

A BROWNIAN WEAK LIMIT FOR THE LEAST COMMON MULTIPLE OF A RANDOM M -TUPLE OF INTEGERS

DARIUSZ BURACZEWSKI, ALEXANDER IKSANOV, AND ALEXANDER MARYNYCH

ABSTRACT. Let $B_n(m)$ be a set picked uniformly at random among all m -elements subsets of $\{1, 2, \dots, n\}$. We provide a pathwise construction of the collection $(B_n(m))_{1 \leq m \leq n}$ and prove that the logarithm of the least common multiple of the integers in $(B_n(\lfloor mt \rfloor))_{t \geq 0}$, properly centered and normalized, converges to a Brownian motion when both m, n tend to infinity. Our approach consists of two steps. First, we show that the aforementioned result is a consequence of a multidimensional central limit theorem for the logarithm of the least common multiple of m independent random variables having uniform distribution on $\{1, 2, \dots, n\}$. Second, we offer a novel approximation of the least common multiple of a random sample by the product of the elements of the sample with neglected multiplicities in their prime decompositions.

1. INTRODUCTION

The analysis of divisibility properties of *random integers* is a classical problem in the probabilistic number theory going back to pioneer works by Dirichlet [13] and Cesàro [9, 10, 11]. Among other results, Cesàro in [11] has proved that the expected least common multiple (lcm) of two integers picked uniformly at random from the set $[n] := \{1, 2, \dots, n\}$ is asymptotically, as $n \rightarrow \infty$, equal to their product multiplied by the constant $\zeta(3)/\zeta(2)$, where ζ is the Riemann zeta function. In the modern probabilistic language, this result can be stated as

$$(1) \quad \lim_{n \rightarrow \infty} \mathbb{E} \left(\frac{\text{lcm}(U_1^{(n)}, U_2^{(n)})}{U_1^{(n)} U_2^{(n)}} \right) = \lim_{n \rightarrow \infty} \mathbb{E} \left(\frac{1}{\text{gcd}(U_1^{(n)}, U_2^{(n)})} \right) = \frac{\zeta(3)}{\zeta(2)},$$

where $U_1^{(n)}, U_2^{(n)}$ are independent copies of a random variable $U^{(n)}$ with distribution

$$(2) \quad \mathbb{P}\{U^{(n)} = k\} = 1/n, \quad k \in [n],$$

and gcd denotes the greatest common divisor.

There are various ways to generalize (1) and some of them have received attention in probabilistic as well as number theoretic literature. For example, one can ask

2010 *Mathematics Subject Classification*. Primary: 11K65, 60F05; Secondary: 11A05.

Key words and phrases. Brownian motion, central limit theorem, coupon collector's problem, least common multiple.

about the asymptotic distribution (instead of the asymptotic average) of sequences of random variables

$$\left(\frac{\text{lcm}(U_1^{(n)}, U_2^{(n)}, \dots, U_m^{(n)})}{U_1^{(n)} U_2^{(n)} \dots U_m^{(n)}} \right)_{n \in \mathbb{N}} \quad \text{or} \quad \left(\frac{\text{lcm}(U_1^{(n)}, U_2^{(n)}, \dots, U_m^{(n)})}{n^m} \right)_{n \in \mathbb{N}}$$

where $m \in \mathbb{N}$ is a fixed integer. This problem has been solved in [8], see also [16, 18], by showing that both fractions converge in distribution to proper random variables, see Eq. (15) and (16) in [8] and also Proposition 3.3 below. Another way towards generalization of (1) is to replace the random set $\{U_1^{(n)}, U_2^{(n)}, \dots, U_m^{(n)}\}$ by a more sophisticated (and/or with cardinality depending on n) random subset of $[n]$. In [12] the following model has been proposed: fix $\theta \in (0, 1)$, remove every element $j \in [n]$, independently of the other elements, with probability $1 - \theta$ and denote the remaining subset by $A_n(\theta)$. This model has been intensively analyzed in [1], see also [21], where the authors proved various limit theorems, including a strong law of large numbers for $L_n(\theta) := \log \text{lcm}(A_n(\theta))$, functional limit theorems for the process $(L_{\lfloor nt \rfloor}(\theta))_{t \in [0, 1]}$ and Poisson limit theorems when $\theta = \theta_n$ is either close to zero or one.

Another class of examples is related to the theory of random permutations. Let \mathfrak{S}_n be the symmetric group of all permutations of $[n]$. If $\sigma \in \mathfrak{S}_n$ is a random permutation picked according to some probability distribution on \mathfrak{S}_n , then the collection of cycle lengths of σ is a random subset of $[n]$ and its least common multiple is equal to the order $\text{ord}(\sigma)$ of the permutation σ . This object has been studied for Ewens' permutations (including the case of the uniform distribution on \mathfrak{S}_n), see [4], and also for more general probability measures on \mathfrak{S}_n , see [17, 22]. For example, the famous Erdős–Turán law [15] states that $\log \text{ord}(\sigma)$ is asymptotically normal, as $n \rightarrow \infty$, if σ is picked uniformly at random from \mathfrak{S}_n .

In the present paper we investigate another model of choosing a random subset of $[n]$. Let $B_n(m)$ be a subset chosen uniformly at random among all subsets of $[n]$ containing exactly m elements. To the best of our knowledge, the asymptotic behavior of $\log \text{lcm}(B_n(m))$ has only been (partly) investigated in [12]. Theorem 1.2 therein states that for $m = m_n < n$,

$$\frac{(1 - m/n) \log \text{lcm}(B_n(m))}{m \log(n/m)} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 1,$$

where $\xrightarrow[n \rightarrow \infty]{\mathbb{P}}$ denotes convergence in probability. The purpose of our work is to shed more light on this model. We provide here an explicit pathwise construction of the whole random collection $(B_n(m))_{1 \leq m \leq n}$ and then prove new limit theorems for the

properly normalized sequence of stochastic processes

$$\left(\log \operatorname{lcm}(B_n(\lfloor tm \rfloor)) \right)_{t \geq 0}$$

assuming that $m = o(n)$ and $m, n \rightarrow \infty$ (Theorem 3.5). As an important ingredient of our approach we also derive limit theorems for accompanying random processes

$$\left(\log \operatorname{lcm}(U_1^{(n)}, \dots, U_{\lfloor mt \rfloor}^{(n)}) \right)_{t \geq 0},$$

where $(U_j^{(n)})_{j \geq 1}$ is a sequence of independent copies of a random variable $U^{(n)}$ with distribution (2), and both $m, n \rightarrow \infty$ (Theorem 3.6). The case of fixed m follows essentially from the aforementioned results proved in [8], however the case when $m_n \rightarrow \infty$ requires a completely different and novel approach, whilst existing techniques fail for this problem.

The paper has the following structure. In Section 2 we provide an explicit construction of the collection $(B_n(m))_{1 \leq m \leq n}$ and relate it to the classical coupon collector problem. Section 2 culminates with formula (4), which defines all $B_n(m)$, $1 \leq m \leq n$ on the same probability space via the accompanying sequence $(U_j^{(n)})_{j \geq 1}$ and an appropriate sequence of stopping times $(\tau^{(n)}(m))_{m \geq 1}$. The main results are presented in Subsections 3.1 (Proposition 3.3 and Theorem 3.4) and 3.2 (Theorems 3.5 and 3.6). Our strategy of the proof of the main results is discussed in Subsection 3.3. In Section 4 some elementary facts about the sequence $(\tau^{(n)}(m))_{m \geq 1}$, used later in the proofs, are collected. The proofs of Theorems 3.5 and 3.6 are scattered across Sections 5, 6 and 7. The proof of Proposition 3.3 is given in the Appendix.

2. EXPLICIT CONSTRUCTION OF $(B_n(m))_{1 \leq m \leq n}$

Let $(U_j^{(n)})_{j \in \mathbb{N}}$ be a sequence of independent copies of a random variable $U^{(n)}$ with the uniform distribution on $[n]$, that is, the distribution given by (2). For $n \in \mathbb{N}$, $1 \leq m \leq n$, put

$$\tau^{(n)}(m) := \inf \{ j \in \mathbb{N} : \text{there are exactly } m \text{ different values among } U_1^{(n)}, U_2^{(n)}, \dots, U_j^{(n)} \}.$$

Note that $\mathbb{P}\{m \leq \tau^{(n)}(m) < \infty\} = 1$. The variable $\tau^{(n)}(m)$ can be interpreted in terms of the well-known coupon collector problem. Assume that there are infinitely many coupons and that each coupon has one of n different types. Then $\tau^{(n)}(m)$ is the number of coupons a collector needs to buy in order to have coupons of exactly m different types. In particular, $\tau^{(n)}(n)$ is the number of purchases needed to acquire the entire collection. We put $\tau^{(n)}(0) := 0$ and $\tau^{(n)}(m) := \tau^{(n)}(n)$ for $m > n$.

The following representation is immediate. For $i = 1, \dots, n$, let $X_{i,n}$ be the number of purchases needed to buy a coupon of a new type given that the number of different types in the current collection is $i - 1$. It is straightforward to check that the variables $(X_{i,n})_{i=1, \dots, n}$ are mutually independent and

$$\mathbb{P}\{X_{i,n} = j\} = \left(\frac{i-1}{n}\right)^{j-1} \left(1 - \frac{i-1}{n}\right), \quad j \in \mathbb{N}.$$

Furthermore,

$$(3) \quad \tau^{(n)}(m) = X_{1,n} + X_{2,n} + \dots + X_{m,n}, \quad 1 \leq m \leq n.$$

We have the following result.

Lemma 2.1. *For every fixed $1 \leq m \leq n$, the random set*

$$\{U_1^{(n)}, U_2^{(n)}, \dots, U_{\tau^{(n)}(m)}^{(n)}\} = \{U_{\tau^{(n)}(1)}^{(n)}, U_{\tau^{(n)}(2)}^{(n)}, \dots, U_{\tau^{(n)}(m)}^{(n)}\}$$

is uniformly distributed among all subsets of $[n]$ containing exactly m elements.

Proof. It is enough to show that

$$\mathbb{P}\{\{U_{\tau^{(n)}(1)}^{(n)}, U_{\tau^{(n)}(2)}^{(n)}, \dots, U_{\tau^{(n)}(m)}^{(n)}\} = \{b_1, b_2, \dots, b_m\}\} = 1/\binom{n}{m}, \quad 1 \leq m \leq n.$$

for every fixed set $\{b_1, b_2, \dots, b_m\} \subset [n]$ of pairwise distinct integers. We argue by induction on m . For $m = 1$, the claim is obvious because $\tau^{(n)}(1) = 1$ by definition.

We have, for $2 \leq m \leq n$,

$$\begin{aligned} \mathbb{P}\{\{U_{\tau^{(n)}(1)}^{(n)}, U_{\tau^{(n)}(2)}^{(n)}, \dots, U_{\tau^{(n)}(m)}^{(n)}\} = \{b_1, b_2, \dots, b_m\}\} \\ = \sum_{j=1}^m \mathbb{P}\{U_{\tau^{(n)}(m)}^{(n)} = b_j | U_{\tau^{(n)}(1)}^{(n)}, U_{\tau^{(n)}(2)}^{(n)}, \dots, U_{\tau^{(n)}(m-1)}^{(n)} = \{b_1, b_2, \dots, b_m\} \setminus \{b_j\}\} \\ \times \mathbb{P}\{U_{\tau^{(n)}(1)}^{(n)}, U_{\tau^{(n)}(2)}^{(n)}, \dots, U_{\tau^{(n)}(m-1)}^{(n)} = \{b_1, b_2, \dots, b_m\} \setminus \{b_j\}\}. \end{aligned}$$

The first probability under the sum is equal to $(n - m + 1)^{-1}$ and the second probability, by the induction assumption, is $\binom{n}{m-1}^{-1}$. Therefore,

$$\begin{aligned} \mathbb{P}\{\{U_{\tau^{(n)}(1)}^{(n)}, U_{\tau^{(n)}(2)}^{(n)}, \dots, U_{\tau^{(n)}(m)}^{(n)}\} = \{b_1, b_2, \dots, b_m\}\} \\ = \sum_{j=1}^m \left((n - m + 1) \binom{n}{m-1}\right)^{-1} = m \left((n - m + 1) \binom{n}{m-1}\right)^{-1} = 1/\binom{n}{m}, \end{aligned}$$

as wanted. \square

The explicit construction in Lemma 2.1 allows us to construct the whole collection $(B_n(m))_{1 \leq m \leq n}$ in a consistent way from the sequence $(U_j^{(n)})_{j \in \mathbb{N}}$. Thus, from now on we *redefine* the sets $B_n(m)$ by putting

$$(4) \quad B_n(m) := \{U_1^{(n)}, U_2^{(n)}, \dots, U_{\tau^{(n)}(m)}^{(n)}\}, \quad 1 \leq m \leq n.$$

Now we can speak about the distribution of $(B_n(i), B_n(j))$, $i \neq j$ and more generally about finite-dimensional distributions of the collection $(B_n(m))_{1 \leq m \leq n}$.

3. LIMIT THEOREMS FOR THE LEAST COMMON MULTIPLE OF $B_n(m)$

Denote by \mathcal{P} the set of prime numbers. Also, let $\lambda_p(n)$ denote the multiplicity of a prime number $p \in \mathcal{P}$ in the unique decomposition of $n \in \mathbb{N}$ into prime factors, that is,

$$n = \prod_{p \in \mathcal{P}} p^{\lambda_p(n)}.$$

In what follows we tacitly assume that all products and sums with indices p, q, r, s only extend over prime numbers. We also stipulate that “const” is a constant whose value is of no importance and may change from one appearance to another. Also, all unspecified limit relations are assumed to hold as $n \rightarrow \infty$.

