LIMIT THEOREMS FOR THE LEAST COMMON MULTIPLE OF A RANDOM SET OF INTEGERS

GEROLD ALSMEYER, ZAKHAR KABLUCHKO, AND ALEXANDER MARYNYCH

Abstract. Let L_n be the least common multiple of a random set of integers obtained from $\{1,\dots,n\}$ by retaining each element with probability $\theta \in (0,1)$ independently of the others. We prove that the process $(\log L_{\lfloor nt \rfloor})_{t\in[0,1]}$, after centering and normalization, converges weakly to a certain Gaussian process that is not Brownian motion. Further results include a strong law of large numbers for $\log L_n$ as well as Poisson limit theorems in regimes when θ depends on n in an appropriate way.

1. Introduction and main results

For $n \in \mathbb{N}$, let $[n]$ denote the set $\{1,2,\dots,n\}$. Fixing a number $0<\theta<1$, remove each element in $[n]$ with probability $1-\theta$, independently of all other elements in the set. Denote by A_n the random subset of remaining elements and by $L_n := LCM(A_n)$ their least common multiple. In a recent article, Cilleruelo et al. [4, Thm. 1.1] proved the following weak law of large numbers: As $n\to\infty$,

$$\frac{\log L_n}{n} \overset{\mathbb{P}}{\to} \frac{\theta \log(1/\theta)}{1-\theta}, \tag{1.1}$$

where $\overset{\mathbb{P}}{\to}$ means convergence in probability. The result remains valid in the limiting case $\theta = 1$ when defining the right-hand side of (1.1) as 1 as well, thus

$$\lim_{n\to\infty} \frac{\log LCM([n])}{n} = 1.$$

This is in fact a well-known consequence of the prime number theorem.

On the other hand, the derivation of results beyond (1.1), like a strong law of large numbers or a central limit theorem for $\log L_n$, seem to be open problems to our best knowledge. The purpose of this article is to not only provide limit theorems of this kind for both fixed θ and when θ varies with n, but also prove a functional limit theorem for the stochastic process

$$t \;\mapsto\; L_{\lfloor nt \rfloor}, \quad t \in [0,1]$$

as $n\to\infty$. This latter result will actually be presented first and then yield a central limit theorem for $\log L_n$ as an immediate consequence (Corollary 1.5).

2010 *Mathematics Subject Classification.* Primary: 60F05; secondary 11N37, 60F15. Keywords: random set of integers, least common multiple, law of large numbers, central limit theorem, functional limit theorem, Gaussian process, von Mangoldt function, Chebyshev functions.

GA and ZK were partially supported by the Deutsche Forschungsgemeinschaft (SFB 878), AM was partially supported by the Alexander von Humboldt Foundation.

1.1. **A functional central limit theorem.** In order to state the main result, we define the function

$$g(z) \ := \ \sum_{k\geq 1}\frac{z^k(1-z^k)}{k(k+1)} \tag{1.2}$$

for $|z|<1$. It can be provided in closed form which is done in Remark 1.3 below.

Theorem 1.1. *As $n\to\infty$, the following weak convergence holds true in the Skorokhod space $D[0,1]$ of càdlàg functions endowed with the J_1-topology:*

$$\left(\frac{\log L_{\lfloor nt\rfloor}-\mathbb{E}\log L_{\lfloor nt\rfloor}}{\sqrt{n\log n}}\right)_{t\in[0,1]} \ \overset{J_1}{\Longrightarrow}\ (G(t))_{t\in[0,1]}, \tag{1.3}$$

where $(G(t))_{t\in[0,1]}$ is a centered Gaussian process with covariance function

$$\mathbb{E}[G(t)G(s)] = \sum_{k\geq 1}\left(\frac{t}{k}\wedge\frac{s}{k}\right)p_k - \sum_{k,l\geq 1}\left(\frac{t}{k}\wedge\frac{s}{l}\right)p_kp_l \tag{1.4}$$

for $0\leq s,t\leq 1$, where $p_k:=\theta(1-\theta)^{k-1}$ for $k\in\mathbb{N}$. In particular,

$$\mathrm{Var}[G(t)] = g(1-\theta)t. \tag{1.5}$$

The process $(G(t))_{t\in[0,1]}$ a.s. has continuous paths.

A distributional property as well as a probabilistic representation of the limit process $(G(t))_{t\in[0,1]}$ are given in the subsequent proposition.

Proposition 1.2. *(a) Let $\mathcal{G}_\theta$ be a random variable with geometric distribution on $\mathbb{N}$, viz.*

$$\mathbb{P}\{\mathcal{G}_\theta=k\} \ = \ p_k \ = \ \theta(1-\theta)^{k-1}, \quad k\in\mathbb{N},$$

and $B=(B(t))_{t\in[0,1]}$ an independent standard Brownian motion. Then

$$(G(t)+\mathbb{E}[B(t/\mathcal{G}_\theta)|B])_{t\in[0,1]} \ \overset{d}{=} \ \left(B(t\,\mathbb{E}\mathcal{G}_\theta^{-1})\right)_{t\in[0,1]}, \tag{1.6}$$

where $\mathbb{E}[\cdot|B]$ denotes the conditional expectation and the process $(G(t))_{t\in[0,1]}$ is independent of $(B(t))_{t\in[0,1]},\mathcal{G}_\theta$ on the left-hand side.

(b) If $B_1,B_2,\ldots$ denote independent standard Brownian motions, then

$$(G(t))_{t\in[0,1]} \ \overset{d}{=} \ \left(\sqrt{\theta(1-\theta)}\sum_{i\geq 1}(1-\theta)^{(i-1)/2} \times\left(B_i\left(\frac{t}{i}\right)-\sum_{k\geq i+1}\theta(1-\theta)^{k-i-1}B_i\left(\frac{t}{k}\right)\right)\right)_{t\in[0,1]}. \tag{1.7}$$

Note that

$$\mathbb{E}\mathcal{G}_\theta^{-1} \ = \ \frac{\theta\log(1/\theta)}{1-\theta}.$$

Three realisations of the process G, simulated by using the representation (1.7), are shown in the right panel of Figure 1.

Remark 1.3. As already mentioned, the function g in (1.2) can be found explicitly, namely

$$g(z) \;=\; \frac{z-1}{z^2}\big(\log(1-z) + (1+z)\log(1+z)\big), \quad |z|<1.$$

The graph of the variance $g(1-\theta)$, thus the variance of $G(t)/t^{1/2}$ for any $0 < t \le 1$, is shown in the left panel of Figure 1. Indeed, it follows from (1.2) that $g(z) = h(z) - h(z^2)$, where

$$h(z) \;=\; \sum_{k\ge1}\frac{z^k}{k(k+1)} \;=\; \sum_{k\ge1}\frac{z^k}{k} - \frac{1}{z}\sum_{k\ge1}\frac{z^{k+1}}{k+1} \;=\; 1 + \frac{1-z}{z}\log(1-z)$$

for $|z|<1$.

Remark 1.4. It is known, see [4, Prop. 2.1 and Cor. 2.1], that

$$\begin{aligned}\mathbb{E}\log L_n \;&=\; \theta\sum_{k\ge1}\psi\left(\frac{n}{k}\right)(1-\theta)^k \\ &=\; \frac{n\theta\log(1/\theta)}{1-\theta}\left(1+O\left(\exp\{-C\sqrt{\log(\theta n)}\}\right)\right),\end{aligned} \tag{1.8}$$

where ψ is the second Chebyshev function, see (2.3) below. Assuming the Riemann hypothesis the O-term can be substantially improved to

$$\mathbb{E}\log L_n = \frac{n\theta\log(1/\theta)}{1-\theta}\left(1+O\left(\frac{\log^2 n}{\sqrt{n}}\right)\right), \tag{1.9}$$

see formula (6.2) in [8]. However, even (1.9) does not allow one to replace $\mathbb{E}\log L_{\lfloor nt\rfloor}$ in (1.3) by $nt\theta(1-\theta)^{-1}\log(1/\theta)$.

The following central limit theorem is an immediate consequence of Theorem 1.1.

Corollary 1.5. *As $n\to\infty$,*

$$\frac{\log L_n - \mathbb{E}\log L_n}{\sqrt{n\log n}} \;\overset{d}{\to}\; \sqrt{g(1-\theta)}\,\mathcal{N}(0,1),$$

where $\mathcal{N}(0,1)$ is a standard normal random variable.

