MEMBERSHIP IN RANDOM RATIO SETS

CARLO SANNA†

Abstract. Let $A$ be a random set constructed by picking independently each element of \{1, \ldots, n\} with probability $\alpha \in (0, 1)$. We give a formula for the probability that a rational number $q$ belong to the random ratio set $A/A := \{a/b : a, b \in A\}$. This generalizes a previous result of Cilleruelo and Guijarro-Ordóñez. Moreover, we make some considerations about formulas for the probability of the event $\bigvee_{i=1}^{k}(q_i \in A/A)$, where $q_1, \ldots, q_k$ are rational numbers, showing that they are related to the study of the connected components of certain graphs. In particular, we give formulas for the probability that $q^e \in A/A$ for some $e \in E$, where $E$ is a finite or cofinite set of positive integers with $1 \in E$.

1. Introduction

For every positive integer $n$ and for every $\alpha \in (0, 1)$, let $B(n, \alpha)$ denote the probabilistic model in which a random set $A \subseteq \{1, \ldots, n\}$ is constructed by picking independently every element of \{1, \ldots, n\} with probability $\alpha$. Several authors studied number-theoretic objects involving random sets in this probabilistic model, including: the least common multiple $\text{lcm}(A)$ [1, 4] (see also [8]), the product set $AA := \{ab : a, b \in A\}$ [3, 6, 7], and the ratio set $A/A := \{a/b : a, b \in A\}$ [2, 3].

Regarding random ratio sets, Cilleruelo and Guijarro-Ordóñez [2] proved the following:

Theorem 1.1. Let $A$ be a random set in $B(n, \alpha)$. Then, for $\alpha$ fixed and $n \to +\infty$, we have

$$|A/A| \sim \frac{6}{\pi^2} \cdot \frac{\alpha^2 \text{Li}_2(1-\alpha^2)}{1-\alpha^2} \cdot n^2,$$

with probability $1 - o(1)$, where $\text{Li}_2(z) := \sum_{k=1}^\infty z^k/k^2$ is the dilogarithm function.

A fundamental step in the proof of Theorem 1.1 is determining a formula for the probability that certain rational numbers belong to $A/A$. Precisely, Cilleruelo and Guijarro-Ordóñez [2, Eq. (2)] showed that for all positive integers $r < s$, with $(r, s) = 1$ and $s > n^{1/2}$, we have

$$\mathbb{P}(r/s \in A/A) = 1 - \left(1 - \alpha^2\right)^{\lfloor n/s \rfloor}.$$

Note that the assumption $r < s$ is not restrictive, since $r/s \in A/A$ if and only if $s/r \in A/A$, while the assumption $s > n^{1/2}$ is indeed a restriction.

Our first result is a general formula for the probability that a rational number belongs to the ratio set $A/A$.

Theorem 1.2. Let $A$ be a random set in $B(n, \alpha)$. Then we have

$$\mathbb{P}(r/s \in A/A) = 1 - \prod_{i=1}^{\left\lfloor\log \frac{n}{\log s}\right\rfloor} \gamma_i^{[n/s^i]},$$

for all positive integers $r < s$ with $(r, s) = 1$, where $\gamma_i := \beta_i - \beta_i/\beta_i^{2}$ with $\beta_0 := 1$, $\beta_1 := 1$, and $\beta_{i+1} := (1-\alpha)\beta_i + \alpha(1-\alpha)\beta_{i-1}$, for all integers $i \geq 1$.

Remark 1.1. If $\alpha = 1/2$ then for all integers $i \geq 0$ we have $\beta_i = F_{i+2}/2^i$, where $\{F_i\}_{i=0}^\infty$ is the sequence of Fibonacci numbers, defined recursively by $F_0 := 0$, $F_1 := 1$, and $F_{i+2} := F_{i+1} + F_i$.

2010 Mathematics Subject Classification. Primary: 11N25, Secondary: 11K99.

Key words and phrases. probability, random set, ratio set.

† C. Sanna is a member of GNSAGA of INdAM and of CryptTO, the group of Cryptography and Number Theory of Politecnico di Torino.

