# A Sum and Product Game

Yuanyuan Shen

**Abstract**

A sum-and-product game involves two numbers $2 \leq p, q \leq n$ for fixed $n$ as well as two participants, a sum person who knows $p+q$ and a product person who knows $pq$. Starting from the sum person, the two participants alternatively answer the dichotomous question of whether they know $p$ and $q$. By identifying a game with a graph, this paper examines various properties of the sum-and-product game, eventually proving that a certain independence conjecture implies the conjecture that there are infinitely many $n$ where an observer can determine $p, q$ after hearing exactly 4 NO before a YES.

## 1 Introduction

In a sum-and-product game of $n$, two numbers, not necessarily distinct, are chosen from the range of positive integers greater than 1 and not greater than $n$. The sum of the two numbers is given to a sum person, and the product of the two numbers is given to a product person. Starting from the sum person, the two participants alternatively answer the question of whether they know the two numbers.

The game can be identified with a bipartite graph $G(n)$ whose vertices consist of all possible sums and products, and where each edge, representing a possible pair of numbers, connects their sum with their product. We deduce the necessary and sufficient conditions on the structure of the graph centered around the sum node $p+q$ for a game with the pair of numbers $(p, q)$ to involve a certain number of 'NO' before the first 'YES.'

We prove additional properties of $G(n)$, and hence of the corresponding game. First, there is no path of length greater than one starting from a sum node $k$ if $1 + 2n - \sqrt{1+4n} < k \leq 2n$. Secondly, if a pair of numbers involves $l$ 'NO' before the first 'YES' in a game of $n$, then for every $l' < l$, there is a pair of numbers involving $l'$ 'NO' before the first 'YES' in a game of $n$. Thirdly, the pair of $(4, 4)$ involves 4 'NO' before the first 'YES' in a game of $n$ if and only if $n \geq 8$.

An observer is able to determine the pair of numbers $(p, q)$ after hearing 4 'No' before the first 'YES' if and only if there is exactly one pair of numbers involving 4 'No' before the first 'YES'. Equivalently, when $n \geq 8$, there is no other pair than $(4, 4)$ involving 4 'No' before the first 'YES'. We prove there are infinitely many $n$ such that an observer cannot determine the pair of numbers $(p, q)$ after hearing 4 'No' before the first 'YES' using an explicit construction. Under certain assumptions of independence, we also prove there are infinitely many $n$ such that an observer can determine the pair of numbers $(p, q)$ after hearing 4 'No' before the first 'YES'.

## 2 Properties of the sum-and-product game

A sum-and-product game of $n$ can be identified with a graph $G(n)$. In the example of $G(12)$ in Figure 1, each edge represents a possible pair $(p, q)$, $2 \leq p, q \leq n$, and connects a square node of their product with a circle node of their sum. Starting with definitions relating to the graph, we prove various properties of the graph and of its related game.

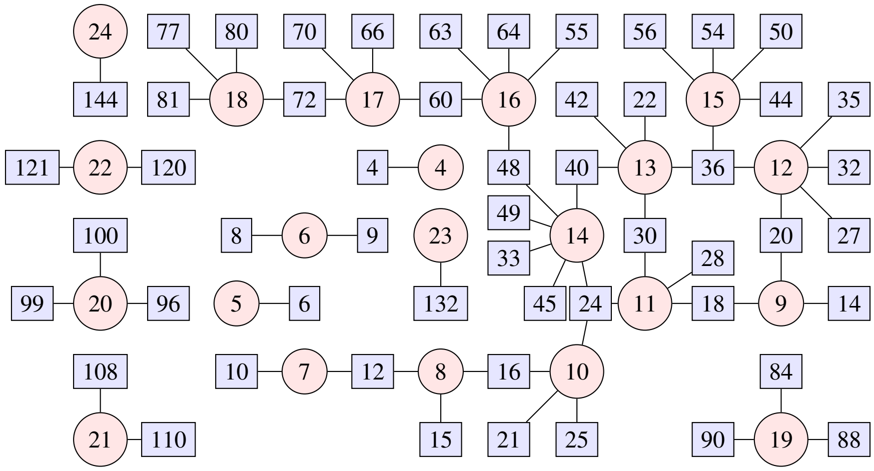

Figure 1: The graph $G(12)$

**Definition 1.** *A sum path $P$ of length $l = l(P)$ in $G(n)$ is two length $l$ sequences $a_i, b_i$ such that $(\forall i)$ $a_i \neq a_{i+1}$, $a_i \geq b_i$ , and $(\forall 1 \leq i \leq \frac{l}{2})$ $a_{2i}b_{2i} = a_{2i-1}b_{2i-1}$, $(\forall 1 \leq i \leq \frac{l-1}{2})$ $a_{2i} + b_{2i} = a_{2i+1} + b_{2i+1}$.*

**Definition 2.** *A product path $P$ of length $l = l(P)$ in $G(n)$ is two length $l$ sequences $a_i, b_i$ such that $(\forall i)$ $a_i \neq a_{i+1}$, $a_i \geq b_i$, and $(\forall 1 \leq i \leq \frac{l}{2})$ $a_{2i} + b_{2i} = a_{2i-1} + b_{2i-1}$, $(\forall 1 \leq i \leq \frac{l-1}{2})$ $a_{2i}b_{2i} = a_{2i+1}b_{2i+1}$.*

**Definition 3.** *A path $P$ of length $l$ in $G(n)$ is either a sum path of length $l$ or a product path of length $l$.*

**Definition 4.** *A cycle $C$ of length $l$ is a path of length $l(C)$ satisfying $a_1b_1 = a_lb_l$ or $a_1 + b_1 = a_l + b_l$.*