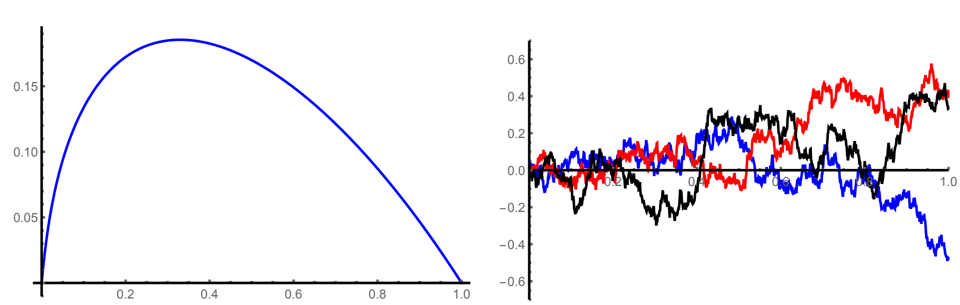

FIGURE 1. The graph of $g(1-\theta)$, $0\le\theta\le1$ (left) and three realisations of the limit Gaussian process $(G(t))_{t\in[0,1]}$ for $\theta=1/2$ (right).

Expansion (1.8) for the mean of $\log L_n$ in combination with an estimate of its variance provided in [4] will also allow us to prove the following strong version of (1.1).

Theorem 1.6. *As* $n \to \infty$,

$$\frac{\log L_n}{n} \ \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} \ \frac{\theta \log(1/\theta)}{1-\theta}. \tag{1.10}$$

1.2. **Poisson limit theorems.** Two further theorems deal with the case when θ varies with n. In the first one it tends to zero at an appropriate speed, namely $\theta = \theta(n) \simeq \frac{\lambda}{n}$ as $n \to \infty$ for some $\lambda > 0$. Since the number of points retained in A_n is binomial with parameters n and θ and thus, for large n, approximately Poissonian with mean λ in the regime just defined, it should not surprise that the limit in the subsequent result is also Poisson. Let $\Pi(\lambda)$ denote a Poisson random variable with mean λ.

Theorem 1.7. *Suppose that, as* $n \to \infty$, $\theta = \theta(n) \simeq \frac{\lambda}{n}$ *for some* $\lambda > 0$. *Then*

$$\frac{\log L_n}{\log n} \ \overset{d}{\to} \ \Pi(\lambda)$$

as $n \to \infty$.

Another, in a sense antipodal regime in which the Poisson distribution appears is when $\theta = \theta(n) \to 1$ at an appropriate speed.

Theorem 1.8. *Suppose that, as* $n \to \infty$, $\theta = \theta(n) = 1 - \frac{\lambda + o(1)}{n} \log n$ *for some* $\lambda > 0$. *Then*

$$\frac{1}{\log n}\left(\sum_{k=1}^{n} \Lambda(k) - \log L_n\right) \ \overset{d}{\to} \ \Pi(\lambda/2)$$

as $n \to \infty$, *where* Λ *denotes the von Mangoldt function defined by formula* (2.1) *below.*

2. Preliminaries

In what follows, we let $\mathcal{P}$ denote the set of prime numbers and $m\mathbb{N}$ the set $\{m, 2m, 3m, \ldots\}$ of integral multiples of $m \in \mathbb{N}$. Recall that the von Mangoldt function $\Lambda : \mathbb{N} \mapsto \mathbb{R}$ is defined as

$$\Lambda(n) = \begin{cases} \log p, & \text{if } n = p^k \text{ for some } k \in \mathbb{N} \text{ and } p \in \mathcal{P}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

We will also use the two Chebyshev functions ϑ and ψ. The first Chebyshev function $\vartheta : \mathbb{R} \mapsto \mathbb{R}$ is defined as

$$\vartheta(x) \ = \sum_{p \in \mathcal{P}:\, p \le x} \log p, \tag{2.2}$$

and the second Chebyshev function $\psi : \mathbb{R} \mapsto \mathbb{R}$ as

$$\psi(x) \ = \ \sum_{k \le x} \Lambda(k). \tag{2.3}$$

Recalling the identity

$$\log LCM([n]) \ = \ \psi(n),$$

we state the following result taken from [4], see Lemma 2.1 therein.

Lemma 2.1. *Let A be an arbitrary set of positive integers and $LCM(A)$ the least common multiple of the elements of A. Then*

$$\log LCM(A) \;=\; \sum_m \Lambda(m) I_A(m),$$

where

$$I_A(m) \;:=\; \begin{cases} 1, & \text{if } A\cap m\mathbb{N} \neq \varnothing, \\ 0, & \text{otherwise.} \end{cases}$$

Proof. Since

$$\log n \;=\; \sum_{p\in\mathcal{P}} \log p \sum_{k\in\mathbb{N}} \mathbb{1}_{\{p^k|n\}} \;=\; \sum_{\substack{p\in\mathcal{P},k\in\mathbb{N}\\ p^k|n}} \log p,$$

we further have

$$\begin{aligned} \log LCM(A) \;&=\; \sum_{\substack{p\in\mathcal{P},k\in\mathbb{N}\\ p^k|LCM(A)}} \log p \\ &=\; \sum_{p\in\mathcal{P},k\in\mathbb{N}} \log p\, I_A(p^k) \;=\; \sum_m \Lambda(m) I_A(m), \end{aligned}$$

where

$$p^k|LCM(A) \quad\Longleftrightarrow\quad A\cap\{p^k,2p^k,3p^k,\ldots,\} \;=\; A\cap p^k\mathbb{N} \;\neq\; \varnothing$$

should be observed for the second equality. □

3. Proof of Theorem 1.1

The proof is divided into four steps. The first two steps provide that $\log L_n$ is well approximated by a sum of independent random variables. The third step will be to check that finite-dimensional distributions of the approximating sum converge to finite-dimensional distributions of the Gaussian process $(G(t))_{t\in[0,1]}$. In the fourth step, this will be improved to give the asserted functional limit theorem.

Step 1. By Lemma 2.1,

$$\begin{aligned} \log L_{\lfloor nt\rfloor} \;&=\; \log LCM(A_{\lfloor nt\rfloor}) \;=\; \sum_m \Lambda(m) I_{A_{\lfloor nt\rfloor}}(m) \\ &=\; \sum_{p\in\mathcal{P}} \log p \sum_{k\geq 1} I_{A_{\lfloor nt\rfloor}}(p^k) \\ &=\; \sum_{p\in\mathcal{P}} \log p\, I_{A_{\lfloor nt\rfloor}}(p) \;+\; \sum_{p\in\mathcal{P}} \log p \sum_{k\geq 2} I_{A_{\lfloor nt\rfloor}}(p^k) \\ &=:\; S_1(\lfloor nt\rfloor) \;+\; S_2(\lfloor nt\rfloor). \end{aligned}$$

We will show first that, as $n\to\infty$,

$$\frac{\sup_{t\in[0,1]} S_2(\lfloor nt\rfloor)}{\sqrt{n\log n}} \;\overset{\mathbb{P}}{\to}\; 0$$

which, by monotonicity of $t\mapsto S_2(\lfloor nt\rfloor)$, is equivalent to

$$\frac{\sum_{p\in\mathcal{P}} \log p \sum_{k\geq 2} I_{A_n}(p^k)}{\sqrt{n\log n}} \;\overset{\mathbb{P}}{\to}\; 0.$$

By Markov's inequality, it suffices to verify

$$\frac{\sum_{p\in\mathcal{P}}\log p\sum_{k\geq 2}\mathbb{E}I_{A_n}(p^k)}{\sqrt{n\log n}}\ \to\ 0 \tag{3.1}$$

as $n\to\infty$. To this end, use Boole's inequality to obtain

$$\mathbb{E}I_{A_n}(p^k)\ =\ \mathbb{P}\{A_n\cap p^k\mathbb{N}\neq\varnothing\}\ \leq\ \left(\sum_{m\leq n/p^k}\mathbb{P}\{mp^k\in A_n\}\right)\wedge 1\ \leq\ \frac{n\theta}{p^k}\wedge 1.$$

Fix $k\geq 2$ and write

$$\begin{aligned}\sum_{p\in\mathcal{P}}\log p\,\mathbb{E}I_{A_n}(p^k)\ &=\ \sum_{p\in\mathcal{P}:p^k\leq n}\log p\,\mathbb{E}I_{A_n}(p^k)\\ &\leq\ \sum_{p\in\mathcal{P}:p^k\leq n}\log p\left(\frac{n\theta}{p^k}\wedge 1\right)\\ &= n\theta\sum_{p\in\mathcal{P}:(n\theta)^{1/k}<p\leq n^{1/k}}\frac{\log p}{p^k}\ +\ \sum_{p\in\mathcal{P}:p\leq(n\theta)^{1/k}}\log p\\ &\leq n\theta\sum_{p\in\mathcal{P}:p>(n\theta)^{1/k}}\frac{\log p}{p^k}\ +\ \sum_{p\in\mathcal{P}:p\leq(n\theta)^{1/k}}\log p.\end{aligned}$$

