THE LEAST COMMON MULTIPLE OF RANDOM SETS OF POSITIVE INTEGERS

JAVIER CILLERUELO, JUANJO RUÉ, PAULIUS ŠARKA, AND ANA ZUMALACÁRREGUI

Abstract. We study the typical behavior of the least common multiple of the elements of a random subset $A \subset \{1, \dots, n\}$. For example we prove that $\operatorname{lcm}\{a : a \in A\} = 2^{n(1+o(1))}$ for almost all subsets $A \subset \{1, \dots, n\}$.

1. Introduction

The function $\psi(n) = \log \operatorname{lcm}\{m : 1 \le m \le n\}$ was introduced by Chebyshev in his study on the distribution of the prime numbers. It is a well known fact that the asymptotic relation $\psi(n) \sim n$ is equivalent to the Prime Number Theorem, which was proved independently by J. Hadamard and C.J. de la Vallée Poussin.

In the present paper, instead of considering the whole set $\{1, \dots, n\}$, we study the typical behavior of the quantity $\psi(A) := \log \operatorname{lcm}\{a : a \in A\}$ for a random set A in $\{1, \dots, n\}$ when $n \to \infty$. We define $\psi(\emptyset) = 0$. We consider two natural models.

In the first one, denoted by $B(n, \delta)$, each element in A is chosen independently at random in $\{1, \dots, n\}$ with probability $\delta = \delta(n)$, typically a function of n.

Theorem 1.1. *If $\delta = \delta(n) < 1$ and $\delta n \to \infty$ then*

$$\psi(A) \sim n \frac{\delta \log(\delta^{-1})}{1 - \delta}$$

asymptotically almost surely in $B(n, \delta)$ when $n \to \infty$.

The case $\delta = 1$ corresponds to the classical Chebyshev function and its asymptotic estimate appears as the limiting case, as δ tends to 1, in Theorem 1.1, since $\lim_{\delta \to 1} \frac{\delta \log(\delta^{-1})}{1-\delta} = 1$.

When $\delta = 1/2$ all the subsets $A \subset \{1, \dots, n\}$ are chosen with the same probability and Theorem 1.1 gives the following result.

Corollary 1.1. *For almost all sets $A \subset \{1, \dots, n\}$ we have that*

$$\operatorname{lcm}\{a : a \in A\} = 2^{n(1+o(1))}.$$

For a given positive integer $k = k(n)$, again typically a function of n, we consider the second model, where each subset of k elements is chosen uniformly at random among all sets of size k in $\{1, \dots, n\}$. We denote this model by $S(n, k)$.

When $\delta = k/n$ the heuristic suggests that both models are quite similar. Indeed, this is the strategy we follow to prove Theorem 1.2.

Theorem 1.2. *For $k = k(n) < n$ and $k \to \infty$ we have*

$$\psi(A) = k\frac{\log(n/k)}{1 - k/n}\left(1 + O(e^{-C\sqrt{\log k}})\right)$$

almost surely in $S(n,k)$ when $n \to \infty$ for some positive constant C.

The case $k = n$, which corresponds to Chebyshev's function, is also obtained as a limiting case in Theorem 1.2 in the sense that $\lim_{k/n\to 1} \frac{\log(n/k)}{1-k/n} = 1$.

This work has been motivated by a result of the first author about the asymptotic behavior of $\psi(A)$ when $A = A_{q,n} := \{q(m) : \ 1 \le q(m) \le n\}$ for a quadratic polynomial $q(x) \in \mathbb{Z}[x]$. We wondered if that behavior was typical among the sets $A \subset \{1, \dots, n\}$ of similar size. We analyze this issue in the last section.

2. Chebyshev's function for random sets in $B(n,\delta)$. Proof of Theorem 1.1

The following lemma provides us with an explicit expression for $\psi(A)$ in terms of the Mangoldt function

$$\Lambda(m) = \begin{cases} \log p & \text{if } m = p^k \text{ for some } k \ge 1 \\ 0, & \text{otherwise.} \end{cases}$$

Lemma 2.1. *For any set of positive integers A we have $\psi(A) = \sum_m \Lambda(m) I_A(m)$, where Λ denotes the classical Von Mangoldt function and*

$$I_A(m) = \begin{cases} 1 & \textit{if } A \cap \{m, 2m, 3m, \dots\} \neq \emptyset, \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. We observe that for any positive integer l, the number $\log l$ can be written as $\log l = \sum_{p^k | l} \log p$, where the sum is taken over all the powers of primes. Thus, using that $p^k |\, \mathrm{lcm}\{a : a \in A\}$ if and only if $A \cap \{p^k, 2p^k, 3p^k, \dots\} \neq \emptyset$, we get

$$\log \mathrm{lcm}(a : \ a \in A) = \sum_{p^k | \mathrm{lcm}(a:\ a\in A)} \log p = \sum_{p^k} (\log p) I_A(p^k) = \sum_m \Lambda(m) I_A(m).$$

□

Note that if $A = \{1, \dots, n\}$ then $\psi(A) = \sum_{m \le n} \Lambda(m)$ is the classical Chebychev function $\psi(n)$.