1
As a consequence of Theorem 1.2, we obtain the following corollary:

**Corollary 1.1.** Let $\mathcal{A}$ be a random set in $\mathcal{B}(n, \alpha)$. Then we have

$$\mathbb{P}(r/s \in \mathcal{A}/\mathcal{A}) = 1 - \exp(-\delta(s)n + O_n(1)),$$

for all positive integers $r < s \leq n$ with $(r, s) = 1$, where

$$\delta(s) := \sum_{i=1}^{\infty} \log(1/\gamma_i/s^i)$$

is an absolutely convergent series.

It is natural to ask if Theorem 1.2 can be generalized to a formula for the probability of the event $\bigvee_{i=1}^{k} (r_i/s_i \in \mathcal{A}/\mathcal{A})$, where $r_1/s_1, \ldots, r_k/s_k$ are rational numbers. The answer should be “yes”, but the task seems very complex (see Section 6 for more details).

However, we proved the following result concerning powers of the same rational number.

**Theorem 1.3.** Let $\mathcal{A}$ be a random set in $\mathcal{B}(n, \alpha)$, and let $\mathcal{E}$ be a finite or cofinite set of positive integers with $1 \in \mathcal{E}$. Then we have

$$\mathbb{P}\left( \bigvee_{e \in \mathcal{E}} ((r/s)^e \in \mathcal{A}/\mathcal{A}) \right) = 1 - \prod_{i=1}^{\left\lfloor \log n / \log s \right\rfloor} \left( \gamma_i^{(\mathcal{E})} \right)^{\left\lfloor n/s^i \right\rfloor},$$

for all positive integers $r < s$ with $(r, s) = 1$, where $\gamma_i^{(\mathcal{E})} := \beta_i^{(\mathcal{E})}{(\mathcal{E})} / (\beta_i^{(\mathcal{E})})^2$, for all integers $i \geq 1$, and $\{\beta_i^{(\mathcal{E})}\}_{i=0}^{\infty}$ is a linear recurrence depending only on $\mathcal{E}$ and $\alpha$. In particular, if $\mathcal{E}$ is cofinite then $\gamma_i^{(\mathcal{E})}$ is a rational function of $i$, for all sufficiently large $i$.

As a matter of example, we provide the following:

**Example 1.1.** $\beta_0^{(1,2)} = 1$, $\beta_1^{(1,2)} = 1$, $\beta_2^{(1,2)} = 1 - \alpha^2$, and

$$\beta_i^{(1,2)} = (1 - \alpha)\beta_{i-1}^{(1,2)} + \alpha(1 - \alpha)^2 \beta_{i-3}^{(1,2)}$$

for all integers $i \geq 3$.

**Example 1.2.** $\beta_0^{(1,3)} = 1$, $\beta_1^{(1,3)} = 1$, $\beta_2^{(1,3)} = 1 - \alpha^2$, $\beta_3^{(1,3)} = 1 - 2\alpha^2 + \alpha^3$, and

$$\beta_i^{(1,3)} = (1 - \alpha)\beta_{i-1}^{(1,3)} + \alpha(1 - \alpha)^2 \beta_{i-2}^{(1,3)} - \alpha(1 - \alpha)^3 \beta_{i-3}^{(1,3)} + \alpha(1 - \alpha)^3 \beta_{i-4}^{(1,3)}$$

for all integers $i \geq 4$.

**Example 1.3.** $\beta_i^{(N)} = (1 - \alpha)^{i-1}((1 - \alpha) + i\alpha)$ and

$$\gamma_i^{(N)} = 1 - \frac{1}{(i + \alpha^{-1} - 1)^2}$$

for all integers $i \geq 1$.

**Example 1.4.** $\beta_i^{(N\setminus2)} = (1 - \alpha)^{i-2}((1 - \alpha)^2 + i\alpha(1 - \alpha) + (i - 2)\alpha^2)$ and

$$\gamma_i^{(N\setminus2)} = 1 - \frac{1}{(i + \alpha^{-1} - \alpha - 2)^2}$$

for all integers $i \geq 3$.

2. **Notation**

We employ the Landau–Bachmann “Big O” notation $O$ with its usual meaning. Any dependence of implied constants is explicitly stated or indicated with subscripts. Greek letters are reserved for quantities that depends on $\alpha$. 

