^{r_i}] = E[\widetilde{X}_p] = \sigma_p \ll 1/p$, hence, $E[\widetilde{S}_N^r] \ll (\log \log N)^r$. Hence,

$$\max_{0 \leqslant k \leqslant r-1} E[\widetilde{S}_{t_0(B)}^{2(r-k)}]^{1/2} \, E[\widetilde{S}_{t_1(B)}^{2k}]^{1/2} \ll \max_{0 \leqslant k \leqslant r-1} (\log \log t_0(B))^{r-k} (\log \log t_1(B))^k,$$

which is $\ll (\log \log t_0(B))(\log \log t_1(B))^{r-1}$. Hence,

$$\left| E_{x \in \widetilde{\Omega}_B}[\widetilde{\omega}_{z,B}(x)^r] - E[\widetilde{S}_{t_1(B)}^r] \right| \ll_r \frac{(\log \log t_1(B))^r}{\min\left\{ \frac{\log \log t_1(B)}{\log \log t_0(B)}, e^s, t_0(B) \right\}}.$$

Therefore,

$$\left| E_{\widetilde{\Omega}_B}\left[\left(\frac{\widetilde{\omega}_{z,B}(x) - \widetilde{c}_{t_1(B)}}{\widetilde{s}_{t_1(B)}} \right)^k \right] - E\left[\left(\frac{S_N - \widetilde{c}_{t_1(B)}}{\widetilde{s}_{t_1(B)}} \right)^k \right] \right|$$

is

$$\ll (\widetilde{s}_{t_1(B)})^{-k} \sum_{r=0}^{k} \left| \widetilde{c}_{t_1(B)} \right|^{k-r} \left(E_{\widetilde{\Omega}_B}[\widetilde{\omega}_{z,B}^r] - E[S_{t_1(B)}^r] \right),$$

which, by (3.2)–(3.3) is

$$\ll (\log \log t_1(B))^{-r} \frac{(\log \log t_1(B))^r}{\min\left\{ \frac{\log \log t_1(B)}{\log \log t_0(B)}, e^s, t_0(B) \right\}} \ll \frac{1}{\min\left\{ \frac{\log \log t_1(B)}{\log \log t_0(B)}, e^s, t_0(B) \right\}}.$$

This vanishes asymptotically as long as we assume that

$$\log \log t_0(B) = o(\log \log t_1(B)). \tag{3.9}$$

This is due to (3.7) which implies that

$$s = \frac{\log D}{\log y} = \frac{1}{20 \exp(z)} \frac{\log B}{\log t_1(B)} \to +\infty.$$

We have therefore shown that, subject to (3.4)–(3.7)–(3.9), one has

$$E_{\widetilde{\Omega}_B} \left[\left(\frac{\widetilde{\omega}_{z,B}(x) - \widetilde{c}_{t_1(B)}}{\widetilde{s}_{t_1(B)}} \right)^k \right] \to \mu_k.$$

The concluding arguments follow those in Lemma 2.10, the only difference being dealing with primes $p \leqslant t_0(B)$. Recall from [14, Lemma 3.2, part (2)] that there exists a constant $A > 0$ and a homogeneous $F \in \mathbb{Z}[x_0, \ldots, x_n]$ (both of which depend only on f) with the property that for all primes p and $x \in \mathbb{P}^n(\mathbb{Q})$ with $f^{-1}(x)$ smooth and $f^{-1}(x)(\mathbb{Q}_p) = \emptyset$, one has $p \mid F(x)$. Then

$$0 \leqslant \omega_{f,z}(x) - \widetilde{\omega}_{z,B}(x) \leqslant \sum_{p \leqslant t_0(B)} 1 + \sum_{\substack{p > t_1(B) \\ f^{-1}(x)(\mathbb{Q}_p) = \emptyset}} 1 \leqslant t_0(B) + \sharp\{p \mid F(x) : p > t_1(B)\}.$$