2.1. **Expectation.** First of all we give an explicit expression for the expected value of the random variable $X = \psi(A)$ where A is a random set in $B(n,\delta)$.

Proposition 2.1. *For the random variable $X = \psi(A)$ in $B(n,\delta)$ we have*

$$\mathbb{E}(X) = n\frac{\delta\log(\delta^{-1})}{1-\delta} + \delta\sum_{r\ge 1} R\left(\frac{n}{r}\right)(1-\delta)^{r-1},$$

where $R(x) = \psi(x) - x$ denotes the error term in the Prime Number Theorem.

Proof. The ambiguous case $\delta = 1$ must be understood as the limit as $\delta \to 1$, which recovers the equality $\psi(n) = n + R(n)$. In the following we assume that $\delta < 1$. By linearity of the expectation, Lemma 2.1 clearly implies

$$\mathbb{E}(X) = \sum_{m \le n} \Lambda(m)\mathbb{E}(I_A(m)).$$

Since $\mathbb{E}(I_A(m)) = \mathbb{P}(A \cap \{m, 2m, \dots\} \neq \emptyset) = 1 - \prod_{r \le n/m} \mathbb{P}(rm \notin A) = 1 - (1-\delta)^{\lfloor n/m \rfloor}$, we obtain

$$\mathbb{E}(X) = \sum_{m \le n} \Lambda(m) \left(1 - (1-\delta)^{\lfloor n/m \rfloor}\right). \tag{1}$$

We observe that $\lfloor n/m \rfloor = r$ whenever $\frac{n}{r+1} < m \le \frac{n}{r}$, so we split the sum into intervals $J_r = (\frac{n}{r+1}, \frac{n}{r}]$, obtaining

$$\begin{aligned}
\mathbb{E}\left(X\right) &= \sum_{r \ge 1} (1 - (1-\delta)^r) \sum_{m \in J_r} \Lambda(m) \\
&= \sum_{r \ge 1} (1 - (1-\delta)^r) \left(\psi\Big(\frac{n}{r}\Big) - \psi\Big(\frac{n}{r+1}\Big)\right) \\
&= \delta \sum_{r \ge 1} \psi\Big(\frac{n}{r}\Big)(1-\delta)^{r-1} \\
&= \delta n \sum_{r \ge 1} \frac{(1-\delta)^{r-1}}{r} + \delta \sum_{r \ge 1} R\left(\frac{n}{r}\right)(1-\delta)^{r-1}. \\
&= n \frac{\delta \log(\delta^{-1})}{1-\delta} + \delta \sum_{r \ge 1} R\left(\frac{n}{r}\right)(1-\delta)^{r-1}.
\end{aligned}$$

□

Corollary 2.1. *If $\delta = \delta(n) < 1$ and $\delta n \to \infty$ then*

$$\mathbb{E}\left(X\right) = n \frac{\delta \log(\delta^{-1})}{1-\delta} \left(1 + O\left(e^{-C\sqrt{\log(\delta n)}}\right)\right).$$

for some constant $C > 0$.

Proof. We estimate the absolute value of sum appearing in Proposition 2.1. For any positive integer T and using that $|R(y)| < 2y$ for all $y > 0$ we have

$$\begin{aligned}
\sum_{r \ge 1} |R\left(n/r\right)|(1-\delta)^{r-1} &= \sum_{1 \le r \le T} |R\left(n/r\right)|(1-\delta)^{r-1} + \sum_{r \ge T+1} |R\left(n/r\right)|(1-\delta)^{r-1} \\
&\le n \sum_{1 \le r \le T} \frac{|R\left(n/r\right)|}{(n/r)} \frac{(1-\delta)^{r-1}}{r} + 2n \sum_{r \ge T+1} \frac{(1-\delta)^{r-1}}{r} \\
&\le n \left(\max_{x \ge n/T} \frac{|R(x)|}{x}\right) \sum_{1 \le r \le T} \frac{(1-\delta)^{r-1}}{r} + 2n \sum_{r \ge T+1} \frac{(1-\delta)^{r-1}}{r} \\
&\le n \frac{\log(\delta^{-1})}{(1-\delta)} \left(\max_{x \ge n/T} \frac{|R(x)|}{x}\right) + \frac{2n}{T+1} \frac{(1-\delta)^T}{\delta}
\end{aligned}$$