3.1. The case of fixed m . Let $((\mathcal{G}_k(2), \mathcal{G}_k(3), \mathcal{G}_k(5), \dots))_{k \in \mathbb{N}}$ be a sequence of independent copies of an infinite vector $(\mathcal{G}(2), \mathcal{G}(3), \mathcal{G}(5), \dots)$ with mutually independent coordinates having a geometric distribution

$$(5) \quad \mathbb{P}\{\mathcal{G}(p) \geq j\} = p^{-j}, \quad j \in \mathbb{N}_0, \quad p \in \mathcal{P}.$$

The importance of these geometric variables stems from the following lemma which has a long history, see, for instance, [20, Formulas (2.5)-(2.7)] and [5], and is presented here in the form borrowed from [8].

The distribution given in (2) is a discrete uniform distribution. We recall that there also exists a continuous uniform distribution μ , say on $[0, 1]$ defined by $\mu(dx) = \mathbb{1}_{(0,1)}(x)dx$. We shall write $\xrightarrow[n \rightarrow \infty]{d, \infty}$ and $\xrightarrow[n \rightarrow \infty]{d}$ to denote convergence in distribution in \mathbb{R}^∞ endowed with the product topology and in \mathbb{R} , respectively.

Lemma 3.1. *Let*

$$U^{(n)} = \prod_{p \in \mathcal{P}} p^{\lambda_p(U^{(n)})}$$

be the decomposition of $U^{(n)}$ with distribution (2) into prime factors. Then

(i)

$$(\lambda_p(U^{(n)}))_{p \in \mathcal{P}} \xrightarrow[n \rightarrow \infty]{d, \infty} (\mathcal{G}(p))_{p \in \mathcal{P}};$$

(ii)

$$(n^{-1}U^{(n)}, (\lambda_p(U^{(n)}))_{p \in \mathcal{P}}) \xrightarrow[n \rightarrow \infty]{d, \infty} (U, (\mathcal{G}(p))_{p \in \mathcal{P}}),$$

with U being uniformly distributed on $[0, 1]$ and independent of $(\mathcal{G}(p))_{p \in \mathcal{P}}$;

(iii) for $p, q \in \mathcal{P}$, $p \neq q$ and $k_p, k_q \in \mathbb{N}_0$,

$$\mathbb{P}\{\lambda_p(U^{(n)}) = k_p, \lambda_q(U^{(n)}) = k_q\} = (1 - p^{-1})(1 - q^{-1})p^{-k_p}q^{-k_q} + O(n^{-1}),$$

where the constant in the O -term does not depend on (p, q, k_p, k_q) .

With the help of this lemma the following result has been proved in [8]. See also [16] for the cases $m = 2, 3$.

Proposition 3.2 (Formula (16) in [8]). *For every fixed $m \in \mathbb{N}$,*

$$\begin{aligned} & \log \operatorname{lcm}(U_1^{(n)}, U_2^{(n)}, \dots, U_m^{(n)}) - m \log n \\ & \xrightarrow[n \rightarrow \infty]{d} \sum_{j=1}^m \log U_j + \sum_p \log p \cdot \left(\max_{1 \leq k \leq m} \mathcal{G}_k(p) - \sum_{k=1}^m \mathcal{G}_k(p) \right), \end{aligned}$$

where $(U_j)_{j=1, \dots, m}$ are independent random variables with the uniform distribution on $[0, 1]$ which are also independent of $(\mathcal{G}_k(p))_{k \in \mathbb{N}, p \in \mathcal{P}}$.

Using the same techniques as in [8] Proposition 3.2 can be strengthened as follows.

Proposition 3.3.

$$\begin{aligned} & \left(\log \operatorname{lcm}(U_1^{(n)}, U_2^{(n)}, \dots, U_m^{(n)}) - m \log n \right)_{m \in \mathbb{N}} \\ & \xrightarrow[n \rightarrow \infty]{d, \infty} \left(\sum_{j=1}^m \log U_j + \sum_p \log p \cdot \left(\max_{1 \leq k \leq m} \mathcal{G}_k(p) - \sum_{k=1}^m \mathcal{G}_k(p) \right) \right)_{m \in \mathbb{N}}, \end{aligned}$$

where $(U_j)_{j \geq 1}$ are independent random variables with the uniform distribution on $[0, 1]$ which are also independent of $(\mathcal{G}_k(p))_{k \in \mathbb{N}, p \in \mathcal{P}}$.

We shall give a short proof of Proposition 3.3 in the Appendix. Since, for every fixed m ,

$$\lim_{n \rightarrow \infty} \mathbb{P}\{\tau^{(n)}(m) = m\} = 1,$$

see formula (12) below, Proposition 3.3 immediately yields the following.

Theorem 3.4.

$$\begin{aligned} & \left(\log \operatorname{lcm}(B_n(m)) - m \log n \right)_{m \in \mathbb{N}} \\ & \xrightarrow[n \rightarrow \infty]{d, \infty} \left(\sum_{j=1}^m \log U_j + \sum_p \log p \cdot \left(\max_{1 \leq k \leq m} \mathcal{G}_k(p) - \sum_{k=1}^m \mathcal{G}_k(p) \right) \right)_{m \in \mathbb{N}}. \end{aligned}$$

3.2. **The case $m = m_n \rightarrow \infty$ and $m_n = o(n)$.** Theorem 3.4 dealing with the case of fixed m follows, for the most part, from the previously known results. The case $m_n \rightarrow \infty$ turns out to be more intriguing and requires a different approach.

As usual, we write $(\mathcal{Z}_n(t))_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\text{f.d.d.}} (\mathcal{Z}(t))_{t \geq 0}$ to denote weak convergence of finite-dimensional distributions, that is, for any $k \in \mathbb{N}$ and any $0 \leq t_1 < t_2 < \dots < t_k < \infty$, $(\mathcal{Z}_n(t_1), \dots, \mathcal{Z}_n(t_k))$ converges in distribution to $(\mathcal{Z}(t_1), \dots, \mathcal{Z}(t_k))$ as $n \rightarrow \infty$. For every fixed $n \in \mathbb{N}$ and $y \geq 0$, put

$$(6) \quad c_n(y) := \sum_{p \leq n} \log p (1 - (1 - n^{-1} \lfloor n/p \rfloor)^y).$$

We distinguish two cases:

- (A) $m_n \leq n^{1/2}$ for all sufficiently large n and $\lim_{n \rightarrow \infty} m_n = \infty$;
- (B) $m_n > n^{1/2}$ for all sufficiently large n and $m_n = o(n)$ as $n \rightarrow \infty$.

Here is our first main result.

Theorem 3.5. *Let $(B(t))_{t \geq 0}$ be a standard Brownian motion.*

(i) *If (A) holds, then*

$$\left(\frac{\log \text{lcm}(B_n(\lfloor m_n t \rfloor)) - c_n(\lfloor m_n t \rfloor)}{\sqrt{2^{-1} m_n \log m_n}} \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\text{f.d.d.}} (B(t))_{t \geq 0}.$$

(ii) *If (B) holds and $m_n = O(n(\log n)^{-1})$, then*

$$\left(\frac{\log \text{lcm}(B_n(\lfloor m_n t \rfloor)) - c_n(-n \log(1 - (m_n t)/n))}{\sqrt{2^{-1} m_n (\log n - \log m_n) (3 \log m_n - \log n)}} \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\text{f.d.d.}} (B(t))_{t \geq 0}.$$

Put

$$Y_n(m) := \log \text{lcm}(U_1^{(n)}, U_2^{(n)}, \dots, U_m^{(n)}), \quad 1 \leq m \leq n$$

and note that

$$(7) \quad \begin{aligned} \log \text{lcm}(B_n(\lfloor m_n t \rfloor)) &= \log \text{lcm}(U_1^{(n)}, U_2^{(n)}, \dots, U_{\tau^{(n)}(\lfloor m_n t \rfloor)}^{(n)}) \\ &= Y_n(\tau^{(n)}(\lfloor m_n t \rfloor)), \quad t \geq 0. \end{aligned}$$

We deduce Theorem 3.5 from the following result, a counterpart of Proposition 3.3 for diverging m_n , which is interesting on its own.

Theorem 3.6. *Let $(B(t))_{t \geq 0}$ be a standard Brownian motion.*

(i) *If (A) holds, then*

$$\left(\frac{Y_n(\lfloor m_n t \rfloor) - c_n(\lfloor m_n t \rfloor)}{\sqrt{2^{-1} m_n \log m_n}} \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\text{f.d.d.}} (B(t))_{t \geq 0}.$$

(ii) *If (B) holds, then*

$$\left(\frac{Y_n(\lfloor m_n t \rfloor) - c_n(\lfloor m_n t \rfloor)}{\sqrt{2^{-1} m_n (\log n - \log m_n) (3 \log m_n - \log n)}} \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\text{f.d.d.}} (B(t))_{t \geq 0}.$$

3.3. Strategy of proof of Theorem 3.6. For some samples $\{c_1^{(n)}, \dots, c_{m_n}^{(n)}\}$ of random integers taking values in $[n]$ the logarithm of the least common multiple

$$(8) \quad \log \text{lcm}(\{c_1^{(n)}, \dots, c_{m_n}^{(n)}\}) = \sum_{p \leq n} \log p \max_{1 \leq k \leq m_n} \lambda_p(c_k^{(n)})$$

may be well-approximated by $\log \prod_{k=1}^{m_n} c_k^{(n)}$. For instance, this is known to be the case when $\{c_1^{(n)}, \dots, c_{m_n}^{(n)}\}$ are the cycle lengths (with m_n being the total number of cycles) of a wide class of random permutations including Ewens' permutations, see [17, 22]. Intuitively, such an approximation is successful provided that 'most' of the values among $c_1^{(n)}, \dots, c_{m_n}^{(n)}$ are distinct and 'most' of the positive multiplicities $\lambda_p(c_k^{(n)})$, $p \in \mathcal{P}$, $k = 1, \dots, m_n$ are ones. Of course, many samples do not enjoy these properties and particularly neither do $\{U_1^{(n)}, \dots, U_{m_n}^{(n)}\}$ that we are focused on.

Roughly speaking, the previous approximation argument is based on comparison of $\max_{1 \leq k \leq m_n} \lambda_p(c_k^{(n)})$ and $\sum_{k=1}^{m_n} \lambda_p(c_k^{(n)})$. However, it seems that in many cases $\max_{1 \leq k \leq m_n} \lambda_p(c_k^{(n)})$ should be closer to $\sum_{k=1}^{m_n} \mathbb{1}_{\{\lambda_p(c_k^{(n)}) \geq 1\}}$ rather than to $\sum_{k=1}^{m_n} \lambda_p(c_k^{(n)})$, and our strategy is to exploit this line of reasoning. We shall show in Lemma 6.1 that $Y_n(\lfloor m_n t \rfloor) = \log \text{lcm}(U_1^{(n)}, \dots, U_{\lfloor m_n t \rfloor}^{(n)})$ is well-approximated by $\sum_{p \leq n} \log p \mathbb{1}_{\{\max_{1 \leq k \leq \lfloor m_n t \rfloor} \lambda_p(U_k^{(n)}) \geq 1\}}$. Furthermore, we shall prove in Lemmas 6.2 and 6.3, respectively, that 'small' primes $p \leq m_n$ do not give significant contribution to $Y_n(\lfloor m_n t \rfloor)$ and that in the range of 'large' primes $p > m_n$ the indicators $\mathbb{1}_{\{\max_{1 \leq k \leq \lfloor m_n t \rfloor} \lambda_p(U_k^{(n)}) \geq 1\}}$ can be safely replaced by $\sum_{k=1}^{\lfloor m_n t \rfloor} \mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}}$. Summarizing, we are going to approximate $Y_n(\lfloor m_n t \rfloor)$ by

$$\sum_{m_n < p \leq n} \log p \sum_{k=1}^{\lfloor m_n t \rfloor} \mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}} = \sum_{k=1}^{\lfloor m_n t \rfloor} \sum_{m_n < p \leq n} \log p \cdot \mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}}$$

which is the sum of independent random variables. A limit theorem for the approximating processes is given in Lemma 6.5.

We close this section by a discussion of inefficiency of a tempting alternative approach. A specialization of formula (8) reads

$$Y_n(\lfloor m_n t \rfloor) = \sum_{p \leq n} \log p \max_{1 \leq k \leq \lfloor m_n t \rfloor} \lambda_p(U_k^{(n)}), \quad t \geq 0.$$

As far as a proof of Theorem 3.6 is concerned, a naive idea justified in part by Lemma 3.1 is to replace the terms $\lambda_p(U_k^{(n)})$ with their limits $\mathcal{G}_k(p)$ and to approximate $(Y_n(\lfloor m_n t \rfloor))_{t \geq 0}$ by $(\widehat{Y}_n(\lfloor m_n t \rfloor))_{t \geq 0}$, where

$$(9) \quad \widehat{Y}_n(\lfloor m_n t \rfloor) = \sum_{p \leq n} \log p \max_{1 \leq k \leq \lfloor m_n t \rfloor} \mathcal{G}_k(p), \quad t \geq 0.$$

The right-hand side is the sum of independent random variables which ensures that the analysis of $\widehat{Y}_n(\lfloor m_n t \rfloor)$ is simple. It turns out that unless $\log n / \log m_n \rightarrow 1$, that is, m_n is rather close to n , this intuition is *wrong* in a sense that the limit relation in Theorem 3.6 does not hold with Y_n replaced by \widehat{Y}_n . The details can be found in Proposition 8.1 given in the Appendix. This is an unexpected and peculiar phenomenon because most of the known results in probabilistic number theory involving discrete uniform random variables can be proved using the formal substitution $\lambda_p(U^{(n)}) \mapsto \mathcal{G}(p)$. The list includes:

- (i) the Hardy–Ramanujan central limit theorem for the number of prime divisors;
- (ii) the Erdős–Kac central limit theorem for strongly additive functions [14];
- (iii) the functional central limit theorem for the counts of prime factors [5];
- (iv) the Kubilius theorems on convergence to infinitely divisible laws [20, Section 5];
- (v) Proposition 3.3 of the present paper.