For the first term in the previous line, Lemma 7.1 in the Appendix provides the upper bound $Cn^{1/k}$ for all $n\geq 1$ and some $C>0$. For the second sum we use the bound

$$\sum_{p\in\mathcal{P}:p\leq(n\theta)^{1/k}}\log p\ \leq\ Cn^{1/k}$$

for all n and some C which follows from $\sum_{p\in\mathcal{P}:p\leq x}\log p\simeq x$ as $x\to\infty$, an equivalent form of the prime number theorem. In both estimates, the constant C does not depend on k. Summarizing, we arrive at the inequality

$$\sum_{p\in\mathcal{P}}\log p\,\mathbb{E}I_{A_n}(p^k)\ \leq\ Cn^{1/k} \tag{3.2}$$

for all $n\geq 1$ and some positive constant C. Returning to (3.1) and noting that summands in the numerator are nonzero only for $k\leq\log_2 n$, (3.2) implies

$$\begin{aligned}\sum_{p\in\mathcal{P}}\log p\sum_{k\geq 2}\mathbb{E}I_{A_n}(p^k)\ &\leq\ C\sum_{k=2}^{\lceil\log_2 n\rceil}n^{1/k}\\ &\leq\ C(\sqrt{n}+n^{1/3}\log_2 n)=o(\sqrt{n\log n}),\end{aligned}$$

as $n\to\infty$, and this proves (3.1).

Step 2. We start with the decomposition

$$S_1(\lfloor nt\rfloor)\ =\ \sum_{p\in\mathcal{P}}\log p\,I_{A_{\lfloor nt\rfloor}}(p)\ =\ S_1^{(1,n)}(t)\ +\ S_1^{(2,n)}(t),$$

where

$$S_1^{(1,n)}(t)\ :=\ \sum_{p\in\mathcal{P}:p\leq\sqrt{n}}\log p\,I_{A_{\lfloor nt\rfloor}}(p)$$

and

$$S_1^{(2,n)}(t) \ := \sum_{p\in\mathcal{P}:p>\sqrt{n}} \log p\, I_{A_{\lfloor nt\rfloor}}(p).$$

For the first sum, we then proceed as follows. Using the prime number theorem,

$$\sup_{t\in[0,1]} S_1^{(1,n)}(t) \ \le \sum_{p\in\mathcal{P}:p\le\sqrt{n}} \log p \ \simeq \ \sqrt{n}$$

and therefore

$$\frac{\sup_{t\in[0,1]} S_1^{(1,n)}(t)}{\sqrt{n\log n}} \ \overset{\mathbb{P}}{\to} \ 0$$

as well as

$$\frac{\mathbb{E}S_1^{(1,n)}(1)}{\sqrt{n\log n}} \ \to \ 0$$

as $n\to\infty$.

In view of what has been shown so far, it remains to prove the asserted limit theorem for $(S_1^{(2,n)}(t))_{t\in[0,1]}$, i.e.

$$\left(\frac{S_1^{(2,n)}(t)-\mathbb{E}S_1^{(2,n)}(t)}{\sqrt{n\log n}}\right)_{t\in[0,1]} \ \overset{J_1}{\Longrightarrow} \ (G(t))_{t\in[0,1]},$$

which is possible because the processes $(I_{A_{\lfloor nt\rfloor}}(p))_{t\in[0,1]}$ and $(I_{A_{\lfloor nt\rfloor}}(q))_{t\in[0,1]}$ are independent for distinct primes $p,q>\sqrt{n}$. For the latter, just observe that the sets $p\mathbb{N}\cap[n]$ and $q\mathbb{N}\cap[n]$ are disjoint for such p,q.

Step 3. Our aim is to show that, as $n\to\infty$,

$$\left(\frac{\sum_{p\in\mathcal{P}:p>\sqrt{n}}\log p\left(I_{A_{\lfloor nt\rfloor}}(p)-\mathbb{E}I_{A_{\lfloor nt\rfloor}}(p)\right)}{\sqrt{n\log n}}\right)_{t\in[0,1]} \ \overset{f.d.d.}{\Longrightarrow} \ (G(t))_{t\in[0,1]}. \tag{3.3}$$

First we show convergence of the covariances. For $0<s\le t\le 1$, we have

$$\begin{aligned}
&\mathrm{Cov}[S_1^{(2,n)}(t),S_1^{(2,n)}(s)]\\
&\quad=\sum_{p\in\mathcal{P}\cap(\sqrt{n},nt]}\ \sum_{q\in\mathcal{P}\cap(\sqrt{n},ns]}\log p\log q\ \mathrm{Cov}[I_{A_{\lfloor nt\rfloor}}(p),I_{A_{\lfloor ns\rfloor}}(q)]\\
&\quad=\sum_{p\in\mathcal{P}\cap(\sqrt{n},ns]}\log^2 p\ \mathrm{Cov}[I_{A_{\lfloor nt\rfloor}}(p),I_{A_{\lfloor ns\rfloor}}(p)]\\
&\quad=\sum_{p\in\mathcal{P}\cap(\sqrt{n},ns]}\log^2 p\,\mathbb{E}I_{A_{\lfloor ns\rfloor}}(p)\big(1-\mathbb{E}I_{A_{\lfloor nt\rfloor}}(p)\big)\\
&\quad=\sum_{p\in\mathcal{P}\cap(\sqrt{n},ns]}\log^2 p\left(1-(1-\theta)^{\lfloor ns/p\rfloor}\right)(1-\theta)^{\lfloor nt/p\rfloor},
\end{aligned}$$

where the independence of $(I_{A_{\lfloor nt\rfloor}}(p))_{t\in[0,1]}$ and $(I_{A_{\lfloor nt\rfloor}}(q))_{t\in[0,1]}$ enters when passing to the second equality. By invoking Lemma 7.3 in the Appendix, we infer for

$0 < s \le t \le 1$

$$\begin{aligned}\lim_{n\to\infty} \mathrm{Cov}\left[\frac{S_1^{(2,n)}(t)}{\sqrt{n\log n}}, \frac{S_1^{(2,n)}(s)}{\sqrt{n\log n}}\right] &= \sum_{j\ge 1}(1-(1-\theta)^j) \sum_{i\,\in\,\left(\frac{tj}{s}-1,\, \frac{tj}{s}+\frac{t}{s}\right)} (1-\theta)^i \left(\frac{t}{i}\wedge\frac{s}{j} - \frac{t}{i+1}\vee\frac{s}{j+1}\right)\\ &=: C_1(t,s).\end{aligned}$$

In order to prove formula (1.4) for $C_1(t,s)$, write the latter in the form

$$\begin{aligned}C_1(t,s) &= \sum_{i,j\ge 1}(1-(1-\theta)^j)(1-\theta)^i\left(\frac{t}{i}\wedge\frac{s}{j} - \frac{t}{i+1}\vee\frac{s}{j+1}\right)^+\\ &= \sum_{i,j\ge 1}\left(\sum_{k=1}^{j} p_k\right)\left(\sum_{l\ge i+1} p_l\right)\left(\frac{t}{i}\wedge\frac{s}{j} - \frac{t}{i+1}\vee\frac{s}{j+1}\right)^+\\ &= \sum_{k\ge 1}\sum_{l\ge 2} p_k p_l \sum_{j\ge k}\sum_{i=1}^{l-1}\left(\frac{t}{i}\wedge\frac{s}{j} - \frac{t}{i+1}\vee\frac{s}{j+1}\right)^+.\end{aligned}$$