Taking into account that $(1-\delta)^T < e^{-\delta T}$ and the known estimate

$$\max_{x > y} \frac{|R(x)|}{x} \ll e^{-C_1\sqrt{\log y}}$$

for the error term in the PNT, we have

$$\sum_{r\geq 1} |R\,(n/r)\,|(1-\delta)^{r-1} \ll n\frac{\log(\delta^{-1})}{(1-\delta)} e^{-C_1\sqrt{\log(n/T)}} + n\frac{e^{-\delta T}}{\delta T}.$$

Thus we have proved that for any positive integer T we have

$$\mathbb{E}(X) = n\frac{\delta\log(\delta^{-1})}{1-\delta}\left(1 + O\left(e^{-C_1\sqrt{\log(n/T)}}\right) + O\left(\frac{1-\delta}{\log(\delta^{-1})}\frac{e^{-\delta T}}{\delta T}\right)\right).$$

We take $T \asymp \delta^{-1}\sqrt{\log(\delta n)}$ to minimize the error term. To estimate the first error term we observe that $\log(n/T) \gg \log(\delta n/\sqrt{\log(\delta n)}) \gg \log(\delta n)$, so $e^{-C_1\sqrt{\log(n/T)}} \ll e^{-C\sqrt{\log(\delta n)}}$ for some constant C. To bound the second error term we simply observe that $\delta T > 1$ and that $\frac{1-\delta}{\log(\delta^{-1})} \leq 1$ and we get a similar upper bound. $\square$

2.2. **Variance.**

Proposition 2.2. *For the random variable $X = \psi(A)$ in $B(n,\delta)$ we have*

$$V(X) \ll \delta n \log^2 n.$$

Proof. By linearity of expectation we have that

$$\begin{aligned} V(X) &= \mathbb{E}\left(X^2\right) - \mathbb{E}^2\left(X\right) \\ &= \sum_{m,l\leq n} \Lambda(m)\Lambda(l)\left(\mathbb{E}\left(I_A(m)I_A(l)\right) - \mathbb{E}\left(I_A(m)\right)\mathbb{E}\left(I_A(l)\right)\right). \end{aligned}$$

We observe that if $\Lambda(m)\Lambda(l) \neq 0$ then $l \mid m,\ m \mid l$ or $(m,l) = 1$. Let us now study the term $\mathbb{E}(I_A(m)I_A(l))$ in these cases.

(i) If $l \mid m$ then

$$\mathbb{E}(I_A(m)I_A(l)) = 1 - (1-\delta)^{\lfloor n/m\rfloor}.$$

(ii) If $(l,m) = 1$ then

$$\mathbb{E}(I_A(m)I_A(l)) = 1 - (1-\delta)^{\lfloor n/m\rfloor} - (1-\delta)^{\lfloor n/l\rfloor} + (1-\delta)^{\lfloor n/m\rfloor + \lfloor n/l\rfloor - \lfloor n/ml\rfloor}.$$

Both of these relations are subsumed in

$$\mathbb{E}(I_A(m)I_A(l)) = 1 - (1-\delta)^{\lfloor n/m\rfloor} - (1-\delta)^{\lfloor n/l\rfloor} + (1-\delta)^{\lfloor n/m\rfloor + \lfloor n/l\rfloor - \lfloor n(m,l)/ml\rfloor}.$$

Therefore, it follows from (1) that for each term in the sum we have

$$\begin{aligned} &\Lambda(m)\Lambda(l)\left(\mathbb{E}\left(I_A(m)I_A(l)\right) - \mathbb{E}\left(I_A(m)\right)\mathbb{E}\left(I_A(l)\right)\right) \\ &= \Lambda(m)\Lambda(l)(1-\delta)^{\lfloor n/m\rfloor + \lfloor n/l\rfloor - \lfloor n(m,l)/ml\rfloor}\left(1 - (1-\delta)^{\lfloor n(m,l)/ml\rfloor}\right). \end{aligned}$$

Finally, by using the inequality $1-(1-x)^r \leq rx$ we have

$$\Lambda(m)\Lambda(l)\left(\mathbb{E}\left(I_A(m)I_A(l)\right) - \mathbb{E}\left(I_A(m)\right)\mathbb{E}\left(I_A(l)\right)\right) \leq \delta n\frac{\Lambda(l)}{l}\frac{\Lambda(m)}{m}(m,l),$$

and therefore:

$$V(X) \leq 2\delta n \sum_{1\leq l\leq m\leq n} \frac{\Lambda(l)}{l}\frac{\Lambda(m)}{m}(m,l).$$