A detailed discussion, proofs and further examples can be found in [2] and in Section 1.2 of [3]. In particular, an optimal coupling between $(\lambda_p(U^{(n)}))_{p \in \mathcal{P}}$ and $(\mathcal{G}(p))_{p \in \mathcal{P}}$ is constructed in [2]. A partial explanation of the inefficiency of this approach in our situation is that the cumulative error caused by replacing $\lambda_p(U_k^{(n)})$ by $\mathcal{G}_k(p)$, $k = 1, \dots, m_n$, which is negligible when the number m_n of such replacements is bounded (as in examples (i)–(v) above), becomes significant with a growth of the sample.

4. SOME AUXILIARY RESULTS RELATED TO THE COUPON COLLECTOR PROBLEM

In this section we discuss the asymptotic behaviour of the stopping time $\tau^{(n)}(m_n)$ as $n \rightarrow \infty$. According to formula (7), this information is of principal importance for deducing Theorem 3.5 from Theorem 3.6. Using (3) we infer

$$(10) \quad \mathbb{E}\tau^{(n)}(m) = \sum_{k=1}^m \mathbb{E}X_{k,n} = \sum_{i=1}^m n(n-i+1)^{-1} = n(H_n - H_{n-m}),$$

where $H_n := \sum_{k=1}^n k^{-1}$ is the n th harmonic number. Moreover,

$$\begin{aligned} \text{Var } \tau^{(n)}(m) &= \sum_{k=1}^m \text{Var } X_{k,n} = \sum_{k=1}^m ((k-1)/n)((n-k+1)/n)^{-2} \\ &= n \sum_{k=n-m+1}^n k^{-2}(n-k) = n^2(H_{n,2} - H_{n-m,2}) - n(H_n - H_{n-m}), \end{aligned}$$

where $H_{n,2} := \sum_{k=1}^n k^{-2}$.

Assume now that $m = m_n < n$ depends on n in such a way that $m_n = o(n)$. By using the standard expansions

$$H_n = \log n + \gamma + (2n)^{-1} + O(n^{-2}), \quad H_{n,2} = \zeta(2) - n^{-1} + 2^{-1}n^{-2} + O(n^{-3}),$$

where γ is the Euler-Mascheroni constant, we obtain

$$(11) \quad \mathbb{E}\tau^{(n)}(m_n) = m_n + O(m_n^2/n), \quad \text{Var } \tau^{(n)}(m_n) = O(m_n^2/n).$$

In particular, if $m_n = o(n^{1/2})$, then

$$\tau^{(n)}(m_n) - m_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

or, in other words,

$$(12) \quad \lim_{n \rightarrow \infty} \mathbb{P}\{\tau^{(n)}(m_n) = m_n\} = 1.$$

5. ASYMPTOTICS OF THE CENTRAL MOMENTS OF $\sum_{m_n < p \leq n} \log p \cdot \mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}}$

For $n \in \mathbb{N}$ and $1 \leq m < n$, put

$$\tilde{U}^{(n,m)} = \prod_{m < p \leq n} p^{\mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}}.$$

In this section we investigate the behavior of $\mathbb{E}\left(\log \tilde{U}^{(n,m_n)} - \mathbb{E} \log \tilde{U}^{(n,m_n)}\right)^{2s}$ for $s = 1, 2$ as $n \rightarrow \infty$. The results obtained here are an important ingredient in the proof of Theorem 3.6.

5.1. Auxiliary results. We start with several auxiliary facts.

Lemma 5.1. *For $s \in \mathbb{N}$,*

$$\lim_{n \rightarrow \infty} \mathbb{E}\left(\log U^{(n)} - \mathbb{E} \log U^{(n)}\right)^{2s} = \mathbb{E}(\mathcal{E}_1 - \mathbb{E}\mathcal{E}_1)^{2s} = \mathbb{E}(\mathcal{E}_1 - 1)^{2s},$$

where \mathcal{E}_1 is a random variable with the exponential distribution of unit mean.

Proof. To justify this it is tempting to use distributional convergence of $\log n - \log U^{(n)}$ to \mathcal{E}_1 in combination with a uniform integrability argument. However, we find it simpler to exploit a more analytic approach based on a direct calculation of the moments: for $r \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}(\log U^{(n)})^r &= n^{-1} \sum_{k=1}^n \log^r k = n^{-1} \int_1^n \log^r x \, dx + O(n^{-1} \log^r n) \\ &= \sum_{i=0}^r (-1)^i i! \binom{r}{i} \log^{r-i} n + O(n^{-1} \log^r n). \end{aligned}$$

This follows from the Euler–Maclaurin summation formula. An application of the binomial theorem completes the proof. \square

As usual, $a_n \sim b_n$ as $n \rightarrow \infty$ means that $\lim_{n \rightarrow \infty} (a_n/b_n) = 1$.

Lemma 5.2. *Let $s \in \mathbb{N}$ and $(X_n)_{n \geq 1}$ and $(Y_n)_{n \geq 1}$ be arbitrarily dependent sequences of random variables with finite moments of order $2s$. If $\mathbb{E}X_n^{2s} = O(1)$ and $\lim_{n \rightarrow \infty} \mathbb{E}Y_n^{2s} = \infty$, then $\mathbb{E}(X_n - Y_n)^{2s} \sim \mathbb{E}Y_n^{2s}$ as $n \rightarrow \infty$.*

Proof. We start with a representation

$$\mathbb{E}(X_n - Y_n)^{2s} = \sum_{k=0}^{2s} (-1)^k \binom{2s}{k} \mathbb{E}X_n^{2s-k} Y_n^k.$$

For $k = 1, \dots, 2s - 1$, an application of Hölder's inequality yields

$$|\mathbb{E}X_n^{2s-k} Y_n^k| \leq (\mathbb{E}X_n^{2s})^{1-k/(2s)} (\mathbb{E}Y_n^{2s})^{k/(2s)} = o(\mathbb{E}Y_n^{2s})$$

because $k < 2s$. Thus, we have proved that $\mathbb{E}(X_n - Y_n)^{2s} = \mathbb{E}Y_n^{2s} + o(\mathbb{E}Y_n^{2s})$. \square

Lemma 5.3. *The following asymptotic relations hold:*

$$(13) \quad K_n := \mathbb{E} \left(\log U^{(n)} - \log \tilde{U}^{(n)} \right)^4 = O(1), \quad n \rightarrow \infty,$$

where $\tilde{U}^{(n)} = \prod_{p \leq n} p^{\mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}}$ for $n \in \mathbb{N}$, and

$$(14) \quad \mathbb{E} \left(\log U^{(n)} - \log \tilde{U}^{(n)} - \mathbb{E} \log U^{(n)} + \mathbb{E} \log \tilde{U}^{(n)} \right)^4 = O(1), \quad n \rightarrow \infty.$$

Proof. We write with the help of the binomial theorem

$$\begin{aligned} K_n &= \mathbb{E} \left(\sum_{p \leq n} \log p \cdot (\lambda_p(U^{(n)}) - \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}) \right)^4 \\ &= \sum_{p \leq n} \log^4 p \times \mathbb{E}(\lambda_p(U^{(n)}) - 1)_+^4 \\ &\quad + 4 \sum_{p \neq q \leq n} \log^3 p \log q \times \mathbb{E}(\lambda_p(U^{(n)}) - 1)_+^3 (\lambda_q(U^{(n)}) - 1)_+ \\ &\quad + 6 \sum_{p < q \leq n} \log^2 p \log^2 q \times \mathbb{E}(\lambda_p(U^{(n)}) - 1)_+^2 (\lambda_q(U^{(n)}) - 1)_+^2 \\ &\quad + 12 \sum_{p \neq q \neq r \leq n} \log^2 p \log q \log r \times \mathbb{E}(\lambda_p(U^{(n)}) - 1)_+^2 (\lambda_q(U^{(n)}) - 1)_+ (\lambda_r(U^{(n)}) - 1)_+ \\ &\quad + 24 \sum_{p < q < r < s \leq n} \log p \log q \log r \log s \times \\ &\quad \mathbb{E}(\lambda_p(U^{(n)}) - 1)_+ (\lambda_q(U^{(n)}) - 1)_+ (\lambda_r(U^{(n)}) - 1)_+ (\lambda_s(U^{(n)}) - 1)_+ \\ &= \sum_{i=1}^5 K_i(n). \end{aligned}$$

For any $k \in \mathbb{N}$, any positive integers v_1, \dots, v_k and any distinct prime numbers p_1, \dots, p_k , we have

$$\begin{aligned}
& \mathbb{E} \left(\prod_{i=1}^k (\lambda_{p_i}(U^{(n)}) - 1)_+^{v_i} \right) \\
&= \sum_{j_1, \dots, j_k \geq 2} (j_1 - 1)^{v_1} \cdots (j_k - 1)^{v_k} \mathbb{P} \{ \lambda_{p_1}(U^{(n)}) = j_1, \dots, \lambda_{p_k}(U^{(n)}) = j_k \} \\
&\leq \sum_{j_1, \dots, j_k \geq 2} (j_1 - 1)^{v_1} \cdots (j_k - 1)^{v_k} \mathbb{P} \{ \lambda_{p_1}(U^{(n)}) \geq j_1, \dots, \lambda_{p_k}(U^{(n)}) \geq j_k \} \\
&= \sum_{j_1, \dots, j_k \geq 2} (j_1 - 1)^{v_1} \cdots (j_k - 1)^{v_k} \lfloor n / (p_1^{j_1} \cdots p_k^{j_k}) \rfloor n^{-1} \\
&\leq \sum_{j_1, \dots, j_k \geq 2} j_1^{v_1} p_1^{-j_1} \cdots j_k^{v_k} p_k^{-j_k} \leq \text{const} \cdot \prod_{i=1}^k p_i^{-2}.
\end{aligned}$$

Now, for $i = 1, \dots, 5$, the asymptotic estimate $K_i(n) = O(1)$ follows from the preceding bound, $\sum_p p^{-2} \log^4 p < \infty$ and the following inequalities:

$$\begin{aligned}
K_2(n) &\leq \text{const} \cdot \sum_{p \neq q \leq n} (p^{-2} \log^3 p)(q^{-2} \log q) \\
&\leq \text{const} \cdot \left(\sum_p p^{-2} \log^3 p \right) \left(\sum_p p^{-2} \log p \right);
\end{aligned}$$

$$K_3(n) \leq \text{const} \cdot \sum_{p < q \leq n} (p^{-2} \log^2 p)(q^{-2} \log^2 q) \leq \text{const} \cdot \left(\sum_p p^{-2} \log^2 p \right)^2;$$

$$\begin{aligned}
K_4(n) &\leq \text{const} \cdot \sum_{p \neq q \neq r \leq n} (p^{-2} \log^2 p)(q^{-2} \log q)(r^{-2} \log r) \\
&\leq \text{const} \cdot \left(\sum_p p^{-2} \log^2 p \right) \left(\sum_p p^{-2} \log p \right)^2;
\end{aligned}$$

$$\begin{aligned}
K_5(n) &\leq \text{const} \cdot \sum_{p < q < r < s \leq n} (p^{-2} \log p)(q^{-2} \log q)(r^{-2} \log r)(s^{-2} \log s) \\
&\leq \text{const} \cdot \left(\sum_p p^{-2} \log p \right)^4.
\end{aligned}$$

Thus, (13) holds. Inequality (14) follows immediately, because for any random variable X with $\mathbb{E}X^4 < \infty$ we have $\mathbb{E}(X - \mathbb{E}X)^4 \leq 8\mathbb{E}X^4$. \square

5.2. Asymptotics of the variance. The function π defined by $\pi(x) := \sum_{p \leq x} 1$ for $x \geq 0$ is called a prime counting function. Recall that the prime number theorem (see, for instance, Theorem 6.2.1 in [7]) states that

$$(15) \quad \pi(x) \sim x / \log x, \quad x \rightarrow \infty.$$

For later multiple use, we note that (15) in combination with integration by parts entails

$$(16) \quad \int_{I_x} f(y) d\pi(y) \sim \int_{I_x} (f(y)/\log y) dy, \quad x \rightarrow \infty$$

where $f(x) = x^{-a} \log^b x$ and either (a) $I_x = [2, x]$, $a \leq 1$, $b > 0$ or (b) $I_x = (x, \infty)$, $a > 1$, $b \geq 0$.