We claim that the inner double sum equals $(s/k - t/l)^+$. Consider two intervals $(t/l, t]$ and $(0, s/k]$. Cover the first interval by the disjoint subintervals $(t/(i+1), t/i]$, $i = 1, \ldots, l-1$ and, analogously, the second interval by the disjoint subintervals $(s/(j+1), s/j]$, $j \ge k$. Then $\left(\frac{t}{i}\wedge\frac{s}{j} - \frac{t}{i+1}\vee\frac{s}{j+1}\right)^+$ equals the length of the intersection of $(t/(i+1), t/i]$ and $(s/(j+1), s/j]$ and is zero if they are disjoint. The total sum of these lengths equals the length of the intersection of the original intervals $(t/l, t]$ and $(0, s/k]$, thus $(s/k - t/l)^+$. Consequently,

$$C_1(t,s) = \sum_{k\ge 1}\sum_{l\ge 2} p_k p_l\left(\frac{s}{k}-\frac{t}{l}\right)^+ = \sum_{k\ge 1}\sum_{l\ge 1} p_k p_l\left(\frac{s}{k}-\frac{t}{l}\right)^+,$$

where the second equality holds because $(s/k - t)^+ = 0$. Let η_1, η_2 be two independent geometric random variables on $\mathbb{N}$, viz.

$$\mathbb{P}\{\eta_1 = k\} = \mathbb{P}\{\eta_2 = k\} = \theta(1-\theta)^{k-1}, \quad k\in\mathbb{N}.$$

Then

$$\begin{aligned}C_1(t,s) &= \mathbb{E}\left(\frac{s}{\eta_1}-\frac{t}{\eta_2}\right)^+ = \mathbb{E}\left(\frac{s}{\eta_1}-\frac{s}{\eta_1}\wedge\frac{t}{\eta_2}\right)\\ &= \sum_{k\ge 1}\frac{s}{k}p_k - \sum_{k,l\ge 1}\left(\frac{s}{k}\wedge\frac{t}{l}\right)p_k p_l\end{aligned}$$

which is the asserted result as $s < t$.

To complete the proof of (3.3), it remains to verify the Lindeberg condition

$$\lim_{n\to\infty}\sum_{p\in\mathcal{P}\cap(\sqrt{n},nt]}\mathbb{E}\left[|V_{n,p}(t)|^2\mathbb{1}_{\{|V_{n,p}(t)|>\varepsilon\}}\right] = 0. \tag{3.4}$$

for any $t \in [0,1]$ and $\varepsilon > 0$, where

$$V_{n,p}(t) \;:=\; \frac{\log p \left(I_{A_{\lfloor nt \rfloor}}(p) - \mathbb{E} I_{A_{\lfloor nt \rfloor}}(p)\right)}{\sqrt{n \log n}}$$

for $p \in \mathcal{P} \cap (\sqrt{n}, n]$. But this is obvious because $|V_{n,p}(t)| \le \sqrt{\frac{\log n}{n}}$ for all such p and $t \in [0,1]$, $n \in \mathbb{N}$.

Step 4. In order to finally show the functional limit theorem, we will apply Theorem 10.6 from [7] that provides general conditions for the convergence of triangular arrays of row-wise independent processes to a Gaussian limit. Actually, this theorem yields convergence in the sense of convergence in the space of bounded functions with the usual supremum-norm, which is stronger. Convergence in $D[0,1]$ with the J_1-topology follows as a direct consequence.

To conform with the notation in [7], put

$$f_{n,p}(t) := \frac{\log p \, I_{A_{\lfloor nt \rfloor}}(p)}{\sqrt{n \log n}} \quad \text{and} \quad F_{n,p} := \frac{\log p \, I_{A_n}(p)}{\sqrt{n \log n}}$$

for $p \in \mathcal{P} \cap (\sqrt{n}, n]$. Conditions (ii) and (iv) of Theorem 10.6 in [7] were checked in Step 3. Condition (iii) is obvious. The manageability of the family $(f_{n,p}(\cdot))_p$ (Condition (i) of Theorem 10.6 in [7]) follows from the monotonicity of $(f_{n,p}(t))_{t \in [0,1]}$ in t for every fixed n and p, and the observation in the paragraph just before Theorem 11.17 in [6, p. 221]. It remains to verify condition (v). To this end, introduce the function

$$\rho_n(s,t) \;:=\; \left(\frac{\sum_{p \in \mathcal{P} \cap (\sqrt{n}, n]} \log^2 p \; \mathbb{E}|I_{A_{\lfloor nt \rfloor}}(p) - I_{A_{\lfloor ns \rfloor}}(p)|^2}{n \log n} \right)^{1/2}$$

for $0 \le s, t \le 1$. Note that

$$\begin{aligned}
\mathbb{E}|I_{A_{\lfloor nt \rfloor}}(p) - I_{A_{\lfloor ns \rfloor}}(p)|^2 \;&=\; \mathbb{P}\{I_{A_{\lfloor nt \rfloor}}(p) - I_{A_{\lfloor ns \rfloor}}(p) = 1\} \\
&=\; 1 - \mathbb{P}\{I_{A_{\lfloor nt \rfloor}}(p) - I_{A_{\lfloor ns \rfloor}}(p) = 0\} \\
&=\; (1-\theta)^{\lfloor ns/p \rfloor} - (1-\theta)^{\lfloor nt/p \rfloor}.
\end{aligned}$$

and therefore

$$\rho_n(s,t) := \left(\frac{\sum_{p \in \mathcal{P} \cap (\sqrt{n}, n]} \log^2 p \left((1-\theta)^{\lfloor ns/p \rfloor} - (1-\theta)^{\lfloor nt/p \rfloor}\right)}{n \log n} \right)^{1/2}$$

for $0 < s \le t \le 1$. Decomposing the numerator on the right-hand side as

$$\begin{aligned}
&\sum_{p \in \mathcal{P} \cap (\sqrt{n}, n]} \log^2 p \; \left((1-\theta)^{\lfloor ns/p \rfloor} - (1-\theta)^{\lfloor nt/p \rfloor}\right) \\
&\qquad = \sum_{p \in \mathcal{P} \cap (\sqrt{n}, ns]} \log^2 p \, (1-\theta)^{\lfloor ns/p \rfloor} \;+\; \sum_{p \in \mathcal{P} \cap (ns, n]} \log^2 p \\
&\qquad\quad - \sum_{p \in \mathcal{P} \cap (\sqrt{n}, nt]} \log^2 p \, (1-\theta)^{\lfloor nt/p \rfloor} \;-\; \sum_{p \in \mathcal{P} \cap (nt, n]} \log^2 p,
\end{aligned}$$

and applying Lemma 7.2 in conjunction with formula (7.2) in the Appendix, we deduce

$$\lim_{n \to \infty} \rho_n(s,t) = \sqrt{(1 - h(1-\theta))(t-s)}$$

for $0 < s \leq t \leq 1$. Now let $(s_n)_{n\geq 1}$ and $(t_n)_{n\geq 1}$ be two deterministic sequences in $[0,1]$ such that $s_n - t_n \to 0$ as $n \to \infty$. We must show that

$$\lim_{n\to\infty} \rho_n(s_n, t_n) = 0,$$

or, equivalently,

$$\lim_{n\to\infty} \frac{\sum_{p\in\mathcal{P}\cap(\sqrt{n},n]} \log^2 p \left|(1-x)^{\lfloor ns_n/p\rfloor} - (1-x)^{\lfloor nt_n/p\rfloor}\right|}{n\log n} \ = \ 0.$$

Putting $u_n := s_n \wedge t_n$ and $v_n := s_n \vee t_n$, this follows if

$$\lim_{n\to\infty} \frac{\sum_{p\in\mathcal{P}:p\leq n} \log p \left((1-\theta)^{\lfloor nu_n/p\rfloor} - (1-\theta)^{\lfloor nv_n/p\rfloor}\right)}{n} \ = \ 0. \tag{3.5}$$

Using Lemma 7.4, we find that, for a suitable constant $C > 0$,

$$\begin{aligned}
&\frac{1}{n}\sum_{p\in\mathcal{P}:p\leq n} \log p \left((1-\theta)^{\lfloor nu_n/p\rfloor} - (1-\theta)^{\lfloor nv_n/p\rfloor}\right) \\
&\qquad \leq \ (v_n - u_n)h(1-\theta) \ + \ \frac{Cu_n}{\log(nu_n+2)} \ + \ \frac{Cv_n}{\log(nv_n+2)} \\
&\qquad \leq \ (v_n - u_n)h(1-\theta) \ + \ \frac{2C}{\log(n+2)},
\end{aligned}$$

and the last line converges to 0 because $v_n - u_n = |s_n - t_n| \to 0$, as $n \to \infty$.

