#A6 INTEGERS 11 (2011)

MAXIMUM GCD AMONG PAIRS OF RANDOM INTEGERS

R. W. R. Darling
Mathematics Research Group, National Security Agency, Fort George G. Meade, Maryland, USA

E. E. Pyle
Mathematics Research Group, National Security Agency, Fort George G. Meade, Maryland, USA

Received: 11/13/09, Revised: 9/3/10, Accepted: 11/23/10, Published: 1/13/11

Abstract

Fix $\alpha > 0$, and sample N integers uniformly at random from $\{1, 2, \ldots, \lfloor e^{\alpha N} \rfloor\}$. Given $\eta > 0$, the probability that the maximum of the pairwise GCDs lies between $N^{2-\eta}$ and $N^{2+\eta}$ converges to 1 as $N \to \infty$. More precise estimates are obtained. This is a Birthday Problem: two of the random integers are likely to share some prime factor of order $N^2 / \log(N)$. The proof generalizes to any arithmetical semigroup where a suitable form of the Prime Number Theorem is valid.

1. Main Result

The distribution of the sizes of the prime divisors of a random integer has been well studied—see portions of Billingsley [1]. Diaconis and Erdös [3] compute the probability distribution of the GCD of two random integers; moments of this distribution were estimated earlier by Cesàro [2]; the GCD of k random integers was treated by Nymann [5]. However the authors are unaware of any published results on the pairwise Greatest Common Divisors (GCD) among a large collection of random integers. Theorem 1 establishes probabilistic upper and lower bounds for the maximum of these pairwise GCDs.

Theorem 1. *Suppose $\alpha > 0$, and $T_1, \ldots, T_N$ is a random sample, drawn with replacement, from the integers $\{n \in \mathbb{N} : n \leq e^{\alpha N}\}$. Let $\Gamma_{j,k}$ denote the Greatest Common Divisor of T_j and T_k. For any $\eta > 0$,*

$$\lim_{N \to \infty} \mathbb{P}\left[N^{2-\eta} < \max_{1 \leq j < k \leq N} \{\Gamma_{j,k}\} < N^{2+\eta} \right] = 1. \tag{1}$$

Indeed there are more precise estimates: for all $s \in (0, 1)$, and $b > 0$, the right side *of* (2) *is finite, and*

$$\mathbb{P}\left[\max_{1\leq j<k\leq N}\{\Gamma_{j,k}\}\geq N^{2/s}b^{1/s}\right]\leq\frac{1}{2b}\prod_{p\in\mathcal{P}}\left(1+\frac{p^s-1}{p^2-p^s}\right), \tag{2}$$

where $\mathcal{P}$ *denotes the rational primes; while if* $\Lambda_{j,k}$ *denotes the largest common prime factor of* T_j *and* T_k*, then for all* $\theta>0$*,*

$$\lim_{N\to\infty}\mathbb{P}\left[\max_{1\leq k<j\leq N}\{\Lambda_{j,k}\}<\frac{N^2}{\log\left(N^\theta\right)}\right]\leq e^{-\theta/8}. \tag{3}$$

Remark. There is an upper bound, similar to (2), for the radical (i.e the largest square-free divisor) $\mathrm{rad}\,(\Gamma_{j,k})$ of the GCD:

$$\mathbb{P}\left[\max_{1\leq j<k\leq N}\{\mathrm{rad}\,(\Gamma_{j,k})\}\geq N^{2/s}b^{1/s}\right]\leq\frac{1}{2b}\prod_{p\in\mathcal{P}}\left(1-p^{-2}+p^{s-2}\right). \tag{4}$$

The proof, which is omitted, uses methods similar to those of Proposition 2, based upon a Bernoulli model for occurrence of prime divisors, instead of a Geometric model for prime divisor multiplicities. For example, when $s=0.999$, the product on the right side of (4) is approximately 12.44; for the right side of (2), it is approximately 17.64.