While investigating the asymptotics of $\text{Var}(\log \tilde{U}^{(n, m_n)})$ we treat the two cases (A) $m_n \leq n^{1/2}$ for large n and $m_n \rightarrow \infty$ and (B) $m_n > n^{1/2}$ for large n and $m_n = o(n)$ separately in Theorems 5.4 and 5.5. Here is a brief explanation of such a case distinction. In case (B), the variance of $\log \tilde{U}^{(n, m_n)}$ is given by the sum of two terms which can exhibit different first order asymptotics (formula (26)). This necessitates us to provide several terms expansions for each. In the case considered the m 's are reasonably large and kill the influence of nonprincipal terms of these expansions (in particular, it turns out that two terms expansions suffice). In case (A), the aforementioned reasoning does not help. We state, without going into details, that

$$(17) \quad \text{Var}(\log \tilde{U}^{(n, m_n)}) = 2^{-1} \log^2 m_n + O(\log n \log^{-2} m_n),$$

where, to the best of our knowledge, the big-oh term cannot be improved. Hence, formula (17) as it stands only provides the correct asymptotics $\text{Var}(\log \tilde{U}^{(n, m_n)}) \sim 2^{-1} \log^2 m_n$ for m_n satisfying $m_n e^{-\log^{1/4} n} \rightarrow \infty$. For smaller m formula (17) is useless for us. In view of these issues we offer an alternative approach for the case (A). We show that $\text{Var}(\log \tilde{U}^{(n, m_n)}) \sim \text{Var}\left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}\right)$. The variance on the right-hand side which is given by the sum of three terms is much easier to deal with, for the first order asymptotics of the terms is enough. On the other hand, this simple reasoning is not applicable in case (B) because the terms of the sum representing the variance may exhibit different first order behavior. For instance, (22) still holds true in case (B), whereas

$$2n^{-1} \sum_{p < q \leq m_n} \log p \log q \lfloor n/(pq) \rfloor \sim 2^{-1} \log^2 n - \log^2(n/m_n).$$

Theorem 5.4. *Assume that $m_n \leq n^{1/2}$ for all n large enough and $m_n \rightarrow \infty$ as $n \rightarrow \infty$. Then*

$$\text{Var}\left(\log \tilde{U}^{(n, m_n)}\right) \sim 2^{-1} \log^2 m_n, \quad n \rightarrow \infty.$$

Proof. From Lemma 5.1 with $s = 1$ we know that

$$(18) \quad \text{Var}\left(\sum_{p \leq n} \log p \cdot \lambda_p(U^{(n)})\right) = \text{Var}(\log U^{(n)}) \rightarrow 1.$$

Further, relation (14) ensures that

$$(19) \quad \text{Var}(\log U^{(n)} - \log \tilde{U}^{(n)}) = O(1)$$

because $\text{Var} X \leq \sqrt{\mathbb{E}(X - \mathbb{E}X)^4}$ for any random variable X with $\mathbb{E}X^4 < \infty$. These asymptotic relations in combination with the inequality $(x + y)^2 \leq 2(x^2 + y^2)$, $x, y \in \mathbb{R}$ entail

$$\text{Var}\left(\sum_{p \leq n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}\right) = \text{Var}\left(\log \tilde{U}^{(n)}\right) = O(1).$$

Further, by Lemma 5.2 applied with $s = 1$,

$$\begin{aligned} X_n &:= \log \tilde{U}^{(n)} - \mathbb{E} \log \tilde{U}^{(n)} \\ &= \sum_{p \leq n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} - \mathbb{E}\left(\sum_{p \leq n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}\right), \end{aligned}$$

and

$$Y_n := \sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} - \mathbb{E}\left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}\right)$$

we have

$$\begin{aligned} \text{Var}(\log \tilde{U}^{(n, m_n)}) &= \text{Var}\left(\log \tilde{U}^{(n)} - \sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}\right) \\ &\sim \text{Var}\left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}\right) \end{aligned}$$

provided the right-hand side diverges to infinity. Thus, it is enough to prove that

$$(20) \quad \text{Var}\left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}\right) \sim 2^{-1} \log^2 m_n.$$

As a preparation, write

$$\begin{aligned} (21) \quad &\text{Var}\left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}\right) \\ &= \mathbb{E}\left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}\right)^2 - \left(\mathbb{E}\left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}\right)\right)^2 \\ &= n^{-1} \sum_{p \leq m_n} \log^2 p [n/p] + 2n^{-1} \sum_{p < q \leq m_n} \log p \log q [n/(pq)] \\ &\quad - \left(n^{-1} \sum_{p \leq m_n} \log p [n/p]\right)^2 \end{aligned}$$

Here, the equalities $\mathbb{P}\{\lambda_p(U^{(n)}) \geq 1\} = n^{-1} [n/p]$ and, for $p \neq q$,

$$\mathbb{P}\{\lambda_p(U^{(n)}) \geq 1, \lambda_q(U^{(n)}) \geq 1\} = n^{-1} [n/(pq)]$$

have to be recalled.

We start by analyzing the first term

$$\sum_{p \leq m_n} p^{-1} \log^2 p - n^{-1} \sum_{p \leq m_n} \log^2 p \leq n^{-1} \sum_{p \leq m_n} \log^2 p \lfloor n/p \rfloor \leq \sum_{p \leq m_n} p^{-1} \log^2 p,$$

so that

$$(22) \quad n^{-1} \sum_{p \leq m_n} \log^2 p \lfloor n/p \rfloor \sim 2^{-1} \log^2 m_n$$

is a consequence of

$$\begin{aligned} \sum_{p \leq m_n} p^{-1} \log^2 p &\sim 2^{-1} \log^2 m_n \quad \text{and} \\ n^{-1} \sum_{p \leq m_n} \log^2 p &\sim n^{-1} m_n \log m_n = o(\log m_n). \end{aligned}$$

The latter limit relations are justified by (16) with $f(x) = x^{-1} \log^2 x$ and $f(x) = \log^2 x$, respectively. Similarly,

$$(23) \quad \left(n^{-1} \sum_{p \leq m_n} \log p \lfloor n/p \rfloor \right)^2 \sim \log^2 m_n$$

follows from

$$\sum_{p \leq m_n} p^{-1} \log p - n^{-1} \sum_{p \leq m_n} \log p \leq n^{-1} \sum_{p \leq m_n} \log p \lfloor n/p \rfloor \leq \sum_{p \leq m_n} p^{-1} \log p$$

and

$$\sum_{p \leq m_n} p^{-1} \log p \sim \log m_n, \quad n^{-1} \sum_{p \leq m_n} \log p \sim n^{-1} m_n = o(1).$$

Finally, we use

$$\begin{aligned} \sum_{p < q \leq m_n} (p^{-1} \log p)(q^{-1} \log q) - n^{-1} \sum_{p < q \leq m_n} \log p \log q \\ \leq n^{-1} \sum_{p < q \leq m_n} \log p \log q \lfloor n/(pq) \rfloor \leq \sum_{p < q \leq m_n} (p^{-1} \log p)(q^{-1} \log q) \end{aligned}$$

together with

$$2 \sum_{p < q \leq m_n} (p^{-1} \log p)(q^{-1} \log q) = \left(\sum_{p \leq m_n} p^{-1} \log p \right)^2 - \sum_{p \leq m_n} p^{-2} \log^2 p \sim \log^2 m_n$$

and

$$2n^{-1} \sum_{p < q \leq m_n} \log p \log q \leq n^{-1} \left(\sum_{p \leq m_n} \log p \right)^2 \sim n^{-1} m_n^2 = O(1)$$

(recall that $m_n \leq n^{1/2}$ by assumption) to obtain

$$(24) \quad 2n^{-1} \sum_{p < q \leq m_n} \log p \log q \lfloor n/(pq) \rfloor \sim \log^2 m_n.$$

A combination of (22), (23) and (24) proves (20). \square

Theorem 5.5. *Assume that $m_n > n^{1/2}$ for all n large enough and $m_n = o(n)$ as $n \rightarrow \infty$. Then*

$$(25) \quad \text{Var}(\log \tilde{U}^{(n, m_n)}) \sim 2^{-1}(\log n - \log m_n)(3 \log m_n - \log n), \quad n \rightarrow \infty.$$

Proof. Note that

$$\log \tilde{U}^{(n, m_n)} - \mathbb{E} \log \tilde{U}^{(n, m_n)} = \sum_{m_n < p \leq n} \log p (\mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} - \mathbb{P}\{\lambda_p(U^{(n)}) \geq 1\})$$

whence

$$(26) \quad \begin{aligned} \text{Var}(\log \tilde{U}^{(n, m_n)}) &= n^{-1} \sum_{m_n < p \leq n} \log^2 p \lfloor n/p \rfloor - \left(n^{-1} \sum_{m_n < p \leq n} \log p \lfloor n/p \rfloor \right)^2 \\ &\quad + 2n^{-1} \sum_{m_n < p < q \leq n} \log p \log q \lfloor n/(pq) \rfloor \\ &= n^{-1} \sum_{m_n < p \leq n} \log^2 p \lfloor n/p \rfloor - \left(n^{-1} \sum_{m_n < p \leq n} \log p \lfloor n/p \rfloor \right)^2 \end{aligned}$$

which is a counterpart of (21). Observe that the last equality follows from the fact that $\lfloor n/(pq) \rfloor = 0$ whenever $p > q > m_n > n^{1/2}$. We claim that

$$(27) \quad n^{-1} \sum_{m_n < p \leq n} \log^2 p \lfloor n/p \rfloor = 2^{-1} \log^2 n - 2^{-1} \log^2 m_n + o(\log^2 n - \log^2 m_n)$$

and

$$(28) \quad \left(n^{-1} \sum_{m_n < p \leq n} \log p \lfloor n/p \rfloor \right)^2 = \log^2(n/m_n) + o(\log^2(n/m_n)).$$

To prove (27), write

$$\begin{aligned} \frac{1}{n} \sum_{m_n < p \leq n} \log^2 p \lfloor n/p \rfloor &= \sum_{m_n < p \leq n} p^{-1} \log^2 p + O(\log n) \\ &= 2^{-1} \log^2 n - 2^{-1} \log^2 m_n + o(\log^2 n - \log^2 m_n) + O(\log n) \\ &= 2^{-1} \log^2 n - 2^{-1} \log^2 m_n + o(\log^2 n - \log^2 m_n), \end{aligned}$$

where the first equality follows from the trivial estimate $n/p - 1 < \lfloor n/p \rfloor \leq n/p$, the second is a consequence of (16) with $f(x) = x^{-1} \log^2 x$ and $I_x = (m_n, n]$, and the third is implied by

$$0 \leq \log n / (\log^2 n - \log^2 m_n) \leq 1 / \log(n/m_n) \rightarrow 0.$$

Formula (28) follows along similar lines from

$$n^{-1} \sum_{m_n < p \leq n} \log p \lfloor n/p \rfloor = \sum_{m_n < p \leq n} p^{-1} \log p + O(1) = \log(n/m_n) + o(\log(n/m_n)),$$

where the second equality results from (16) with $f(x) = x^{-1} \log x$ and the term $O(1)$ is killed by $o(\log(n/m_n))$.

Subtracting (28) from (27) and using that $\log^2(n/m_n) \leq \log^2 n - \log^2 m_n$ we arrive at

$$\text{Var}(\log \tilde{U}^{(n, m_n)}) = 2^{-1}(\log n - \log m_n)(3 \log m_n - \log n) + \hat{R}_n,$$

where

$$\lim_{n \rightarrow \infty} |\hat{R}_n| / (\log^2 n - \log^2 m_n) = 0.$$

This implies (25) because

$$\begin{aligned} \frac{|\hat{R}_n|}{(\log n - \log m_n)(3 \log m_n - \log n)} &= \frac{|\hat{R}_n|}{\log^2 n - \log^2 m_n} \frac{\log n + \log m_n}{3 \log m_n - \log n} \\ &\leq 4 \frac{|\hat{R}_n|}{\log^2 n - \log^2 m_n} \rightarrow 0, \end{aligned}$$

where we have used $n^{1/2} < m_n \leq n$ for large n for the inequality. \square

5.3. The big-oh estimate for the fourth central moment. The proof of Lemma 5.6 is very similar to the proof of Theorem 5.1 but is essentially simpler for it only provides an upper estimate rather than the exact rate of growth.

Lemma 5.6. *Assume that $m_n \rightarrow \infty$ and $m_n = o(n)$ as $n \rightarrow \infty$. Then*

$$(29) \quad \mathbb{E} \left(\log \tilde{U}^{(n, m_n)} - \mathbb{E} \log \tilde{U}^{(n, m_n)} \right)^4 = O(\log^4 m_n), \quad n \rightarrow \infty.$$

Proof. We divide the proof into two steps. The purpose of Step 1 is to demonstrate that (29) is implied by

$$(30) \quad \mathbb{E} \left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} - \mathbb{E} \sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} \right)^4 = O(\log^4 m_n).$$

Step 2 is devoted to showing that (30) holds true.