It remains to note that Theorem 10.6 in [7] guarantees that the limit process a.s. has uniformly continuous paths. This completes the proof of Theorem 1.1. □

4. Proof of Proposition 1.2

(a) Since $(G(t))_{t\in[0,1]}$ defined in Theorem 1.1 is Gaussian, the same holds true for the process $(\mathbb{E}[B(t/\mathcal{G}_\theta)|B])_{t\in[0,1]}$ as one can readily see from the representation

$$\begin{aligned}
\mathbb{E}[B(t/\mathcal{G}_\theta)|B] \ &= \ \sum_{k\geq 1} p_k B(t/k) \ = \sum_{k\geq 1} p_k \int_0^\infty \mathbb{1}_{\{z\leq t/k\}} \ \mathrm{d}B(z) \\
&= \ \int_0^\infty \left(\sum_{k\geq 1} p_k \mathbb{1}_{\{z\leq t/k\}}\right) \ \mathrm{d}B(z).
\end{aligned}$$

By the independence assumption, it is enough to check the equality of covariances which follows immediately from the identities

$$\begin{aligned}
\mathrm{Cov}[\mathbb{E}[B(t/\mathcal{G}_\theta)|B], \mathbb{E}[B(s/\mathcal{G}_\theta)|B]] \ &= \ \mathrm{Cov}\left[\sum_{k\geq 1} p_k B(t/k), \sum_{l\geq 1} p_l B(s/l)\right] \\
&= \ \sum_{k,l\geq 1} p_k p_l \left(\frac{t}{k} \wedge \frac{s}{l}\right)
\end{aligned}$$

and

$$\mathrm{Cov}[B(t(\mathbb{E}\mathcal{G}_\theta^{-1})), B(s(\mathbb{E}\mathcal{G}_\theta^{-1}))] \ = \ (t\wedge s)\mathbb{E}\mathcal{G}_\theta^{-1} \ = \ \sum_{k\geq 1}\left(\frac{s}{k}\wedge\frac{t}{k}\right) p_k$$

for $0 \leq s, t \leq 1$.

(b) Since the series on the right-hand side of (1.7) is a centered Gaussian process, it suffices again to check the equality of covariances. We have

$$\sqrt{\theta(1-\theta)}\sum_{i\geq 1}(1-\theta)^{(i-1)/2}\left(B_i\left(\frac{t}{i}\right)-\sum_{k\geq i+1}\theta(1-\theta)^{k-i-1}B_i\left(\frac{t}{k}\right)\right) =: \sum_{i\geq 1}\sum_{k\geq i}a_{ik}B_i\left(\frac{t}{k}\right),$$

where

$$a_{ik} := \begin{cases} \theta^{1/2}(1-\theta)^{i/2}, & \text{if } k=i,\\ -\theta^{3/2}(1-\theta)^{k-i/2-1}, & \text{if } k>i. \end{cases}$$

Using independence of $B_1, B_2, \dots$ and the formula $\mathrm{Cov}[B_i(s), B_i(t)] = s\wedge t$, we obtain for its covariance function

$$\begin{aligned}\rho(s,t) &:= \mathrm{Cov}\left[\sum_{i\geq 1}\sum_{k\geq i}a_{ik}B_i\left(\frac{t}{k}\right), \sum_{j\geq 1}\sum_{l\geq j}a_{jl}B_j\left(\frac{s}{l}\right)\right]\\ &= \sum_{i\geq 1}\sum_{k\geq i}\sum_{l\geq i}a_{ik}a_{il}\left(\frac{t}{k}\wedge\frac{s}{l}\right) = \sum_{k\geq 1}\sum_{l\geq 1}\left(\frac{t}{k}\wedge\frac{s}{l}\right)\sum_{i=1}^{k\wedge l}a_{ik}a_{il}.\end{aligned}$$

If $k=l$, then

$$\begin{aligned}\sum_{i=1}^{k\wedge l}a_{ik}a_{il} &= \sum_{i=1}^{k}a_{ik}^2 = \theta(1-\theta)^k+\theta^3\sum_{i=1}^{k-1}(1-\theta)^{2k-i-2}\\ &= \theta(1-\theta)^k+\theta^2(1-\theta)^{k-1}(1-(1-\theta)^{k-1})\\ &= p_k-p_k^2 = p_k-p_kp_l,\end{aligned}$$

while for $k<l$

$$\begin{aligned}\sum_{i=1}^{k\wedge l}a_{ik}a_{il} &= a_{kk}a_{kl}+\sum_{i=1}^{k-1}a_{ik}a_{il}\\ &= -\theta^2(1-\theta)^{l-1}+\theta^3\sum_{i=1}^{k-1}(1-\theta)^{k+l-i-2} = -p_kp_l.\end{aligned}$$

A combination of these results yields

$$\rho(s,t) = \sum_{k\geq 1}\frac{s\wedge t}{k}p_k - \sum_{k,l\geq 1}\left(\frac{t}{k}\wedge\frac{s}{l}\right)p_kp_l,$$

which shows the desired equality of distributions and completes the proof of Proposition 1.2. □

5. Proof of Theorem 1.6

For $\varepsilon>0$ and $n\in\mathbb{N}$, we define the events

$$A_n(\varepsilon) := \{|\log L_n - \mathbb{E}\log L_n| > \varepsilon\,\mathbb{E}\log L_n\}.$$

Proposition 2.2 from [4] provides us with

$$\mathrm{Var}[\log L_n] = O(n\log^2 n) \tag{5.1}$$

as $n \to \infty$ which in combination with the expansion of $\mathbb{E}\log L_n$ in (1.8) implies that

$$\mathbb{P}\{A_n(\varepsilon)\} \;\le\; \frac{C\log^2(n+1)}{n} \tag{5.2}$$

for all $n \ge 1$ and some constant $C = C(\varepsilon, \theta) > 0$. Putting $n_k := [k\log^4 k]$ for $k \ge 1$, it follows from (5.2) that $\sum_{k\ge 2}\mathbb{P}\{A_{n_k}(\varepsilon)\} < \infty$ for any $\varepsilon > 0$ and thus

$$\lim_{k\to\infty}\frac{\log L_{n_k}}{\mathbb{E}\log L_{n_k}} \;=\; 1 \quad \text{a.s.} \tag{5.3}$$

by the Borel-Cantelli lemma. For arbitrary $n \in \mathbb{N}$, let $k(n)$ be such that $n_{k(n)} \le n < n_{k(n)+1}$ and notice that, as a trivial consequence of (1.8),

$$\lim_{k\to\infty}\frac{\mathbb{E}\log L_{n_{k+1}}}{\mathbb{E}\log L_{n_k}} \;=\; 1.$$

The proof of the theorem is now completed by a combination of the latter fact with (5.3) and the inequalities

$$\frac{\log L_{n_k}}{\mathbb{E}\log L_{n_k}}\cdot\frac{\mathbb{E}\log L_{n_k}}{\mathbb{E}\log L_{n_{k+1}}} \;\le\; \frac{\log L_n}{\mathbb{E}\log L_n} \;\le\; \frac{\log L_{n_{k+1}}}{\mathbb{E}\log L_{n_{k+1}}}\cdot\frac{\mathbb{E}\log L_{n_{k+1}}}{\mathbb{E}\log L_{n_k}},$$

valid for any $n \ge 1$. □

6. Proof of Theorems 1.7 and 1.8

6.1. **Proof of Theorem 1.7.** Let $R(n)$ be the number of remaining integers in $[n]$, i.e. $R(n) := |A_n|$, and note that $R(n)$ has a binomial distribution with parameters n and $\theta(n)$. Since $\theta(n) \simeq \lambda/n$, the classical Poisson limit theorem gives

$$R(n) \;\stackrel{d}{\to}\; \Pi(\lambda)$$

as $n \to \infty$. Let $(X_1^{(n)}, X_2^{(n)}, \ldots, X_{R(n)}^{(n)})$ denote the ordered sample of remaining integers which, conditioned upon $R(n) = k$, has the same distribution as an ordered k-sample without replacement from the set $[n]$. In order to show that, as $n \to \infty$,

$$\frac{\log LCM(X_1^{(n)}, X_2^{(n)}, \ldots, X_{R(n)}^{(n)})}{\log n} \;\stackrel{d}{\to}\; \Pi(\lambda),$$

it is enough to show that, conditioned upon $R(n) = k$ for any fixed $k \in \mathbb{N}$,

$$\frac{\log LCM(X_1^{(n)}, X_2^{(n)}, \ldots, X_k^{(n)})}{\log n} \;\stackrel{\mathbb{P}}{\to}\; k \tag{6.1}$$

as $n \to \infty$. Given any finite set of positive integers $\{n_1, n_2, \ldots, n_k\}$, we have that

$$\frac{n_1 n_2 \cdots n_k}{\prod_{1\le i<j\le k} GCD(n_i, n_j)} \;\le\; LCM(n_1, n_2, \ldots, n_k) \;\le\; n_1 n_2 \cdots n_k.$$