We now split the sum according to $l \mid m$ or $(l,m)=1$ and estimate each one separately.

$$\sum_{\substack{1\le l\le m\le n\\ l\mid m}} \frac{\Lambda(l)}{l}\frac{\Lambda(m)}{m}(m,l) \le \sum_{p\le n}\sum_{1\le j\le i} \frac{\log p}{p^j}\frac{\log p}{p^i}p^j = \sum_{p\le n}\sum_{1\le i} \frac{i\log^2 p}{p^i} \ll \log^2 n,$$

$$\sum_{\substack{1\le l\le m\le n\\ (l,m)=1}} \frac{\Lambda(l)}{l}\frac{\Lambda(m)}{m}(m,l) \le \left(\sum_{1\le l\le n}\frac{\Lambda(l)}{l}\right)\left(\sum_{1\le m\le n}\frac{\Lambda(m)}{m}\right) \ll \log^2 n,$$

as we wanted to prove. $\square$

We finish the proof of Theorem 1.1 by observing that $V(X)=o(\mathbb{E}(X)^2)$ when $\delta n\to\infty$, so $X\sim\mathbb{E}(X)$ asymptotically almost surely.

3. Chebyshev's function for random sets in $S(n,k)$. Proof of Theorem 1.2

Let us consider again the random variable $X=\psi(A)$, but in the model $S(n,k)$. From now on $\mathbb{E}_k(X)$ and $V_k(X)$ will denote the expected value and the variance of X in this probability space. Clearly, for $s=1,2$ we have

$$\begin{aligned}\mathbb{E}_k(X^s) &= \frac{1}{\binom{n}{k}}\sum_{|A|=k}\psi^s(A)\\ V_k(X) &= \frac{1}{\binom{n}{k}}\sum_{|A|=k}\left(\psi(A)-\mathbb{E}_k(X)\right)^2\end{aligned}$$

Lemma 3.1. *For $s=1,2$ and $1\le j<k$ we have that*

$$\mathbb{E}_j(X^s)\le\mathbb{E}_k(X^s)\le\mathbb{E}_j(X^s)+(k^s-j^s)\log^s n.$$

Proof. In order to prove the lower bound it is enough to consider the case $j=k-1$. Observe that the function ψ is monotone with respect to inclusion, i.e. $\psi(A\cup\{a\})\ge\psi(A)$ for any $A,\{a\}\subseteq[n]$. Using this we get

$$\sum_{|A|=k-1}\psi^s(A)\le\frac{1}{n-k+1}\sum_{a\in[n]\setminus A}\psi^s(A\cup\{a\})=\frac{k}{(n-k+1)}\sum_{|A'|=k}\psi^s(A').$$

Inequality then follows from $\binom{n}{k-1}=\frac{k}{(n-k+1)}\binom{n}{k}$.

For the second inequality we observe that for any set $A\in\binom{[n]}{k}$ and any partition into two sets $A=A'\cup A''$ with $|A'|=j,\ |A''|=k-j$ we have that $\psi(A)\le\psi(A')+\psi(A'')\le\psi(A')+(k-j)\log n$. Similarly,

$$\begin{aligned}\psi^2(A) &\le (\psi(A')+(k-j)\log n)^2\\ &= \psi^2(A')+2\psi(A')(k-j)\log n+(k-j)^2\log^2 n\\ &\le \psi^2(A')+2j(k-j)\log^2 n+(k-j)^2\log^2 n\\ &= \psi^2(A')+(k^2-j^2)\log^2 n.\end{aligned}$$

Thus, for $s = 1, 2$ we have

$$\begin{aligned}\psi^s(A) &\le \binom{k}{j}^{-1} \sum_{\substack{A'\subset A\\|A'|=j}} (\psi^s(A') + (k^s - j^s)\log^s n)\\ &\le \binom{k}{j}^{-1} \Big(\sum_{\substack{A'\subset A\\|A'|=j}} \psi^s(A') \Big) + (k^s - j^s)\log^s n.\end{aligned}$$