STEP 1. From Lemma 5.1 with $s = 2$ we know that

$$\lim_{n \rightarrow \infty} \mathbb{E}(\log U^{(n)} - \mathbb{E} \log U^{(n)})^4 = 9.$$

This together with (14) yields

$$\lim_{n \rightarrow \infty} \mathbb{E}(\log \tilde{U}^{(n)} - \mathbb{E} \log \tilde{U}^{(n)})^4 = O(1).$$

The latter in combination with Lemma 5.2 applied with $s = 2$,

$$X_n = \log \tilde{U}^{(n)} - \mathbb{E} \log \tilde{U}^{(n)} = \sum_{p \leq n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} - \mathbb{E} \sum_{p \leq n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}$$

and

$$Y_n = \sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} - \mathbb{E} \sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}}$$

enables us to conclude that (29) is implied by (30) and

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} - \mathbb{E} \sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} \right)^4 = \infty.$$

The latter holds true in view of

$$\begin{aligned} \mathbb{E} \left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} - \mathbb{E} \sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} \right)^4 \\ \geq \left(\text{Var} \left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} \right) \right)^2 \rightarrow \infty. \end{aligned}$$

Here, divergence is justified by (20) when $m_n \leq n^{1/2}$ with $m_n \rightarrow \infty$ and by Theorem 5.5 together with

$$\text{Var} \left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} \right) \sim \text{Var} (\log \tilde{U}^{(n, m_n)})$$

(which holds true by another application of Lemma 5.2) when $m_n > n^{1/2}$ with $m_n = o(n)$.

STEP 2. Passing to the proof of (30) we first note that, for $a > 0$,

$$(31) \quad \sum_{p \leq m_n} p^{-1} \log^a p = \int_{[2, m_n]} x^{-1} \log^a x \, d\pi(x) \sim a^{-1} \log^a m_n,$$

where the asymptotic equivalence follows from (16) with $f(x) = x^{-1} \log^a x$. With this at hand, we obtain

$$\begin{aligned} & \mathbb{E} \left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\lambda_p(U^{(n)}) \geq 1\}} \right)^4 \\ &= \sum_{p \leq m_n} \log^4 p \lfloor n/p \rfloor n^{-1} \\ &+ 4 \sum_{p \neq q \leq m_n} \log^3 p \log q \times \lfloor n/(pq) \rfloor n^{-1} \\ &+ 6 \sum_{p < q \leq m_n} \log^2 p \log^2 q \times \lfloor n/(pq) \rfloor n^{-1} \\ &+ 12 \sum_{p \neq q \neq r \leq m_n} \log^2 p \log q \log r \times \lfloor n/(pqr) \rfloor n^{-1} \\ &+ 24 \sum_{p < q < r < s \leq m_n} \log p \log q \log r \log s \times \lfloor n/(pqrs) \rfloor n^{-1}. \end{aligned}$$

Each summand is $O(\log^4 m_n)$, by (31). For example, for the fourth summand this is a consequence of

$$\begin{aligned} \sum_{p \neq q \neq r \leq m_n} \log^2 p \log q \log r \times \lfloor n/(pqr) \rfloor n^{-1} &\leq \sum_{p, q, r \leq m_n} (pqr)^{-1} \log^2 p \log q \log r \\ &= \left(\sum_{p \leq m_n} p^{-1} \log^2 p \right) \left(\sum_{q \leq m_n} q^{-1} \log q \right)^2 = O(\log^4 m_n). \end{aligned}$$

Since $\mathbb{E}(X - \mathbb{E}X)^4 \leq 8\mathbb{E}X^4$ for any random variable X with $\mathbb{E}X^4 < \infty$, (30) follows. \square

6. PROOF OF THEOREM 3.6

We prove Theorem 3.6 via the sequence of lemmas. For $n \in \mathbb{N}$ and $t \geq 0$, put

$$(32) \quad Z_n(t) := \sum_{p \leq n} \log p \cdot \mathbb{1}_{\{\max_{1 \leq k \leq \lfloor m_n t \rfloor} \lambda_p(U_k^{(n)}) \geq 1\}}.$$

Lemma 6.1. *Assume that $m_n \rightarrow \infty$ and $m_n = o(n)$ as $n \rightarrow \infty$. Then, for all $T > 0$,*

$$\mathbb{E} \left(\sup_{t \in [0, T]} (Y_n(\lfloor m_n t \rfloor) - Z_n(t)) \right) = O(m_n^{1/2}), \quad n \rightarrow \infty.$$

Proof. Fix any $T > 0$. Then, for all $t \in [0, T]$,

$$\begin{aligned} 0 &\leq Y_n(\lfloor m_n t \rfloor) - Z_n(t) \\ &= \sum_{p \leq n} \log p \left(\max_{1 \leq k \leq \lfloor m_n t \rfloor} \lambda_p(U_k^{(n)}) - \mathbb{1}_{\{\max_{1 \leq k \leq \lfloor m_n t \rfloor} \lambda_p(U_k^{(n)}) \geq 1\}} \right) \\ &\leq \sum_{p \leq m_n^{1/2}} \log p \max_{1 \leq k \leq \lfloor m_n t \rfloor} \lambda_p(U_k^{(n)}) \\ &\quad + \sum_{m_n^{1/2} < p \leq n} \log p \max_{1 \leq k \leq \lfloor m_n t \rfloor} \lambda_p(U_k^{(n)}) \mathbb{1}_{\{\max_{1 \leq k \leq \lfloor m_n t \rfloor} \lambda_p(U_k^{(n)}) \geq 2\}} \\ &:= I_n + J_n. \end{aligned}$$

It suffices to check that

$$(33) \quad \mathbb{E}I_n = O(m_n^{1/2}) \quad \text{and} \quad \mathbb{E}J_n = O(m_n^{1/2}).$$

To prove the first relation, write

$$\begin{aligned}
\mathbb{E}I_n &= \sum_{p \leq m_n^{1/2}} \log p \cdot \mathbb{E}(\max_{1 \leq k \leq \lfloor m_n T \rfloor} \lambda_p(U_k^{(n)})) \\
&\leq \sum_{p \leq m_n^{1/2}} \log p \sum_{j \geq 1} \mathbb{P}\left\{ \max_{1 \leq k \leq \lfloor m_n T \rfloor} \lambda_p(U_k^{(n)}) \geq j \right\} \\
&= \sum_{p \leq m_n^{1/2}} \log p \sum_{j \geq 1} (1 - (1 - n^{-1} \lfloor n/p^j \rfloor)^{\lfloor m_n T \rfloor}) \\
&\leq \sum_{p \leq m_n^{1/2}} \log p \sum_{j \geq 1} (1 - (1 - p^{-j})^{\lfloor m_n T \rfloor}).
\end{aligned}$$

Note that

$$\sum_{j \geq 1} (1 - (1 - p^{-j})^{\lfloor m_n T \rfloor}) = \mathbb{E}(\max_{1 \leq k \leq \lfloor m_n T \rfloor} \mathcal{G}_k(p)),$$

where $(\mathcal{G}_k(p))_{k \in \mathbb{N}}$ are mutually independent random variables with geometric distribution (5). Denote by $\mathcal{E}_1, \mathcal{E}_2, \dots$ independent copies of a random variable having the exponential distribution of unit mean. Using the distributional equality

$$\max_{1 \leq k \leq \lfloor m_n T \rfloor} \mathcal{G}_k(p) \stackrel{d}{=} \max_{1 \leq k \leq \lfloor m_n T \rfloor} \lfloor \log^{-1} p \mathcal{E}_k \rfloor = \lfloor \log^{-1} p \max_{1 \leq k \leq \lfloor m_n T \rfloor} \mathcal{E}_k \rfloor$$

we conclude that

$$\begin{aligned}
\sum_{j \geq 1} (1 - (1 - p^{-j})^{\lfloor m_n T \rfloor}) &= \mathbb{E}(\max_{1 \leq k \leq \lfloor m_n T \rfloor} \mathcal{G}_k(p)) = \mathbb{E} \lfloor \log^{-1} p \max_{1 \leq k \leq \lfloor m_n T \rfloor} \mathcal{E}_k \rfloor \\
&\leq \mathbb{E}(\log^{-1} p \max_{1 \leq k \leq \lfloor m_n T \rfloor} \mathcal{E}_k) = \log^{-1} p \sum_{k=1}^{\lfloor m_n T \rfloor} k^{-1} \leq \log^{-1} p (1 + \log(m_n T)).
\end{aligned}$$

Hence,

$$\mathbb{E}I_n \leq \sum_{p \leq m_n^{1/2}} (1 + \log(m_n T)) = (1 + \log(m_n T)) \pi(m_n^{1/2}),$$

where π is the prime counting function. An application of (15) proves the first relation in (33).

We are now passing to the second relation in (33):

$$\begin{aligned}
\mathbb{E}J_n &\leq \sum_{m_n^{1/2} < p \leq n} \log p \cdot \mathbb{E}(\max_{1 \leq k \leq \lfloor m_n T \rfloor} \lambda_p(U_k^{(n)})) \mathbb{1}_{\{\max_{1 \leq k \leq \lfloor m_n T \rfloor} \lambda_p(U_k^{(n)}) \geq 2\}} \\
&\leq 2 \sum_{m_n^{1/2} < p \leq n} \log p \cdot \sum_{j \geq 2} \mathbb{P}\left\{ \max_{1 \leq k \leq \lfloor m_n T \rfloor} \lambda_p(U_k^{(n)}) \geq j \right\} \\
&= 2 \sum_{m_n^{1/2} < p \leq n} \log p \sum_{j \geq 2} (1 - (1 - n^{-1} \lfloor n/p^j \rfloor)^{\lfloor m_n T \rfloor}) \\
&\leq 2 \sum_{m_n^{1/2} < p \leq n} \log p \sum_{j \geq 2} (1 - (1 - p^{-j})^{\lfloor m_n T \rfloor}).
\end{aligned}$$

Using the inequality $1 - (1 - x)^{\lfloor m_n T \rfloor} \leq x \lfloor m_n T \rfloor \leq x m_n T$, $x \in [0, 1]$ we infer

$$\begin{aligned} \mathbb{E}J_n &\leq 2m_n T \sum_{m_n^{1/2} < p \leq n} \log p \sum_{j \geq 2} p^{-j} \\ &= 2m_n T \sum_{m_n^{1/2} < p \leq n} (p^2 - p)^{-1} \log p \leq 4m_n T \sum_{m_n^{1/2} < p} p^{-2} \log p = O(m_n^{1/2}) \end{aligned}$$

having utilized (16) with $f(x) = x^{-2} \log x$ for the last equality. \square

Note that

$$\begin{aligned} \mathbb{E}Z_n(t) &= \mathbb{E} \left(\sum_{p \leq n} \log p \cdot \mathbb{1}_{\{\max_{1 \leq k \leq \lfloor m_n t \rfloor} \lambda_p(U_k^{(n)}) \geq 1\}} \right) \\ &= \sum_{p \leq n} \log p (1 - (1 - n^{-1} \lfloor n/p \rfloor)^{\lfloor m_n t \rfloor}) = c_n(\lfloor m_n t \rfloor). \end{aligned}$$

This equality and Lemma 6.1 demonstrate that Theorem 3.6 follows once we can show that

$$(34) \quad \left(\frac{Z_n(t) - \mathbb{E}Z_n(t)}{\sqrt{a_n}} \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\text{f.d.d.}} (B(t))_{t \geq 0},$$

where

$$(35) \quad a_n = \begin{cases} 2^{-1} m_n \log^2 m_n, & \text{in case (A),} \\ 2^{-1} m_n (\log n - \log m_n) (3 \log m_n - \log n), & \text{in case (B).} \end{cases}$$

For later use recall that in view of Theorems 5.4 and 5.5

$$(36) \quad a_n \sim m_n \text{Var}(\log \tilde{U}^{n, m_n}).$$

As a preparation for the proof of (34) we need a couple of lemmas.

Lemma 6.2. *Assume that $m_n \rightarrow \infty$ and $m_n = o(n)$ as $n \rightarrow \infty$. Then*

$$(37) \quad \text{Var} \left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}} \right) = O(m_n \log m_n), \quad n \rightarrow \infty.$$

Proof. We start with

$$\begin{aligned} &\text{Var} \left(\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}} \right) \leq \sum_{p \leq m_n} \log^2 p \\ &+ 2 \sum_{p < q \leq m_n} \log p \log q \cdot \text{Cov} \left(\mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}}, \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_q(U_k^{(n)}) \geq 1\}} \right). \end{aligned}$$

In view of (16) with $f(x) = \log^2 x$ the first term is asymptotically equivalent to $m_n \log m_n$. Further,

$$\begin{aligned}
& \left| \text{Cov} \left(\mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}}, \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_q(U_k^{(n)}) \geq 1\}} \right) \right| \\
&= \left| \text{Cov} \left(\mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) = 0\}}, \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_q(U_k^{(n)}) = 0\}} \right) \right| \\
&= \left| \mathbb{P}^{m_n} \{ \lambda_p(U^{(n)}) = 0, \lambda_q(U^{(n)}) = 0 \} - \mathbb{P}^{m_n} \{ \lambda_p(U^{(n)}) = 0 \} \mathbb{P}^{m_n} \{ \lambda_q(U^{(n)}) = 0 \} \right| \\
&\leq m_n \left| \mathbb{P} \{ \lambda_p(U^{(n)}) = 0, \lambda_q(U^{(n)}) = 0 \} - \mathbb{P} \{ \lambda_p(U^{(n)}) = 0 \} \mathbb{P} \{ \lambda_q(U^{(n)}) = 0 \} \right| \\
&= m_n |n^{-1} \lfloor n/(pq) \rfloor - n^{-2} \lfloor n/p \rfloor \lfloor n/q \rfloor|,
\end{aligned}$$

where the equalities

$$(38) \quad \mathbb{P} \{ \lambda_p(U^{(n)}) = 0 \} = 1 - n^{-1} \lfloor n/p \rfloor,$$

$$(39) \quad \mathbb{P} \{ \lambda_p(U^{(n)}) = 0, \lambda_q(U^{(n)}) = 0 \} = 1 - n^{-1} \lfloor n/p \rfloor - n^{-1} \lfloor n/q \rfloor + n^{-1} \lfloor n/(pq) \rfloor$$

which hold for prime $p \neq q$ have been utilized. For later use, we note that whenever $q > p > n^{1/2}$ we have

$$\begin{aligned}
(40) \quad & \text{Cov} \left(\mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}}, \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_q(U_k^{(n)}) \geq 1\}} \right) \\
&= (1 - n^{-1} \lfloor n/p \rfloor - n^{-1} \lfloor n/q \rfloor)^{m_n} \\
&\quad - (1 - n^{-1} \lfloor n/p \rfloor - n^{-1} \lfloor n/q \rfloor + n^{-2} \lfloor n/p \rfloor \lfloor n/q \rfloor)^{m_n} \leq 0
\end{aligned}$$

as a consequence of $\lfloor n/(pq) \rfloor = 0$. The inequalities

$$|n^{-1} \lfloor n/(pq) \rfloor - n^{-2} \lfloor n/p \rfloor \lfloor n/q \rfloor| \leq n^{-1}$$

readily imply that

$$\begin{aligned}
2 \left| \sum_{p < q \leq m_n} \log p \log q \text{Cov} \left(\mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}}, \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_q(U_k^{(n)}) \geq 1\}} \right) \right| \\
\leq 2n^{-1} m_n \sum_{p < q \leq m_n} \log p \log q \leq n^{-1} m_n \left(\sum_{p \leq m_n} \log p \right)^2 \sim n^{-1} m_n^3,
\end{aligned}$$

where the last asymptotic relation is a consequence of (16) with $f(x) = \log x$. Thus, if $m_n \leq n^{1/2}$ and $m_n \rightarrow \infty$, the proof of (37) is complete, for the right-hand side of the last centered formula is $O(m_n)$.

