POISSON–DIRICHLET APPROXIMATION FOR COUNTING INTEGERS WITH DIVISORS IN AN INTERVAL

TONY HADDAD

ABSTRACT. We give a simple inequality that compares the laws of two random variables taking values in a convex set of any normed vector space. By combining it with Arratia's coupling, recently refined by Koukoulopoulos and the author, we present a general strategy to reduce the problem of finding an asymptotic formula for the number of integers whose prime factorization falls into any given subset of $\ell^1(\mathbb{R})$, to finding upper bounds for two key probabilities measuring proximity to the boundary of the subset in question.

We apply this strategy to give an asymptotic formula for counting integers in $[1, x]$ that have a divisor in an interval (y, z) with $z/y \to \infty$ as $x \to \infty$.

1. INTRODUCTION

Let $H(x, y, z)$ be the number of integers in $[1, x]$ having a divisor in the open interval (y, z). The size of this function was studied in various ranges over the last century. In 1934, Besicovitch [2] showed that

$$\liminf_{y\to\infty} \lim_{x\to\infty} \frac{H(x, y, z)}{x} = 0.$$

when $z = 2y$. In 1935, Erdős [4] replaced the $\liminf$ by $\lim$ in the statement above, and he later showed in [5] a similar statement for any $z = z(y)$ which satisfies $\frac{\log z/y}{\log y} \to 0$ as $y \to \infty$. It was later established by Erdős in 1960 that

$$\lim_{x\to\infty} \frac{H(x, y, 2y)}{x} = (\log y)^{-\delta+o(1)}$$

with $\delta := 1 - \frac{1+\log_2 2}{\log 2} = 0.086\ldots$ as $y \to \infty$.

In 1980, Tenenbaum [11] proved that the following limit always converges for all $0 \leqslant u < v \leqslant 1$:

$$h(u, v) := \lim_{x\to\infty} \frac{H(x, x^u, x^v)}{x}. \tag{1}$$

In 1984, Tenenbaum [12] gave precise upper and lower bounds for $H(x, y, z)$ in various ranges of (x, y, z), and in 2008, Ford [7] was able to determine the order of magnitude of $H(x, y, z)$ for all (x, y, z) satisfying $2 \leqslant y < z \leqslant x$. This result in one key range obtained in [7] is the following:

$$H(x, y, z) \asymp x w^{\delta} (\log 2/w)^{-3/2} \tag{2}$$

whenever $2y \leqslant z \leqslant \min\{y^2, x\}$, $100 \leqslant y \leqslant x^{2/3}$ and x is sufficiently large and with $w := \frac{\log z/y}{\log y}$.

In (1), we have an asymptotic formula for $H(x, x^u, x^v)$ as $x \to \infty$ when u and v stay fixed, but if we were to allow u and v to vary with x, all we have is the order of magnitude. The main theorem of this paper gives the range of (x, u, v) for which the density $h(u, v)$ approximates well the function $H(x, x^u, x^v)$:

Date: December 15, 2025.

Theorem 1. *Let* $\delta := 1 - \frac{1+\log_2 2}{\log 2}$*. We have*

$$H(x,y,z) = x \cdot h\big(\tfrac{\log y}{\log x}, \tfrac{\log z}{\log x}\big) + O\left(\frac{x}{(\log y)^{\delta}(\log_2 y)^{3/2}}\right)$$

for all $3 \leqslant y < z \leqslant \sqrt{x}$*.*

By using (2), this theorem gives an asymptotic when $x \to \infty$ if y and z are functions of x and $z/y \to \infty$.

In the same paper where Tenenbaum proved that the limit $h(u,v)$ exists [11], he also gave an explicit formula for it in the range $(u,v) \in [0, \frac{2}{5}] \times \{\frac{1}{2}\}$. We give an explicit formula for $h(u,v)$ for any $u, v \in [0,1]$ satisfying $u < v$:

Theorem 2. *For any fixed* $u, v \in [0,1]$ *satisfying* $u < v$*, let* $k := \lfloor \frac{1}{v-u} \rfloor$*, and let* Δ^k *be the simplex* $\{(x_1, \dots, x_k) \in \mathbb{R}^k : x_1 \geqslant \cdots \geqslant x_k \geqslant 0$ *and* $\sum_i x_i \leqslant 1\}$*. Let* $\rho \colon [0,\infty) \to \mathbb{R}$ *be the Dickman function defined by the delay differential equation* $u\rho'(u) = -\rho(u-1)$ *for* $u > 1$*, and the initial condition* $\rho([0,1]) = \{1\}$*. For any family* $\mathcal{E}$ *in the power set* $2^{2^{\mathbb{N}\cap[1,k]}}$*, we also define the compact convex polytope* $P_{u,v}(\mathcal{E})$ *as*

$$P_{u,v}(\mathcal{E}) := \left\{ \mathbf{x} \in \Delta^k : \sum_{i \in E} x_i \geqslant v \text{ for all } E \in \mathcal{E}, \text{ and } \sum_{i \in E} x_i \geqslant 1 - u \text{ for all } E \in 2^{\mathbb{N}\cap[1,k]} \setminus \mathcal{E} \right\}.$$

Then we have

$$h(u,v) = 1 - \sum_{\mathcal{E} \in 2^{2^{\mathbb{N}\cap[1,k]}}} \int \cdots \int_{(t_1,\dots,t_k) \in P_{u,v}(\mathcal{E})} \rho\left(\frac{1 - \sum_{i=1}^{k} t_i}{t_k}\right) \frac{\mathrm{d}t_1 \cdots \mathrm{d}t_k}{t_1 \cdots t_k}$$

for every $u, v \in [0,1]$ *with* $u < v$*.*

It is not a coincidence that the integral in this theorem involves the joint density of the first k terms of a Poisson–Dirichlet process. In fact, the function $H(x,y,z)$ counts integers with some particular property on the size of all its prime factors, and as we will see, the Poisson–Dirichlet process is a good model for the prime factorization of a random integer.

At the end of the next section, we will have reduced Theorem 1 and Theorem 2 into a few key lemmas. We will first explain a general strategy to make this reduction possible on a family of problems which includes the theorems above.

Notation. In some parts of the paper, we will use the unique factorization of positive integers $n = n^\flat n^\sharp$ with $n^\flat$ being squarefree, $n^\sharp$ being squarefull, and $\gcd(n^\flat, n^\sharp) = 1$. The variable p is always reserved for prime numbers unless stated otherwise. We let $\log_j$ denote the j-iteration of the natural logarithm, meaning that $\log_1 = \log$ and $\log_j = \log \circ \log_{j-1}$ for $j \geqslant 2$. The norm $\|\cdot\|$ will always be the ℓ^1-norm, except in the statement and proof of Proposition 2.1 which considers $\|\cdot\|$ to be any arbitrary norm. If P is some proposition, then the indicator $\mathbb{1}_P$ will be equal to 1 if P is true, and 0 if P is false. To describe various estimates, we use Vinogradov's notation $f(x) \ll g(x)$ or Landau's notation $f(x) = O(g(x))$ to mean that $|f(x)| \leqslant C \cdot g(x)$ for a positive constant C. If C depends on a parameter α, we write $f(x) \ll_\alpha g(x)$ or $f(x) = O_\alpha(g(x))$.

Acknowledgments. The author thanks Dimitris Koukoulopoulos, Kaisa Matomäki and Gérald Tenenbaum for helpful comments on this project. The author was supported by the Courtois Chair II in fundamental research at the start of the project, and is now supported by the Research Council of Finland grant no. 364214.

2. Reduction to boundary lemmas

2.1. **Theorem 1 and 2 under a general framework.** Let $\ell^1(\mathbb{R})$ take its usual definition: It is the Banach space of real sequences $\mathbf{x} = (x_i)_{i\geqslant 1}$ with the property that $\sum_{i\geqslant 1}|x_i| < \infty$. We equip it with the ℓ^1-norm defined by $\|\mathbf{x}\| := \sum_{i\geqslant 1}|x_i|$. We define Δ to be the set of sequences $\mathbf{x}\in\ell^1(\mathbb{R})$ satisfying both $x_1 \geqslant x_2 \geqslant \cdots \geqslant 0$ and $\|\mathbf{x}\| \leqslant 1$.