**Definition 5.** *A sum tail of length $l$ is a sum path $T = (a_i, b_i)$ of length $l$ such that for every other sum path $\bar{T} = (\bar{a_i}, \bar{b_i})$ of length $\bar{l}$, where $(\bar{a_1}, \bar{b_1}) = (a_1, b_1)$, $\bar{l} \leq l$.*

**Definition 6.** *A product tail of length $l$ is a product path $T = (a_i, b_i)$ of length $l$ such that for every other product path $\bar{T} = (\bar{a_i}, \bar{b_i})$ of length $\bar{l}$, where $(\bar{a_1}, \bar{b_1}) = (a_1, b_1)$, $\bar{l} \leq l$.*

**Lemma 1.** *The length of a sum tail is odd. The length of a product tail is one or even.*

*Proof.* Suppose to the contrary that the length of a sum tail $T = (a_i, b_i)$ is $l = 2m$, then $a_{l-1}b_{l-1} = a_lb_l$. If $a_l + b_l = 4$ or $a_l + b_l = 2n$, then the pair has length one. Otherwise $\exists$ $(a_{l+1}, b_{l+1}) \neq (a_l, b_l)$ such that $a_l + b_l = a_{l+1} + b_{l+1}$, and $T' = (a_i, b_i)$, $1 \leq i \leq l + 1$ is a sum path of length $l + 1$. Similarly the length of a product tail is one or even. □

**Definition 7.** *A pair of numbers $(p, q)$, $q \leq p \leq n$, has length $l$ in the sum-and-product game of $n$ if $(p, q)$ involves $l$ NO before the first YES in the game of $n$. Let $C_{l,n}$ denote the set of pairs of numbers of length $l$ in the game of $n$.*

**Theorem 2.1.** *A pair of numbers $(p, q)$ has length $l = 2m - 1$ if and only if*

- *There is one sum tail of length $l$ with $(a_1, b_1) = (p, q)$*
- *There is at least one other sum path of length $l' \geq l$ with $a'_1 + b'_1 = p + q$, $a_1 \neq p$*

*A pair of numbers $(p, q)$ has length $l = 2m$ if and only if*

- *There is at least one sum tail of length $l - 1$ with $a_1 + b_1 = p + q$, $a_1 \neq p$*
- *There is one sum path of length $l' > l - 1$ with $(a'_1, b'_1) = (p, q)$*
- *There is no sum path of length $l'' > l - 1$ with $a''_1 + b''_1 = p + q$, $a''_1 \neq p$*

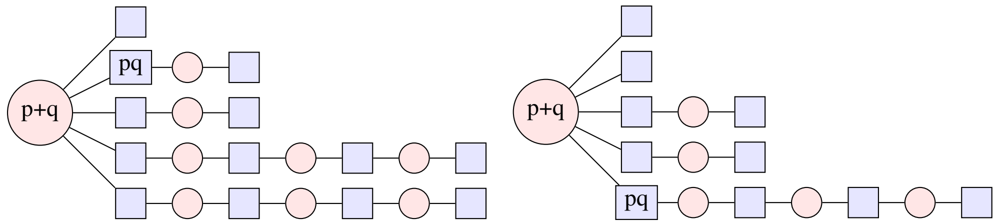


Figure 2: Examples of a pair $(p, q)$ of length three (left) and a pair $(p, q)$ of length four (right)

*Proof.* Let the sum person be named Alice and the product person be named Bob.

When $l = 1$, Alice cannot differentiate $(a_1, b_1)$ from $(p, q)$ and says NO. Bob only has one way to decompose his product and says YES. This results in a pair of length one. In the other direction, suppose there is no sum tail of length 1 with $(a_1, b_1) = (p, q)$, then Bob has more than one way to decompose his product and says the second NO. Otherwise, suppose there is no sum path of length $l' \geq l$ with $a_1 + b_1 = p + q$, $a_1 \neq p$. Then Alice only has one way to decompose her sum, resulting in a pair of length zero.

When $l = 2$, Alice cannot differentiate $(a_1, b_1)$ from $(p, q)$ and says NO. Bob cannot differentiate $(a'_1, b'_1)$ from $(a'_2, b'_2)$ and says NO. Alice knows the numbers must be $(p, q)$, or else Bob only has one way to decompose his product and would have said YES. This results in a pair of length two. In the other direction, suppose there is no sum path of length $l' > 1$ with $(a'_1, b'_1) = (p, q)$, then Bob only has one way to decompose his product, resulting in a pair of length one. Suppose there is another sum path of length $l'' > 1$ with $a''_1 + b''_1 = p + q$, $a''_1 \neq p$, then after two NO Alice cannot differentiate between $(p, q)$ and $a''_1, b''_1$, resulting in a pair of length more than two. Suppose there is no sum tail of length 1 with $a_1 + b_1 = p + q$, $a_1 \neq p$, then Alice only has one way to decompose her sum, resulting in a pair of length zero.

Suppose the statement is true for all $l \leq 2M - 2$. When $l = 2M - 1$, the pair has length at least $2M - 1$ by the induction hypothesis. After Alice says the $(2M - 2)th$ NO, Bob knows the numbers must be $(p, q)$, or else Alice would have said YES by the induction hypothesis. This results in a pair of length $2M - 1$.

In the other direction, suppose there is no sum path of length $l' > l$ with $a_1 + b_1 = p + q$, $a_1 \neq p$, and at most one sum path of length $l$ with $a_1 + b_1 = p + q$, then by the induction hypothesis the pair has length smaller than $2M - 1$. Otherwise, suppose there is no sum tail of length $l$ with $(a_1, b_1) = (p, q)$. If there is no sum path of length $l$ with $(a_1, b_1) = (p, q)$, then by the induction hypothesis the pair has length smaller than $2M - 1$. If there is a sum path of length $l'' > l$ with $(a''_1, b''_1) = (p, q)$, then Bob cannot differentiate $(a''_2, b''_2)$ from $(p, q)$ at the $(2M - 1)th$ step.