For $z > 1$ and $m \in \mathbb{N}$, we have $\sharp\{p \mid m : p > z\} \leqslant (\log m)/(\log z)$. For $x \in \widetilde{\Omega}_B$, we have $H(x) \leqslant B$, thus, $\log |F(x)| \ll \log B$. In particular,

$$\omega_{f,z}(x) = \widetilde{\omega}_{z,B}(x) + O\left(t_0(B) + \frac{\log B}{\log t_1(B)} \right),$$

where the implied constant is independent of B, z and x. Combined with arguments similar to the ones in Lemma 2.10, we obtain

$$\lim_{B \to \infty} \frac{\max\left\{ \left| (\omega_{f,z}(x) - \Delta(f) e^{-z\Delta(f)} \log \log B) - (\widetilde{\omega}_{z,B}(x) - \widetilde{c}_{t_1(B)}) \right| : x \in \widetilde{\Omega}_B \right\}}{\sqrt{\log \log B}} = 0,$$

as long as

$$t_0(B) = o\left(\sqrt{\log \log B} \right) \quad \text{and} \quad \frac{\log B}{\log t_1(B)} = o\left(\sqrt{\log \log B} \right). \tag{3.10}$$

The proof of Theorem 3.2 concludes by adapting the arguments in Section 2.4 to the current setting. This can be achieved as long as we assume that

$$\frac{\log \log t_1(B)}{\log \log B} \to 1 \quad \text{and} \quad \log \log B - \log \log t_1(B) = o(\sqrt{\log \log B}) \tag{3.11}$$

and it now remains to find functions $t_0(B)$ and $t_1(B)$ that satisfy all assumptions (3.4)–(3.7)–(3.9)–(3.10)–(3.11). This can be done by choosing $t_0(B)$ and $t_1(B)$ so that

$$t_0(B) = \log \log \log B \quad \text{and} \quad \frac{\log t_1(B)}{\log B} = \frac{\log \log \log B}{\sqrt{\log \log B}}.$$

References

[1] P. Billingsley, On the central limit theorem for the prime divisor functions, *Ann. Probab.* **2** (1974), 749–791.

[2] S. Chan, P. Koymans, D. Milovic and C. Pagano, On the negative Pell equation. arXiv:1908.01752.

[3] J.-M. de Koninck and J. Galambos, The intermediate prime divisors of integers, *Proc. Am. Math. Soc.* **101** (1987), 213–216.

[4] R. de la Bretèche and G. Tenenbaum, On the gap distribution of prime factors. arXiv:2107.02055.

[5] P. D. T. A. Elliott, *Probabilistic Number Theory. II*, Grundlehren der Mathematischen Wissenschaften, vol. 240 (Springer-Verlag, New York-Berlin, 1980).

[6] P. Erdős, Some remarks about additive and multiplicative functions, *Bull. Am. Math. Soc.* **52** (1946), 527–537.

[7] P. Erdős and M. Kac, The Gaussian law of errors in the theory of additive number theoretic functions, *Am. J. Math.* **62** (1940), 738–742.

[8] J. B. Friedlander and H. Iwaniec, *Opera de Cribro*, vol. 57 (American Mathematical Society Colloquium Publications, 2010), xx+527.

[9] J. Galambos, Extensions of some extremal properties of prime divisors to Poisson limit theorems. A tribute to Emil Grosswald: number theory and related analysis, *Contemp. Math.* **143** (1993), 363–369.

[10] A. Granville, Prime divisors are Poisson distributed, *Int. J. Number Theory* **3** (2007), 1–18.

[11] A. J. Harper, Two new proofs of the Erdős-Kac theorem, with bound on the rate of convergence, by Stein's method for distributional approximations, *Math. Proc. Cambridge Philos. Soc.* **147** (2009), 95–114.

[12] E. Kowalski and A. Nikeghbali, Mod-Poisson convergence in probability and number theory, *Int. Math. Res. Not. IMRN* **18** (2010), 3549–3587.

[13] D. Loughran and A. Smeets, Fibrations with few rational points, *Geom. Funct. Anal.* **26** (2016), 1449–1482.

[14] D. Loughran and E. Sofos, An Erdős-Kac law for local solubility in families of varieties, *Selecta Mathematica* (to appear), arXiv:1711.08396

[15] A. Rényi and P. Turán, On a theorem of Erdős-Kac, *Acta Arith.* **4** (1958), 71–84.

[16] J.-P. Serre, Spécialisation des éléments de $Br_2(\mathbf{Q}(T_1, \cdots, T_n))$, *C. R. Acad. Sci. Paris Sér. I Math.* **311** (1990), 397–402.

[17] A. N. Skorobogatov, Descent on fibrations over the projective line, *Am. J. Math.* **118** (1996), 905–923.

[18] C. Stein, *Approximate Computation of Expectations*, Institute of Mathematical Statistics Lecture Notes-Monograph Series, vol. 7 (Hayward, CA, 1986).

[19] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic *Mathematics* (Cambridge University Press, Cambridge, 1998).