For any positive integer n, we set $(p_i(n))_{i\geqslant 1}$ to be the unique sequence satisfying the factorization $n = \prod_{i=1}^{\infty} p_i(n)$, with $p_1(n) \geqslant p_2(n) \geqslant \cdots$ all being primes or ones. The following sequence of relative sizes of prime factors

$$\mathbf{Primes}(n,x) := \left(\frac{\log p_i(n)}{\log x}\right)_{i\geqslant 1}$$

is in the set Δ for any n, x satisfying $1 \leqslant n \leqslant x$. We assign two probability measures on the measurable space $(\Delta, \mathcal{B}(\Delta))$ with $\mathcal{B}(\Delta)$ being the Borel σ-algebra of Δ: For any fixed $x \geqslant 2$, let ν_x be the probability measure defined by

$$\nu_x(A) := \frac{1}{\lfloor x\rfloor}\cdot\#\Big\{1 \leqslant n \leqslant x : \mathbf{Primes}(n,x)\in A\Big\}.$$

for any set $A \in \mathcal{B}(\Delta)$. Let μ be the Poisson–Dirichlet distribution of parameter 1. One way to define this probability measure is with its finite-dimensional distributions[1]:

$$\mu(Q_k(\mathbf{u})) := \int_{u_k}^{\infty}\cdots\int_{u_1}^{\infty}\frac{\mathbb{1}_{0\leqslant t_1\leqslant\cdots\leqslant t_k}\cdot\mathbb{1}_{t_1+\cdots+t_k\leqslant 1}}{t_1\cdots t_k}\rho\left(\frac{1-\sum_{i=1}^k t_i}{t_k}\right)\,\mathrm{d}t_1\cdots\mathrm{d}t_k$$

for all $k \geqslant 1$, $\mathbf{u} = (u_1,\ldots,u_k)\in\mathbb{R}^k$ and with $Q_k(\mathbf{u})$ being the set of real sequences $(x_i)_{i\geqslant 1}$ with $x_i \geqslant u_i$ for all $1 \leqslant i \leqslant k$. In the above definition, the function $\rho\colon [0,\infty)\to\mathbb{R}$ is again the Dickman function as defined in Theorem 2. It is important to keep in mind that this measure is supported on $\Delta^* := \{\mathbf{x}\in\Delta : \|\mathbf{x}\| = 1\}$. Other equivalent definitions of μ are given in [6, Chapter 2].

For $0 \leqslant u < v \leqslant 1$, we consider the set $D(u,v)$ of sequences $\mathbf{x}\in\Delta$ that have a subsum in the open interval (u,v). It becomes clear that this set is open in Δ if we rewrite it as a union of the preimage of continuous functions over open intervals

$$D(u,v) = \bigcup_{E\subseteq\mathbb{N}}\phi_E^{-1}\big((u,v)\big), \tag{3}$$

with each $\phi_E\colon \Delta\to\mathbb{R}$ being the continuous functions defined by $\phi_E(\mathbf{x}) := \sum_{i\in E} x_i$. By the definition of the measure ν_x, we have

$$H(x,y,z) = \lfloor x\rfloor\cdot\nu_x\Big(D\big(\tfrac{\log y}{\log x},\tfrac{\log z}{\log x}\big)\Big), \tag{4}$$

and as we will prove at the end of the section, we have

$$h(u,v) = \mu\big(D(u,v)\big). \tag{5}$$

[1]Note that finite-dimensional distributions of stochastic processes allow us to compute probabilities of events in cylindrical σ-algebras. In the case of $\ell^1(\mathbb{R})$, any closed balls of the form $\overline{B}(\mathbf{y},r) := \{\mathbf{x}\in\ell^1(\mathbb{R}) : \|\mathbf{x}-\mathbf{y}\| \leqslant r\}$ for some $\mathbf{y}\in\ell^1(\mathbb{R})$ and $r > 0$ can be generated by the countable intersection $\bigcap_{k=1}^{\infty} E_k(\mathbf{y},r)$ with $E_k(\mathbf{y},r)$ being the cylindrical set of real sequences $(x_i)_{i\geqslant 1}$ satisfying $\sum_{i=1}^k |x_i - y_i| \leqslant r$. Since $\ell^1(\mathbb{R})$ is a separable space, any open set of $\ell^1(\mathbb{R})$ can be generated by closed balls, and therefore finite-dimensional distributions of the probability measure μ makes it well-defined on $\mathcal{B}(\Delta)$.

We will show how to appropriately bound $|\nu_x(A) - \mu(A)|$ for any Borel set A.

2.2. **Comparing laws of random variables on normed vector spaces.** Before moving on to approximating $\nu_x(A)$, we first present a simple inequality which will be fundamental in our approach:

Proposition 2.1. *Let $\mathcal{W}$ be a convex set in any normed vector space. For any subset $A \subseteq \mathcal{W}$, let ∂A be its boundary with respect to the subspace topology of $\mathcal{W}$. For any $\eta \geqslant 0$, we write $\|w - \partial A\| := \inf_{z \in \partial A} \|w - z\|$ and*

$$\partial_\eta A := \Big\{ w \in \mathcal{W} : \|w - \partial A\| < \eta \Big\}.$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be any probability space. For any three random variables X, Y, R on this space with both X and Y taking values in $\mathcal{W}$, and with R taking values in $\mathbb{R}_{\geqslant 0}$, we have

$$\Big| \mathbb{P}\big[X \in A\big] - \mathbb{P}\big[Y \in A\big] \Big| \leqslant \mathbb{P}\Big[\{X, Y\} \subseteq \partial_R A\Big] + \mathbb{P}\Big[\|X - Y\| \geqslant R\Big],$$

for any Borel sets $A \subseteq \mathcal{W}$.

Proof. First, we have

$$\begin{aligned} \Big| \mathbb{P}\big[X \in A\big] - \mathbb{P}\big[Y \in A\big] \Big| &\leqslant \Big| \mathbb{P}\big[X \in A, Y \notin A\big] - \mathbb{P}\big[X \notin A, Y \in A\big] \Big| \\ &\leqslant \max \Big\{ \mathbb{P}\big[X \in A, Y \notin A\big],\ \mathbb{P}\big[X \notin A, Y \in A\big] \Big\}. \end{aligned}$$

For any two points of $a, b \in \mathcal{S}$, let $\mathfrak{L}(a, b)$ be the line segment between a, b. Suppose that exactly one point of $\{a, b\}$ belong to some set $A \subseteq \mathcal{S}$, then $\mathfrak{L}(a, b)$ must intersect ∂A since any line segment is connected. Thus, there exists $c \in \mathfrak{L}(a, b) \cap \partial A$ satisfying

$$\|a - \partial A\| + \|b - \partial A\| \leqslant \|a - c\| + \|c - b\| = \|a - b\|.$$

Therefore,

$$\Big| \mathbb{P}\big[X \in A\big] - \mathbb{P}\big[Y \in A\big] \Big| \leqslant \mathbb{P}\Big[\|X - \partial A\| + \|Y - \partial A\| \leqslant \|X - Y\|\Big].$$

The proposition follows directly from this. ☐

2.3. **Approximating prime factorizations by a Poisson–Dirichlet process.** In 1972, Billingsley [3] proved that for any fixed $\mathbf{u} = (u_1, \ldots, u_k) \in [0, 1]^k$, we have

$$\nu_x\big(\Delta \cap Q_k(\mathbf{u})\big) = \mu\big(\Delta \cap Q_k(\mathbf{u})\big) + o(1)$$

as $x \to \infty$, where $Q_k(\mathbf{u})$ is again the set of real sequences $(x_i)_{i \geqslant 1}$ with $x_i \geqslant u_i$ for all $1 \leqslant i \leqslant k$. In other words, the k largest prime factors of a random integer uniformly distributed in $\mathbb{N} \cap [1, x]$ converge in distribution to the first k components of a Poisson–Dirichlet process of parameter 1. Later, research has been done on understanding the speed of convergence in Billingsley's Theorem. In 1976, Knuth and Trabb Pardo [10] studied the case where we fix $u_1 = \cdots = u_{k-1} = 0$ and $u_k \in [0, 1]$ (corresponding to the distribution of the k^{th} largest prime factor). In 2000, Tenenbaum [13] gave the following asymptotic expansion for the error term in Billingsley's Theorem: There exists a sequence of functions $(\varphi_j)_{j \geqslant 1}$ such that

$$\nu_x\big(\Delta \cap Q_k(\mathbf{u})\big) = \mu\big(\Delta \cap Q_k(\mathbf{u})\big) + \sum_{j=1}^{m} \frac{\varphi_j(\mathbf{u})}{(\log x)^j} + O_{\mathbf{u}, m}\bigg(\frac{1}{(\log x)^{m+1}}\bigg) \tag{6}$$

for all fixed m and $\mathbf{u} \in [0,1]^k$.[2]

In 2002, Arratia went in different direction to quantify Billinglsey's Theorem. He proved in [1] that there exists a coupling between a random integer N_x uniformly distributed in $\mathbb{N} \cap [1,x]$, and a Poisson–Dirichlet process $\mathbf{V} = (V_1, V_2, \dots)$ of parameter 1 such that the ℓ^1-norm $\|\mathbf{Primes}(N_x, x) - \mathbf{V}\|$ is in expectation $O(\frac{\log_2 x}{\log x})$. In 2025, Koukoulopoulos and the author [8] modified his coupling to improve this expectation to $O(\frac{1}{\log x})$. Note that in these couplings, the law of $\mathbf{Primes}(N_x, x)$ is ν_x and the law of $\mathbf{V}$ is μ.