When $l = 2M$, the pair has length at least $2M$ by the induction hypothesis. After Bob says the $(2M - 1)th$ NO, Alice knows the numbers must be $(p, q)$, or else Bob would have said YES by the induction hypothesis. This results in a pair of length $2M$.

In the other direction, suppose there is no sum path of length $l' > l - 1$ with $(a'_1, b'_1) = (p, q)$, then by the induction hypothesis the pair has length smaller than $2M - 1$. Suppose there is a sum path of length $l'' > l - 1$ with $a''_1 + b''_1 = p + q$, $a''_1 \neq p$, then Alice cannot differentiate $(a''_1, b''_1)$ from $(p, q)$ at the $2Mth$ step. Suppose there is no sum tail of length $l - 1$ with $a_1 + b_1 = p + q$, $a_1 \neq p$, then by the induction hypothesis the pair has length smaller than $2M - 1$.

□

**Corollary 1.** *A pair of numbers $(p, q)$ has length $l = 2m$ if and only if*

- *There is one product tail of length $l$ with $(a_1, b_1) = (p, q)$*
- *There is at least one other product path of length $l' \geq l - 1$ with $a'_1 b'_1 = pq$, $a'_1 \neq p$*

**Lemma 2.** *A sum path $P = (a_i, b_i)$ of length two, where $a_1 + b_1 = \sum - \delta < \sum = a_2 + b_2$, satisfies $\sum \leq 2a_2 + \delta - 2\sqrt{\delta a_2}$*

*Proof.* Consider

$$(a_1+b_1)^2 \geq 4a_1a_2 = 4a_2b_2$$
$$(2a_2 - \sum + \delta)^2 = 4a_2^2 + (a_1+b_1)^2 - 4a_2(a_1+b_1) \geq 4a_2(a_2+b_2-a_1-b_1) = 4\delta a_2$$
$$2a_2 + \delta - 2\sqrt{\delta a_2} \geq \sum$$

□

**Lemma 3.** *Given $b$, there is no sum path $P = (a_i, b_i)$ of length greater than one with $(a_1, b_1) = (b+k, b-k)$ in $G(n)$, $n < b + \sqrt{b}$.*

*Proof.* Suppose $(b+k)(b-k) = (b+k_1)(b+k_2)$, and without loss of generality $k_1 \geq k_2$. Then $k_1 \neq 0$, or else $b \mid k^2 < b$, and $k_1 \neq k$, $k_1 \neq -k_2$, or else $k_1 = -k_2 = \pm k$.

If $k_1 < k < \sqrt{b}$, then $k_1 > 0$. Suppose to the contrary $-\sqrt{b} < -k < k_2 < k_1 < 0$, then $b^2 - k^2 > b^2 - b$ and $b^2 + (k_1+k_2)b + k_1k_2 < b^2 - b$, contradicting $(b+k)(b-k) = (b+k_1)(b+k_2)$. Then $0 < k_1 < k \leq \sqrt{b}$, $-k_2 > k_1$ and $(b+k_1)(b+k_2) < b^2 - b - (-k_2 - k_1 - 1)b - k_1k_2 < b^2 - b$, contradicting $(b+k)(b-k) > b^2 - b$.

If $0 < k < k_1 \leq \sqrt{b}$, then $-k_2 < k_1$ and by Lemma 2

$$\begin{aligned} 2b + k_1 + k_2 &\leq 2(b+k_1) + k_1 + k_2 - 2\sqrt{(b+k_1)(k_1+k_2)} \\ (b+k_1)(k_1+k_2) &\leq k_1^2 \\ 0 < k_1 + k_2 &\leq -\frac{k_1k_2}{b} < 1 \end{aligned}$$

which is a contradiction. □

**Lemma 4.** *Given $b$, there is no sum path of length greater than one $P = (a_i, b_i)$ with $(a_1, b_1) = (b+k+1, b-k)$ in $G(n)$, $n \leq b + \sqrt{b} - 1$*

*Proof.* Suppose $(b+k+1)(b-k) = (b+k_1)(b+k_2)$, and without loss of generality $k_1 \geq k_2$. Then $k_1 + k_2 > 0$, or else

$$\begin{aligned} (b+k+1)(b-k) &= b^2 + b - k - k^2 \\ &\geq b^2 + b - (\sqrt{b}-1) - (\sqrt{b}-1)^2 \\ &= b^2 + \sqrt{b} \\ &> b^2 \\ &\geq (b+k_1)(b+k_2) \end{aligned}$$

In addition $k_1 + k_2 \neq 1$, or else $k_1 = -k$ or $k_1 = k+1$.

Suppose $k_1 < k + 1 \leq \sqrt{b} - 1$. Therefore $k_1 + k_2 - 1 > 0$, and $(b+k_1) + (b+k_2) > (b+1) + b$. By Lemma 2

$$\begin{aligned} 2b + k_1 + k_2 &\leq 2(b+k_1) + k_1 + k_2 - 1 - 2\sqrt{(k_1+k_2-1)(b+k_1)} \\ 4(k_1+k_2-1)(b+k_1) &\leq 4k_1^2 + 1 - 4k_1 \\ 0 < k_1 + k_2 - 1 &\leq \frac{1 - 4k_1k_2}{4b} \end{aligned}$$

Moreover,

$$\begin{aligned} \frac{1-4k_1k_2}{4b} &< \frac{1 - 4(\sqrt{b}-1)(2-\sqrt{b})}{4b} \\ &< 1 \end{aligned}$$

Therefore

$$0 < k_1 + k_2 - 1 \leq \frac{1 - 4k_1 k_2}{4b} < 1$$

which is a contradiction. $\square$

**Theorem 2.2.** *Given $n$, there is no sum path of length greater than one $P = (a_i, b_i)$ with $(a_1, b_1) = (b + k, b - k)$ in $G(n)$, $\frac{1+2n-\sqrt{1+4n}}{2} < b \leq n$, or $P = (a_i, b_i)$ with $(a_1, b_1) = (b + k + 1, b - k)$ in $G(n)$, $\frac{3+2n-\sqrt{5+4n}}{2} \leq b \leq n$.*

