SEQUENCES CONSTRUCTED BY A MODIFIED INCLUSION-EXCLUSION PRINCIPLE

TATSUO KONNO

Abstract. Defining an average behavior of primes, we construct a family of sequences using a modified inclusion-exclusion principle and investigate whether these sequences have the same asymptotic property as the primes.

1. Introduction

By the principle of inclusion and exclusion [4], the probability that a natural number is not divisible by either 2 or 3 is

$$\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)=\frac{1}{3}.$$

It follows that, on average, every third number is not divisible by either 2 or 3. It is natural that the prime 5 exists between 3 and 6 $(=3+3)$. In the same way, the probability that a natural number is not divisible by any of 2, 3, or 5 is

$$\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\left(1-\frac{1}{5}\right)=\frac{4}{15}.$$

It also follows that, on average, one out of every 3.75 $(=15/4)$ natural numbers is not divisible by 2, 3, or 5. The prime 7 exists between 5 and 8.75 $(=5+3.75)$, as expected.

Generally, the probability that a natural number is not divisible by $2,3,5,\ldots,$ or p_n (p_n: the nth prime number) is

$$\prod_{k=1}^{n}\left(1-\frac{1}{p_k}\right).$$

We consider θ_n $(n=1,2,\ldots)$ satisfying

$$p_{n+1}=p_n+\theta_n\left\{\prod_{k=1}^{n}\left(1-\frac{1}{p_k}\right)\right\}^{-1}.$$

For instance,

$$5=3+\theta_2\left\{\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\right\}^{-1},\qquad \theta_2\approx 0.67,$$

$$7=5+\theta_3\left\{\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\left(1-\frac{1}{5}\right)\right\}^{-1},\qquad \theta_3\approx 0.53.$$

n	1	2	3	4	5	6	7	8	9	10	11	$\cdots$
θ_n	0.50	0.67	0.53	0.91	0.42	0.77	0.36	0.68	0.98	0.32	0.92	$\cdots$

The above recurrence relations are considered to be an analytic expression of the sieve of Eratosthenes. Motivated by this expression, we construct a family of sequences $\{a_n\}$ by the recurrence relations

$$a_{n+1} = a_n + \theta \left\{ \prod_{k=1}^{n} \left(1 - \frac{1}{a_k}\right) \right\}^{-1} \tag{1.1}$$

for arbitrarily fixed $a_1 > 1$ and $\theta > 0$, where the average behavior of primes is considered.

Example 1.1. *The arithmetic mean of θ_n for $1 \leq n \leq 11$ is approximately 0.64. When $a_1 = 2$ and $\theta = 0.64$, $\{a_n\}$ is compared with $\{p_n\}$ $(1 \leq n \leq 12)$ in the table below. We can observe that $\{p_n\}$ is interwoven with $\{a_n\}$.*

n	1	2	3	4	5	6	7	8	9	10	11	12
p_n	2	3	5	7	11	13	17	19	23	29	31	37
a_n	2.0	3.3	5.1	7.4	10.1	13.0	16.2	19.6	23.1	26.9	30.8	34.8

Concerning the distribution of primes, the asymptotic formula

$$\lim_{n\to\infty} \frac{n \log_e p_n}{p_n} = 1$$

is well-known as the prime number theorem [1]. The family of sequences $\{a_n\}$ also has the same asymptotic property as the primes.

Theorem 1.2. *Let $a_1 > 1$ and $\theta > 0$ be constants. The sequence $\{a_n\}$ defined by recurrence formula (1.1) satisfies*

$$\lim_{n\to\infty} \frac{n \log_e a_n}{a_n} = 1. \tag{1.2}$$

Sequences are commonly investigated using sieve processes (e.g., [2, 3, 5]), and in the present article we have defined a type of sieve from the viewpoint of an average behavior of θ_n. This suggests the mechanism of the prime number theorem. More precisely, this is written as the following statements.

Conjecture 1.3.

(i) *There exist positive constants a, b such that $a < \sum_{k=1}^{n} \theta_k/n$ and $\theta_n < b$ hold for all n.*

(ii) *Let a', b' $(a' < b')$ be any positive constants. For any sequence $\{\theta_{n'}\}$ which satisfies $a' < \sum_{k=1}^{n} \theta_{k'}/n$ and $0 < \theta_{n'} < b'$ $(n = 1, 2, \ldots)$, formula (1.2) holds by replacing θ by $\theta_{n'}$ in recurrence (1.1).*

To prove the propositions (i) and (ii) is equivalent to giving a new proof of the prime number theorem.