For (6.1), it hence suffices to verify that, given $R(n) = k$,

$$\frac{\sum_{i=1}^k \log X_i^{(n)}}{\log n} \;\stackrel{\mathbb{P}}{\to}\; k \tag{6.2}$$

and

$$\frac{\sum_{1\le i<j\le k}\log GCD(X_i^{(n)}, X_j^{(n)})}{\log n} \;\stackrel{\mathbb{P}}{\to}\; 0. \tag{6.3}$$

Let $(U_k^{(n)})_{k\in\mathbb{N}}$ be a sequence of i.i.d. random variables with a uniform distribution on $[n]$. Then, as already stated above, the conditional law of $(X_1^{(n)},\dots,X_k^{(n)})$ given $R(n)=k$ for any fixed $k\in\mathbb{N}$, is the same as the conditional law of $(U_{(1)}^{(n)},\dots,U_{(k)}^{(n)})$, the order statistics of $(U_1^{(n)},\dots,U_k^{(n)})$, given the event

$$A_{n,k} \;:=\; \left\{U_i^{(n)}\neq U_j^{(n)} : i,j=1,\dots,k, i\neq j,\right\}.$$

Since $\mathbb{P}\{A_{n,k}\}$ tends to 1 for any $k\in\mathbb{N}$ and $n\to\infty$, (6.2) and (6.3) are equivalent to

$$\frac{\sum_{i=1}^k \log U_i^{(n)}}{\log n} \;\overset{\mathbb{P}}{\to}\; k \tag{6.4}$$

and

$$\frac{\log GCD(U_1^{(n)},U_2^{(n)})}{\log n} \overset{\mathbb{P}}{\to} 0, \tag{6.5}$$

respectively. Assertion (6.4) follows directly from

$$\lim_{n\to\infty}\mathbb{P}\left\{1-\varepsilon<\frac{\log U_1^{(n)}}{\log n}\le 1\right\} \;=\; \lim_{n\to\infty}\mathbb{P}\{n^{1-\varepsilon}<U_1^{(n)}\le n\} \;=\; 1$$

for any $\varepsilon\in(0,1)$ and Slutsky's lemma.

For (6.5), we will in fact prove the stronger result that, as $n\to\infty$,

$$\log GCD(U_1^{(n)},U_2^{(n)}) \;\overset{d}{\to}\; \xi \tag{6.6}$$

for some proper nondegenerate random variable ξ to be defined below. Writing

$$U_1^{(n)} \;=\; \prod_{p\in\mathcal{P}} p^{\lambda_p(U_1^{(n)})} \quad\text{and}\quad U_2^{(n)} \;=\; \prod_{p\in\mathcal{P}} p^{\lambda_p(U_2^{(n)})},$$

where $\lambda_p(m)\ge 0$ is the power of prime p in the prime decomposition of $m\in\mathbb{N}$, we have

$$\log GCD(U_1^{(n)},U_2^{(n)}) \;=\; \sum_{p\in\mathcal{P}}\Big(\lambda_p(U_1^{(n)})\wedge\lambda_p(U_2^{(n)})\Big)\log p.$$

It is a simple fact, see for example the last display on p. 28 in [2], that

$$\Big(\lambda_p(U_1^{(n)})\Big)_{p\in\mathcal{P}} \;\overset{d}{\to}\; (Z_{1,p})_{p\in\mathcal{P}}\,, \tag{6.7}$$

where $(Z_{1,p})_{p\in\mathcal{P}}$ forms a sequence of independent random variables and $Z_{1,p}$ has a geometric distribution on $\{0,1,2,\dots,\}$ with parameter $1-\frac{1}{p}$. Likewise,

$$\Big(\lambda_p(U_2^{(n)})\Big)_{p\in\mathcal{P}} \;\overset{d}{\to}\; (Z_{2,p})_{p\in\mathcal{P}}\,,$$

where $(Z_{1,p})_{p\in\mathcal{P}}\overset{d}{=}(Z_{2,p})_{p\in\mathcal{P}}$, $(Z_{1,p})_{p\in\mathcal{P}}$ and $(Z_{2,p})_{p\in\mathcal{P}}$ are independent. The series

$$\xi \;:=\; \sum_{p\in\mathcal{P}}(Z_{1,p}\wedge Z_{2,p})\log p$$

converges a.s. because it has finite mean, viz.

$$\sum_{p\in\mathcal{P}}\mathbb{E}(Z_{1,p}\wedge Z_{2,p})\log p \;=\; \sum_{p\in\mathcal{P}}\frac{\log p}{p^2-1}<\infty. \tag{6.8}$$

We note in passing that the explicit form of the distribution of ξ may be found in [5]. According to Theorem 3.2 in [3], a sufficient condition for (6.6) is that

$$\lim_{m\to\infty}\limsup_{n\to\infty}\mathbb{P}\left\{\sum_{p\in\mathcal{P}:p\geq m}\Big(\lambda_p(U_1^{(n)})\wedge\lambda_p(U_2^{(n)})\Big)\log p\geq\varepsilon\right\} = 0 \tag{6.9}$$

for any $\varepsilon>0$. We will show the in fact stronger condition (by Markov's inequality)

$$\lim_{m\to\infty}\limsup_{n\to\infty}\sum_{p\in\mathcal{P}:p\geq m}\mathbb{E}\Big(\lambda_p(U_1^{(n)})\wedge\lambda_p(U_2^{(n)})\Big)\log p = 0. \tag{6.10}$$

To this end, note that

$$\begin{aligned}\mathbb{E}\Big(\lambda_p(U_1^{(n)})\wedge\lambda_p(U_2^{(n)})\Big) &= \sum_{i\geq 1}\mathbb{P}\{\lambda_p(U_1^{(n)})\geq i,\lambda_p(U_2^{(n)})\geq i\}\\ &= \sum_{i\geq 1}\left(\frac{1}{n}\left\lfloor\frac{n}{p^i}\right\rfloor\right)^2 \leq \sum_{i\geq 1}\frac{1}{p^{2i}} = \frac{1}{p^2-1}.\end{aligned}$$

Relation (6.10) now follows from (6.8), thus completing the proof of Theorem 1.7. □

6.2. **Proof of Theorem 1.8.** Using Lemma 2.1, we can write

$$\sum_{k=1}^{n}\Lambda(k)-\log L_n = \sum_{k=1}^{n}\Lambda(k)\,(1-I_{A_n}(k))$$

and infer

$$\sum_{k\leq n/2}\Lambda(k)\,(1-I_{A_n}(k)) \xrightarrow{\mathbb{P}} 0 \quad (n\to\infty)$$

from

$$\begin{aligned}\mathbb{P}\left\{\sum_{k\leq n/2}\Lambda(k)\,(1-I_{A_n}(k))>0\right\} &\leq \mathbb{P}\{I_{A_n}(k)=0 \text{ for some } k\leq n/2\}\\ &\leq \sum_{k\leq n/2}\mathbb{P}\{I_{A_n}(k)=0\} \leq \sum_{k\leq n/2}\mathbb{P}\{k\notin A_n, 2k\notin A_n\}\\ &\leq \frac{n}{2}\,\frac{(\lambda+o(1))^2\log^2 n}{n^2}.\end{aligned}$$

Left with the sum $\sum_{n/2<k\leq n}\Lambda(k)\,(1-I_{A_n}(k))$, we note that the random variables $\{1-I_{A_n}(k):n/2<k\leq n\}$ are independent indicators satisfying

$$\mathbb{P}\{1-I_{A_n}(k)=1\} = \mathbb{P}\{k\notin A_n\} = 1-\theta(n) \simeq \frac{\lambda\log n}{n}$$

as $n\to\infty$. By definition of the von Mangoldt function Λ, we have

$$\begin{aligned}\sum_{n/2<k\leq n}\Lambda(k)\,(1-I_{A_n}(k)) &= \sum_{p\in\mathcal{P}}\log p\,(1-I_{A_n}(p))\,\mathbb{1}_{\{n/2<p\leq n\}}\\ &\quad+\sum_{p\in\mathcal{P}}\sum_{l\geq 2}\log p\,\big(1-I_{A_n}(p^l)\big)\,\mathbb{1}_{\{n/2<p^l\leq n\}}.\end{aligned}$$