In [8], we actually gave the following stronger result: Let $(\Omega_\star, \mathcal{F}_\star, \mathbb{P}_\star)$ be the coupling constructed in [8]. We factorize N_x into its unique factorization $N_x = N_x^\flat N_x^\sharp$, with $N_x^\flat$ being squarefree, with $N_x^\sharp$ being squarefull and $\gcd(N_x^\flat, N_x^\sharp) = 1$. We also consider the deterministic function $r\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ defined in [8, Equation (2.3)] (we only need to keep in mind for our discussion that it is a function satisfying the bound $r(t) \ll \min\{t, t^{-2}\}$ for all $t > 0$). As a direct consequence of [8, Lemma 2.1 and Proposition 2.4], we deduce

$$\mathbb{P}_\star\Big[\big\|\mathbf{Primes}(N_x, x) - \mathbf{V}\big\| \geqslant \max\{S_x, T_x\}\Big] \ll \frac{1}{\log x}, \tag{7}$$

with

$$S_x := \frac{5}{\log x}\log(x/N_x^\flat) \qquad \text{and} \qquad T_x := \frac{5}{\log x}\sum_{i\geqslant 1} r(V_i \log x). \tag{8}$$

By applying Proposition 2.1 with (7), we directly obtain the following theorem:

Theorem 3. *Let $x \geqslant 2$, $\eta := \frac{1}{\log x}$. We have*

$$\Big|\nu_x(A) - \mu(A)\Big| \ll \mathbb{P}_\star\Big[\big\{\mathbf{Primes}(N_x, x),\ \mathbf{V}\big\} \subseteq \partial_{\max\{S_x, T_x\}} A\Big] + \frac{1}{\log x},$$

for any $A \in \mathcal{B}(\Delta)$.

As described in [8, Lemmas 2.2 and 2.3], the random variables S_x and T_x are both concentrated around $\frac{1}{\log x}$. Thus, we should expect the following from this theorem:

$$\text{RHS of Theorem 3} \approx \min\big\{\nu_x(\partial_\eta A), \mu(\partial_\eta A)\big\} + \frac{1}{\log x} \tag{9}$$

with $\eta = \frac{1}{\log x}$. This heuristic only becomes a problem if either S_x or T_x is positively correlated with the event that $\mathbf{Primes}(N_x, x)$ and $\mathbf{V}$ are close to the boundary of A.

However, if we do not have any clues on how to compute the probability on the right-hand side of Theorem 3, then we can always use the following inequality

$$\mathbb{P}_\star\Big[\|\mathbf{Primes}(N_x, x) - \mathbf{V}\| \geqslant \eta\Big] \ll_\varepsilon x^{-\frac{\eta}{4+\varepsilon}} + \frac{1}{\log x}.$$

for any $\varepsilon, \eta > 0$ and $x \geqslant 2$, which can be directly deduced from Chernoff's inequality with [8, Propositions 2.4 and 2.5]. Therefore, we immediately arrive at the following theorem by applying again Proposition 2.1.

[2]Actually, the dependence in $\mathbf{u}$ in the implicit constant of the error term was given in [13]. It is however more complicated to state in one line.

Theorem 4. *Fix $\varepsilon > 0$. For $x \geqslant 2$ and $\frac{1}{\log x} \leqslant \eta \leqslant \frac{(4+\varepsilon)\log_2 x}{\log x}$, we have*

$$\Big|\nu_x(A) - \mu(A)\Big| \ll_\varepsilon \min\Big\{\nu_x(\partial_\eta A),\ \mu(\partial_\eta A)\Big\} + x^{-\frac{\eta}{4+\varepsilon}},$$

for any set $A \in \mathcal{B}(\Delta)$.

For most choices of set A, Theorem 4 leads to a weaker result than Theorem 3, as discussed in (9). However, the ratio between the bounds obtain in those two theorems should be relatively small, as we will see later in (15) when applying this theorem to the set $D(u, v)$.

2.4. **Reducing Poisson–Dirichlet approximation problems to two lemmas.** All we need to get an appropriate bound using Theorem 3 are the following two lemmas:

- A *Number Theory Boundary Lemma* (NTBL), which is an upper bound on the probability that either $\mathbf{Primes}(N_x, x)$ or $\mathbf{V}$ is close to the boundary of A by S_x (this last random variable is a deterministic function of number theoretic random variables in the coupling, see (8)). More precisely, it is a bound of the form

$$\min\Big\{\mathbb{P}_\star\big[\mathbf{Primes}(N_x, x) \in \partial_{S_x} A\big],\ \mathbb{P}_\star\big[\mathbf{V} \in \partial_{S_x} A\big]\Big\} \ll (\text{NTBL for the set } A) \tag{10}$$

  for all $x \geqslant 2$.
- A *Poisson–Dirichlet Boundary Lemma* (PDBL), which is an upper bound on the probability that either $\mathbf{Primes}(N_x, x)$ or $\mathbf{V}$ is close to the boundary of A by T_x (this last random variable is a deterministic function of the Poisson–Dirichlet process in the coupling, see (8) again). More precisely, it is a bound of the form

$$\min\Big\{\mathbb{P}_\star\big[\mathbf{Primes}(N_x, x) \in \partial_{T_x} A\big],\ \mathbb{P}_\star\big[\mathbf{V} \in \partial_{T_x} A\big]\Big\} \ll (\text{PDBL for the set } A) \tag{11}$$

  for all $x \geqslant 2$.

If we do have a NTBL and a PDBL for some set A, then we directly deduce

$$\begin{aligned}\mathbb{P}_\star\Big[\big\{\mathbf{Primes}(N_x, x),\ \mathbf{V}\big\} \subseteq \partial_{\max\{S_x, T_x\}} A\Big] \ll (&\text{NTBL for the set } A) + \\ &(\text{PDBL for the set } A).\end{aligned} \tag{12}$$

We note that if the bounds in (10) and (11) are sharp, then we should not expect a loss in the upper bound (12) because we expect two random variables close to each other to be either both close to the boundary or both far from the boundary.

The minima in (10) and (11) offer a choice of which quantity to bound to whoever tries to apply this strategy. In practice, we have one of three options depending on if we have only good number-theoretic information, only good Poisson–Dirichlet information, or both, about proximity to the boundary: Let $I := [\frac{1}{\log x}, \frac{100\log_2 x}{\log x}]$.

1) Suppose that we know how to compute both the bounds $\nu_x(\partial_\eta A)$ and $\mu(\partial_\eta A)$ for all $\eta \in I$. This is the ideal situation. We would then want to attempt to bound the probabilities $\mathbb{P}_\star\big[\mathbf{Primes}(N_x, x) \in \partial_{S_x} A\big]$ and $\mathbb{P}_\star\big[\mathbf{V} \in \partial_{T_x} A\big]$. This is because the events $\big\{\mathbf{Primes}(N_x, x) \in \partial_{S_x}\big\}$ and $\big\{\mathbf{V} \in \partial_{T_x}\big\}$ belong to the σ-algebras generated by the random variables N_x and $\mathbf{V}$ respectively. Therefore, this approach has the benefit of completely avoiding going through the coupling to compute the desired probabilities.