2. Proof of Theorem 1.2

Let

$$b_n = \prod_{k=1}^{n} \left(1 - \frac{1}{a_k}\right). \tag{2.1}$$

By (1.1) and (2.1), the ratio b_n/b_{n+1} is written as

$$\frac{a_{n+2} - a_{n+1}}{a_{n+1} - a_n} = \frac{a_{n+1}}{a_{n+1} - 1}, \tag{2.2}$$

and we get the recurrence formula

$$a_{n+2} = a_{n+1}\left(2 - \frac{a_n - 1}{a_{n+1} - 1}\right) .$$

For the sequence $\{a_n\}$, we have the following facts.

$$a_{n+1} > a_n + \theta \quad (n = 1,\ 2, \ldots), \tag{2.3}$$

$$\lim_{n\to\infty} a_n = \infty, \tag{2.4}$$

$$a_{n+2} - a_{n+1} > a_{n+1} - a_n \ (n = 1,\ 2, \ldots), \tag{2.5}$$

$$\lim_{n\to\infty} \frac{a_{n+2} - a_{n+1}}{a_{n+1} - a_n} = 1. \tag{2.6}$$

Since $b_n < 1$, we know (2.3) from (1.1). Equation (2.4) is obvious from (2.3) and (2.5) is given by

$$a_{n+2} - a_{n+1} = \frac{\theta}{b_{n+1}} > \frac{\theta}{b_n} = a_{n+1} - a_n.$$

Equation (2.6) is an immediate consequence of (2.2) and (2.4). Moreover, we have, by (2.2), that

$$\frac{a_{n+2} - 2a_{n+1} + a_n}{a_{n+1} - a_n} = \frac{1}{a_{n+1} - 1} . \tag{2.7}$$

By (2.2) and (2.7), we have

$$\frac{a_{n+3} - 2a_{n+2} + a_{n+1}}{a_{n+2} - 2a_{n+1} + a_n} = \frac{a_{n+1}}{a_{n+2} - 1} . \tag{2.8}$$

If $a_{n+2} - a_{n+1} > 1$, we have, by (2.8), that

$$a_{n+3} - 2a_{n+2} + a_{n+1} < a_{n+2} - 2a_{n+1} + a_n. \tag{2.9}$$

Lemma 2.1. *For the sequence* $\{a_n\}$, *we have*
(i) $\lim_{n\to\infty}(a_{n+1} - a_n) = \infty$,
(ii) $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = 1$.

Proof. If (i) does not hold, the differences $\{a_{n+1} - a_n\}$ are bounded above. Therefore, the monotonic property of (2.5) means that the differences converge to some value a (> 0). In this case, there exists a constant b such that the line $y = ax + b$ is the asymptote of the point set $\{(n, a_n)\}$. Then, $a_k \le ak - a + a_1$ holds for any natural number k, and

$$\frac{\theta}{a} = \lim_{n\to\infty} b_n \le \prod_{k=1}^{\infty}\left(1 - \frac{1}{ak - a + a_1}\right).$$

Taking the logarithms of both sides, we have

$$\log_e \frac{\theta}{a} \le \sum_{k=1}^{\infty} \log_e\left(1 - \frac{1}{ak - a + a_1}\right).$$

Evaluating the right-hand side of the inequality by an integral, we obtain

$$\log_e \frac{\theta}{a} < \frac{1}{a}\int_{a_1}^{\infty} \log_e\left(1 - \frac{1}{x}\right) dx = -\infty.$$

However, this inequality contradicts the fact that a is finite. Therefore, the differences $\{a_{n+1} - a_n\}$ are not bounded. The first part (i) of Lemma 2.1 is proved. Next, we shall prove the second part (ii) of Lemma 2.1. Setting $c_n = a_{n+1}/a_n$ $(n = 1,\ 2, \ldots)$, we get

$$c_n > 1 + \frac{\theta}{a_n} > 1$$

from (2.3), and we know that $\{c_n\}$ is bounded below. From (2.2) and (i), we have

$$\frac{c_{n+1} - 1}{c_n - 1} = \frac{a_n}{a_{n+1} - 1} < 1$$

for any sufficiently large n. So, we also know that $\{c_n\}$ is a decreasing sequence. Let γ denote the limit of $\{c_n\}$. From (2.6), γ satisfies the equation

$$\gamma^2 - 2\gamma + 1 = 0,$$

and $\gamma = 1$. The second part of Lemma 2.1 is proved. □

Applying differential calculus we state the next lemma.