Assume that $m_n > n^{1/2}$ and $m_n = o(n)$. We shall use a representation

$$\begin{aligned}
\sum_{p \leq m_n} \log p \cdot \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}} &= \sum_{p \leq n^{1/2}} \log p \cdot \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}} \\
&\quad + \sum_{n^{1/2} < p \leq m_n} \log p \cdot \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}}.
\end{aligned}$$

Repeating verbatim the previous argument yields

$$\text{Var} \left(\sum_{p \leq n^{1/2}} \log p \cdot \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}} \right) = O(n^{1/2} \log n) = O(m_n \log m_n).$$

Finally, we infer with the help of (40) that

$$\begin{aligned} 0 &\leq \text{Var} \left(\sum_{n^{1/2} < p \leq m_n} \log p \cdot \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}} \right) \\ &\leq \sum_{n^{1/2} < p \leq m_n} \log^2 p \text{Var}(\mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}}) \leq \sum_{p \leq m_n} \log^2 p = O(m_n \log m_n). \end{aligned}$$

□

Observe that $a_n/(m_n \log m_n) \rightarrow \infty$. In case (A) this is obvious. In case (B) this is a consequence of

$$\begin{aligned} a_n/(m_n \log m_n) &= 2^{-1}(\log n - \log m_n)(3 \log m_n - \log n)/\log m_n \\ &\geq 2^{-1}(\log n - \log m_n) \rightarrow \infty. \end{aligned}$$

Thus, using (37) in combination with Chebyshev's inequality we conclude that (34) is equivalent to

$$(41) \quad \left(a_n^{-\frac{1}{2}} \sum_{m_n < p \leq n} \log p (\mathbb{1}_{\{\max_{1 \leq k \leq \lfloor m_n t \rfloor} \lambda_p(U_k^{(n)}) \geq 1\}} - \mathbb{P}\{\max_{1 \leq k \leq \lfloor m_n t \rfloor} \lambda_p(U_k^{(n)}) \geq 1\}) \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\text{f.d.d.}} (B(t))_{t \geq 0}.$$

Lemma 6.3. *Assume that $m_n \rightarrow \infty$ and $m_n = o(n)$ as $n \rightarrow \infty$. Then*

$$(42) \quad \text{Var} \left(\sum_{m_n < p \leq n} \log p \left(\sum_{k=1}^{m_n} \mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}} - \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}} \right) \right) = O(m_n \log m_n), \quad n \rightarrow \infty.$$

For the proof we need a technical result.

Lemma 6.4. *Let $\text{Bin}(m, \theta)$ be a random variable having a binomial distribution with parameters $m \in \mathbb{N}$ and $\theta \in (0, 1)$, that is,*

$$\mathbb{P}\{\text{Bin}(m, \theta) = k\} = \binom{m}{k} \theta^k (1 - \theta)^{m-k}, \quad k = 0, 1, \dots, m.$$

Then

$$\text{Var}((\text{Bin}(m, \theta) - 1)_+) = \text{Var}(\text{Bin}(m, \theta) - \mathbb{1}_{\{\text{Bin}(m, \theta) \geq 1\}}) \leq (m\theta)^2.$$

Proof. This follows from

$$\begin{aligned}
& \text{Var}(\text{Bin}(m, \theta) - \mathbb{1}_{\{\text{Bin}(m, \theta) \geq 1\}}) \\
&= \text{Var}(\text{Bin}(m, \theta)) - 2\text{Cov}(\text{Bin}(m, \theta), \mathbb{1}_{\{\text{Bin}(m, \theta) \geq 1\}}) + \text{Var}(\mathbb{1}_{\{\text{Bin}(m, \theta) \geq 1\}}) \\
&= m\theta(1 - \theta) - 2m\theta(1 - \theta)^m + (1 - \theta)^m(1 - (1 - \theta)^m) \\
&\leq m\theta(1 - \theta) - 2m\theta(1 - \theta)^m + m\theta(1 - \theta)^m \\
&= m\theta(1 - (1 - \theta)^m) - m\theta^2 \leq (m\theta)^2.
\end{aligned}$$

Here, we have used twice the inequality $1 - (1 - \theta)^m \leq \theta m$. \square

Proof of Lemma 6.3. In view of

$$\sum_{k=1}^{m_n} \mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}} - \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}} = \left(\sum_{k=1}^{m_n} \mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}} - 1 \right)_+,$$

relation (42) is equivalent to

$$\text{Var} \left(\sum_{m_n < p \leq n} \log p \left(\sum_{k=1}^{m_n} \mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}} - 1 \right)_+ \right) = O(m_n \log m_n).$$

We represent the left-hand side as follows:

$$\begin{aligned}
& \sum_{m_n < p \leq n} \log^2 p \cdot \text{Var} \left(\left(\sum_{k=1}^{m_n} \mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}} - 1 \right)_+ \right) + 2 \sum_{m_n < p < q \leq n} \log p \log q \times \\
& \text{Cov} \left(\left(\sum_{k=1}^{m_n} \mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}} - 1 \right)_+, \left(\sum_{k=1}^{m_n} \mathbb{1}_{\{\lambda_q(U_k^{(n)}) \geq 1\}} - 1 \right)_+ \right) =: A_1(n) + A_2(n).
\end{aligned}$$

Since the sum $\sum_{k=1}^{m_n} \mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}}$ has the binomial distribution with parameters m_n and $\mathbb{P}\{\lambda_p(U_k^{(n)}) \geq 1\} = n^{-1} \lfloor n/p \rfloor$, an application of Lemma 6.4 gives

$$(43) \quad A_1(n) \leq \sum_{p > m_n} \log^2 p (n^{-1} m_n \lfloor n/p \rfloor)^2 \leq m_n^2 \sum_{p > m_n} p^{-2} \log^2 p = O(m_n \log m_n),$$

where we have used (16) with $f(x) = x^{-2} \log^2 x$ for the last step.

Passing to the analysis of $A_2(n)$ we have to estimate the covariance: for prime $p \neq q$,

$$\begin{aligned}
 C_n(p, q) &:= \text{Cov}\left(\left(\sum_{k=1}^{m_n} \mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}} - 1\right)_+, \left(\sum_{k=1}^{m_n} \mathbb{1}_{\{\lambda_q(U_k^{(n)}) \geq 1\}} - 1\right)_+\right) \\
 &= \sum_{i,j=1}^{m_n} \text{Cov}\left(\mathbb{1}_{\{\lambda_p(U_i^{(n)}) \geq 1\}}, \mathbb{1}_{\{\lambda_q(U_j^{(n)}) \geq 1\}}\right) \\
 &\quad - \sum_{i=1}^{m_n} \text{Cov}\left(\mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}}, \mathbb{1}_{\{\lambda_q(U_i^{(n)}) \geq 1\}}\right) \\
 &\quad - \sum_{j=1}^{m_n} \text{Cov}\left(\mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_q(U_k^{(n)}) \geq 1\}}, \mathbb{1}_{\{\lambda_p(U_j^{(n)}) \geq 1\}}\right) \\
 &\quad + \text{Cov}\left(\mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) \geq 1\}}, \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_q(U_k^{(n)}) \geq 1\}}\right) \\
 &= \sum_{i,j=1}^{m_n} \text{Cov}\left(\mathbb{1}_{\{\lambda_p(U_i^{(n)}) \geq 1\}}, \mathbb{1}_{\{\lambda_q(U_j^{(n)}) \geq 1\}}\right) \\
 &\quad + \sum_{i=1}^{m_n} \text{Cov}\left(\mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) = 0\}}, \mathbb{1}_{\{\lambda_q(U_i^{(n)}) \geq 1\}}\right) \\
 &\quad + \sum_{j=1}^{m_n} \text{Cov}\left(\mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_q(U_k^{(n)}) = 0\}}, \mathbb{1}_{\{\lambda_p(U_j^{(n)}) \geq 1\}}\right) \\
 &\quad + \text{Cov}\left(\mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_p(U_k^{(n)}) = 0\}}, \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \lambda_q(U_k^{(n)}) = 0\}}\right).
 \end{aligned}$$

For typographical simplicity we shall use the following abbreviations until the end of the proof of (42):

$$a := n^{-1} \lfloor n/p \rfloor, \quad b := n^{-1} \lfloor n/q \rfloor, \quad \Delta := n^{-1} \lfloor n/(pq) \rfloor - n^{-2} \lfloor n/p \rfloor \lfloor n/q \rfloor.$$

Using (38) and (39) we obtain

$$\begin{aligned}
 C_n(p, q) &= m_n \Delta - m_n \Delta (1-a)^{m_n-1} - m_n \Delta (1-b)^{m_n-1} \\
 &\quad + ((1-a)(1-b) + \Delta)^{m_n} - (1-a)^{m_n} (1-b)^{m_n} \\
 &= m_n \Delta (1 - (1-a)^{m_n-1}) (1 - (1-b)^{m_n-1}) \\
 &\quad + \sum_{k=2}^{m_n} \binom{m_n}{k} \Delta^k (1-a)^{m_n-k} (1-b)^{m_n-k}.
 \end{aligned}$$

Therefore,

$$\begin{aligned}
|A_2(n)| &= \left| 2 \sum_{m_n < p < q \leq n} \log p \log q \cdot C_n(p, q) \right| \\
(44) \quad &\leq 2m_n \sum_{m_n < p < q \leq n} |\Delta| \log p \log q \cdot (1 - (1 - a)^{m_n - 1})(1 - (1 - b)^{m_n - 1}) \\
&\quad + 2 \sum_{k=2}^{m_n} \binom{m_n}{k} \sum_{m_n < p < q \leq n} \log p \log q |\Delta|^k.
\end{aligned}$$

The summands are bounded from above, respectively, by

$$2m_n^3 \sum_{m_n < p < q \leq n} p^{-2} q^{-2} \log p \log q \leq m_n^3 \left(\sum_{p > m_n} p^{-2} \log p \right)^2 = O(m_n)$$

and

$$\begin{aligned}
&2 \sum_{k=2}^{m_n} \binom{m_n}{k} \sum_{m_n < p < q \leq n} p^{-k} q^{-k} \log p \log q \leq \sum_{k=2}^{m_n} \binom{m_n}{k} \left(\sum_{p > m_n} p^{-k} \log p \right)^2 \\
&\leq \text{const} \times \sum_{k=2}^{m_n} \binom{m_n}{k} m_n^{2(1-k)} = \text{const} \times m_n^2 ((1 + m_n^{-2})^{m_n} - 1 - m_n^{-1}) = O(1),
\end{aligned}$$

where we have used $|\Delta| \leq (pq)^{-1}$, $a < p^{-1}$, $b < q^{-1}$ and the fact that

$$\sum_{p > m_n} p^{-k} \log p \leq C m_n^{1-k},$$

for some constant $C > 0$ which does not depend on k . The last estimate is justified by Lemma 7.1 in [1]. \square

Now we conclude with the help of (42), $a_n/(m_n \log m_n) \rightarrow \infty$ and Chebyshev's inequality that (41) is equivalent to (45).

For $1 \leq m < n$ and $k \in \mathbb{N}$, put

$$\tilde{U}_k^{(n)} = \prod_{p \leq n} p^{\mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}}} \quad \text{and} \quad \tilde{U}_k^{(n,m)} = \prod_{m < p \leq n} p^{\mathbb{1}_{\{\lambda_p(U_k^{(n)}) \geq 1\}}}.$$

Lemma 6.5. *Assume that $m_n \rightarrow \infty$ and $m_n = o(n)$ as $n \rightarrow \infty$. Then*

$$(45) \quad \left(a_n^{-1/2} \sum_{k=1}^{\lfloor m_n t \rfloor} \left(\log \tilde{U}_k^{(n,m_n)} - \mathbb{E} \log \tilde{U}_k^{(n,m_n)} \right) \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\text{f.d.d.}} (B(t))_{t \geq 0}$$

with a_n defined by (35).