2) Suppose now that we have tools to bound $\nu_x(\partial_\eta A)$ for all $\eta \in I$, but we do not know how to compute bounds on $\mu(\partial_\eta A)$. Then we should aim at computing bounds on the probabilities $\mathbb{P}_\star\big[\mathbf{Primes}(N_x, x) \in \partial_{S_x} A\big]$ and $\mathbb{P}_\star\big[\mathbf{Primes}(N_x, x) \in \partial_{T_x} A\big]$. Note that we have no choice here to understand $\mathbb{P}_\star\big[\mathbf{Primes}(N_x, x) \in \partial_{T_x} A\big]$ within the structure of the coupling.

3) Similarly, if we know how to bound $\mu(\partial_\eta A)$, but we have poor understanding of $\nu_x(\partial_\eta A)$ for all $\eta \in I$, then we should aim for $\mathbb{P}_\star\big[\mathbf{V} \in \partial_{S_x} A\big]$ and $\mathbb{P}_\star\big[\mathbf{V} \in \partial_{T_x} A\big]$. Again, the probability $\mathbb{P}_\star\big[\mathbf{V} \in \partial_{S_x} A\big]$ will have to be studied within the coupling.

Remarks. Here are a few remarks about Theorems 3 and 4 and the strategy presented above:

a) If it is not immediately clear what the boundary ∂A is, we have available the identity

$$\big\|\mathbf{z} - \partial A\big\| = \max\Big\{\big\|\mathbf{z} - A\big\|,\ \big\|\mathbf{z} - \Delta \setminus A\big\|\Big\} \qquad \text{for all } \mathbf{z} \in \Delta \text{ and } A \subseteq \Delta \tag{13}$$

to compute the measure of $\partial_\eta A$. This identity is true simply from the fact that Δ is a convex set. The proof is elementary and we leave it to the reader.

b) For any integer $n \geqslant 2$, we could also study the sequence

$$\mathbf{Primes}^*(n) := \left(\frac{\log p_i(n)}{\log n}\right),$$

instead of $\mathbf{Primes}(n, x)$. Now this sequence is in the set Δ^* defined as all the sequences $\mathbf{x} \in \Delta$ satisfying $\|\mathbf{x}\| = 1$. If we define the measure

$$\nu_x^*(A) := \frac{1}{\lfloor x \rfloor - 1} \cdot \#\Big\{2 \leqslant n \leqslant x : \mathbf{Primes}^*(n) \in A\Big\},$$

then we can obtain analogous versions of Theorems 3 and 4 simply from the fact that

$$\big\|\mathbf{Primes}(n, x) - \mathbf{Primes}^*(n)\big\| = \frac{\log(x/n)}{\log x}$$

for all $n \geqslant 2$.

c) In [8, Theorem 2], Koukoulopoulos and the author also studied the convergence in law of random k-factorizations of a random integer to a Dirichlet law using the coupling $(\Omega_\star, \mathcal{F}_\star, \mathbb{P}_\star)$. It is important to note that we cannot write this problem as $\nu_x(A)$ for some set A, and thus, we cannot apply directly Theorems 3 or 4.

However, a very similar strategy as the one explained earlier was used for this problem: First, we extended the coupling $\Omega_\star$ to include the randomization coming from the k-factorization. Then, we introduce two random vectors in $\mathbb{R}^k$ defined inside this extended coupling: the size of the random k-factorization $\boldsymbol{\delta}_{f,x}$, and a Dirichlet-distributed vector $\mathbf{Z}$. Our Lemma 9.1 in [8] proves that these vectors are close in $\mathbb{R}^k$ to each other with high probability, and Lemma 9.2 reduces the problem into something resembling a NTBL and PDBL lemma. Finally, since we had good number-theoretic and Poisson–Dirichlet methods to study the boundary events, we did not need to dive into the coupling at all to prove these two boundary lemmas.

2.5. **The boundary lemmas for the set $D(u,v)$.** We will use Theorems 3 and 4 to estimate $H(x,y,z)$. To this end, if $\mathbf{z} \in \partial_\eta D(u,v)$ for some $\eta \geqslant 0$, then both $\|\mathbf{z}-\mathbf{w}\|$ and $\|\mathbf{z}-\mathbf{y}\|$ are $< \eta$ for some $\mathbf{w} \in D(u,v)$ and some $\mathbf{y} \notin D(u,v)$ by using (13). Thus, there must exist a subset $E \subseteq \mathbb{N}$ satisfying $\phi_E(\mathbf{w}) \in (u,v)$. It follows from the fact that ϕ_E is 1-Lipschitz that $\phi_E(\mathbf{z}) \in (u-\eta, u+\eta) \cup (v-\eta, v+\eta)$. Therefore,

$$\partial_\eta D(u,v) \subseteq D(u-\eta, u+\eta) \cup D(v-\eta, v+\eta). \tag{14}$$

Applying directly (2), (4), (14) and Theorem 4 with $\varepsilon = 1$ and $\eta = \frac{5}{\log x}\left(\delta \log_2 y + (\frac{3}{2}-\delta)\log_3 y\right)$, we immediately arrive at the estimate

$$H(x,y,z) = x \cdot h\left(\tfrac{\log y}{\log x}, \tfrac{\log z}{\log x}\right) + O\left(\frac{x}{(\log y)^\delta (\log_2 y)^{3/2-\delta}}\right) \tag{15}$$

whenever $3 \leqslant y < z \leqslant \sqrt{x}$, and it proves that $h(u,v) = \mu(D(u,v))$ (which was suggested earlier in (5)). Since μ is supported in Δ^*, which was defined in an earlier remark as the subset of $\mathbf{x} \in \Delta$ satisfying $\|\mathbf{x}\| = 1$, we reduce Theorem 2 to proving the following lemma:

Main Lemma 1. *Let $0 \leqslant u < v \leqslant 1$, let $k = \lfloor \frac{1}{v-u} \rfloor$, and let $\pi_k \colon \Delta \to \mathbb{R}^k$ be the projection map defined as $\pi_k(\mathbf{x}) \coloneqq (x_1, \ldots, x_k)$. We also define $P_{u,v}(\mathcal{E})$ as in Theorem 2. Then*

$$\Delta^* \cap D(u,v)^c = \Delta^* \cap \pi_k^{-1}\left(\bigsqcup_{\mathcal{E} \in 2^{2^{\mathbb{N}\cap[1,k]}}} P_{u,v}(\mathcal{E})\right).$$

To obtain Theorem 1, all we need is a $(\log_2 y)^\delta$ saving from the error term in (15). However, this will require significantly more work since we are using Theorem 3 instead of Theorem 4. Here are the two boundary lemmas that we need to prove in order to directly deduce Theorem 1:

Main Lemma 2 (NTBL for the set $D(u,v)$)**.** *We can factorize uniquely any positive integer n into $n = n^\flat n^\sharp$ with $n^\flat$ being squarefree, $n^\sharp$ being squarefull and $\gcd(n^\flat, n^\sharp) = 1$. We have*

$$\#\left\{ n \leqslant x : \exists\, d | n \text{ satisfying } d \in \left(y(x/n^\flat)^{-5},\ y(x/n^\flat)^5\right)\right\} \ll \frac{x}{(\log y)^\delta (\log_2 y)^{3/2}}$$

for all $3 \leqslant y \leqslant x$.

Main Lemma 3 (PDBL for the set $D(u,v)$)**.** *Let $(\Omega_\star, \mathcal{F}_\star, \mathbb{P}_\star)$ be the coupling constructed in* [8], *and let T_x be defined as in* (8)*. We have*

$$\mathbb{P}_\star\left[\exists\, d | N_x \text{ satisfying } d \in (yx^{-T_x}, yx^{T_x})\right] \ll \frac{1}{(\log y)^\delta (\log_2 y)^{3/2}}$$

for all $3 \leqslant y \leqslant x$.

All that is left is to prove Main Lemmas 1, 2 and 3. The proof of Main Lemma 3 will be the longest part since we will have to dive into the structure of the coupling.

3. Short proofs Main Lemmas 1 & 2

The only thing in common between the proofs of Main Lemmas 1 and 2 is the fact that they can be done quickly. It is why we collect them in this section.