The expectation of the last term on the right-hand side equals

$$\begin{aligned}(1-\theta(n))\sum_{p\in\mathcal{P}}\sum_{l\geq 2}\log p\mathbb{1}_{\{n/2<p^l\leq n\}} &= \frac{(\lambda+o(1))\log n}{n}\sum_{p\in\mathcal{P}}\sum_{l\geq 2}\log p\,\mathbb{1}_{\{n/2<p^l\leq n\}}\\ &\leq \frac{(\lambda+o(1))\log n}{n}\sum_{p\in\mathcal{P}}\sum_{l\geq 2}\log p\,\mathbb{1}_{\{p^l\leq n\}}\\ &= \frac{(\lambda+o(1))\log n}{n}(\psi(n)-\vartheta(n)),\end{aligned}$$

where ψ and ϑ are the Chebyshev functions, see Formulae (2.2) and (2.3). By Theorem 4.1 in [1],

$$\psi(n)-\vartheta(n) \;=\; O(\sqrt{n}\log^2 n)$$

as $n\to\infty$, and this in combination with Markov's inequality implies

$$\sum_{p\in\mathcal{P}}\sum_{l\geq 2}\log p\left(1-I_{A_n}(p^l)\right)\mathbb{1}_{\{n/2<p^l\leq n\}} \overset{\mathbb{P}}{\to} 0$$

as $n\to\infty$. It remains to show that

$$\frac{\sum_{p\in\mathcal{P}}\log p\,(1-I_{A_n}(p))\,\mathbb{1}_{\{n/2<p\leq n\}}}{\log n} \overset{d}{\to} \Pi(\lambda/2).$$

In view of the obvious inequalities

$$\begin{aligned}\log(n/2)\sum_{p\in\mathcal{P}}(1-I_{A_n}(p))\,\mathbb{1}_{\{n/2<p\leq n\}} &\leq \sum_{p\in\mathcal{P}}\log p\,(1-I_{A_n}(p))\,\mathbb{1}_{\{n/2<p\leq n\}}\\ &\leq \log n\sum_{p\in\mathcal{P}}(1-I_{A_n}(p))\,\mathbb{1}_{\{n/2<p\leq n\}},\end{aligned}$$

the claim of the theorem follows from the observation that

$$\sum_{p\in\mathcal{P}\cap(n/2,n]}(1-I_{A_n}(p)) \;=\; \sum_{p\in\mathcal{P}\cap(n/2,n]}\mathbb{1}_{\{p\notin A_n\}} \overset{d}{\to} \Pi(\lambda/2),\quad n\to\infty.$$

The latter convergence holds by the classic Poisson limit theorem for independent indicators, here Bernoulli variables with parameter $1-\theta(n)$. The factor $1/2$ in the parameter of the Poisson random variable appears because, with $\pi(x)$ denoting the number of primes $\leq x$, the number of summands is $\pi(n)-\pi(n/2)\sim\frac{n}{2\log n}$, as $n\to\infty$ by the prime number theorem. The proof of Theorem 1.8 is complete. □

7. Appendix

We have used the following estimate for the tails of convergent series involving primes.

Lemma 7.1. *There exists a positive constant C such that, for all $n\in\mathbb{N}$ and $k\geq 2$,*

$$\sum_{p\in\mathcal{P}:p\geq n}\frac{\log p}{p^k} \;\leq\; \frac{C}{n^{k-1}}. \tag{7.1}$$

Proof. For any $n \geq 2$ and with $\pi(x)$ as above, integration by parts yields

$$\begin{aligned}\sum_{p\in\mathcal{P}:p\geq n}\frac{\log p}{p^k} &= \int_{[n,\infty)}\frac{\log x}{x^k}\,\mathrm{d}\pi(x)\\ &= \frac{\log x}{x^k}\pi(x)\Big|_n^\infty + \int_n^\infty \pi(x)\frac{kx^{k-1}\log x - x^{k-1}}{x^{2k}}\,\mathrm{d}x\\ &\leq \int_n^\infty \pi(x)\frac{kx^{k-1}\log x}{x^{2k}}\,\mathrm{d}x.\end{aligned}$$

By the prime number theorem, $\pi(x) \leq C_1 x/\log x$ for some constant $C_1 > 0$ and all $x \geq 2$. Consequently,

$$\sum_{p\in\mathcal{P}:p\geq n}\frac{\log p}{p^k} \leq C_1 k\int_n^\infty \frac{\mathrm{d}x}{x^k} = \frac{C_1 k}{k-1}n^{1-k} \leq \frac{2C_1}{n^{k-1}}$$

and thus (7.1) holds with $C := 2C_1$. □

Lemma 7.2. *For any fixed* $x, t \in (0,1)$,

$$\sum_{p\in\mathcal{P}:p\in(\sqrt{n},nt]}\log^2 p\,(1-x)^{\lfloor nt/p\rfloor} \simeq t\cdot n\log n\cdot h(1-x),$$

where $h(x) = \sum_{k\geq 1}\frac{x^k}{k(k+1)}$.

Proof. Let us first show that

$$\sum_{p\in\mathcal{P}:p\leq x}\log^2 p \simeq x\log x. \tag{7.2}$$

as $x \to \infty$. To this end, note that

$$\sum_{p\leq x,p\in\mathcal{P}}\log^2 p = \int_{[2,x]}\log^2 z\,\mathrm{d}\pi(z) = \pi(z)\log^2 z\Big|_2^x - \int_2^x \frac{2\pi(z)\log z}{z}\,\mathrm{d}z.$$

By another appeal to the prime number theorem, $\pi(x)\log^2 x \simeq x\log x$ and the integrand is bounded, i.e., the integral itself is $O(x)$ as $x \to \infty$. This proves (7.2) which in turn further provides us with the relation

$$\sum_{p\in\mathcal{P}\cap\,(nt/(k+1),nt/k]}\log^2 p \simeq \frac{nt\log n}{k(k+1)} \quad (n\to\infty) \tag{7.3}$$

for any $k \in \mathbb{N}$ and $t \in (0,1]$.

Now fix an arbitrary $m \in \mathbb{N}$ and write

$$\begin{aligned}\sum_{p\in\mathcal{P}\cap(\sqrt{n},nt]}(1-x)^{\lfloor nt/p\rfloor}\log^2 p &= \sum_{k=1}^{\lfloor\sqrt{n}t\rfloor}(1-x)^k\sum_{p\in\mathcal{P}\cap\,(nt/(k+1),nt/k]}\log^2 p\\ &= \sum_{k=1}^{m}(1-x)^k\sum_{p\in\mathcal{P}\cap\,(nt/(k+1),nt/k]}\log^2 p\\ &\quad + \sum_{k=m+1}^{\lfloor\sqrt{n}t\rfloor}(1-x)^k\sum_{p\in\mathcal{P}\cap\,(nt/(k+1),nt/k]}\log^2 p\\ &=: A_1(n,m) + A_2(n,m).\end{aligned}$$

By (7.3), we have

$$\lim_{n\to\infty} \frac{A_1(n,m)}{n\log n} \;=\; t\sum_{k=1}^{m} \frac{(1-x)^k}{k(k+1)},$$

and for $A_2(n,m)$, the estimate

$$\begin{aligned} A_2(n,m) \;&\leq\; \sum_{k\geq m+1} (1-x)^k \sum_{p\in\mathcal{P}:p\leq n} \log^2 p \\ &\leq\; Cn\log n \sum_{k\geq m+1} (1-x)^k \;=\; Cx^{-1}(1-x)^{m+1} n\log n \end{aligned}$$

for some $C > 0$ follows as a consequence of (7.2). Hence,

$$\limsup_{n\to\infty} \frac{A_2(n,m)}{n\log n} \leq Cx^{-1}(1-x)^{m+1}.$$

By combining these facts, we obtain

$$\begin{aligned} t\sum_{k=1}^{m} \frac{(1-x)^k}{k(k+1)} \;&\leq\; \liminf_{n\to\infty} \frac{\sum_{p\in\mathcal{P}\cap(\sqrt{n},nt]}(1-x)^{\lfloor nt/p\rfloor}\log^2 p}{n\log n} \\ \leq\; \limsup_{n\to\infty} \frac{\sum_{p\in\mathcal{P}\cap(\sqrt{n},nt]}(1-x)^{\lfloor nt/p\rfloor}\log^2 p}{n\log n} \;&\leq\; t\sum_{k=1}^{m} \frac{(1-x)^k}{k(k+1)} + Cx^{-1}(1-x)^{m+1} \end{aligned}$$

for any fixed $m \in \mathbb{N}$. Sending $m \to \infty$ yields the assertion and completes the proof. □

The next lemma forms an extension of the previous one and an important ingredient in the proof of our main theorem (convergence of covariances).