Proof. The left-hand side in (45) is the sum of independent random variables. Thus, the proof of (45) boils down to showing convergence of covariances

$$(46) \quad a_n^{-1} \text{Cov} \left(\sum_{k=1}^{\lfloor m_n t \rfloor} \log \tilde{U}_k^{(n,m_n)}, \sum_{k=1}^{\lfloor m_n s \rfloor} \log \tilde{U}_k^{(n,m_n)} \right) \rightarrow \min(s, t)$$

for $s, t \geq 0$ and checking the Lindeberg–Feller condition, see Theorem 4.12 in [19],

$$(47) \quad \mathbb{E}(\log \tilde{U}_1^{(n, m_n)} - \mathbb{E} \log \tilde{U}_1^{(n, m_n)})^2 \mathbb{1}_{\{|\log \tilde{U}_1^{(n, m_n)} - \mathbb{E} \log \tilde{U}_1^{(n, m_n)}| > \varepsilon \sqrt{a_n}\}} \\ = o(a_n/m_n)$$

for all $\varepsilon > 0$. Formula (46) follows from

$$\text{Cov}\left(\sum_{k=1}^{\lfloor m_n t \rfloor} \log \tilde{U}_k^{(n, m_n)}, \sum_{k=1}^{\lfloor m_n s \rfloor} \log \tilde{U}_k^{(n, m_n)}\right) = \text{Var}\left(\sum_{k=1}^{\lfloor m_n \min(t, s) \rfloor} \log \tilde{U}_k^{(n, m_n)}\right) \\ = \lfloor m_n \min(t, s) \rfloor \text{Var}(\log \tilde{U}_1^{(n, m_n)}) \sim \min(t, s) a_n.$$

Further, we use the Cauchy–Schwartz inequality and the Markov inequality to obtain

$$\mathbb{E}(\log \tilde{U}_1^{(n, m_n)} - \mathbb{E} \log \tilde{U}_1^{(n, m_n)})^2 \mathbb{1}_{\{|\log \tilde{U}_1^{(n, m_n)} - \mathbb{E} \log \tilde{U}_1^{(n, m_n)}| > \varepsilon \sqrt{a_n}\}} \\ \leq \left(\mathbb{E}(\log \tilde{U}_1^{(n, m_n)} - \mathbb{E} \log \tilde{U}_1^{(n, m_n)})^4\right)^{1/2} \\ \times \left(\mathbb{P}\{|\log \tilde{U}_1^{(n, m_n)} - \mathbb{E} \log \tilde{U}_1^{(n, m_n)}| > \varepsilon \sqrt{a_n}\}\right)^{1/2} \\ \leq \left(\mathbb{E}(\log \tilde{U}_1^{(n, m_n)} - \mathbb{E} \log \tilde{U}_1^{(n, m_n)})^4\right)^{1/2} \varepsilon^{-1} a_n^{-1/2} (\text{Var}(\log \tilde{U}_1^{(n, m_n)}))^{1/2} \\ = O(m_n^{-1/2} \log^2 m_n) = o(1)$$

which proves (47) because $a_n/m_n \rightarrow \infty$. Here, the next to the last equality is justified by Lemma 5.6 and (36).

We note in passing that (47) holds trivially whenever

$$(48) \quad \lim_{n \rightarrow \infty} (a_n / \log^2 n) = \infty$$

which is particularly the case when m_n grows faster than $\log^2 n$. Observe that $|\log \tilde{U}_1^{(n, m_n)} - \mathbb{E} \log \tilde{U}_1^{(n, m_n)}| \leq 2 \log n$ a.s. as a consequence of $\log \tilde{U}_1^{(n, m_n)} \leq \log U_1^{(n)} \leq \log n$ a.s. Thus, under (48), the indicator in (47) is equal to 0 for large n , whence (47). \square

The proof of Theorem 3.6 is complete.

7. PROOF OF THEOREM 3.5

PROOF OF PART (I). We start by noting that whenever $m_n = o(n^{1/2})$, Theorem 3.6 immediately implies Theorem 3.5 because (12) yields

$$\lim_{n \rightarrow \infty} \mathbb{P}\{\log \text{lcm}(B_n(\lfloor m_n t \rfloor)) = Y_n(\lfloor m_n t \rfloor)\} = 1$$

for each $t \geq 0$.

In the situation that $m_n \neq o(n^{1/2})$ our proof relies on the following inequality

$$(49) \quad |Y_n(k) - Y_n(l)| \leq |k - l| \log n \quad \text{a.s.}$$

To prove it, we assume, without loss of generality, that $k > l$. Using the fact that, for finite sets $A, B \subset \mathbb{N}$,

$$\text{lcm}(A \cup B) \leq \text{lcm}(A) \times \text{lcm}(B),$$

we conclude that

$$\begin{aligned} 0 \leq Y_n(k) - Y_n(l) &= \log \text{lcm}(U_1^{(n)}, U_2^{(n)}, \dots, U_k^{(n)}) - \log \text{lcm}(U_1^{(n)}, U_2^{(n)}, \dots, U_l^{(n)}) \\ &\leq \log \text{lcm}(U_{l+1}^{(n)}, U_2^{(n)}, \dots, U_k^{(n)}) \leq \log \prod_{j=l+1}^k U_j^{(n)} \leq (k - l) \log n. \end{aligned}$$

Applying (49) with $k = \tau^{(n)}(\lfloor m_n t \rfloor)$ and $l = \lfloor m_n t \rfloor$ (note that $k \geq l$ in this case) and taking expectations we arrive at

$$\begin{aligned} \frac{\mathbb{E}(Y_n(\tau^{(n)}(\lfloor m_n t \rfloor)) - Y_n(\lfloor m_n t \rfloor))}{\sqrt{m_n} \log m_n} &\leq \frac{\mathbb{E}(\tau^{(n)}(\lfloor m_n t \rfloor) - \lfloor m_n t \rfloor) \log n}{\sqrt{m_n} \log m_n} \\ &= O\left(\frac{m_n^{3/2} \log n}{n \log m_n}\right), \end{aligned}$$

where the last equality follows from (11). It remains to note that in case (A) $m_n \leq n^{1/2}$ for all sufficiently large n and $m_n \rightarrow \infty$, so that the right-hand side of the last centered formula converges to zero in view of

$$\frac{m_n^{3/2} \log n}{n \log m_n} \leq n^{-1/4} \log n.$$

Thus,

$$0 \leq \frac{Y_n(\tau^{(n)}(\lfloor m_n t \rfloor)) - Y_n(\lfloor m_n t \rfloor)}{\sqrt{m_n} \log m_n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$$

which in combination with part (i) of Theorem 3.6 proves part (i) of Theorem 3.5.

PROOF OF PART (II). We appeal to (49) once again but now with $k = \tau^{(n)}(\lfloor m_n t \rfloor)$ and $l = \lfloor \mathbb{E} \tau^{(n)}(\lfloor m_n t \rfloor) \rfloor$. This gives

$$\begin{aligned} &\mathbb{E}|Y_n(\tau^{(n)}(\lfloor m_n t \rfloor)) - Y_n(\lfloor \mathbb{E} \tau^{(n)}(\lfloor m_n t \rfloor) \rfloor)| \\ &\leq \log n \mathbb{E}|\tau^{(n)}(\lfloor m_n t \rfloor) - \lfloor \mathbb{E} \tau^{(n)}(\lfloor m_n t \rfloor) \rfloor| \\ &\leq \log n + \log n (\text{Var}(\tau^{(n)}(\lfloor m_n t \rfloor)))^{1/2} = O(n^{-1/2} m_n \log n), \end{aligned}$$

where we have used (11) and $m_n > n^{1/2}$ for all large enough n . Therefore,

$$a_n^{-\frac{1}{2}} (Y_n(\tau^{(n)}(\lfloor m_n t \rfloor)) - Y_n(\lfloor \mathbb{E} \tau^{(n)}(\lfloor m_n t \rfloor) \rfloor)) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

by Markov's inequality as a consequence of

$$\frac{m_n^2 \log^2 n}{na_n} = \frac{2m_n \log^2 n}{n(\log n - \log m_n)(3 \log m_n - \log n)} \leq \frac{4m_n \log n}{n} \frac{1}{\log(n/m_n)}.$$

The first term is bounded because $m_n = O(n(\log n)^{-1})$ by assumption and the second tends to zero in view of $m_n = o(n)$.

It remains to prove the convergence

$$(50) \quad \left(a_n^{-1/2} (Y_n(\lfloor \mathbb{E}\tau^{(n)}(\lfloor m_n t \rfloor) \rfloor) - c_n(-n \log(1 - (m_n t)/n))) \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\text{f.d.d.}} (B(t))_{t \geq 0}.$$

Since $\lfloor \mathbb{E}\tau^{(n)}(\lfloor m_n t \rfloor) \rfloor = O(m_n)$, Lemma 6.1 implies that

$$\mathbb{E} \left(\sup_{t \in [0, T]} \left(Y_n(\lfloor \mathbb{E}\tau^{(n)}(\lfloor m_n t \rfloor) \rfloor) - Z_n(m_n^{-1} \mathbb{E}\tau^{(n)}(\lfloor m_n t \rfloor)) \right) \right) = O(m_n^{1/2})$$

for every fixed $T > 0$, where the definition of Z_n , see (32), has to be recalled.

Thus, (50) follows if we can check

$$(51) \quad \left(a_n^{-1/2} (Z_n(m_n^{-1} \mathbb{E}\tau^{(n)}(\lfloor m_n t \rfloor)) - c_n(-n \log(1 - (m_n t)/n))) \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\text{f.d.d.}} (B(t))_{t \geq 0}.$$

For fixed n , the function c_n defined in (6) is increasing and subadditive on $(0, \infty)$, whence

$$\begin{aligned} 0 \leq c_n(x+y) - c_n(x) &\leq c_n(y) \leq \sum_{p \leq n} \log p (1 - (1 - n^{-1} \lfloor n/p \rfloor)^y) \\ &\leq y \sum_{p \leq n} p^{-1} \log p \leq \text{const} \times y \log n \end{aligned}$$

and, see (10),

$$\lfloor \mathbb{E}\tau^{(n)}(\lfloor m_n t \rfloor) \rfloor = \lfloor n(H_n - H_{n - \lfloor m_n t \rfloor}) \rfloor = -n \log(1 - (m_n t)/n) + O(1).$$

Therefore, the equality $\mathbb{E}Z_n(t) = c_n(\lfloor m_n t \rfloor)$ shows that convergence (51) is equivalent to

$$(52) \quad \left(a_n^{-1/2} (Z_n(m_n^{-1} \mathbb{E}\tau^{(n)}(\lfloor m_n t \rfloor)) - \mathbb{E}Z_n(m_n^{-1} \mathbb{E}\tau^{(n)}(\lfloor m_n t \rfloor))) \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\text{f.d.d.}} (B(t))_{t \geq 0}.$$

Using once again the estimate $\lfloor \mathbb{E}\tau^{(n)}(\lfloor m_n t \rfloor) \rfloor = O(m_n)$, we can apply Lemmas 6.2 and 6.3 to deduce that (52) is implied by

$$(53) \quad \left(a_n^{-1/2} \times \sum_{k=1}^{\lfloor \mathbb{E}\tau^{(n)}(\lfloor m_n t \rfloor) \rfloor} (\log \tilde{U}_k^{(n, m_n)} - \mathbb{E} \log \tilde{U}_k^{(n, m_n)}) \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\text{f.d.d.}} (B(t))_{t \geq 0}.$$

The latter relation follows from Lemma 6.5 and

$$\begin{aligned} & \text{Var} \left(\sum_{k=\lfloor m_n t \rfloor + 1}^{\lfloor \mathbb{E} \tau^{(n)}(\lfloor m_n t \rfloor) \rfloor} (\log \tilde{U}_k^{(n, m_n)} - \mathbb{E} \log \tilde{U}_k^{(n, m_n)}) \right) \\ & \leq (\mathbb{E} \tau^{(n)}(\lfloor m_n t \rfloor) - \lfloor m_n t \rfloor) \text{Var}(\log \tilde{U}_1^{(n, m_n)}) = O(n^{-1} m_n^2 \text{Var}(\log \tilde{U}_1^{(n, m_n)})) \end{aligned}$$

which, in view of Chebyshev's inequality, (36) and $m_n = o(n)$, ensures

$$a_n^{-1/2} \times \sum_{k=\lfloor m_n t \rfloor + 1}^{\lfloor \mathbb{E} \tau^{(n)}(\lfloor m_n t \rfloor) \rfloor} (\log \tilde{U}_k^{(n, m_n)} - \mathbb{E} \log \tilde{U}_k^{(n, m_n)}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

8. APPENDIX

We intend to show that the normalization in a limit theorem for $\hat{Y}_n(\lfloor m_n t \rfloor)$ defined in (9) is different from that in Theorem 3.6 unless $\lim_{n \rightarrow \infty} (\log n / \log m) = 1$. We confine ourselves with the one-dimensional convergence.

Proposition 8.1. *Assume that $m_n \rightarrow \infty$ and $m_n = o(n)$ as $n \rightarrow \infty$. Then*

$$(54) \quad \frac{\hat{Y}_n(m_n) - \sum_{p \leq n} \log p (1 - (1 - p^{-1})^{m_n})}{\sqrt{2^{-1} m_n (\log^2 n - \log^2 m_n)}} \xrightarrow[n \rightarrow \infty]{\text{d}} B(1),$$

where $B(1)$ has a standard normal distribution.