Proof of Main Lemma 1. We first need to show that for any $E \subseteq \mathbb{N}$, we have

$$D(u,v)^c \cap \phi_E^{-1}(\mathbb{R}_{\geqslant v}) = D(u,v)^c \cap \phi_{E\cap[1,k]}^{-1}(\mathbb{R}_{\geqslant v}). \tag{16}$$

The proof of $\supseteq$ is trivial, so we will only show the inclusion $\subseteq$. Let $\mathbf{x} \in D(u,v)^c \cap \phi_E^{-1}(\mathbb{R}_{\geqslant v})$ for any set $E \subseteq \mathbb{N}$. We construct a sequence $(\gamma_j)_{j\geqslant 0}$ by defining $\gamma_0 := 0$ and $\gamma_j := \phi_{E\cap[1,j]}(\mathbf{x})$ for all $j \geqslant 1$. Since the sequence is non-decreasing, $\lim_{j\to\infty} \gamma_j \geqslant v$ and the set $\{\gamma_j : j \geqslant 0\}$ is disjoint from the interval (u,v), then there must exist a unique index j^* satisfying $\gamma_{j^*-1} \leqslant u < v \leqslant \gamma_{j^*}$. It follows that

$$x_{j^*} = \gamma_{j^*} - \gamma_{j^*-1} \geqslant v - u > \frac{1}{k+1} \geqslant \frac{1}{k+1}\sum_{i=1}^{k+1} x_i \geqslant x_{k+1}.$$

We proved that $x_{j^*} > x_{k+1}$, hence $j^* \leqslant k$ and therefore $\mathbf{x} \in \phi_{E\cap[1,k]}^{-1}(\mathbb{R}_{\geqslant v})$. We conclude that (16) is true.

Second, for any $E \subseteq \mathbb{N}$, we have

$$\Delta^* \cap \phi_E^{-1}\big((u,v)^c\big) = \Delta^* \cap \Big(\phi_E^{-1}\big(\mathbb{R}_{\geqslant v}\big) \sqcup \phi_{E^c}^{-1}\big(\mathbb{R}_{\geqslant 1-u}\big)\Big). \tag{17}$$

Combining (3), (16) and (17), we get

$$\begin{aligned}\Delta^* \cap D(u,v)^c &= \bigcap_{E\subseteq\mathbb{N}} \Big(\Delta^* \cap D(u,v)^c \cap \phi_E^{-1}((u,v)^c)\Big)\\ &= \Delta^* \cap \bigcap_{E\subseteq\mathbb{N}\cap[1,k]} \Big(\phi_E^{-1}\big(\mathbb{R}_{\geqslant v}\big) \sqcup \phi_{E^c}^{-1}\big(\mathbb{R}_{\geqslant 1-u}\big)\Big).\end{aligned}$$

Using distributivity laws of the intersection and properties of the preimage, we prove the lemma. □

Proof of Main Lemma 2. Let $3 \leqslant y \leqslant x$ and positive integers b, j, with b being squarefull. We consider the sets

$$\begin{aligned}\mathcal{G}(y,x) &:= \{n \in \mathbb{N}\cap[1,x] : \exists\, d|n \text{ satisfying } d \in (y(x/n^\flat)^{-5}, y(x/n^\flat)^5)\},\\ \mathcal{K}(b,j,x) &:= \{n \in \mathbb{N}\cap[1,x] : n^\sharp = b \text{ and } n^\flat \in (\tfrac{x}{2^j b}, \tfrac{x}{2^{j-1}b}]\}.\end{aligned}$$

The cardinality of the set $\mathcal{G}(y,x)$ is what we need for the bound in Main Lemma 2. Since $\#\mathcal{K}(b,j,x) \leqslant \frac{x}{2^{j-1}b}$, we have

$$\sum_{\substack{b,j\geqslant 1:\ b \text{ squarefull}\\ 2^j b > \log y}} \#\mathcal{K}(b,j,x) \leqslant \frac{2x}{(\log y)^{1/3}} \sum_{\substack{b\geqslant 1\\ b \text{ squarefull}}} \frac{1}{b^{2/3}} \sum_{j\geqslant 1} \frac{1}{2^{2j/3}}.$$

The two series on the right-hand side are convergent, hence

$$\#\mathcal{G}(y,x) \ll \sum_{\substack{b,j\geqslant 1:\ b \text{ squarefull}\\ 2^j b \leqslant \log y}} \#\big(\mathcal{G}(y,x)\cap\mathcal{K}(b,j,x)\big) + \frac{x}{(\log y)^{1/3}}. \tag{18}$$

Consider the map $\psi\colon \mathcal{G}(y,x)\cap\mathcal{K}(b,j,x) \to \mathbb{N}$ defined by $\psi(n) = n^\flat$. For every integer $n \in \mathcal{G}(y,x)\cap\mathcal{K}(b,j,x)$, there is a divisor $d|n$ in the interval $d \in (y(2^j b)^{-5}, y(2^j b)^5)$, so we have

$$\gcd(d,\psi(n)) = \frac{d}{\gcd(d,b)} \in (y(2^j b)^{-6}, y(2^j b)^5).$$

Therefore, any $m \in \psi\big(\mathcal{G}(y,x) \cap \mathcal{K}(b,j,x)\big)$ must satisfy $m \in \mathbb{N} \cap [1, \frac{x}{2^{j-1}b}]$ and has a divisor in $(y(2^jb)^{-6}, y(2^jb)^5)$. Since ψ is injective, we must have

$$\begin{aligned}\#\big(\mathcal{G}(y,x) \cap \mathcal{K}(b,j,x)\big) &= \#\psi\big(\mathcal{G}(y,x) \cap \mathcal{K}(b,j,x)\big) \\ &\leqslant H\Big(\tfrac{x}{2^{j-1}b},\ y(2^jb)^{-6},\ y(2^jb)^5\Big) \\ &\ll \frac{(j+\log b)^\delta}{2^jb} \cdot \frac{x}{(\log y)^\delta (\log_2 y)^{3/2}}\end{aligned}$$

whenever $2^jb \leqslant \log y$ by using (2) in the last inequality. We insert this bound in (18), and Main Lemma 2 immediately follows. □

4. The coupling

The proof of Main Lemma 3 require for us to understand further the structure of the coupling in [8].

Here is one way to construct this coupling. Let $(\Omega_\star, \mathcal{F}_\star, \mathbb{P}_\star)$ be any ambient probability space containing the following random objects:

- A Poisson point process $\mathscr{R}$ in $\mathbb{R}_{>0} \times \mathbb{R}_{>0}$ with intensity $e^{-wy}\,\mathrm{d}w\,\mathrm{d}y$.[3]
- Three i.i.d. random variables U_1, U_2, U_3 that are uniformly distributed in $(0,1)$, and which are also independent of $\mathscr{R}$.

From these random objects, we explain how to extract a random integer N_x and Poisson–Dirichlet process $\mathbf{V}$. All the following random variables were introduced in [1] or in [8], except M_x^*, $\Theta_\infty^{(1)}$ and $\Theta_x^{(2)}$. We will write $\mathbb{E}_\star$ as the expectation inside this coupling.

Definition 4.1 (x-labelling of $\mathscr{R}$)**.** Almost surely, we may find a unique sequence of random points $(W_i, Y_i)_{i\in\mathbb{Z}}$ satisfying

- $\mathscr{R} = \big\{(W_i, Y_i) : i \in \mathbb{Z}\big\}$;
- $W_i < W_{i+1}$ for all $i \in \mathbb{Z}$;
- if we let $S_i := \sum_{\ell \geqslant i} Y_\ell$ for all $i \in \mathbb{Z}$, then we have $S_1 \leqslant \log x < S_0$.

We refer to this sequence as the *x-labelling* of the points of $\mathscr{R}$.

Definition 4.2 (The processes $\mathbf{L}$ and $\mathbf{V}$)**.** Let $x \geqslant 2$, let $(W_i, Y_i)_{i\in\mathbb{Z}}$ be the x-labelling of $\mathscr{R}$, and let $S_i := \sum_{\ell\geqslant i} Y_\ell$ for all $i \in \mathbb{Z}$. We define the process $\mathbf{L} = (L_1, L_2, \ldots)$ by $L_1 := 1 - \frac{S_1}{\log x}$, and $L_i := \frac{Y_{i-1}}{\log x}$ for all $i \geqslant 2$. We also define the process $\mathbf{V}$ as the rearrangement of the terms of the process $\mathbf{L}$ in non-increasing order.