Then,

$$\begin{aligned}\sum_{|A|=k} \psi^s(A) &\le \binom{k}{j}^{-1} \sum_{|A|=k} \sum_{\substack{A'\subset A\\|A'|=j}} \psi^s(A') + \binom{n}{k}(k^s - j^s)\log^s n\\ &= \binom{k}{j}^{-1} \sum_{|A'|=j} \psi^s(A') \sum_{\substack{A'\subset A\\|A|=k}} 1 + \binom{n}{k}(k^s - j^s)\log^s n\\ &= \binom{k}{j}^{-1} \binom{n-j}{k-j} \sum_{|A'|=j} \psi^s(A') + \binom{n}{k}(k^s - j^s)\log^s n\\ &= \frac{\binom{n}{k}}{\binom{n}{j}} \sum_{|A'|=j} \psi^s(A') + \binom{n}{k}(k^s - j^s)\log^s n,\end{aligned}$$

and the second inequality holds. $\square$

Proposition 3.1. *For $s = 1, 2$ we have that*

$$\mathbb{E}_k(X^s) = \mathbb{E}(X^s) + O(k^{s-1/2}\log^s n)$$

where $\mathbb{E}(X^s)$ denotes the expectation of X^s in $B(n, k/n)$ and $\mathbb{E}_k(X^s)$ the expectation in $S(n, k)$.

Proof. Observe that for $s = 1, 2$ we have

$$\begin{aligned}\mathbb{E}(X^s) - \mathbb{E}_k(X^s) &= -\mathbb{E}_k(X^s) + \sum_{j=0}^{n} \left(\frac{k}{n}\right)^j \left(1 - \frac{k}{n}\right)^{n-j} \sum_{|A|=j} \psi^s(A)\\ &= -\mathbb{E}_k(X^s) + \sum_{j=0}^{n} \left(\frac{k}{n}\right)^j \left(1 - \frac{k}{n}\right)^{n-j} \binom{n}{j} \mathbb{E}_j(X^s)\\ &= \sum_{j=0}^{n} \left(\frac{k}{n}\right)^j \left(1 - \frac{k}{n}\right)^{n-j} \binom{n}{j} \left(\mathbb{E}_j(X^s) - \mathbb{E}_k(X^s)\right),\end{aligned}$$

for $s = 1, 2$. Using Lemma 3.1 we get

$$|\mathbb{E}_k(X^s) - \mathbb{E}(X^s)| \le \log^s n \sum_{j=0}^{n} \left(\frac{k}{n}\right)^j \left(1 - \frac{k}{n}\right)^{n-j} \binom{n}{j} |j^s - k^s|. \tag{2}$$

The sum in (2) for $s = 1$ is $\mathbb{E}(|Y - \mathbb{E}(Y)|)$, where $Y \sim \mathrm{Bin}(n, k/n)$ is the binomial distribution of parameters n and k/n. Cauchy–Schwarz inequality for the expectation implies that this quantity is bounded by the standard deviation of the binomial distribution.

$$\sum_{j=0}^{n} \left(\frac{k}{n}\right)^j \left(1 - \frac{k}{n}\right)^{n-j} \binom{n}{j} |j - k| \le \sqrt{n(k/n)(1 - k/n)} \le \sqrt{k}, \tag{3}$$

which proves Proposition 3.1 for $s = 1$.

To estimate the sum in (2) for $s = 2$, we split the expression in two terms: the sum indexed by $j \le 2k$ and the one with $j > 2k$. We use (3) to get

$$\begin{aligned}\sum_{j\le 2k}\left(\frac{k}{n}\right)^j\left(1-\frac{k}{n}\right)^{n-j}\binom{n}{j}|j^2-k^2| &\le 3k\sum_{j=0}^{n}\left(\frac{k}{n}\right)^j\left(1-\frac{k}{n}\right)^{n-j}\binom{n}{j}|j-k|\\ &\le 3k^{3/2}.\end{aligned}$$

On the other hand,

$$\begin{aligned}\sum_{j>2k} &\left(\frac{k}{n}\right)^j\left(1-\frac{k}{n}\right)^{n-j}\binom{n}{j}|j^2-k^2|\\ \le &\sum_{l\ge 2}(l+1)^2k^2\sum_{lk<j\le(l+1)k}\left(\frac{k}{n}\right)^j\left(1-\frac{k}{n}\right)^{n-j}\binom{n}{j}\\ \le &\sum_{l\ge 2}(l+1)^2k^2\ \mathbb{P}(Y>lk)\end{aligned}$$

where, once again, $Y \sim \mathrm{Bin}(n, k/n)$. Chernoff's Theorem implies that for any $\epsilon > 0$ we have

$$\mathbb{P}(Y > (1+\epsilon)k) \le e^{-\epsilon^2 k/3}.$$

Applying this inequality to $\mathbb{P}(Y > lk)$ we get

$$\begin{aligned}\sum_{j>2k} &\left(\frac{k}{n}\right)^j\left(1-\frac{k}{n}\right)^{n-j}\binom{n}{j}|j^2-k^2|\\ \le &\sum_{l\ge 2}(l+1)^2k^2e^{-(l-1)^2k/3} \ll k^2e^{-k/3} \ll k^{3/2}.\end{aligned}$$

□

The next corollary proves the first part of Theorem 1.2.