1.1. Overview of the Proof of Theorem 1

Let Z_i^k be a Bernoulli random variable, which takes the value 1 when prime p_i divides T_k. As a first step towards the proof, imagine proving a comparable result in the case where $\left\{Z_i^k, 1\leq k\leq N, i\geq 1\right\}$ were independent, and $\mathbb{P}\left[Z_i^k=1\right]=1/p_i$. The harder parts of the proof arise in dealing with the reality that, for fixed k, $\left\{Z_i^k, i\geq 1\right\}$ are negatively associated, and change with N. Convergence of the series

$$\sum_{p\in\mathcal{P}}p^{-2}\log(p)<\infty$$

ensures that the parameter α, which governs the range of integers being sampled, appears in none of the bounds (1), (2), nor (3). However the proof for the lower bound depends crucially on an exponential (in N) rate of growth in the range, in order to moderate the dependence among $\left\{Z_i^k, i\geq 1\right\}$ for fixed k.

Consider primes as labels on a set of urns; the random variable T_j contributes a ball to the urn labelled p if prime p divides T_j. The lower bound comes from showing that, with asymptotic probability at least $1-e^{-\theta/8}$, some urn with a label $p>N^2/\log\left(N^\theta\right)$ contains more than one ball; in that case prime p is a common divisor of two distinct members of the list $T_1,\ldots,T_N$. The upper bound comes from a first moment estimate: multiply the number of pairs by the probability that a specific pair has a GCD above some threshold.

If $T_1, \dots, T_N$ were sampled uniformly without replacement from the integers from 1 to N^2, the lower bound (3) would fail; see the analysis in [1] of the distribution of the largest prime divisor of a random integer. In the case of sampling from integers from 1 to N^r, where $r \geq 3$, the upper bound (2) remains valid, but we do not know whether the lower bound (3) holds or not.

1.2. Generalizations to Arithmetical Semigroups

Although details will not be given, the techniques used to prove Theorem 1 will be valid in the more general context of a commutative semigroup G with identity element 1, containing a countably infinite subset $\mathcal{P} = \{p_1, p_2, \dots\}$ called the *primes* of G, such that every element $a \neq 1$ of G has a unique factorization of the form

$$a = \prod_{i \geq 1} p_i^{e_i}, \quad (e_1, e_2, \dots) \in \mathbb{Z}_+^\infty$$

where all but finitely many (e_i) are zero. Assume in addition that G is an *arithmetical semigroup* in the sense of [4], meaning that there exists a real-valued norm $|\cdot|$ on G such that:

- $|1| = 1$, $|p_i| > 1$ for all $p_i \in \mathcal{P}$.
- $|ab| = |a||b|$ for all $a, b \in G$.
- The set $\pi_G(x) = \{i \geq 1 : |p_i| \leq e^x\}$ is finite, for each real $x > 0$.

The only analytic condition needed is an abstract form of the Prime Number Theorem (see [4, Chapter 6]):

$$\lim_{x \to \infty} x e^{-x} \left|\pi_G(x)\right| = 1,$$

used in the proof of Proposition 5. This in turn will imply convergence of series such as:

$$\sum_{p \in \mathcal{P}} \log\left(1 + |p|^{s-2}\right), \ s < 1,$$

which appear (in an exponentiated form) in the bound (2). For example, Landau's Prime Ideal Theorem provides such a result in the case where G is the set of integral ideals in an algebraic number field, $\mathcal{P}$ is the set of prime ideals, and $|a|$ is the norm of a. Knopfmacher [4] also studies a more general setting where, for some $\delta > 0$,

$$\lim_{x \to \infty} x e^{-\delta x} \left|\pi_G(x)\right| = \delta.$$

The authors have not attempted to modify Theorem 1 to fit this case.

1.3. Future Lines of Enquiry

Test of Arithmetic Randomness. The authors do not know whether

$$N^{-2} \max_{1 \le j < k \le N} \{\Gamma_{j,k}\}$$

has a limit in distribution as $N \to \infty$. An anonymous referee points out that, if it does, then a test for the arithmetic randomness of a sequence of N integers would result: namely, compute the maximum of the pairwise GCDs, divide by N^2, and compute a p-value.