3. Proof of Theorem 1.2

Let us consider the directed graph $G(n;r,s)$ having vertices $1, \ldots, n$ and edges $rt \to st$, for all positive integers $t \leq n/s$. For an example, see Figure 1. Note that two edges $rt \to st$ and $rt' \to st'$, with $t < t'$, have a common vertex if and only if $st = rt'$. In such a case, recalling that $r$ and $s$ are relatively prime, it follows that $t = ru$ and $t' = su$, for some positive integer $u$, and consequently the two edges form a directed path $r^j u \to rsu \to s^j u$. Indeed, iterating this reasoning, it follows that all paths of $i + 1$ vertices are of the type

$$r^i u \to r^{i-1} su \to \cdots \to rs^{i-1} u \to s^i u,$$

for some positive integer $u \leq n/s^i$. Moreover, it is easy to check that each vertex of $G(n;r,s)$ is incident to at most two edges. Therefore, all the connected components of $G(n;r,s)$ are directed path graphs of at most $k := \lfloor \log n/\log s \rfloor + 1$ vertices.

![Figure 1. The directed graph $G(30;2,3)$.](image)

Let $c_i$, respectively $d_i$, be the number of connected components, respectively directed paths, of $G(n;r,s)$ consisting of exactly $i$ vertices (considering each isolated vertex as a directed path with one vertex). On the one hand, from the previous reasonings, we have that $d_i = \lfloor n/s^{i-1} \rfloor$ for each positive integer $i$. On the other hand, since each connected component of $i$ vertices contains exactly $i - j + 1$ directed paths of $j$ vertices, for all positive integers $j \leq i$, we have

$$\sum_{i=j}^{k} (i - j + 1) c_i = d_j, \quad j = 1, \ldots, k. \quad (5)$$

The linear system (5) in unknowns $c_1, \ldots, c_k$ can be solved by subtracting to each equation the next one, and then again subtracting to each equation the next one. This yields

$$c_i = d_i - 2d_{i+1} + d_{i+2}, \quad i = 1, \ldots, k. \quad (6)$$

(Note that $d_i = 0$ for every integer $i > k$.)

Now we have that $r/s \in A/A$ if and only if there exists a positive integer $t \leq n/s$ such that $rt \in \mathcal{A}$ and $st \in \mathcal{A}$. Therefore,

$$\mathbb{P}(r/s \notin A/A) = \mathbb{P}\left( \bigcap_{t \leq n/s} E(rt, st) \right)$$

where $E(a,b)$ denotes the event $-(a \in \mathcal{A} \land b \in \mathcal{A})$. Clearly, there is a natural correspondence between the events $E(rt, st)$ and the edges $rt \to st$ of $G(n;r,s)$. In particular, events corresponding to edges of different connected components of $G(n;r,s)$ are independent. Furthermore, if $m_1 \to \cdots \to m_i$ is a connected component of $G(n;r,s)$, then

$$\mathbb{P}(E(m_1,m_2) \land \cdots \land E(m_{i-1},m_i))$$

is equal to the probability that the string $m_1, \ldots, m_i$ has no consecutive elements in $\mathcal{A}$. In turn, this is easily seen to be equal to $\beta_i$. Indeed, for $i = 0,1$ the claim is obvious, since $\beta_0 = 1$ and $\beta_1 = 1$; while for $i \geq 2$ we have that $m_1, \ldots, m_i$ contains no consecutive elements in $\mathcal{A}$ if and only if either $m_i \notin \mathcal{A}$ and $m_1, \ldots, m_{i-1}$ has no consecutive elements in $\mathcal{A}$, or $m_i \in \mathcal{A}$, $m_{i-1} \notin \mathcal{A}$, and $m_1, \ldots, m_{i-2}$ has no consecutive elements in $\mathcal{A}$, so that the claim follows from the recursion $\beta_i = (1-\alpha)\beta_{i-1} + \alpha(1-\alpha)\beta_{i-2}$. 