*Proof.* By inverting the inequality in Lemma 3 and Lemma 4. $\square$

**Lemma 5.** *For a sum tail $T = (a_i, b_i)$ of length $l$, every $T' = (a'_i, b'_i)$, $2k + 1 \leq i \leq l$ is a sum tail of length $l - 2k$ and every $T' = (a'_i, b'_i)$, $2k \leq i \leq l$ is a product tail of length $l + 1 - 2k$, where $1 \leq k \leq \frac{l}{2}$.*

*Proof.* Consider $T' = (a'_i, b'_i)$, $2k + 1 \leq i \leq l$, a sum tail of length $l - 2k$ where $1 \leq k \leq \frac{l}{2}$. Suppose to the contrary that there is a sum path $\bar{T} = (\bar{a}_i, \bar{b}_i)$ of length $\bar{l}$, where $(\bar{a}_1, \bar{b}_1) = (a'_1, b'_1)$ and $\bar{l} > l$. Then $T'' = (a''_i, b''_i)$, $(a''_i, b''_i) = (a_i, b_i)$ for $1 \leq j \leq 2k$, $(a''_i, b''_i) = (\bar{a}_{i-2k}, \bar{b}_{i-2k})$ for $2k + 1 \leq j \leq l + 1$ is a sum path with length $l'' > l$, $(a_1, b_1) = (a''_1, b''_1)$, contradicting that $T$ is a sum tail. Similarly every $T' = (a'_i, b'_i)$, $2k \leq i \leq l$ is a product tail of length $l + 1 - 2k$, where $1 \leq k \leq \frac{l}{2}$. $\square$

**Lemma 6.** *For a product tail $T = (a_i, b_i)$ of length $l$, every $T' = (a_j, b_j)$, $2k \leq j \leq l$ is a sum tail of length $l + 1 - 2k$ and every $T' = (a_j, b_j)$, $2k - 1 \leq j \leq l$ is a product tail of length $l + 2 - 2k$, where $1 \leq k \leq \frac{l-1}{2}$.*

*Proof.* Similar to the above. $\square$

**Theorem 2.3.** *If $G(n)$ has a pair $(p, q)$ of length $l$, it has another pair of length $l'$ for all $l' < l$.*

*Proof.* Suppose $l$ is odd, and let $P = (a_i, b_i)$ be the sum path of length $l$ with $(a_1, b_1) = (p, q)$, then $(a_{l+1-l'}, b_{l+1-l'})$ is a pair of length $l' < l$ by Lemma 5 and Theorem 2.1. Suppose $l$ is even, and let $P = (a_i, b_i)$ be the product path of length $l$ with $(a_1, b_1) = (p, q)$, then $(a_{l+1-l'}, b_{l+1-l'})$ is a pair of length $l' < l$ by Lemma 6 and Theorem 2.1. $\square$

**Lemma 7.** *For all $n \geq 12$, $(6, 4)$ is a pair of length six in $G(n)$.*

*Proof.* Consider the sum path $T$ of length five with $(a_1, b_1) = (8, 2)$, $(a_2, b_2) = (4, 4)$, $(a_1, b_1) = (6, 2)$, $(a_2, b_2) = (4, 3)$, $(a_3, b_3) = (5, 2)$, with $a_1 + b_1 = 8 + 2 = 7 + 3 = 6 + 4 = 5 + 5$, $(a_1, b_1) = (8, 2) \neq (6, 4)$. Then $T$ is a sum tail of length five since $16 = 8 \times 2 = 4 \times 4$, $12 = 6 \times 2 = 4 \times 3$, $10 = 5 \times 2$ have no other factorization, $7 = 4 + 3$ has no other partition, and $8 = 6 + 2 = 5 + 3 = 4 + 4$, $15 = 5 \times 3$ has no other factorization.

In addition, there is a sum path $P'$ of length six with $(a'_1, b'_1) = (6, 4)$, $(a'_2, b'_2) = (8, 3)$, $(a'_3, b'_3) = (9, 2)$, $(a'_4, b'_4) = (6, 3)$, $(a'_5, b'_5) = (5, 4)$, $(a'_6, b'_6) = (10, 2)$. Consider all partitions of $10 = 8 + 2 = 7 + 3 = 6 + 4 = 5 + 5$. For any other sum path $P''$ of length $l''$ with $a''_1 + b''_1 = 6 + 4$, $a''_1 \neq 4$, it follows that $(a''_1 + b''_1) = (7, 3)$ or $(5, 5)$. Since $21 = 7 \times 3$, $25 = 5 \times 5$ have no other factorization, $l'' = 1$. $\square$

**Corollary 2.** *For all $n \geq 8$, $(4, 4)$ is a pair of length four in $G(n)$.*

**Definition 8.** *Let $O$ denote an observer of the game, then $O(n, r) = Y$ if the observer is able to determine $(p, q)$ where there are $r$ NO before a YES in a sum-and-product game of $n$ and $O(n, r) = N$ if the observer is not able to determine $(p, q)$ where there are $r$ NO before a YES in a sum-and-product game of $n$.*