Efficient Computation. How might the maximum of the pairwise GCDs of N large random integers be computed efficiently? Perhaps smaller prime factors could be removed by a sieve. Is there an efficient way to detect a squared prime of size about $N^2 \big/ \log(N)$ in the product of all the integers? To detect the largest common prime factor among all pairs of integers, is it better to compute for $k = 1, \ldots, N-1$ the GCD of the product $T_1 T_2 \cdots T_k$ with T_{k+1}, rather than to compute each of the pairwise GCDs?

2. Pairwise Minima in a Geometric Probability Model

2.1. Geometric Random Vectors

Let $\mathcal{P} = \{p_1, p_2, \ldots\}$ denote the rational primes $\{2, 3, 5, \ldots\}$ in increasing order. Let $\mathcal{I}$ denote the set of non-negative integer vectors $(e_1, e_2, \ldots)$ for which $\sum e_i < \infty$. Let $X_1, X_2, \ldots$ be (possibly dependent) positive integer random variables, whose joint law has the property that, for every $k \in \mathbb{N}$, and every $(e_1, e_2, \ldots) \in \mathcal{I}$ for which $e_k = 0$,

$$\mathbb{P}\left[X_k \ge m \mid \bigcap_{i \ne k} \{X_i = e_i\}\right] \le \left(\frac{1}{p_k}\right)^m. \tag{5}$$

Consider $X_1, X_2, \ldots$ as a general model for prime multiplicities in the prime factorization of a random integer, without specifying exactly how that integer will be sampled. Let ζ denote the random vector:

$$\zeta = (X_1, X_2, \ldots) \in \mathbb{N}^{\mathbb{N}}. \tag{6}$$

Let $\zeta^{(1)}, \zeta^{(2)}, \ldots, \zeta^{(N)}$ be independent random vectors, all having the same law as ζ in (6). Write $\zeta^{(k)}$ as $\left(X_1^k, X_2^k, \ldots\right)$. Then

$$L_{j,k} = \sum_i \min\left\{X_i^k, X_i^j\right\} \log\left(p_i\right)$$

is a model for the log of the GCD of two such random integers. We shall now derive an upper bound for

$$\Delta_N = \max_{1\leq k<j\leq N} \{L_{j,k}\},$$

which models the log maximum of the pairwise GCD among a set of N "large, random" integers.

Proposition 2. *Assume the joint law of the components of ζ satisfies* (5)*. Then*

(i) For every $s \in (0,1)$, the following expectation is finite:

$$\mathbb{E}\left[e^{sL_{k,j}}\right] < \prod_i \left(1 + \frac{p_i^s - 1}{p_i^2 - p_i^s}\right) = C_s < \infty. \tag{7}$$

(ii) For any $s \in (0,1)$ and $b > C_s/2$, with C_s as in (7)*, there is an upper bound:*

$$\mathbb{P}\left[\Delta_N \geq \log(N^{2/s}) + s^{-1}\log(b)\right] \leq \frac{C_s}{2b} < 1. \tag{8}$$

Proof. Consider first the case where $X_1, X_2, \ldots$ are independent Geometric random variables, and

$$\mathbb{P}[X_k \geq m] = \left(\frac{1}{p_k}\right)^m, \ m = 1, 2, \ldots$$

It is elementary to check that, for $s \in (0,1)$, and any $p \in \mathcal{P}$, if X'', X' are independent Geometric random variables with

$$\mathbb{P}[X'' \geq m] = p^{-m} = \mathbb{P}[X' \geq m], \ m = 1, 2, \ldots,$$

then their minimum is also a Geometric random variable, which satisfies

$$\mathbb{E}\left[p^{s\min\{X'',X'\}}\right] = 1 + \frac{p^s - 1}{p^2 - p^s} < 1 + p^{s-2}.$$

It follows from the independence assumption that

$$\mathbb{E}\left[e^{sL_{k,j}}\right] = \mathbb{E}\left[\prod_i p_i^{s\min\{X_i^k, X_i^j\}}\right] = \prod_i \left(1 + \frac{p_i^s - 1}{p_i^2 - p_i^s}\right) = C_s.$$

This verifies the assertion (7).