Definition 4.3 (The deterministic functions h and r)**.** Let $(\lambda_j)_{j\geqslant 0}$ be the increasing sequence of positive real numbers defined by $\lambda_0 := e^{-\gamma}$ and

$$\lambda_j := \exp\bigg(-\gamma + \sum_{1\leqslant i\leqslant j} \frac{1}{v_i q_i}\bigg) \text{ for } j \geqslant 1,$$

[3] The coupling in [8] was actually defined as any ambient space with a GEM process and three i.i.d uniform random variables that are also independent of the GEM. As we will see later in Proposition 4.10, the GEM process can be extracted from the Poisson process $\mathscr{R}$.

with γ being the Euler-Mascheroni constant and $q_i = p_i^{v_i}$ being the i^{th} smallest prime power, i.e., $(q_i)_{i\geqslant 1}$ is the sequence $2, 3, 2^2, 5, 7, 2^3, 3^2, \ldots$ We define the step-function $h\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ by

$$h(t) := \sum_{j\geqslant 1} (\log q_j) \mathbb{1}_{\lambda_{j-1} < t \leqslant \lambda_j},$$

and we define the function $r\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ as

$$r(t) := |h(t) - t|.$$

Definition 4.4 (The Poisson process $\mathscr{R}^*$ and its x-labelling)**.** Let h be the deterministic function from Definition 4.3. Consider map $\psi\colon \mathbb{R}_{>0} \times \mathbb{R}_{>e^{-\gamma}} \to \mathbb{R}_{>0} \times \mathcal{Q}$, with $\mathcal{Q}$ being the set of prime powers, and

$$\psi(w, y) := \Big(\frac{wy}{h(y)}, e^{h(y)}\Big).$$

We define the Poisson process $\mathscr{R}^* := \big\{\psi(W, Y) : (W, Y) \in \mathscr{R} \text{ and } Y > e^{-\gamma}\big\}$. The x-*labelling* of $\mathscr{R}^*$ is the almost surely unique sequence of random points $(T_i^*, Q_i^*)_{i\in\mathbb{Z}_{\leqslant K}}$ along with the random integer K satisfying

- $\mathscr{R}^* = \big\{(T_i^*, Q_i^*) : i \in \mathbb{Z}_{\leqslant K}\big\}$;
- $T_i < T_{i+1}$ for all $i \in \mathbb{Z}_{\leqslant K}$;
- $\prod_{i=1}^K Q_i^* \leqslant x < \prod_{i=0}^K Q_i^*$.

Definition 4.5 (The random integer M_x)**.** Let $(W_i, Y_i)_{i\in\mathbb{Z}}$ be the x-labelling of $\mathscr{R}$ (see Definition 4.1). We define $J_x := \prod_{i\geqslant 1} e^{h(Y_i)}$. We also define the *extra prime* $P_{\mathtt{extra}}$ to be the smallest element of $\{1\} \cup \{\text{primes}\}$ that would satisfy $\theta(P_{\mathtt{extra}}) \geqslant U_1\theta(x/J_x)$, where $\theta(y) = \sum_{p\leqslant y} \log p$ is Chebyshev's function. Finally, we define the random integer $M_x := J_x P_{\mathtt{extra}}$.

Definition 4.6 (The random integer M_x^*)**.** Let $(T_i^*, Q_i^*)_{i\in\mathbb{Z}_{\leqslant K}}$ be the x-labelling of $\mathscr{R}^*$ (see Definition 4.4). We define $J_x^* := \prod_{i\geqslant 1} Q_i^*$. We also define the *extra prime* $P_{\mathtt{extra}}^*$ to be the smallest element of $\{1\} \cup \{\text{primes}\}$ that would satisfy $\theta(P_{\mathtt{extra}}^*) \geqslant U_1\theta(x/J_x^*)$, where $\theta(y) = \sum_{p\leqslant y} \log p$ is Chebyshev's function. Finally, we define the random integer $M_x^* := J_x^* P_{\mathtt{extra}}^*$.

Definition 4.7 (The random integer N_x)**.** For any two probability measures μ_1 and μ_2 supported on $\mathbb{N}$, we define a sequence $(z_j)_{j\geqslant 0}$ with $z_0 := 0$ and

$$z_j := \sum_{1\leqslant i\leqslant j} \frac{(\mu_1(i) - \mu_2(i))^-}{\mathrm{d}_{\mathrm{TV}}(\mu_1, \mu_2)},$$

with $\mathrm{d}_{\mathrm{TV}}(\mu_1, \mu_2) := \sup_{A\subseteq\mathbb{N}} |\mu_1(A) - \mu_2(A)|$ being the total variation distance between μ_1 and μ_2. Let $f_{\mu_1,\mu_2} : \mathbb{N} \times (0, 1)^2 \to \mathbb{N}$ be defined as

$$f_{\mu_1,\mu_2}(m, a, b) = \begin{cases} m & \text{if } a \cdot \mu_1(m) \leqslant \mu_2(m), \\ \sum_{i\geqslant 1} i \cdot \mathbb{1}_{z_{i-1} < b \leqslant z_i} & \text{otherwise.} \end{cases}$$

We define the random integer

$$N_x := f_{\mu_1,\mu_2}(M_x, U_2, U_3)$$

with μ_1 being the law of M_x and μ_2 being the uniform distribution on $\mathbb{N} \cap [1, x]$.

Definition 4.8 (The random variables $\Theta_\infty^{(1)}$ and $\Theta_x^{(2)}$)**.** Let r be the function from Definition 4.3. We define $r_0 := \sup_{t>0} r(t)$ and

$$\Theta_\infty^{(1)} := r_0 + \sum_{\substack{(W,Y)\in\mathscr{R} \\ Y\leqslant e^{-\gamma}}} r(Y).$$

Let $(W_i, Y_i)_{i\in\mathbb{Z}}$ be the x-labelling of $\mathscr{R}$ (see Definition 4.1). We also define $\Theta_x^{(2)}$ as the random variable

$$\Theta_x^{(2)} := \sum_{\substack{i\geqslant 1 \\ Y_i>e^{-\gamma}}} r(Y_i).$$

Here are some properties that these random objects satisfy:

Proposition 4.9 (Distribution of N_x)**.** *The random integer N_x is uniformly distributed in* $\mathbb{N}\cap[1,x]$.

Proof. In general, if X is any random variable with law μ_1 and U, U' are two uniform random variables independent of each other and of X, then it was shown in [1, Section 3.8] that the law of the random variable $f_{\mu_1,\mu_2}(X,U,U')$ is μ_2, with f_{μ_1,μ_2} being defined as in Definition 4.7. This was also reproven in [8, Lemma B.2] using the notation of Definition 4.7. □

Proposition 4.10 (Distribution of $\mathbf{V}$)**.** *The process* $\mathbf{V}$ *follows a Poisson–Dirichlet distirbution.*

Proof. The process $\mathbf{L}$ from Definiton 4.2 follows a GEM distribution (see [8, Proposition 4.3]). If we rearrange the terms of any GEM process in non-increasing order, then we always obtain a Poisson–Dirichlet process (see [6, Theorem 2.7]). □

Proposition 4.11 (Total variation distance between N_x, M_x and M_x^*)**.** *For all $x \geqslant 2$, let $\mathcal{E}(x) := \{N_x = M_x = M_x^*\}$. We have*

$$\mathbb{P}_\star\big[\mathcal{E}(x)^c\big] \ll \frac{1}{\log x}.$$

Proof. Note that if $N_x = M_x$ and $J_x = J_x^*$ (with J_x and J_x^* being defined as in Definitions 4.5 and 4.6), then $\mathcal{E}(x)$ must occur. Therefore,

$$\mathbb{P}_\star\big[\mathcal{E}(x)^c\big] \leqslant \mathbb{P}_\star\big[N_x \neq M_x\big] + \mathbb{P}_\star\big[J_x \neq J_x^*\big].$$

It follows from [8, Propositions 2.4 and 6.5] that the right-hand side of the inequality above is $O(\frac{1}{\log x})$. □

Proposition 4.12 (Bound on $r(t)$)**.** *For all $t > 0$, we have the upper bound $r(t) \ll \min\{t, t^{-2}\}$.*

Proof. This follows from the Prime Number Theorem. See [8, Section 2] for more details. □

Proposition 4.13 (Moment generating function of $\Theta_\infty^{(1)} + \Theta_x^{(2)}$)**.** *Fix $\alpha > 0$, and let $\Theta_\infty^{(1)}$ and $\Theta_x^{(2)}$ be defined as in Definition 4.8. We have*

$$\mathbb{E}_\star\big[e^{\alpha(\Theta_\infty^{(1)}+\Theta_x^{(2)})}\big] \ll_\alpha 1.$$

In particular, we have

$$\mathbb{P}_\star\big[\Theta_\infty^{(1)} + \Theta_x^{(2)} > \alpha\cdot\log_2 y\big] \ll_\alpha \frac{1}{\log y}$$

for any $y \geqslant 2$.