**Lemma 8.** *$O(n, r) = Y$ if and only if exactly one pair numbers has length $r$ in the sum-and-product game of $n$.*

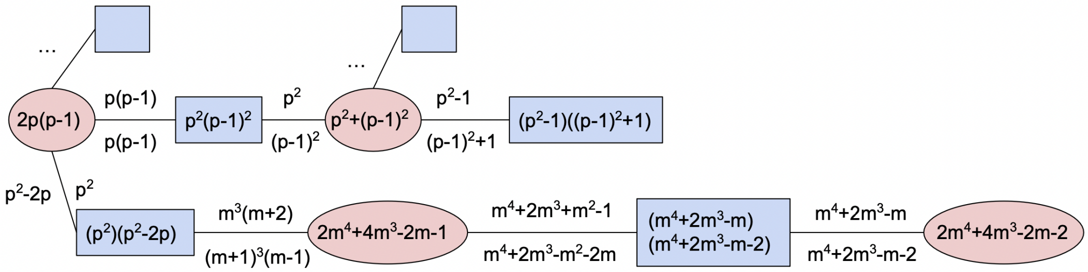


Figure 3: The pair $(p^2, p^2-2p)$ has length four in a game of $n = p^2$, $p = m(m+1)$

**Lemma 9.** *A sum path $P$ given by $(a_1, b_1) = (p(p-1), p(p-1))$, $(a_2, b_2) = (p^2, (p-1)^2)$, $(a_3, b_3) = (p^2-1, (p-1)^2+1)$ is a sum tail of length three in $G(p^2)$*

*Proof.* Consider $a_1b_1 = p^2(p-1)^2 = a_2b_2$, $a_2 + b_2 = p^2 + (p-1)^2 = a_3 + b_3$, then $P$ is a sum path. Consider another sum path $P' = (a'_i, b'_i)$ of length $l'$ with $(a'_1, b'_1) = (a_1, b_1) = (p(p-1), p(p-1))$.

If $(a'_2, b'_2) \neq (a_2, b_2) = (p^2, (p-1)^2)$, then $(p-1)^2 = b_2 < b'_2 < b'_1 = b_1 = p(p-1) = a_1 = a'_1 < a'_2 < a_2 = p^2$. Since $a_1b_1 = a'_2b'_2 = a_2b_2 = p^2(p-1)^2$, we have $2p^2 - 2p = a_1 + b_1 < a'_2 + b'_2 < a_2 + b_2 = 2p^2 - 2p + 1$. This is a contradiction since $a'_2 + b'_2$ is an integer.

If $(a'_2, b'_2) = (a_2, b_2)$, we seek to show $l' \leq 3$. Consider $(a'_3, b'_3) = (p^2 - p + 1 + k, p^2 - p - k)$, $0 \leq k \leq p-2$, $(a'_4, b'_4) = (p^2 - p + \alpha, p^2 - p + \beta)$, $\alpha \leq p$, $(p^2 - p + 1 + k)(p^2 - p - k) \leq (p^2 - p + \alpha)^2$. Then

$$\begin{aligned}(p^2-p)^2 - k^2 + p^2 - p - k &\leq (p^2-p)^2 + \alpha^2 + 2\alpha(p^2-p) \\ 0 &\leq \alpha^2 + 2\alpha(p^2-p) + k^2 - p^2 + p + k \\ \alpha &\geq (p - p^2) + \sqrt{(p^2-p)^2 + p^2 - p - k^2 - k} \\ &\geq 0\end{aligned}$$

Consider $a'_3b'_3 = a'_4b'_4$, then $a'_3 \mid a'_4b'_4$ and

$$p^2 - p + \alpha \quad | \quad (p^2 - p + 1 + k)(p^2 - p - k)$$

Since $\gcd(p^2 - p + \alpha, p^2 - p + 1 + k) \mid (k + 1 - \alpha)$, $\gcd(p^2 - p + \alpha, p^2 - p - k) \mid (k + \alpha)$,

$$p^2 - p + \alpha \mid (k + 1 - \alpha)(k + \alpha) = k^2 + k - \alpha^2 - \alpha$$

Note that if $\alpha \leq p - 1$

$$\begin{aligned}p^2 - p < n\sqrt{(p^2 - p + 1 + k)(p^2 - p - k)} &\leq |p^2 - p + \alpha| \\ |k^2 + k - \alpha^2 - \alpha| &\leq (p-1)^2 + (p-1) = p^2 - p\end{aligned}$$

which is a contradiction. If $\alpha = p$, then

$$p^2 \mid k^2 + k - p$$

This is a contradiction since $|k^2 + k - p| < p^2$ when $0 \leq k \leq p-2$. Therefore $l' \leq 3$ and the sum path $P$ is a sum tail of length three in $G(p^2)$

□

**Theorem 2.4.** *There are infinitely many $n$ such that $O(n, 4) = N$.*

*Proof.* By Lemma 8, it suffices to find another pair $(p, q) \neq (4, 4)$ of length four in $G(n)$ for infinitely many $n$. By Lemma 9, consider the sum tail of length three with $(a_1, b_1) = (p(p-1), p(p-1))$, $p = m(m+1)$ in $G(n)$. Consider another sum path $P'$ of length $l'$ with $a'_1 + b'_2 = 2p(p-1)$, and without loss of generality let $a'_1 = p(p-1) + k$, $b'_1 = p(p-1) - k$, $1 \leq k \leq p$.

Suppose $l' \geq 2$, without loss of generality let $(a'_2, b'_2) = (p(p-1) + \alpha, p(p-1) + \beta)$, $\alpha \leq p$, $(p(p-1) + k)(p(p-1) - k) \leq (p(p-1) + \alpha)^2$. Then

$$-k^2 \leq 2\alpha(p^2 - p) + \alpha^2$$

Since $2\alpha(p^2 - p) + \alpha^2 \leq -2p^2 + 2p + 1 < -p^2$ when $p \geq 3$, it follows that $-p^2 \leq -k^2 \leq 2\alpha(p^2 - p) + \alpha^2 < -p^2$ when $\alpha \leq -1$. This is a contradiction, and $\alpha \geq 0$. Consider

$$p(p-1) + \alpha \mid (p(p-1) + k)(p(p-1) - k)$$

Since $\gcd(p^2 - p + \alpha, p^2 - p + k) \mid (\alpha - k)$, $\gcd(p^2 - p + \alpha, p^2 - p - k) \mid (\alpha + k)$,