Markov's inequality $a\mathbb{P}[X \geq a] \leq \mathbb{E}[X]$ shows that, for any $s \in (0,1)$,

$$C_s \geq e^{st}\mathbb{P}[L_{k,j} \geq t].$$

Furthermore

$$\begin{aligned}\mathbb{P}\left[\max_{1\leq k<j\leq N}\{L_{k,j}\} \geq t\right] &= \mathbb{P}\left[\bigcup_{1\leq k<j\leq N}\{L_{k,j} \geq t\}\right] \\ &\leq \sum_{1\leq k<j\leq N} \mathbb{P}[L_{k,j} \geq t] \qquad = \frac{N(N-1)}{2}\mathbb{P}[L_{k,j} \geq t].\end{aligned}$$

It follows that, for $s \in (0,1)$, $b > 0$, and $t = s^{-1}\log\left(bN^2\right)$,

$$\mathbb{P}\left[\Delta_N \geq \log(N^{2/s}) + s^{-1}\log(b)\right] \leq \frac{N^2}{2}e^{-st}C_s = \frac{C_s}{2b}.$$

It remains to consider the case where $X_1, X_2, \ldots$ satisfies (5) without the independence assumption. This will be easy because, under (5), large values of the X_i are less likely than in the independent Geometric case, so the same upper bound remains valid. A coupling construction will be used to build dependent random variables on the same probability space as the one for independent random variables. By taking products of probability spaces, construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which independent Geometric random variables $X'_1, X'_2, \ldots$ and $X''_1, X''_2, \ldots$ are defined, such that for all $i \geq 1$,

$$\mathbb{P}\left[X''_i \geq m\right] = p_i^{-m} = \mathbb{P}\left[X'_i \geq m\right],\ m = 1, 2, \ldots.$$

We propose to construct $\zeta^{(1)} = \left(X_1^1, X_2^1, \ldots\right)$ and $\zeta^{(2)} = \left(X_1^2, X_2^2, \ldots\right)$ by induction on this probability space $(\Omega, \mathcal{F}, \mathbb{P})$, so that for each $n \geq 1$, $\left\{\left(X_i^1, X_i^2\right)\right\}_{1\leq i\leq n}$ have the correct joint law, and

$$X_i^1 \leq X'_i,\ X_i^2 \leq X''_i,\ i = 1, 2, \ldots.$$

Once this is achieved, monotonicity implies

$$\mathbb{E}\left[e^{sL_{1,2}}\right] \leq \mathbb{E}\left[\prod_i p_i^{s\min\left\{X'_i, X''_i\right\}}\right],$$

so the desired result will follow from the previous one for independent Geometric random variables.

Since $\zeta^{(1)}$ and $\zeta^{(2)}$are independent, it suffices to construct $\zeta^{(1)}$ in terms of $X'_1, X'_2, \ldots$ so that $X_i^1 \leq X'_i$ for all i. Let $(U_{i,j}, i \geq 1, j \geq 0)$ be independent Uniform$(0,1)$ random variables. Suppose either $i = 1$, or else some values $X_1^1 = e_1, X_2^1 = e_2, \ldots, X_{i-1}^1 = e_{i-1}$ have already been determined. By assumption, there exist parameters

$$q_{i,k} = \mathbb{P}\left[X_i \geq k \mid \bigcap_{j<i}\left\{X_j = e_j\right\}\right] \leq \left(\frac{1}{p_i}\right)^k,\ k = 1, 2, \ldots.$$

Use these to construct X'_i and X_i^1 as follows:

$$X'_i = \min\left\{k : U_{i,0}U_{i,1}\ldots U_{i,k} > \left(\frac{1}{p_i}\right)^k\right\}$$
$$X_i^1 = \min\left\{k : U_{i,0}U_{i,1}\ldots U_{i,k} > q_{i,k}\right\} \leq X'_i.$$