Proof. The proof of this proposition is very similar to the proof of [8, Lemma 2.3]: Let Θ_∞ be defined as

$$\Theta_\infty := r_0 + \sum_{(W,Y)\in\mathscr{R}} r(Y).$$

We have $\Theta_\infty^{(1)} + \Theta_x^{(2)} \leqslant \Theta_\infty$. By using Campbell's Theorem (see [9, Section 3.2])

$$\mathbb{E}_\star\Big[e^{\alpha(\Theta_\infty^{(1)}+\Theta_x^{(2)})}\Big] \leqslant \mathbb{E}_\star\Big[e^{\alpha\Theta_\infty}\Big] = \exp\left(\alpha r_0 + \int_0^\infty \frac{e^{\alpha r(y)}-1}{y}\,\mathrm{d}y\right) \ll_\alpha 1,$$

with the convergence of the integral for all fixed $\alpha > 0$ following from $e^{\alpha r(y)} - 1 \ll_\alpha r(y) \ll \min\{y, y^{-2}\}$ (see Proposition 4.12). The bound on the probability $\mathbb{P}_\star[\Theta_\infty^{(1)} + \Theta_x^{(2)} > \log_2 y]$ is then a direct application of Chernoff's inequality. $\square$

5. Proof of Main Lemma 3

Recall the construction of the coupling $(\Omega_\star, \mathcal{F}_\star, \mathbb{P}_\star)$ from the previous section, as well as all the random variables that were introduced there. Recall also the definition of T_x in (8). We define the infinite set $\mathcal{H}(y,\kappa)$ and the function $\xi(y)$ as

$$\mathcal{H}(y,\kappa) := \Big\{n \in \mathbb{N} : \exists\, d|n \text{ such that } d \in (y/\kappa, y\cdot\kappa)\Big\}$$
$$\xi(y) := (\log y)^{-\delta}(\log_2 y)^{-3/2}.$$

We have to keep in mind the monotonicity of the set $\mathcal{H}(y,\kappa)$ with respect to κ, i.e. if $1 < \kappa_1 \leqslant \kappa_2$, then $\mathcal{H}(y,\kappa_1) \subseteq \mathcal{H}(y,\kappa_2)$. We can rewrite what we need to show in Main Lemma 3 as

$$\mathbb{P}_\star\Big[N_x \in \mathcal{H}(y, x^{T_x})\Big] \ll \xi(y). \tag{19}$$

To prove Main Lemma 3, we will need four lemmas:

Lemma 5.1. *If x, y, κ are real numbers satisfying $2 \leqslant \kappa \leqslant (\log y)^{40}$ and $2 \leqslant y \leqslant x^{2/3}$, then*

$$\mathbb{P}_\star\Big[\{N_x, M_x, M_x^*\} \text{ intersects } \mathcal{H}(y,\kappa)\Big] \ll (\log\kappa)^\delta \cdot \xi(y).$$

Proof. Define the event $\mathcal{E}(x) := \{N_x = M_x = M_x^*\}$. If $\{N_x, M_x, M_x^*\}$ intersects $\mathcal{H}(y,\kappa)$, then either $N_x \in \mathcal{H}(y,\kappa)$ or $\mathcal{E}(x)$ does not occur. Therefore, by Proposition 4.11, we have

$$\mathbb{P}_\star\Big[\{N_x, M_x, M_x^*\} \text{ intersects } \mathcal{H}(y,\kappa)\Big] \ll \mathbb{P}_\star\Big[N_x \in \mathcal{H}(y,\kappa)\Big] + (\log x)^{-1}.$$

We have

$$\mathbb{P}_\star\Big[N_x \in \mathcal{H}(y,\kappa)\Big] \ll (\log\kappa)^\delta \cdot \xi(y)$$

by [7, Theorem 1]. $\square$

Lemma 5.2. *Let x, y, κ be real numbers satisfying $2 \leqslant \kappa \leqslant (\log y)^{30}$ and $2 \leqslant y \leqslant \sqrt{x}$. Let d be a positive integer satisfying $d \leqslant \sqrt{y}$. Then*

$$\mathbb{P}_\star\Big[d|M_x \text{ and } M_x \in \mathcal{H}(y,\kappa)\Big] \ll \frac{\tau(d)}{d}\cdot(\log\kappa)^\delta\cdot\xi(y).$$

with $\tau(d)$ being the number of positive divisors of d.

Proof. Using Proposition 4.11, we have

$$\mathbb{P}_\star\Big[d|M_x \text{ and } M_x \in \mathcal{H}(y,\kappa)\Big] = \mathbb{P}_\star\Big[d|N_x \text{ and } N_x \in \mathcal{H}(y,\kappa)\Big] + O\Big(\frac{1}{\log x}\Big).$$

Suppose that a and b are positive integers satisfying $a|db$ and that $a \in (y/\kappa, y\cdot\kappa)$. We must have $\frac{a}{\gcd(a,b)}|d$. Therefore, the integer $\gcd(a,b)$ is in the set $\{a/s : s|d\}$, which is itself a subset of the union of open intervals $\bigcup_{s|d}(\frac{y}{s\kappa}, \frac{y\kappa}{s})$. It follows that if $db \in \mathcal{H}(y,\kappa)$, then $b \in \bigcup_{s|d}\mathcal{H}(y/s,\kappa)$. Thus, we have

$$\begin{aligned}\mathbb{P}_\star\Big[d|N_x \text{ and } N_x \in \mathcal{H}(y,\kappa)\Big] &\leqslant \sum_{s|d}\mathbb{P}_\star\Big[d|N_x \text{ and } N_x/d \in \mathcal{H}(y/s,\kappa)\Big]\\ &= \frac{\lfloor x/d\rfloor}{\lfloor x\rfloor}\sum_{s|d}\mathbb{P}_\star\Big[N_{x/d} \in \mathcal{H}(y/s,\kappa)\Big]\\ &\ll \frac{\tau(d)}{d}\cdot\max_{s|d}\mathbb{P}_\star\Big[N_{x/d} \in \mathcal{H}(y/s,\kappa)\Big].\end{aligned}$$

By applying Lemma 5.1, we finish the proof. □

Lemma 5.3. *For all $2 \leqslant y \leqslant x$, we have*

$$\mathbb{P}_\star\Big[M_x^* \in \mathcal{H}(y, e^{10\cdot\Theta_\infty^{(1)}})\Big] \ll \xi(y).$$

Proof. Let Σ_1 be the σ-algebra generated by U_1 and the restriction of $\mathscr{R}$ on the set $(0,\infty)\times(e^{-\gamma},\infty)$, and let Σ_2 be the σ-algebra generated by the restriction of $\mathscr{R}$ on the set $(0,\infty)\times(0,e^{-\gamma}]$. Note that Σ_1 and Σ_2 are independent σ-algebras. Since M_x^* is Σ_1-measurable and $\Theta_\infty^{(1)}$ is Σ_2-measurable, we conclude that M_x^* and $\Theta_\infty^{(1)}$ are independent. Thus, over all the coupling $\Omega_\star$, we can bound the conditional expectation

$$\mathbb{E}_\star\Big[\mathbb{1}_{M_x^*\in\mathcal{H}(y,e^{10\cdot\Theta_\infty^{(1)}})}\cdot\mathbb{1}_{\Theta_\infty^{(1)}\leqslant\log_2 y}\,\Big|\,\Theta_\infty^{(1)}\Big] \ll (\Theta_\infty^{(1)})^\delta\cdot\xi(y) \leqslant e^{\Theta_\infty^{(1)}}\cdot\xi(y).$$

Therefore,

$$\mathbb{P}_\star\Big[M_x^* \in \mathcal{H}(y, e^{10\cdot\Theta_\infty^{(1)}})\Big] \ll \mathbb{E}_\star\Big[e^{\Theta_\infty^{(1)}}\Big]\cdot\xi(y) + \mathbb{P}_\star\Big[\Theta_\infty^{(1)} > \log_2 y\Big].$$

The lemma follows directly from applying Proposition 4.13. □

Lemma 5.4. *For all $2 \leqslant y \leqslant x$, we have*

$$\mathbb{P}_\star\Big[M_x \in \mathcal{H}(y, e^{10\cdot\Theta_x^{(2)}})\Big] \ll \xi(y).$$