$$p(p-1) + \alpha \mid \alpha^2 - k^2$$

Note that

$$|\alpha^2 - k^2| \leq p^2$$
$$|p(p-1) + \alpha| \geq p^2 - p$$

Therefore $|\alpha^2 - k^2| = |p(p-1) + \alpha|$. Suppose $\alpha^2 - k^2 = p(p-1) + \alpha$, then $p(p-1) = \alpha^2 - \alpha - k^2 < p(p-1)$, which is a contradiction. Otherwise $k^2 - \alpha^2 = p(p-1) + \alpha$. Suppose $k \leq p-1$, then $p^2 - p = k^2 - \alpha^2 - \alpha < p^2 - p$, which is a contradiction. Let $k = p$, then $p = \alpha(\alpha+1)$. Since $\alpha > 0$, it follows that $\alpha = m$, $(a'_1, b'_1) = (p^2, p^2 - 2p)$, $(a'_2, b'_2) = (m^3(m+2), (m+1)^3(m-1))$. To show $l' > 3$, consider $(a'_3, b'_3) = (m^4 + 2m^3 + m^2 - 1, m^4 + 2m^3 - m^2 - 2m)$, $(a'_4, b'_4) = (m^4 + 2m^3 - m, m^4 + 2m^3 - m - 2)$.

□

# 3 Number of Product nodes

To find whether there are infinitely many $n$ such that the observer is able to determine the pair of numbers after hearing 4 'NO' before the first 'YES,' an estimation on the number of product nodes in $G(n)$ is required. By the unique factorization theorem, any positive integer can be uniquely expressed as $p_1 p_2 ... p_m$, where $p_i$ primes, $p_i \leq p_{i+1}$. A product node $p_1 p_2 ... p_m$ exists in $G(n)$ if and only if $p_1 p_2 ... p_m = (p'_1 ... p'_a)(q'_1 ... q'_b)$ where $p'_i$, $q'_i$ primes, $p'_1 ... p'_a$, $q'_1 ... q'_b \leq n$.

In this section, let $\lfloor n \rfloor$ denote the largest prime less than or equal to $n$, and $\sum_a^b$ sum over all primes $p$ with $a \leq p \leq b$. In addition, let $\pi(n)$ denote the number of primes numbers less than or equal to $n$. The number of product nodes is counted by classifying a product node based on the number of primes $m$ in its prime factorization. When $m = 2$, the number of products nodes of the form $p_1 p_2$ is

$$\sum_{p_1=2}^{\lfloor n \rfloor} \sum_{p_2=p_1}^{\lfloor n \rfloor} 1 = \sum_{p_1=2}^{\lfloor n \rfloor} (\pi(n) - \pi(p_1) + 1) = \frac{\pi(n)(\pi(n)+1)}{2}$$

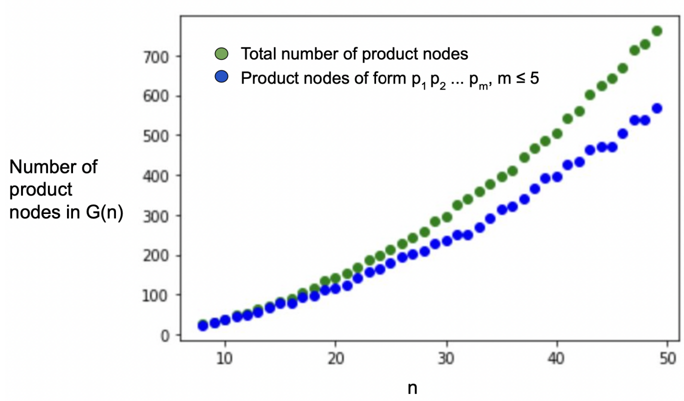


Figure 4: Product nodes of form $p_1p_2...p_m$, $m \leq 5$ in blue and actual number of product nodes in

When $m = 3$, a product node of the form $p_1p_2p_3$ exists in $G(n)$ if and only if $p_1p_2 \leq n$, $p_3 \leq n$. Since $p_1 \leq p_2 \leq p_3$, we obtain $p_1 \leq \lfloor\sqrt{n}\rfloor$, $p_1 \leq p_2 \leq \lfloor n/p_1 \rfloor$, and $p_2 \leq p_3 \leq \lfloor n \rfloor$. Thus the number of product nodes of the form $p_1p_2p_3$ is

$$\sum_{p_1=2}^{\lfloor\sqrt{n}\rfloor} \sum_{p_2=p_1}^{\lfloor n/p_1 \rfloor} \sum_{p_3=p_2}^{\lfloor n \rfloor} 1 = \frac{1}{2} \sum_{p_1=2}^{\lfloor\sqrt{n}\rfloor} [2\pi(n) + 2 - \pi(\frac{n}{p_1}) - \pi(p_1)][\pi(\frac{n}{p_1}) - \pi(p_1) + 1]$$

The number of product nodes when $m = 4$ and $m = 5$ are similarly determined as in Appendix. Figure 4 plots the product nodes of form $p_1p_2...p_m$, $m \leq 5$ and the total number of product nodes.

Alternatively, the number of product nodes can be estimated with the prime number theorem. Since numbers of the form $p_1p_2$ cannot be a product node in $G(n)$, where $p_1 \geq n$ is a prime, the number of product nodes is estimated to be

$$n^2 - \int_n^{n^2} \frac{1}{\ln p} \frac{n^2}{p} dp \approx (1 - \ln 2)n^2$$

# 4 Infinitely many $n$ such that $O(n, 4) = Y$

In this section we prove that under certain assumptions, there are infinitely many $n$ such that an observer is able to determine the pair of numbers $(p, q)$ after hearing 4 'NO' before the first 'YES' in a sum-and-product game of $n$.