This completes the construction and the proof, giving the result (8). □

3. Lower Bound for Largest Collision

3.1. Random Vectors with Independent Components

Let $\mathcal{P} = \{p_1, p_2, \ldots\}$ denote the rational primes $\{2, 3, 5, \ldots\}$ in increasing order, and let $a_j = (\log(p_j))^{1/2}$. Instead of the Geometric model (5), switch to a Bernoulli model in which $Z_1, Z_2, \ldots$ are independent Bernoulli random variables, with

$$\mathbb{P}[Z_j = 1] = \frac{1}{p_j}. \tag{9}$$

Let ξ denote the random vector

$$\xi = (a_1 Z_1, a_2 Z_2, \ldots) \in [0, \infty)^{\mathbb{N}} \tag{10}$$

under this new assumption, and let $\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(N)}$ be independent random vectors, all having the same law as ξ. Note that $\xi^{(1)} \cdot \xi^{(2)}$ is not a suitable model for the GCD of two random integers, because the independence assumption (9) is not realistic. However it is a useful context to develop the techniques which will establish the lower bound in Theorem 1.

Write $\xi^{(k)} = \left(a_1 Z_1^k, a_2 Z_2^k, \ldots\right)$. Informally, we seek a lower bound on the log Δ'_N of the largest prime p_i at which a "collision" occurs; collision means that $Z_i^j = 1 = Z_i^k$ for some j, k. Formally,

$$\Delta'_N = \max_{1 \le k < j \le N} \left\{ \max_i \left\{ Z_i^j Z_i^k \log(p_i) \right\} \le \max_{1 \le k < j \le N} \left\{ \xi^{(k)} \cdot \xi^{(j)} \right\}.$$

Proposition 3. *Given $\delta \in (0, \infty)$, and N such that $N^2 > 8\delta$, we may define $\varphi_N = \varphi_N(\delta)$ implicitly by the identity*

$$\int_{\varphi_N}^{2\varphi_N} \frac{N^2 dx}{2x^2 \log(x)} = \delta. \tag{11}$$

Under the assumption of independence of the components of the random vector (10),

$$\lim_{N \to \infty} \mathbb{P}[\Delta'_N \ge \log(\varphi_N(\delta))] \ge 1 - e^{-\delta}. \tag{12}$$

Remark. Existence of such a φ_N is assured by the fact that the integral of $x^{-2}/\log x$ from 2 to 4 is greater than 1/4. From the integration bounds

$$\frac{1}{2\varphi_N \log(2\varphi_N)} = \frac{1}{\log(2\varphi_N)} \int_{\varphi_N}^{2\varphi_N} \frac{dx}{x^2} < \frac{2\delta}{N^2} < \frac{1}{\log(\varphi_N)} \int_{\varphi_N}^{2\varphi_N} \frac{dx}{x^2} = \frac{1}{2\varphi_N \log(\varphi_N)},$$

it follows that φ_N, defined in (11), satisfies $\varphi_N \log(\varphi_N)/N^2 \to 0.25/\delta$. Hence for all sufficiently large N, $\varphi_N < N^2/2$, and

$$\varphi_N > \frac{N^2}{4\delta\log(2\varphi_N)} > \frac{N^2}{8\delta\log(N)}. \tag{13}$$

The proof uses the following technical lemma, which the reader may treat as a warm-up exercise for the more difficult Proposition 5.

Lemma 4. *Let $\mathcal{P}_N$ denote the set of primes p such that $\varphi_N < p \le 2\varphi_N$. Let $\left\{Z_p^k, p \in \mathcal{P}_N, 1 \le k \le N\right\}$ be independent Bernoulli random variables, where $\mathbb{P}\left[Z_p^k = 1\right] = 1/p$. Take $D_p = Z_p^1 + \ldots + Z_p^N$. Then*

$$\lim_{N\to\infty} \mathbb{P}\left[\bigcup_{p\in\mathcal{P}_N} \{D_p \ge 2\}\right] = 1 - e^{-\delta}. \tag{14}$$