Corollary 3.1. *If $k = k(n) < n$ and $k \to \infty$ then*

$$\mathbb{E}_k(X) = k\frac{\log(n/k)}{1-k/n}\left(1 + O\left(e^{-C\sqrt{\log k}}\right)\right)$$

Proof. Proposition 3.1 for $s = 1$ and Corollary 2.1 imply that

$$\mathbb{E}_k(X) = k\frac{\log(n/k)}{1-k/n}\left(1 + O\left(e^{-C\sqrt{\log k}}\right) + O\left(k^{-1/2}\log n\frac{1-k/n}{\log(n/k)}\right)\right).$$

It is clear that

$$k^{-1/2}\log n\frac{(1-k/n)}{\log(n/k)} = O\left(k^{-1/2}\log k\right) = o\left(e^{-C\sqrt{\log k}}\right),$$

when $k \to \infty$, which concludes the proof. □

To conclude the proof of Theorem 1.2 we combine Proposition 2.2 and Proposition 3.1 to estimate the variance $V_k(X)$ in $S(n,k)$:

$$\begin{aligned}V_k(X) &= \mathbb{E}_k(X^2) - \mathbb{E}_k^2(X)\\ &= V(X) + \left(\mathbb{E}_k(X^2) - \mathbb{E}(X^2)\right) + \left(\mathbb{E}(X) - \mathbb{E}_k(X)\right)\left(\mathbb{E}(X) + \mathbb{E}_k(X)\right)\\ &\ll k\log^2 n + \left(k^{1/2}\log n\right)(k\log n)\\ &\ll k^{3/2}\log^2 n.\end{aligned}$$

The second assertion of Theorem 1.2 is a consequence of the estimate $V_k(X) = o\left(\mathbb{E}_k^2(X)\right)$ when $k\to\infty$.

3.1. **The case when k is constant.** The case when k is constant and $n\to\infty$ is not relevant for our original motivation but we give a brief analysis for the sake of completeness. In this case Fernández and Fernández [3] have proved that $\mathbb{E}_k(\psi(A)) = k\log n + C_k + o(1)$ where $C_k = -k + \sum_{j=2}^k \binom{k}{j}(-1)^j\frac{\zeta'(j)}{\zeta(j)}$. Actually, they consider the probabilistic model with k independent choices in $\{1,\dots,n\}$, but when k is fixed it does not make big differences because the probability of a repetition between the k choices is tiny.

It is easy to prove that with probability $1-o(1)$ we have that $\psi(A)\sim k\log n$. To see this we observe that

$$a_1\cdots a_k\prod_{i<j}(a_i,a_j)^{-1} \le \operatorname{lcm}(a_1,\dots,a_k) \le a_1\cdots a_k \le n^k,$$

so $\sum_{i=1}^k \log a_i - \sum_{i<j}\log(a_i,a_j) \le \psi(A) \le k\log n$.

Now, let us note that $\mathbb{P}(a_i \le n/\log n \text{ for some } i=1,\dots,k) \le k/\log n$ and that $\mathbb{P}((a_i,a_j)\ge \log n) \le \sum_{d>\log n}\mathbb{P}(d\mid a_i,\ d\mid a_j) \le \sum_{d>\log n}\frac{1}{d^2} < \frac{2}{\log n}$. These observations imply that with probability at least $1-\frac{k+2\binom{k}{2}}{\log n} = 1-\frac{2k}{\log n}$ we have that

$$k\log n\left(1 - O\left(\log\log n/\log n\right)\right) \le \psi(A) \le k\log n.$$

The analysis in the model $B(n,\delta)$ when $\delta n\to c$ can be done using again Proposition 2.1.

$$\mathbb{E}\left(\psi(A)\right) = n\frac{\delta\log(\delta^{-1})}{1-\delta} + \delta\sum_{r<n/\log n} R\left(\frac{n}{r}\right)(1-\delta)^{r-1} + \delta\sum_{n/\log n\le r\le n} R\left(\frac{n}{r}\right)(1-\delta)^{r-1}$$