Lemma 7.3. *For any fixed $x \in (0,1)$ and $0 < s \leq t \leq 1$,*

$$\begin{aligned} &\frac{1}{n\log n} \sum_{p\in\mathcal{P}\cap(\sqrt{n},ns]} (1-x)^{\lfloor nt/p\rfloor} \left(1-(1-x)^{\lfloor ns/p\rfloor}\right) \log^2 p \\ &\qquad \simeq\; \sum_{j\geq 1} (1-(1-x)^j) \sum_{i\,\in\,\left(\frac{tj}{s}-1,\,\frac{tj}{s}+\frac{t}{s}\right)} (1-x)^i \left(\frac{t}{i}\wedge\frac{s}{j} - \frac{t}{i+1}\vee\frac{s}{j+1}\right), \end{aligned}$$

as $n \to \infty$.

Proof. We start by noting the equivalence

$$\lfloor nt/p\rfloor = i \quad\text{and}\quad \lfloor ns/p\rfloor = j \quad\Longleftrightarrow\quad p \in \left(\frac{nt}{i+1}\vee\frac{ns}{j+1}, \frac{nt}{i}\wedge\frac{ns}{j}\right],$$

where the interval on the right can be empty. Fix $j \in \mathbb{N}$ and let us find all integers i such that

$$\frac{t}{i+1}\vee\frac{s}{j+1} < \frac{t}{i}\wedge\frac{s}{j}. \tag{7.4}$$

If $t/(i+1) < s/(j+1)$ or, equivalently, $i > t(j+1)/s - 1$, then (7.4) holds iff $i < t(j+1)/s$. If $t/(i+1) \geq s/(j+1)$ or, equivalently, $i \leq t(j+1)/s - 1$, then (7.4) holds iff $i > tj/s - 1$. Therefore, for any fixed $j \in \mathbb{N}$, (7.4) holds iff

$$\frac{tj}{s} - 1 < i < \frac{t(j+1)}{s},$$

and this implies

$$\sum_{p\in\mathcal{P}\cap(\sqrt{n},ns]} \log^2 p(1-x)^{\lfloor nt/p\rfloor}\left(1-(1-x)^{\lfloor ns/p\rfloor}\right)$$
$$= \sum_{j=1}^{\lfloor\sqrt{n}s\rfloor}(1-(1-x)^j)\sum_{i\in\left(\frac{tj}{s}-1,\frac{tj}{s}+\frac{t}{s}\right)}(1-x)^i\sum_{p\in\mathcal{P}\cap\left(\frac{nt}{i+1}\vee\frac{ns}{j+1},\frac{nt}{i}\wedge\frac{ns}{j}\right]}\log^2 p.$$

It remains to note that, for any fixed $j\in\mathbb{N}$ and any integer $i\in\left(\frac{tj}{s}-1,\frac{tj}{s}+\frac{t}{s}\right)$,

$$\sum_{p\in\mathcal{P}\cap\left(\frac{nt}{i+1}\vee\frac{ns}{j+1},\frac{nt}{i}\wedge\frac{ns}{j}\right]}\log^2 p \;\simeq\; \left(\frac{t}{i}\wedge\frac{s}{j}-\frac{t}{i+1}\vee\frac{s}{j+1}\right)n\log n,$$

as $n\to\infty$. Now arguing in the same way as in the last part of the proof of Lemma 7.2 and noting that the series

$$\sum_{j\geq 1}(1-(1-x)^j)\sum_{i\in\left(\frac{tj}{s}-1,\frac{tj}{s}+\frac{t}{s}\right)}(1-x)^i\left(\frac{t}{i}\wedge\frac{s}{j}-\frac{t}{i+1}\vee\frac{s}{j+1}\right)$$

converges, we obtain the assertion of the lemma. □

Lemma 7.4. *For any $x\in(0,1)$, there exists $C=C(x)>0$ such that*

$$\left|\sum_{p\in\mathcal{P}:p\leq t}(1-x)^{\lfloor t/p\rfloor}\log p\;-\;th(1-x)\right|\;\leq\;\frac{Ct}{\log(t+2)}, \tag{7.5}$$

for all $t\geq 0$, where h is as defined in Lemma 7.2.

Proof. We may assume without loss of generality that $t\geq t_0$ for a fixed t_0. Recalling that ϑ denotes the first Chebyshev function defined by (2.2), we can write

$$\sum_{p\in\mathcal{P}:p\leq t}\log p\,(1-x)^{\lfloor t/p\rfloor}\;=\;\sum_{i=1}^{\lfloor t\rfloor}(1-x)^i\sum_{p\in\mathcal{P}\cap(t/(i+1),t/i]}\log p$$
$$=\sum_{i=1}^{\lfloor t\rfloor}(1-x)^i\left(\vartheta\left(\frac{t}{i}\right)-\vartheta\left(\frac{t}{i+1}\right)\right).$$

Using the well-known inequality

$$|\vartheta(z)-z|\;\leq\;\frac{C_1 z}{\log(z+2)},$$

valid for all $z\geq 0$ and some $C_1>0$, we obtain

$$\left|\sum_{p\in\mathcal{P}:p\leq t}(1-x)^{\lfloor t/p\rfloor}\log p\;-\;t\sum_{i=1}^{\lfloor t\rfloor}\frac{(1-x)^i}{i(i+1)}\right|$$
$$\leq\;\frac{2C_1 t}{\log(t+2)}\sum_{i=1}^{\lfloor t\rfloor}(1-x)^i\;\leq\;\frac{2(1-x)C_1 t}{x\log(t+2)}.$$

Finally, since $h(1-x)-\sum_{i=1}^{\lfloor t\rfloor}\frac{(1-x)^i}{i(i+1)}=\sum_{i\geq\lfloor t\rfloor+1}\frac{(1-x)^i}{i(i+1)}$ satisfies the inequality

$$t\sum_{i\geq\lfloor t\rfloor+1}\frac{(1-x)^i}{i(i+1)}\ \leq\ \sum_{i\geq\lfloor t\rfloor+1}\frac{1}{i(i+1)}\ \leq\ \frac{1}{\lfloor t\rfloor+1},$$

we infer (7.5) with $C=C(x)=\frac{2(1-x)C_1}{x}+1$ for all sufficiently large t. □

References

1. Tom M. Apostol, *Introduction to analytic number theory*, Springer-Verlag, New York-Heidelberg, 1976, Undergraduate Texts in Mathematics. MR 0434929
2. Richard Arratia, A. D. Barbour, and Simon Tavaré, *Logarithmic combinatorial structures: a probabilistic approach*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zürich, 2003. MR 2032426 (2004m:60004)
3. Patrick Billingsley, *Convergence of probability measures*, 2^{nd} ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, 1999, A Wiley-Interscience Publication. MR 1700749 (2000e:60008)
4. Javier Cilleruelo, Juanjo Rué, Paulius Šarka, and Ana Zumalacárregui, *The least common multiple of random sets of positive integers*, J. Number Theory **144** (2014), 92–104. MR 3239153
5. Persi Diaconis and Paul Erdös, *On the distribution of the greatest common divisor*, A festschrift for Herman Rubin, IMS Lecture Notes Monogr. Ser., vol. 45, Inst. Math. Statist., Beachwood, OH, 2004, pp. 56–61. MR 2126886
6. Michael R. Kosorok, *Introduction to empirical processes and semiparametric inference*, Springer Series in Statistics, Springer, New York, 2008. MR 2724368
7. David Pollard, *Empirical processes: theory and applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 2, Institute of Mathematical Statistics, Hayward, CA; American Statistical Association, Alexandria, VA, 1990. MR 1089429
8. Lowell Schoenfeld, *Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$. II*, Math. Comp. **30** (1976), no. 134, 337–360, Corrigendum in 136, 900. MR 0457374

Inst. Math. Stochastics, Department of Mathematics and Computer Science, University of Münster, Orléans-Ring 10, D-48149, Münster, Germany
E-mail address: gerolda@uni-muenster.de

Inst. Math. Stochastics, Department of Mathematics and Computer Science, University of Münster, Orléans-Ring 10, D-48149, Münster, Germany
E-mail address: zakhar.kabluchko@uni-muenster.de

Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, 01601 Kyiv, Ukraine
E-mail address: marynych@unicyb.kiev.ua