Proof. Binomial probabilities give:

$$\begin{aligned}
\mathbb{P}[D_p \le 1] &= \left(1 - \frac{1}{p}\right)^N + \frac{N}{p}\left(1 - \frac{1}{p}\right)^{N-1} \\
&= \left(1 - \frac{1}{p}\right)^N \left(1 + \frac{N}{p-1}\right) \\
&= \left(1 - \frac{N}{p} + \frac{N(N-1)}{2p^2} - \ldots\right)\left(1 + \frac{N}{p-1}\right) \\
&= 1 - \frac{N^2}{2p^2} + O\left(\frac{N}{\varphi_N^2}\right) + O\left(\left(\frac{N}{\varphi_N}\right)^3\right).
\end{aligned}$$

Independence of $\left\{Z_p^k, p \in \mathcal{P}_N, 1 \le k \le N\right\}$ implies independence of $\{D_p, p \in \mathcal{P}_N\}$, so

$$\begin{aligned}
\log \mathbb{P}\left[\bigcap_{p\in\mathcal{P}_N} \{D_p \le 1\}\right] &= \sum_{p\in\mathcal{P}_N} \log\left(\mathbb{P}[D_p \le 1]\right) \\
&= \sum_{p\in\mathcal{P}_N} \log\left(1 - \frac{N^2}{2p^2}\right) + O\left(\frac{N|\mathcal{P}_N|}{\varphi_N^2}\right) + O\left(\frac{N^3|\mathcal{P}_N|}{\varphi_N^3}\right).
\end{aligned}$$

Using the estimates $\varphi_N \log\varphi_N = O\left(N^2\right)$, $|\mathcal{P}_N| = O\left(\varphi_N/\log\varphi_N\right)$, and $p/\varphi_N \le 2$, the last expression becomes

$$= -\sum_{p\in\mathcal{P}_N} \frac{N^2}{2p^2} + O\left(\frac{N^2}{\varphi_N^2}\right) + O\left(\frac{N}{\varphi_N\log(\varphi_N)}\right) + O\left(\frac{N^3}{\varphi_N^2\log(\varphi_N)}\right).$$

All terms but the first vanish in the limit, while the Prime Number Theorem ensures that

$$\lim_{N\to\infty} \sum_{p\in\mathcal{P}_N} \frac{N^2}{2p^2} = \delta.$$

Therefore

$$\lim_{N\to\infty} \mathbb{P}\left[\bigcap_{p\in\mathcal{P}_N} \{D_p \le 1\}\right] = e^{-\delta},$$

and the limit (14) follows. □

Proof of Proposition 3. According to our model, if $D_p \ge 2$ for some $p = p_i \in \mathcal{P}_N$, then there are indices $1 \le k < j \le N$ for which $Z_i^j = 1 = Z_i^k$. Since $\log(p_i) \ge \log(\varphi_N(\delta))$,

$$\lim_{N\to\infty} \mathbb{P}\left[\Delta_N' \ge \log(\varphi_N(\delta))\right] \ge \lim_{N\to\infty} \mathbb{P}\left[\bigcup_{p\in\mathcal{P}_N} \{D_p \ge 2\}\right] = 1 - e^{-\delta}.$$

This verifies (12). □

4. Application: Pairwise GCDs of Many Uniform Random Integers

We shall now prove an analogue of Lemma 4 which applies to random integers, dropping the independence assumption for the components of the random vector (10).

Proposition 5. *Suppose $\alpha > 0$, and $T_1, \ldots, T_N$ is a random sample, drawn with replacement, from the integers $\{n \in \mathbb{N} : n \le e^{\alpha N}\}$. Given $\delta \in (0, \infty)$, define $\varphi_N = \varphi_N(\delta)$ implicitly by the identity (11). Let $\mathcal{P}_N$ denote the set of primes p such that $\varphi_N < p \le 2\varphi_N$; for $p \in \mathcal{P}_N$ let D_p denote the number of elements of $\{T_1, \ldots, T_N\}$ which are divisible by p. Then*

$$\lim_{N\to\infty} \mathbb{P}\left[\bigcup_{p\in\mathcal{P}_N} \{D_p \ge 2\}\right] = 1 - e^{-\delta}. \tag{15}$$