Therefore, also using (6), we have

\[ P(r/s \notin A/A) = \prod_{i=1}^{k} \prod_{m_1 \to \cdots \to m_i} P(E(m_1, m_2) \land \cdots \land E(m_{i-1}, m_i)) \]

\[ = \prod_{i=1}^{k} \beta_i^c = \prod_{i=1}^{k} \rho_i^d = \prod_{i=1}^{k-1} \left( \frac{\beta_{i-1} \beta_{i+1}}{\beta_i^2} \right)^{d_{i+1}} = \prod_{i=1}^{[\log n/\log 2]} \gamma_i^{[n/s^i]}, \]

and (1) follows. The proof is complete.

4. Proof of Corollary 1.1

Throughout this section, implied constants may depend on \( \alpha \). Let \( \rho_1, \rho_2 \) be the roots of the characteristic polynomial \( X^2 - (1 - \alpha)X - \alpha(1 - \alpha) \) of the linear recurrence \( \{\beta_i\}_{i=0}^{\infty} \). Recalling that \( \alpha \in (0, 1) \), an easy computation shows that \( |\rho_1| \neq |\rho_2| \). Without loss of generality, assume \( |\rho_1| > |\rho_2| \) and put \( \varrho := |\rho_2/\rho_1| \), so that \( \varrho \in (0, 1) \). Hence, there exist complex numbers \( \zeta_1, \zeta_2 \) such that

\[ \beta_i = \zeta_1 \rho_1^i + \zeta_2 \rho_2^i = \zeta_1 \rho_1^i \left( 1 + O(\varrho^i) \right), \]

for every integer \( i \geq 0 \). Consequently, we have

\[ \gamma_i = \frac{\beta_{i-1} \beta_{i+1}}{\beta_i^2} = \frac{\zeta_1 \rho_1^{i-1} \left( 1 + O(\varrho^{i-1}) \right) \zeta_1 \rho_1^{i+1} \left( 1 + O(\varrho^{i+1}) \right)}{\left( \zeta_1 \rho_1^i \left( 1 + O(\varrho^i) \right) \right)^2} = 1 + O(\varrho^i), \]

and \( \log \gamma_i = O(\varrho^i) \), for every sufficiently large integer \( i \). In particular, it follows that (3) is an absolutely convergent series.

Now put \( \ell := [\log n/\log s] \). From Theorem 1.2, we get that

\[ P(r/s \in A/A) = 1 - e^L, \]

where

\[ L := \sum_{i=1}^{\ell} \frac{[n/s^i]}{s^i} \log \gamma_i = \sum_{i=1}^{\infty} \frac{\log \gamma_i}{s^i} n + O \left( \sum_{i \geq \ell} \frac{\log \gamma_i}{s^i} n \right) + O \left( \sum_{i=1}^{\ell} \frac{\log \gamma_i}{s^i} n \right) \]

\[ = -\delta(s)n + O\left( \frac{n}{s^{\ell+1}} \right) + O \left( \sum_{i=1}^{\ell} \varrho^i \right) = -\delta(s)n + O(1), \]

as desired. The proof is complete.

Remark 4.1. A more detailed analysis shows that \( \left\{ \gamma_i^{(-1)^i} \right\}_{i=1}^{\infty} \) is a strictly decreasing sequence tending to 1. In particular, (3) is an alternating series.

5. Proof of Theorem 1.3

Let us define the directed graph \( G^{(E)}(n; r; s) := \bigcup_{e \in E} G(n; r^e, s^e) \). For an example, see Figure 2.

![Figure 2](image)

Figure 2. The directed graph \( G^{(1,2)}(30; 2, 3) \).

Since \( 1 \in E \), it is easy to check that \( G^{(E)}(n; r; s) \) is the graph obtained from \( G(n; r, s) \) by adding a directed edge \( v_1 \to v_2 \) between each pair \( v_1 < v_2 \) of vertices of \( G(n; r, s) \) that
have distance \( e \), for every \( e \in \mathcal{E} \). In particular, this process connects only vertices that are already connected. Hence, the number of connected components of \( G^\mathcal{E}(n; r, s) \) that have exactly \( i \) vertices is equal to the number of connected components of \( G(n; r, s) \) that have exactly \( i \) vertices, which is the number \( c_i \) that we already determined in the proof of Theorem 1.2. Moreover, the probability that a connected component of \( G^\mathcal{E}(n; r, s) \) having exactly \( i \) vertices has no adjacent vertices both belonging to \( A \) is equal to the probability \( \beta_i^\mathcal{E} \) that the random binary string \( \chi_1 \cdots \chi_i \) does not contain the substring \( 10^{r-1}1 \), for all \( e \in \mathcal{E} \), where \( \{\chi_k\}_{k=1}^\infty \) is a sequence of independent identically distributed random variables in \( \{0, 1\} \) with \( \mathbb{P}(\chi_k = 1) = \alpha \). At this point the same reasonings of (7) yield (4). Let us prove that \( \{\beta_i^\mathcal{E}\}_{i=0}^\infty \) is a linear recurrence.