We use the estimate $R(x)\ll x/\log x$ in the first sum and the estimate $R(x)\ll x$ in the second one. We have

$$\begin{aligned}\mathbb{E}\left(\psi(A)\right) &= c\log n(1+o(1)) + O\left(\frac{c}{\log\log n}\sum_{r<\frac{n}{\log n}}\frac{(1-\delta)^{r-1}}{r}\right) + O\left(c\sum_{\frac{n}{\log n}\le r\le n}\frac{(1-\delta)^{r-1}}{r}\right)\\ &= c\log n + O\left(\frac{c\log\delta}{\log\log n}\right) + O\left(c\log\log n\right)\\ &= c\log n(1+o(1)).\end{aligned}$$

Of course in this model we cannot expect concentration around the expectation because for example the probability that A is the empty set tends to a positive constant, $\mathbb{P}(A=\emptyset)\to e^{-c}$, and then $\mathbb{P}(\psi(A)=0)\to e^{-c}$.

4. The least common multiple of the values of a polynomial

Chebyshev's function could be also generalized to

$$\psi_q(n) = \log \operatorname{lcm} \{q(k) : \ 1 \le k, \ 1 \le q(k) \le n\}$$

for a given polynomial $q(x) \in \mathbb{Z}[x]$ and it is natural to try to obtain the asymptotic behavior for $\psi_q(n)$. Some progress has been made in this direction. While the Prime Number Theorem is equivalent to the asymptotic $\psi_q(n) \sim n$ for $q(x) = x$, Paul Bateman noticed that the Prime Number Theorem for arithmetic progressions could be exploited to obtain the asymptotic estimate when $q(x) = a_1x + a_0$ is a linear polynomial and proposed it as a problem [1] in the American Mathematical Monthly:

$$\psi_q(n) \sim \frac{n}{a_1} \frac{m}{\phi(m)} \sum_{\substack{1 \le l \le m \\ (l,m)=1}} \frac{1}{l},$$

where $m = a_1/(a_1, a_0)$. The first author [2] has extended this result to quadratic polynomials. For a given irreducible quadratic polynomial $q(x) = a_2x^2 + a_1x + a_0$ with $a_2 > 0$ the following asymptotic estimate holds:

$$\psi_q(n) = \frac{1}{2} (n/a_2)^{1/2} \log (n/a_2) + B_q (n/a_2)^{1/2} + o(n^{1/2}), \tag{4}$$

where the constant B_q depends only on q. In the particular case of $q(x) = x^2 + 1$, he got $\psi_q(n) = \frac{1}{2} n^{1/2} \log n + B_q n^{1/2} + o(n^{1/2})$ with

$$B_q = \gamma - 1 - \frac{\log 2}{2} - \sum_{p \ne 2} \frac{(-1)^{\frac{p-1}{2}} \log p}{p-1},$$

where γ is the Euler constant and the sum is considered over all odd prime numbers. It has been proved [4] that the error term in (4) for $q(x) = x^2 + 1$ is $O\left(n^{1/2} (\log n)^{-4/9+\epsilon}\right)$ for each $\epsilon > 0$. When $q(x)$ is a reducible polynomial the behavior is, however, different. In this case it is known (see Theorem 3 in [2]) that:

$$\psi_q(n) \sim cn^{1/2}$$

where c is an explicit constant depending only on q. For example for $q(x) = x^2 - 1$ the constant is $c = 1$.

The asymptotic behavior of $\psi_q(n)$ remains unknown for irreducible polynomials of higher degree.

Conjecture 1 (Cilleruelo [2])**.** *Let $q(x)$ be an irreducible polynomial of degree $d \ge 3$. Then*

$$\psi_q(n) \sim (1 - 1/d) (n/a_d)^{1/d} \log (n/a_d), \tag{5}$$

where $a_d > 0$ is the coefficient of x^d in $q(x)$.

For example, this conjecture would imply $\psi_q(n) \sim \frac{2}{3} n^{1/3} \log n$ for $q(x) = x^3 + 2$.

We observe that $\psi_q(n) = \psi(A_{q,n})$ where $A_{q,n} := \{q(k) : \ 1 \le k, \ 1 \le q(k) \le n\}$. It is natural to wonder whether for a given polynomial $q(x)$ the asymptotic $\mathbb{E}_k(X) \sim \psi_q(n)$ holds, when $n \to \infty$, where $k = |A_{q,n}|$ and $X = \psi(A)$ for a random set A of k elements in $\{1, \ldots, n\}$.