Proof. As noted above, the Prime Number Theorem ensures that

$$\lim_{N\to\infty} \sum_{p\in\mathcal{P}_N} \frac{N^2}{2p^2} = \delta.$$

More generally, the alternating series for the exponential function ensures that there is an even integer $d \ge 1$ such that, given $\epsilon \in (0, 1)$, for all sufficiently large N,

$$1 - e^{-\delta/(1+\epsilon)} < \sum_{r=1}^{d} (-1)^{r+1} I_r < 1 - e^{-\delta/(1-\epsilon)}$$

where, for $\{p_1, \ldots, p_r\} \subset \mathcal{P}_N$

$$I_r = \sum_{p_1<\ldots<p_r} \frac{N^{2r}}{2^r\,(p_1 \ldots p_r)^{\,2}}, \qquad r = 1, 2, \ldots, d.$$

Because $\varphi_N/N^2 \to 0$, it follows that, for every $\{p_1, \ldots, p_d\} \subset \mathcal{P}_N$,

$$\frac{p_1 \ldots p_d}{e^{\alpha N}} < \frac{(\varphi_N)^d}{e^{\alpha N}} < e^{2d\log(N)-\alpha N} \to 0.$$

Suppose that, for this constant value of d, we fix some $\{p_1, \ldots, p_d\} \subset \mathcal{P}_N$; instead of sampling $T_1, \ldots, T_N$ uniformly from integers up to $e^{\alpha N}$, sample $T_1', \ldots, T_N'$ uniformly from integers up to

$$p_1 \ldots p_d \left\lfloor e^{\alpha N} / \left(p_1 \ldots p_d\right) \right\rfloor .$$

From symmetry considerations, the Bernoulli random variables $B_1', \ldots, B_d'$ are independent, with parameters $1\,/p_1\,, \ldots, 1\,/p_d\,$, respectively, where B_i' is the indicator of the event that p_i divides T_1'. By elementary reasoning,

$$\mathbb{P}\left[D_p \geq 2\right] = \frac{N^2}{2p^2} + O\left(\left(N/\varphi_N\right)^3\right);$$

$$\mathbb{P}\left[D_{p_1} \geq 2, \ldots, D_{p_r} \geq 2\right] = \frac{N^{2r}}{2^r\left(p_1 \ldots p_r\right)^{\,2}} + O\left(\left(N\,/\varphi_N\,\right)^{\,2r+1}\right),$$

for $r = 1, 2, \ldots, d$.

If we were to sample $T_1, \ldots, T_N$ instead of $T_1', \ldots, T_N'$, the most that such a probability could change is

$$\mathbb{P}\left[\bigcup_{i=1}^{N}\{T_i \neq T_i'\}\right] \leq \frac{Np_1 \ldots p_d}{e^{\alpha N}} < e^{(2d+1)\log(N)-\alpha N}.$$

The same estimate holds for any choice of $\{p_1, \ldots, p_d\} \subset \mathcal{P}_N$. By the inclusion-exclusion formula, taken to the first d terms,

$$\begin{aligned}\mathbb{P}\left[\bigcup_{p\in\mathcal{P}_N}\{D_p \geq 2\}\right] &\geq \sum_{p\in\mathcal{P}_N}\mathbb{P}\left[D_p \geq 2\right] - \sum_{p_1<p_2}\mathbb{P}\left[D_{p_1} \geq 2, D_{p_2} \geq 2\right] \\ &\qquad + \ldots - \sum_{p_1<\ldots<p_d}\mathbb{P}\left[D_{p_1} \geq 2, \ldots, D_{p_d} \geq 2\right] \\ &= \sum_{r=1}^{d}(-1)^{r+1}I_r + O\left(\left(N\,/\varphi_N\,\right)^{3}\right) + \binom{N}{d}e^{(2d+1)\log(N)-\alpha N}.\end{aligned}$$

So under this simplified model, the reasoning above combines to show that, for all sufficiently large N,

$$1 - e^{-\delta/(1+\epsilon)} < \mathbb{P}\left[\bigcup_{p\in\mathcal{P}_N}\{D_p \geq 2\}\right] < 1 - e^{-\delta/(1-\epsilon)}.$$