**Definition 9.** $C_n = \{(p, q) \mid (p, q)\, involves\, 4\, NO\, before\, the\, first\, YES\, in\, a\, sum\, and\, product\, game\, of\, n,\ (p, q) \neq (4, 4)\}$.

**Definition 10.** *A pair* $(p, q)$ *appears at* $\tau_1(p, q)$ *if* $\tau_1(p, q)$ *is the least integer such that* $(p, q) \in C_{\tau_1(p,q)}$.

**Definition 11.** *A pair* $(p, q)$ *disappears at* $\tau_2(p, q) + 1$ *if* $\tau_2(p, q)$ *is the greatest integer such that* $(p, q) \in C_{\tau_2(p,q)}$.

**Lemma 10.** *Every pair* $(p, q)$ *of length* $l \geq 2$ *apart from* $(6, 4)$, $(8, 2)$, $(4, 4)$, $(6, 2)$ *and* $(4, 3)$ *eventually disappears.*

*Proof.* Suppose the pair has length $l \geq 2$. By Theorem and Corollary, there is either a sum tail or a product tail $T$ of length $l$ with $(a_1, b_1) = (p, q)$. Suppose $m = a_l + b_l \geq 11$, then $m = (4) + (m - 4) = (6) + (m - 6)$, where $4 \neq m - 6$, and $(4)(m - 4) = (2)(2m - 8)$, $(6)(m - 6) = (3)(2m - 12)$. Therefore the pair disappears when $n \geq 2m - 12$.

Suppose a pair $(p, q)$ of length $l \leq 2$ never disappears, then $a_l + b_l < 8$. By observing Figure X, the only pairs $(p, q)$ are $(6, 4)$, $(8, 2)$, $(4, 4)$, $(6, 2)$ and $(4, 3)$. □

**Lemma 11.** *$\tau_2(p,q)$ is well-defined and $(p,q) \in C_\tau$ for every $\tau_1(p,q) \le \tau \le \tau_2(p,q)$.*

*Proof.* Since only new edges are added and no edges are removed as $n$ increases, a pair of a certain length cannot reappear after it disappears. □

**Definition 12.** *The total number of pairs of length four under $n = M$ is*

$$\begin{aligned} TN_M &= \sum_{n=2}^{M-1} \sum_{(p,q)\in C_n} \frac{1}{\tau_2(p,q) - \tau_1(p,q)} \mathbb{1}_{\tau_2(p,q)<M} \\ &= |(C_2 \cup C_3 \cup ... \cup C_{M-1}) \cap \{(p,q) \,|\, \tau_2(p,q) < M\}| \end{aligned}$$

*where $\mathbb{1}_{\tau_2(p,q)<M} = 1$ if $\tau_2(p,q) < M$ and $\mathbb{1}_{\tau_2(p,q)<M} = 0$ otherwise.*

**Definition 13.** *The number of pairs of length four under $n = M$ with another pair of length four at $\tau_2(p,q)+1$ is*

$$AC_M = \sum_{n=2}^{M-1} \sum_{(p,q)\in C_n} \frac{1}{\tau_2(p,q) - \tau_1(p,q)} \mathbb{1}_{C_{\tau_2(p,q)+1} \neq \emptyset}$$

*where $\mathbb{1}_{C_{\tau_2(p,q)+1} \neq \emptyset} = 1$ if $C_{\tau_2(p,q)+1} \neq \emptyset$ and $\mathbb{1}_{C_{\tau_2(p,q)+1} \neq \emptyset} = 0$ otherwise.*

To prove that there are infinitely many $n$ such that $O(4,n) = Y$, it suffices to prove that there are infinitely many pairs $(p,q)$ of length four such that there is no other pair $(p',q')$ of length four where $(p',q') \in C_{\tau_2(p,q)+1}$. Since every pair of length four eventually disappears, it is possible to only consider pairs $(p,q)$ such that there is no other pair $(p',q')$, $\tau_1(p',q') \le \tau_1(p,q)$, $\tau_2(p',q') \ge \tau_2(p,q)$. Equivalently, we hope to prove

$$\lim_{M\to\infty} \frac{AC_M}{TN_M} < 1$$

We first prove a lemma in order to prove the above theorem.

**Lemma 12.** *Let $C_n$ be defined as above, then under certain assumptions of independence $\lim_{n\to\infty} \frac{|C_{n-1} \cap C_n^c|}{|C_{n-1}|} \ge \frac{5}{9}$.*

*Proof.* Consider a pair $(p,q)$ of length four, with its corresponding product tail $T = (a_i, b_i)$ of length four by Corollary 1. Note that $(p,q)$ disappears at $n = \tau_2(p,q)+1$ if and only if $n$ is the smallest integer such that there is a product path $P'$ of length greater than four with $(a'_1, b'_1) = (p,q)$ in $G(n)$. Let $(p,q)$ be a pair of length four in the game of $n-1$ and $P'$ be a product path of length greater than four, then

$$S_{k,n} = \{(p,q) \,|\, \exists\, P' = (a'_i, b'_i) \in G(n),\ (a'_1, b'_1) = (p,q), (a'_j, b'_j) \in G(n-1)\, \forall j < k,\ a'_k = n\}$$

Since $a'_k = n$ for some $2 \le k \le 5$, $C_{n-1} \cap C_n^c = S_{2,n} \cup S_{3,n} \cup S_{4,n} \cup S_{5,n}$. Consider

$$\frac{|C_{n-1} \cap C_n^c|}{|C_{n-1}|} = \frac{|S_{2,n} \cup S_{3,n} \cup S_{4,n} \cup S_{5,n}|}{|C_{n-1}|}$$