However, consider for example the polynomials $q(x) = x^2 - 1$ and $q(x) = x^2 + 1$. In both cases $|A_{q,n}| \sim \sqrt{n}$ but the asymptotic behaviors of $\psi_q(n)$ are distinct:

$$\psi_q(n) \sim \begin{cases} \sqrt{n} & \text{when } q(x) = x^2 - 1 \\ \frac{1}{2}\sqrt{n}\log n & \text{when } q(x) = x^2 + 1. \end{cases}$$

So, what is the typical behavior of $\psi(A)$ when $|A| \sim \sqrt{n}$? Is it like in the reducible case or like in the irreducible one? Maybe neither of them represent the typical behavior of a random set.

This question was the original motivation of this work. Theorem 1.2 with $k = |A_{q,n}| = \sqrt{n/a_2} + O(1)$ gives

$$\mathbb{E}_k(X) = k\frac{\log(n/k)}{1-k/n}\left(1 + O\left(e^{-C\sqrt{\log k}}\right)\right) = \frac{1}{2}(n/a_2)^{1/2}\log(n/a_2) + o\left(n^{1/2}\right).$$

This shows that, when $q(x)$ is an irreducible quadratic polynomial, the asymptotic behavior of $\psi_q(n)$ coincides with $\psi(A)$, for almost all sets of size $|A_{q,n}|$. Theorem 1.2 also supports Conjecture 1 for any $q(x) = a_d x^d + \cdots + a_0$ irreducible polynomial of degree $d \geq 3$.

Nevertheless, there are some differences in the second term. For example, if $q(x) = x^2 + 1$, we have

$$\psi_q(n) = \frac{1}{2}n^{1/2}\log n + B_q n^{1/2} + o(n^{1/2}),$$

for $B = -0.06627563...$ On the other hand, Theorem 1.2 implies that in corresponding model $S(n,k)$ with $k = |A_{q,n}| = \lfloor\sqrt{n-1}\rfloor$ we have that

$$\psi(A) = \frac{1}{2}n^{1/2}\log n + o(n^{1/2})$$

almost surely. In other words, when $q(x)$ is an irreducible quadratic polynomial, the asymptotic behavior of $\psi_q(n)$ is the same that $\psi(A)$ in the corresponding model $S(n,k)$. But, the second term is not typical unless $B_q = 0$. Probably $B_q \neq 0$ for any irreducible quadratic polynomial $q(x)$ but we have not found a proof.

Acknowledgments: This work was supported by grants MTM 2011-22851 of MICINN, ICMAT Severo Ochoa project SEV-2011-0087 and Research Council of Lithuania Grant No. MIP-068/2013/LSS-110000-740. Part of this work was made during third author's visit at Universidad Autónoma de Madrid, he would like to thank the people of the Mathematics Department and especially Javier Cilleruelo for their warm hospitality. The authors are also grateful to Surya Ramana and Javier Cárcamo for useful comments, and especially to Ben Green whose suggestions improved the clarity and exposition of the results.

References

[1] P. Bateman, J. Kalb and A. Stenger, A limit involving Least Common Multiples, *American Mathematical Monthly* 10797, **109** (2002), no. 4, 393-394.

[2] J. Cilleruelo, The least common multiple of a quadratic sequence, *Compositio Mathematica* **147** (2011) no. 4, 1129-1150.

[3] J.L. Fernández and P. Fernández, On the probability distribution of the gcd and lcm of r-tuples of integers. Preprint.

[4] J. Rué, P. Šarka, A. Zumalacárregui, On the error term of the logarithm of the lcm of a quadratic sequence, *Journal de Théorie des Nombres de Bordeaux* **25**(2013), no. 2 , p. 457-470.

J. Cilleruelo: Instituto de Ciencias Matemáticas (CSIC-UAM-UC3M-UCM) and Departamento de Matemáticas, Universidad Autónoma de Madrid, 28049 Madrid, Spain

E-mail address: `franciscojavier.cilleruelo@uam.es`

J. Rué: Institut für Mathematik, Freie Universität Berlin, Arnimallee 3-5, D-14195 Berlin, Germany

E-mail address: `jrue@zedat.fu-berlin.de`

P. Šarka: Institute of Mathematics and Informatics, Akademijos 4, Vilnius LT-08663, Lithuania and Department of Mathematics and Informatics, Vilnius University, Naugarduko 24, Vilnius LT-03225, Lithuania

E-mail address: `paulius.sarka@gmail.com`

A. Zumalacárregui: Instituto de Ciencias Matemáticas (CSIC-UAM-UC3M-UCM) and Departamento de Matemáticas, Universidad Autónoma de Madrid, 28049 Madrid, Spain

E-mail address: `ana.zumalacarregui@uam.es`