Since ϵ can be made arbitrarily small, this verifies the result. □

4.1. Proof of Theorem 1

Suppose $\alpha > 0$, and $T_1, \ldots, T_N$ is a random sample, drawn with replacement, from the integers $\left\{n \in \mathbb{N} : n \leq e^{\alpha N}\right\}$. Let $\Lambda_{j,k}$ denote the largest common prime factor of T_j and T_k. Take

$$\Delta'_N = \max_{1 \leq k < j \leq N} \left\{\log\left(\Lambda_{j,k}\right)\right\}.$$

In the language of Proposition 5, if $D_p \geq 2$ for some $p \in \mathcal{P}_N$, then there are indices $1 \leq k < j \leq N$ for which $\Lambda_{j,k} > \varphi_N$. So inequality (13) and Proposition 5 imply that, for any $\theta = 8\delta > 0$

$$\begin{aligned}\lim_{N\to\infty} \mathbb{P}\left[\Delta'_N \geq 2\log(N) - \log\left(\log\left(N^{\theta}\right)\right)\right] &\geq \lim_{N\to\infty} \mathbb{P}\left[\Delta'_N \geq \log\left(\varphi_N[\theta/8]\right)\right] \\ &\geq \lim_{N\to\infty} \mathbb{P}\left[\bigcup_{p\in\mathcal{P}_N} \{D_p \geq 2\}\right] = 1 - e^{-\theta/8}.\end{aligned}$$

This is precisely the lower bound (3). For any $\eta > 0$, the lower bound in (1) follows from:

$$\lim_{N\to\infty} \mathbb{P}\left[\Delta'_N > (2-\eta)\log(N)\right] = 1.$$

Let $\Gamma_{j,k} \geq \Lambda_{j,k}$ denote the Greatest Common Divisor of T_j and T_k. To obtain the upper bound (2) on $\Gamma_{j,k}$, it suffices by Proposition 2 to check that condition (5) is valid, when X_i denotes the multiplicity to which prime p_i divides T_1. Take any positive integer $r \geq 1$, any prime p_k coprime to r, and any $m \geq 1$. The conditional probability that p_k^m divides T_1, given that r divides T_1, is

$$\frac{\left\lfloor e^{\alpha N}/\left(r p_k^m\right)\right\rfloor}{\left\lfloor e^{\alpha N}/r\right\rfloor} \leq \left(\frac{1}{p_k}\right)^m.$$

So condition (5) holds. Thus (8) holds, which is equivalent to (2).

Finally we derive the upper bound in (1), for an arbitrary $\eta > 0$. Fix $\epsilon \in (0,1)$ and $\eta > 0$. Select $s \in (0,1)$ to satisfy $2/s = 2 + \eta/2$. Then choose $b = C'_s/\epsilon$. According to (8),

$$\mathbb{P}\left[\Delta_N \geq (2+\eta/2)\log(N) + s^{-1}\log(b)\right] \leq \epsilon/2.$$

For any N sufficiently large so that $(\eta/2)\log(N) > s^{-1}\log(b)$,

$$\mathbb{P}\left[\Delta_N \geq (2+\eta)\log(N)\right] \leq \epsilon/2.$$

This yields the desired bound (1). □

References

[1] Patrick Billingsley, *Convergence of Probability Measures*, Wiley, 1999.

[2] Ernest Cesàro, *Étude moyenne du plus grand commun diviseur de deux nombres*, Annali di Matematica Pura e Applicada, 13(2), 233 - 268, 1885.

[3] Persi Diaconis and Paul Erdős, *On the distribution of the greatest common divisor*, A Festschrift for Herman Rubin, 56 - 61, IMS Lecture Notes Monograph Series 45, Institute of Mathematical Statistics, 2004.

[4] John Knopfmacher, *Abstract Analytic Number Theory*, Dover, New York, 1990.

[5] J. E. Nymann, *On the probability that k positive integers are relatively prime*, Journal of Number Theory, 4(5), 469 - 473, 1972.