Suppose that \( \mathcal{E} \) is finite and let \( m := \max(\mathcal{E}) + 1 \). Then \( \beta_i^\mathcal{E} \) can be determined by considering a Markov chain. The states are the binary strings \( x_1 \cdots x_m \in \{0, 1\}^m \) not containing the substring \( 10^{e-1}1 \), for every \( e \in \mathcal{E} \), and one absorbing state. A transition from state \( x_1 \cdots x_m \) to state \( x_2 \cdots x_{m-1}1 \), respectively from state \( x_1 \cdots x_m \) to state \( x_2 \cdots x_{m-1}01 \), happens with probability \( \alpha \), respectively \( 1 - \alpha \), and all the other transitions are to the absorbing state. Finally, the probability of \( x_1 \cdots x_m \) being the initial state is \( \alpha^{x_1+\cdots+x_m}(1 - \alpha)^{m-x_1-\cdots-x_m} \).

Therefore, letting \( u \) be the number of states, we have that \( \beta_i^\mathcal{E} = \pi \Sigma^i(1, 1, \ldots, 1, 0)^t \) for all integers \( i \geq 0 \), where \( \pi \) is a row vector of length \( u \), \( \Sigma \) is a \( u \times u \) stochastic matrix, and \( (1, 1, \ldots, 1, 0)^t \) is column vector of length \( t \), assuming the states are ordered so that the absorbing state is the last one. Consequently, \( \{\beta_i^\mathcal{E}\}_{i=0}^\infty \) is a linear recurrence whose characteristic polynomial is given by the characteristic polynomial of \( \Sigma \).

Now suppose that \( \mathcal{E} \) is cofinite and let \( \ell \) be the minimal positive integer such that \( e \in \mathcal{E} \) for all integers \( e \geq \ell \). If \( \chi_1 \cdots \chi_i \) does not contain \( 10^{\ell-1}1 \), for every \( e \in \mathcal{E} \), then the distance between each pair of 1s in \( \chi_1 \cdots \chi_i \) is less than \( \ell \) positions. In particular, the number of 1s in \( \chi_1 \cdots \chi_i \) is at most \( \ell + 1 \). Therefore, for \( i \geq \ell + 1 \), by elementary probability calculus we can write \( \beta_i^\mathcal{E} \) as a linear combination, whose coefficients do not depend on \( i \), of the power sums \( (i-k+1)(1-\alpha)^{i-k} \) where \( k = 0, \ldots, \ell + 1 \). Consequently, \( \beta_i^\mathcal{E} = (1-\alpha)^{i-\ell-1}B^\mathcal{E}(i) \) for some \( B^\mathcal{E}(X) \in \mathbb{R}[X] \). Hence, \( \{\beta_i^\mathcal{E}\}_{i=0}^\infty \) is a linear recurrence and \( \gamma_i^\mathcal{E} = B(i-1)B(i+1)/B(i)^2 \) is a rational function of \( i \).

The proof is complete.

6. General case

As mentioned in the introduction, providing a general formula for the probability of the event \( \bigvee_{i=1}^k (r_i/s_i \in \mathcal{A}/\mathcal{A}) \), where \( r_1/s_1, \ldots, r_k/s_k \) are rational numbers, seems very complex. In light of the previous reasonings, this task amounts to study the graph \( G := \bigcup_{i=1}^k G(n; r_i, s_i) \). Precisely, one has to classify the connected components of \( G \), and to determine the probability that each of them does not have two adjacent vertices both belonging to \( A \).