Proof. For any prime power p^v, we define $A_{p^v} := \#\{i \geqslant 1 : h(Y_i) = \log p^v\}$. We must have

$$A_{p^v} = \sum_{1\leqslant k<\frac{\log y}{2\log p^v}}\mathbb{1}_{k\leqslant A_{p^v}<\frac{\log y}{2\log p^v}} + A_{p^v}\mathbb{1}_{A_{p^v}\geqslant\frac{\log y}{2\log p^v}} \leqslant \sum_{1\leqslant k<\frac{\log y}{2\log p^v}}\mathbb{1}_{A_{p^v}\geqslant k} + \frac{2\log p^v\cdot A_{p^v}^2}{\log y}.$$

If $A_{p^v} \geqslant k$, then $p^{vk}|M_x$. Therefore,

$$A_{p^v} \leqslant \sum_{1\leqslant k<\frac{\log y}{2\log p^v}}\mathbb{1}_{p^{vk}|M_x} + \frac{2\log p^v\cdot A_{p^v}^2}{\log y}.$$

With this inequality and Proposition 4.12, we have

$$
\begin{aligned}
\Theta_x^{(2)} &\ll \sum_{i\geqslant 1}\frac{1}{h(Y_i)^2} = \sum_{p^v}\frac{A_{p^v}}{(\log p^v)^2} \\
&\leqslant \sum\sum\sum_{\substack{p \text{ prime and } v,k\geqslant 1\\ p^{vk}\leqslant\sqrt{y}}} \frac{\mathbb{1}_{p^{vk}|M_x}}{(\log p)^2} + \frac{2}{\log y}\sum_{p^v}\frac{A_{p^v}^2}{\log p^v}.
\end{aligned}
\tag{20}
$$

Since $A_{p^v} \leqslant \#\{(W,Y)\in\mathscr{R} : h(Y) = \log p^v\}$, and this last random variable is Poisson distributed with paramater $\frac{1}{vp^v}$, then $\mathbb{E}[A_{p^v}^2] \ll \frac{1}{vp^v}$. It follows that

$$
\frac{2}{\log y}\sum_{p^v}\frac{\mathbb{E}[A_{p^v}^2]}{\log p^v} \ll \frac{1}{\log y}.
\tag{21}
$$

Using Proposition 4.13, Lemma 5.1 and the inequalities (20) and (21), we have the decomposition

$$
\begin{aligned}
\mathbb{P}_\star\Big[M_x \in \mathcal{H}(y, e^{10\cdot\Theta_x^{(2)}})\Big] &\leqslant \sum_{1\leqslant j\leqslant \log_3 y} \mathbb{P}_\star\Big[M_x \in \mathcal{H}(y, e^{10\cdot\Theta_x^{(2)}}) \text{ and } \Theta_x^{(2)} \in (e^j, e^{j+1}]\Big] + O\big(\xi(y)\big) \\
&\leqslant \sum_{1\leqslant j\leqslant \log_3 y} e^{-j}\cdot\mathbb{E}_\star\big[\Theta_x^{(2)}\cdot \mathbb{1}_{M_x\in\mathcal{H}(y,e^{10\cdot e^{j+1}})}\big] + O\big(\xi(y)\big) \\
&\ll \sum\sum\sum\sum_{\substack{p \text{ prime and } v,j,k\geqslant 1\\ j\leqslant \log_3 y\\ p^{vk}\leqslant\sqrt{y}}} \frac{\mathbb{P}_\star\big[p^{vk}|M_x \text{ and } \mathbb{1}_{M_x\in\mathcal{H}(y,e^{10\cdot e^{j+1}})}\big]}{e^j(\log p)^2} + O(\xi(y))
\end{aligned}
$$

By applying Lemma (5.2), we arrive at

$$
\mathbb{P}_\star\Big[M_x \in \mathcal{H}(y, e^{10\cdot\Theta_x^{(2)}})\Big] \ll \xi(y)\cdot\sum_{j\geqslant 1} e^{j(\delta-1)}\cdot\sum_p\sum_{v\geqslant 1}\sum_{k\geqslant 1}\frac{vk+1}{p^{vk}(\log p)^2}.
$$

Using the inequalities $vk+1 \leqslant 2vk$ and $p^{vk} \geqslant p2^{v+k-2}$, we have

$$
\mathbb{P}_\star\Big[M_x \in \mathcal{H}(y, e^{10\cdot\Theta_x^{(2)}})\Big] \ll \xi(y)\cdot\left(\sum_{j\geqslant 1} e^{j(\delta-1)}\right)\cdot\left(\sum_p \frac{1}{p(\log p)^2}\right)\cdot\left(\sum_{v\geqslant 1}\frac{v}{2^v}\right)^2.
$$

The series on the right-hand side are all convergent, which proves the lemma. □

Proof of Main Lemma 3. Let $r_0 := \sup_{t>0} r(t)$. From the definition of $\mathbf{L}$, $\mathbf{V}$, $\Theta_\infty^{(1)}$ and $\Theta_x^{(2)}$, we have

$$
T_x(\log x) = 5\sum_{i\geqslant 1} r(V_i\log x) = 5\sum_{i\geqslant 1} r(L_i \log x) \leqslant 5\Big(r_0 + \sum_{i\geqslant 1} r(Y_i)\Big) \leqslant 10\max\{\Theta_\infty^{(1)}, \Theta_x^{(2)}\}.
$$

It follows that

$$
\mathbb{P}_\star[N_x \in \mathcal{H}(y, x^{T_x})] \leqslant \mathbb{P}_\star[M_x \in H(y, e^{10\cdot\Theta_x^{(2)}})] + \mathbb{P}_\star[M_x^* \in H(y, e^{10\cdot\Theta_\infty^{(1)}})] + \mathbb{P}_\star[\mathcal{E}(x)^c].
$$

with $\mathcal{E}(x) = \{N_x = M_x = M_x^*\}$. By combining Proposition 4.11, with Lemmas 5.3 and 5.4, we prove (19) and thus we finish the proof of Main Lemma 3. □

REFERENCES

[1] R. Arratia – “On the amount of dependence in the prime factorization of a random integer”, in *Contemporary Combinatorics*, Bolyai Soc. Math. Stud., vol. 10, János Bolyai Math. Soc., Budapest, 2002, p. 29–91.

[2] A. S. Besicovitch – “On the density of certain sequences of integers”, *Math. Ann.* **110** (1934), p. 336–341.

[3] P. Billingsley – “On the distribution of large prime divisors”, *Period. Math. Hungar.* **2** (1972), p. 283–289.

[4] P. Erdős – “Note on the sequences of integers no one of which is divisible by any other”, *J. London Math. Soc.* **10** (1935), p. 126–128.

[5] P. Erdős – “A generalization of a theorem of Besicovitch”, *J. London Math. Soc.* **11** (1936), p. 92–98.

[6] S. Feng - *The Poisson–Dirichlet distribution and related topics. Models and asymptotic behaviors*, Probability and its Appl., Springer, Heidelberg, 2010.

[7] K. Ford – “The distribution of integers with a divisor in a given interval”, *Ann. of Math. (2)* **168** (2008), no. 2, p. 367–433.

[8] T. Haddad & D. Koukoulopoulos – “On Arratia’s coupling and the Dirichlet law for the factors of a random integer”, *J. Éc. polytech., Math.* **12** (2025), p. 1565–1604.

[9] J. F. C. Kingman – *Poisson processes*, Oxford Studies of Probab., vol. 3, The Clarendon Press, Oxford University Press, New York, 1993.

[10] D. E. Knuth & L. Trabb Pardo – “Analysis of a simple factorization algorithm”, *Theoret. Comput. Sci.* **3** (1976), no. 3, p. 321–348.

[11] G. Tenenbaum – “Lois de répartition des diviseurs, 2”, *Acta Arithm.* **38** (1980), p. 1–36.

[12] G. Tenenbaum – “Sur la probabilité qu’un entier possède un diviseur dans un intervalle donné”, *Compositio Math.* **51** (1984), no. 2, p. 243–263.

[13] G. Tenenbaum – “A rate estimate in Billingsley’s Theorem for the size distribution of large prime factors”, *Q. J. Math.* **51** (2000), no. 3, p. 385–403.

DEPARTEMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF TURKU, 20014 TURKU, FINLAND
Email address: `tohadd@utu.fi`