Lemma 2.2. *For the sequence $\{a_n\}$, there exists a function $f(x)$ $(x \geq 1)$ satisfying $f(n) = a_n$ $(n = 1,\ 2, \ldots)$,*

$$\lim_{n\to\infty} \frac{f'(n)}{a_{n+1} - a_n} = 1 \tag{2.10}$$

and

$$\lim_{n\to\infty} \frac{f''_+(n)}{a_{n+2} - 2a_{n+1} + a_n} = 1, \tag{2.11}$$

where $f''_+(n)$ denotes $\lim\limits_{x\to n^+} f''(x)$.

Proof. We construct a spline by patching quadratic functions. For arbitrary $x_1 > 0$, there exist sequences $\{k_n\},\ \{x_n\}$ $(x_n > 0)$ such that

$$k_n(x_n + 1)^2 - k_n {x_n}^2 = a_{n+1} - a_n \tag{2.12}$$

and

$$k_n(x_n + 1) = k_{n+1}x_{n+1}. \tag{2.13}$$

We define a spline $f(x)$ by

$$a_n + k_n(x - n)(x - n + 2x_n)$$
$$(n \leq x < n + 1,\ n = 1,\ 2, \ldots)$$

Then, $f(n) = a_n$ is obvious. From (2.12), we have

$$\lim_{n\to\infty} \frac{f'(n)}{a_{n+1} - a_n} = \lim_{n\to\infty} \frac{2x_n}{2x_n + 1} = 1,$$

and from (2.12), (2.13), we have

$$k_{n+1} + k_n = a_{n+2} - 2a_{n+1} + a_n, \tag{2.14}$$

therefore,

$$\lim_{n\to\infty} \frac{f''_+(n)}{a_{n+2} - 2a_{n+1} + a_n} = \lim_{n\to\infty} \frac{2k_n}{k_{n+1} + k_n} = 1,$$

where we applied

$$\lim_{n\to\infty} x_n = \infty \tag{2.15}$$

and

$$\lim_{n\to\infty} \frac{k_{n+1}}{k_n} = 1. \tag{2.16}$$

The reason that formulas (2.15) and (2.16) hold is the following. From (2.3) and (2.12), we know that k_n are positive. From (2.9), (i) of Lemma 2.1, and (2.14), we know $k_{n+2} < k_n$ for sufficiently large numbers n, and so the sequence $\{k_n\}$ is bounded. Hence, from (2.12) and (i) of Lemma 2.1 we have formula (2.15). We obtain formula (2.16) from (2.13) and (2.15). □

Now, using $f(x)$ defined in Lemma 2.2, we shall prove formula (1.2).

$$\lim_{n\to\infty} \frac{n \log_e a_n}{a_n} = \lim_{x\to\infty} \frac{x \log_e f(x)}{f(x)} .$$

By l'Hôspital's rule, this limit is reduced to

$$\lim_{n\to\infty} \left(\frac{f'(n)}{f(n) f''_+(n)} + \frac{1}{f'(n)} \right). \tag{2.17}$$

The first term of (2.17) converges to 1 by (2.10), (2.11) and

$$\lim_{n\to\infty} \frac{a_{n+1} - a_n}{a_n \left(a_{n+2} - 2a_{n+1} + a_n\right)} = 1. \tag{2.18}$$

Formula (2.18) holds from (2.4), (2.7) and (ii) of Lemma 2.1. The second term of (2.17) converges to 0 by (2.10) and (i) of Lemma 2.1. Hence, we obtain formula (1.2). Theorem 1.2 is proved.

3. Acknowledgment

The author would like to express his thanks to the referee and the editor for their valuable comments and suggestions.

References

[1] E. Borel, *Les Nombres Premiers*, coll. Q.S.J. n°571, Presses Universitaires de France, 1958.
[2] W. E. Briggs, *Prime-like sequences generated by a sieve process*, Duke Math. J., **30** (1963), 297–311.
[3] D. Hawkins and W. E. Briggs, *The lucky number theorem*, Math. Mag., **31** (1957), 81–84.
[4] C. L. Liu, *Introduction to Combinatorial Mathematics*, McGraw-Hill, 1968.
[5] Hugh L. Montgomery, *Prime-like sequences*, Acta Arith., **49** (1988), 277–280.
[6] P. Ribenboim, *The Book of Prime Number Records*, Springer-Verlag, 1988.

MSC2010: 11B37, 11N64

College of Engineering, Nihon University, Koriyama, Fukushima 963-8642, Japan
E-mail address: konno@ge.ce.nihon-u.ac.jp