![Figure 3](image-url)
Figure 4. The connected components of \( G(30; 2, 3) \cup G(30; 3, 4) \) that have at least 2 vertices. Each horizontal, respectively vertical, edge corresponds to multiply the value of a vertex by 3/2, respectively 4/3.

Figure 5. The directed graph \( G(30; 2, 3) \cup G(30; 4, 5) \).

Figure 6. The connected components of \( G(30; 2, 3) \cup G(30; 4, 5) \) that have at least 2 vertices. Each horizontal, respectively vertical, edge corresponds to multiply the value of a vertex by 3/2, respectively 5/4.

If the multiplicative group generated by \( \{r_i/s_i\}_{i=1}^k \) is cyclic, then the connected components of \( G \) have a somehow “linear” structure, and proving formulas similar to (1) and (4) is doable.

If the generated group has rank \( R > 1 \), then each connected component of \( G \) is isomorphic to a subgraph of the \( R \)-dimensional grid graph. For examples, see Figures 3, 4, 5, and 6.

7. Visible lattice points

Another direction of research can be generalizing ratio sets to sets of visible lattice points. Let \( d \geq 2 \) be an integer. For every \( A \subseteq \mathbb{N} \), a lattice point \( P \in \mathbb{N}^d \) is said to be visible in the lattice \( A^d \) if the line segment from \( 0 \in \mathbb{Z}^d \) to \( P \) intersects \( A^d \) only in \( P \). Let \( \text{vis}(A^d) \) be the set of lattice points visible in \( A^d \). There is a natural bijection between \( \text{vis}(A^2) \) and \( A/A \), given by \( (x_1, x_2) \mapsto x_1/x_2 \). Hence, \( \text{vis}(A^d) \) can be considered as a \( d \)-dimensional generalization of the ratio set \( A/A \) (see also [5] for a similar generalization of ratio sets).

Cilleruelo and Guijarro-Ordóñez [2] gave an asymptotic formula for the cardinality of \( \text{vis}(A^d) \) for \( A \in \mathcal{B}(n, \alpha) \). A natural question is if Theorem 1.2 can be generalized to a formula for \( \mathbb{P}(x_1, \ldots, x_d) \in \text{vis}(A^d) \), where \( (x_1, \ldots, x_d) \in \mathbb{N}^d \). This amount to study the hypergraph \( \mathcal{H}(n; x_1, \ldots, x_d) \) defined as having vertices 1, \ldots, \( n \) and hyperedges \( (x_1t, \ldots, x_dt) \), for every positive integer \( t \leq n/\max(x_1, \ldots, x_d) \). For an example, see Figure 7.

Figure 7. The hypergraph \( \mathcal{H}(28; 2, 3, 4) \).
8. Acknowledgements

The author thanks Paolo Leonetti and Daniele Mastrostefano for suggestions that improved the quality of the article.

References

1. G. Alsmeyer, Z. Kabluchko, and A. Marynych, Limit theorems for the least common multiple of a random set of integers, Trans. Amer. Math. Soc. 372 (2019), no. 7, 4585–4603.
2. J. Cilleruelo and J. Guijarro-Ordóñez, Ratio sets of random sets, Ramanujan J. 43 (2017), no. 2, 327–345.
3. J. Cilleruelo, D. S. Ramana, and O. Ramaré, Quotient and product sets of thin subsets of the positive integers, Proc. Steklov Inst. Math. 296 (2017), 52–64.
4. J. Cilleruelo, J. Rué, P. Šarka, and A. Zumalacárregui, The least common multiple of random sets of positive integers, J. Number Theory 144 (2014), 92–104.
5. P. Leonetti and C. Sanna, Directions sets: a generalisation of ratio sets, Bull. Aust. Math. Soc. 101 (2020), no. 3, 389–395.
6. D. Mastrostefano, On maximal product sets of random sets, J. Number Theory 224 (2021), 13–40.
7. C. Sanna, A note on product sets of random sets, Acta Math. Hungar. 162 (2020), no. 1, 76–83.
8. C. Sanna, On the l.c.m. of random terms of binary recurrence sequences, J. Number Theory 213 (2020), 221–231.

Politecnico di Torino, Department of Mathematical Sciences
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
Email address: carlo.sanna.dev@gmail.com