Proof. Repeating verbatim the argument used in the proof of Lemma 6.1 we obtain a counterpart of the limit relation stated in that lemma

$$\mathbb{E} \sum_{p \leq n} \log p \left(\max_{1 \leq k \leq \lfloor m_n t \rfloor} \mathcal{G}_k(p) - \mathbb{1}_{\{\max_{1 \leq k \leq \lfloor m_n t \rfloor} \mathcal{G}_k(p) \geq 1\}} \right) = O(m_n^{1/2}).$$

Thus, it is enough to prove (54) with $\hat{Y}_n(m_n)$ replaced by

$$\hat{Z}_n(1) = \sum_{p \leq n} \log p \cdot \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \mathcal{G}_k(p) \geq 1\}}.$$

The centering in (54) is just the expectation of $\hat{Z}_n(1)$. To calculate the variance we argue as follows

$$\begin{aligned} \text{Var}(Z_n(1)) &= \sum_{p \leq n} \log^2 p (1 - (1 - p^{-1})^{m_n}) (1 - p^{-1})^{m_n} \\ &= \sum_{m_n < p \leq n} \log^2 p (1 - (1 - p^{-1})^{m_n}) (1 - p^{-1})^{m_n} + O(m_n \log m_n) \\ &= \sum_{m_n < p \leq n} \log^2 p (1 - (1 - p^{-1})^{m_n}) + O(m_n \log m_n), \end{aligned}$$

where we have used (16) to obtain

$$\sum_{p \leq m_n} \log^2 p (1 - (1 - p^{-1})^{m_n})(1 - p^{-1})^{m_n} \leq \sum_{p \leq m_n} \log^2 p = O(m_n \log m_n)$$

and

$$\sum_{m_n < p \leq n} \log^2 p (1 - (1 - p^{-1})^{m_n})^2 \leq m_n^2 \sum_{m_n < p} p^{-2} \log^2 p = O(m_n \log m_n).$$

Further,

$$\begin{aligned} m_n \sum_{m_n < p \leq n} p^{-1} \log^2 p - 2^{-1} m_n (m_n - 1) \sum_{m_n < p \leq n} p^{-2} \log^2 p \\ \leq \sum_{m_n < p \leq n} \log^2 p (1 - (1 - p^{-1})^{m_n}) \leq m_n \sum_{m_n < p \leq n} p^{-1} \log^2 p \end{aligned}$$

and

$$m_n^2 \sum_{m_n < p \leq n} p^{-2} \log^2 p = O(m_n \log m_n)$$

by (43). It remains to note that

$$m_n \sum_{m_n < p \leq n} p^{-1} \log^2 p \sim 2^{-1} m_n (\log^2 n - \log^2 m_n)$$

by (16) and

$$(m_n \log m_n) / (m_n (\log^2 n - \log^2 m_n)) \leq 1 / \log(n/m_n) \rightarrow 0.$$

Put $b_n := 2^{-1} m_n (\log^2 n - \log^2 m_n)$ and

$$V_p(n) := \mathbb{1}_{\{\max_{1 \leq k \leq m_n} \mathcal{G}_k(p) \geq 1\}} - \mathbb{P} \left\{ \max_{1 \leq k \leq m_n} \mathcal{G}_k(p) \geq 1 \right\}$$

for $n \in \mathbb{N}$ and $p \in \mathcal{P}$. The Lindeberg–Feller condition, see Theorem 4.12 in [19],

$$\sum_{p \leq n} \log^2 p \cdot \mathbb{E} V_p^2(n) \mathbb{1}_{\{\log p |V_p(n)| > \varepsilon \sqrt{b_n}\}} = o(b_n)$$

for all $\varepsilon > 0$ is trivial because $\log p |V_p(n)| \leq \log n$ for $p \leq n$ and $m_n (\log^2 n - \log^2 m_n)$ grows faster than $\log^2 n$. The latter follows from

$$m_n (\log^2 n - \log^2 m_n) / \log^2 n \geq m_n \log(n/m_n) / \log n \geq \log(n/m_n) \rightarrow \infty$$

in the case $m_n \geq \lfloor \log n \rfloor$ and $m_n = o(n)$ and

$$m_n (\log^2 n - \log^2 m_n) / \log^2 n \geq m_n \log(n/m_n) / \log n \sim m_n \rightarrow \infty$$

in the case $m_n < \lfloor \log n \rfloor$ and $m_n \rightarrow \infty$. □

Now we use Theorem 3.6 and Proposition 8.1 to conclude that $\lim_{n \rightarrow \infty} (b_n/a_n) = 1$ if, and only if, $\lim_{n \rightarrow \infty} (\log n / \log m_n) = 1$. Recall that a_n is defined in (35). If, for instance, $m_n \leq n^{1/2}$ and $m_n \rightarrow \infty$, then $a_n = 2^{-1} m_n \log^2 m_n$ so that $b_n/a_n \geq 3$. In particular, $\lim_{n \rightarrow \infty} (b_n/a_n) = \infty$ whenever $\lim_{n \rightarrow \infty} (\log n / \log m) = \infty$.

Here is the promised sketch of the proof of Proposition 3.3.

Proof of Proposition 3.3. It is enough to prove the joint convergence

$$\begin{aligned} & \left(\left(\log \operatorname{lcm}(U_1^{(n)}, U_2^{(n)}, \dots, U_m^{(n)}) - \sum_{k=1}^m \log U_k^{(n)} \right)_{m \in \mathbb{N}}, (n^{-1} U_m^{(n)})_{m \in \mathbb{N}} \right) \xrightarrow[n \rightarrow \infty]{d, \infty} \\ & \left(\left(\sum_p \log p \cdot \left(\max_{1 \leq k \leq m} \mathcal{G}_k(p) - \sum_{k=1}^m \mathcal{G}_k(p) \right) \right)_{m \in \mathbb{N}}, (U_m)_{m \in \mathbb{N}} \right). \end{aligned}$$

Note that, for every fixed $M \in \mathbb{N}$,

$$\begin{aligned} & \log \operatorname{lcm}(U_1^{(n)}, U_2^{(n)}, \dots, U_m^{(n)}) - \sum_{k=1}^m \log U_k^{(n)} \\ &= \sum_{p \leq n} \log p \cdot \left(\max_{1 \leq k \leq m} \lambda_p(U_k^{(n)}) - \sum_{k=1}^m \lambda_p(U_k^{(n)}) \right) \\ &= \sum_{p \leq M} \dots + \sum_{M < p \leq n} \dots =: W_1(n, m, M) + W_2(n, m, M). \end{aligned}$$

By the continuous mapping theorem applied to part (ii) of Lemma 3.1,

$$\begin{aligned} & \left((W_1(n, m, M))_{m \in \mathbb{N}}, (n^{-1} U_m^{(n)})_{m \in \mathbb{N}} \right) \xrightarrow[n \rightarrow \infty]{d, \infty} \\ & \left(\left(\sum_{p \leq M} \log p \cdot \left(\max_{1 \leq k \leq m} \mathcal{G}_k(p) - \sum_{k=1}^m \mathcal{G}_k(p) \right) \right)_{m \in \mathbb{N}}, (U_m)_{m \in \mathbb{N}} \right) \end{aligned}$$

for every fixed $M \in \mathbb{N}$. Further, using Theorem 2.3(i) in [8] with $f(x) = x$ yields

$$\begin{aligned} & \lim_{M \rightarrow \infty} \sum_{p \leq M} \log p \cdot \left(\max_{1 \leq k \leq m} \mathcal{G}_k(p) - \sum_{k=1}^m \mathcal{G}_k(p) \right) \\ &= \sum_p \log p \cdot \left(\max_{1 \leq k \leq m} \mathcal{G}_k(p) - \sum_{k=1}^m \mathcal{G}_k(p) \right) \quad \text{a.s.} \end{aligned}$$

It remains to invoke Theorem 3.2 in [6] in conjunction with the relation

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\{|W_2(n, m, M)| > \varepsilon\} = 0$$

for every fixed $\varepsilon > 0$ and $m \in \mathbb{N}$. For the latter, see formula (22) in [8]. \square

ACKNOWLEDGMENTS

The work of AI and AM has received funding from the Ulam Program of the Polish National Agency for Academic Exchange (NAWA), projects no. PPN/ULM/2019/1/00010/DEC/1 and PPN/ULM/2019/1/00004/DEC/1, respectively. DB was partially supported by the National Science Center, Poland (grant number 2019/33/ B/ST1/00207).

REFERENCES

- [1] ALSMEYER, G., KABLUCHKO, Z. AND MARYNYCH, A. (2019). Limit theorems for the least common multiple of a random set of integers. *Trans. Amer. Math. Soc.* **372**(7), 4585–4603.
- [2] ARRATIA, R. (2002). On the amount of dependence in the prime factorization of a uniform random integer. *Contemporary combinatorics*, 29–91, Bolyai Soc. Math. Stud., 10, János Bolyai Math. Soc., Budapest.
- [3] ARRATIA, R., BARBOUR, A. D. AND TAVARÉ, S. (2003). *Logarithmic Combinatorial Structures: a Probabilistic Approach*. EMS Monographs in Mathematics. European Mathematical Society.
- [4] ARRATIA, R. AND TAVARÉ, S. (1992). Limit theorems for combinatorial structures via discrete process approximations. *Random Struct. Alg.* **3**, 321–345.
- [5] BILLINGSLEY, P. (1974). The probability theory of additive arithmetic functions. *Ann. Probab.* **5**(2), 749–791.
- [6] BILLINGSLEY, P. (1999). *Convergence of probability measures*. Second edition. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons.
- [7] BINGHAM, N. H., GOLDIE, C. M. AND TEUGELS, J. L. (1989). *Regular variation*. Encyclopedia of Mathematics and its Applications, 27. Cambridge University Press.
- [8] BOSTAN, A., MARYNYCH, A. AND RASCHEL, K. (2019). On the least common multiple of several random integers. *J. Numb. Theory*, **204**, 113–133.
- [9] CESÀRO, E. (1884). Probabilités de certains faits arithmétiques. *Mathesis*, **4**, 150–151.
- [10] CESÀRO, E. (1885). Sur le plus grand commun diviseur de plusieurs nombres. *Ann. Mat. Pura Appl.*, **13**, 291–294.
- [11] CESÀRO, E. (1885). Étude moyenne du plus grand commun diviseur de deux nombres. *Ann. Mat. Pura Appl.*, **13**, 235–250.

- [12] CILLERUELO, J., RUÉ, J., ŠARKA, P. AND ZUMALACÁRREGUI, A. (2014). The least common multiple of random sets of positive integers. *J. Numb. Theory*, **144**, 92–104.
- [13] DIRICHLET, G. L. (1849). Über die Bestimmung der mittleren Werthe in der Zahlentheorie. *Abhandlungen der Königlich Preussischen Akademie der Wissenschaften*, 69–83. Werke, Zweiter Band, ed. L. KRONECKER, 1897, 51–66. Reprint by Chelsea Publishing Co., 1969.
- [14] ERDŐS, P., AND KAC, M. (1940). The Gaussian law of errors in the theory of additive number theoretic functions. *American Journal of Mathematics*, **62**(1), 738–742.
- [15] ERDŐS, P. AND TURÁN, P. (1967). On some problems of statistical group theory III. *Acta. Math. Acad. Sci. Hungar.* **18**, 309–320.
- [16] FERNÁNDEZ, J. AND FERNÁNDEZ, P. (2013). On the probability distribution of the gcd and lcm of r -tuples of integers. Preprint available at <https://arxiv.org/abs/1305.0536>.
- [17] GNEDIN, A., IKSANOV, A AND MARYNYCH, A. (2012). A generalization of the Erdős-Turán law for the order of random permutation. *Combinatorics, Probability and Computing*. **21**, 715–733.
- [18] HILBERDINK, T. AND TÓTH, L. (2016). On the average value of the least common multiple of k positive integers. *J. Numb. Theory*, **169**, 327–341.
- [19] KALLENBERG, O. (1997). *Foundations of Modern Probability*. Springer: Probability and Its Applications.
- [20] KUBILIUS, J. (1964). *Probabilistic methods in the theory of numbers* (Vol. 11). American Mathematical Soc.
- [21] SANNA, C. (2020). On the lcm of random terms of binary recurrence sequences. *Journal of Number Theory*. In press.
- [22] STORM, J. AND ZEINDLER, D. (2015). The order of large random permutations with cycle weights. *Electronic Journal of Probability*, paper no. 126.

DARIUSZ BURACZEWSKI, MATHEMATICAL INSTITUTE UNIVERSITY OF WROCLAW, PL. GRUNWALDZKI 2/4 50-384 WROCLAW, POLAND

Email address: `dbura@math.uni.wroc.pl`

ALEXANDER IKSANOV, FACULTY OF COMPUTER SCIENCE AND CYBERNETICS, TARAS SHEVCHENKO NATIONAL UNIVERSITY OF KYIV, 01601 KYIV, UKRAINE

Email address: `iksan@univ.kiev.ua`

ALEXANDER MARYNYCH, FACULTY OF COMPUTER SCIENCE AND CYBERNETICS, TARAS SHEVCHENKO NATIONAL UNIVERSITY OF KYIV, 01601 KYIV, UKRAINE

Email address: `marynych@unicyb.kiev.ua`