To approximate $\frac{|S_{3,n}|}{|C_{n-1}|}$ as $n \to \infty$, note that a pair $(p,q) \in C_{n-1}$ satisfies $(p,q) \in S_{3,n}$ only if $n$ is a factor of $a_2 b_2$. Assuming that the product node $a_2 b_2$ is randomly chosen from all product nodes in $G(n-1)$ and considering that $n$ is a factor of approximately $\frac{(1-\ln(2))(n-1)^2}{n}$ in $G(n-1)$, the probability that $n$ is a factor of $a_2 b_2$ is

$$\begin{aligned} \mathbb{P}(n \,|\, a_2 b_2) &\approx \frac{1}{(1-\ln(2))(n-1)^2} \frac{(1-\ln(2))(n-1)^2}{n} \\ &= \frac{1}{n} \end{aligned}$$

Therefore $\frac{|S_{3,n}|}{|C_{n-1}|} \to 0$ as $n \to \infty$, and similarly $\frac{|S_5|}{|C_{n-1}|} \to 0$ as $n \to \infty$. It is therefore appropriate to consider

$$
\begin{aligned}
\lim_{n\to\infty} \frac{|C_{n-1} \cap C_n^c|}{|C_{n-1}|} &= \lim_{n\to\infty} \frac{|S_{2,n} \cup S_{3,n} \cup S_{4,n} \cup S_{5,n}|}{|C_{n-1}|} \\
&\geq \lim_{n\to\infty} \frac{|S_{2,n} \cup S_{4,n}|}{|C_{n-1}|} \\
&= \lim_{n\to\infty} \left(\frac{|S_{2,n}|}{|C_{n-1}|} + \frac{|S_{4,n}|}{|C_{n-1}|} - \frac{|S_{2,n} \cap S_{4,n}|}{|C_{n-1}|}\right)
\end{aligned}
$$

To approximate $\frac{|S_{2,n}|}{|C_{n-1}|}$ as $n \to \infty$, note that a pair $(p,q) \in C_{n-1}$ satisfies $(p,q) \in S_{2,n}$ if and only if the product node $(n)(p+q-n)$ is in $G(n-1)$ and there exists a product path $P'' \in G(n-1)$, $l(P'') \geq 3$ with $(a_1'', b_1'') = (u,v)$, $uv = (n)(p+q-n)$. Let $K(n)$ be the set of product nodes in $G(n)$, and

$$
J(n) = \{x \in K(n) \mid \exists\, P'' \in G(n-1), l(P'') \geq 3, a_1''b_1'' = x\}
$$

where $P''$ is a product path, then $|J(n)| = |K(n)| - o(n)$. Suppose the product node $(n)(p+q-n)$ is randomly chosen from all product nodes that are a multiple of $n$ in $G(n)$, then

$$
\begin{aligned}
\frac{|S_{2,n}|}{|C_{n-1}|} &= \frac{|\{x \mid x \in J(n), n \mid x\}|}{|\{x \mid x \in K(n) \cap K^c(n-1)\} \sqcup \{x \mid x \in K(n-1), n \mid x\}|} \\
&\approx \frac{\frac{(1-\ln(2))n^2}{n} - o(n)}{(1-\ln(2))(n^2-(n-1)^2) + \frac{(1-\ln(2))(n-1)^2}{n}} \\
&\to \frac{1}{3}
\end{aligned}
$$

Similarly to approximate $\frac{|S_{4,n}|}{|C_n|}$, note that a pair $(p,q) \in C_{n-1}$ satisfies $(p,q) \in S_{4,n}$ if the product node $(n)(p+q-n)$ is in $G(n-1)$. Suppose the product node $(n)(p+q-n)$ is randomly chosen from all product nodes that are a multiple of $n$ in $G(n)$,

$$
\begin{aligned}
\frac{|S_{4,n}|}{|C_{n-1}|} &= \frac{|\{x \mid x \in K(n-1), n \mid x\}|}{|\{x \mid x \in K(n) \cap K^c(n-1)\} \sqcup \{x \mid x \in K(n-1), n \mid x\}|} \\
&\approx \frac{\frac{(1-\ln(2))(n-1)^2}{n}}{(1-\ln(2))(n^2-(n-1)^2) + \frac{(1-\ln(2))(n-1)^2}{n}} \\
&\to \frac{1}{3}
\end{aligned}
$$

□

**Conjecture 1.** *$S_{2,n}$ and $S_{4,n}$ are independent, and* $\frac{|S_{2,n} \cap S_{4,n}|}{|C_{n-1}|} = \frac{|S_{2,n}|}{|C_{n-1}|}\frac{|S_{4,n}|}{|C_{n-1}|}$

Assuming independence between $S_{2,n}$ and $S_{4,n}$,

$$
\begin{aligned}
\lim_{n\to\infty} \frac{|C_{n-1} \cap C_n^c|}{|C_{n-1}|} &\geq \lim_{n\to\infty} \frac{|S_{2,n}| + |S_{4,n}| - |S_{2,n} \cap S_{4,n}|}{|C_{n-1}|} \\
&= \frac{1}{3} + \frac{1}{3} - \frac{1}{9} \\
&= \frac{5}{9}
\end{aligned}
$$

**Corollary 3.** *Let $C_n$ be defined as above, then* $\lim_{n\to\infty} \frac{|C_{n-1} \cap C_n|}{|C_{n-1}|} \leq \frac{4}{9}$.

**Theorem 4.1.** *Let $AC_M$, $TN_M$ be defined as above, then* $\lim_{M\to\infty}\frac{AC_M}{TN_M} < 1$.

*Proof.* Given $a < b$, let

$$C_{(a,b)} = C_a^c \cap C_{a+1} \cap ... \cap C_{b-1} \cap C_b^c$$
$$C_{(a,b]} = C_a^c \cap C_{a+1} \cap ... \cap C_{b-1} \cap C_b$$

Consider

$$\frac{AC_M}{TN_M} = \frac{1}{TN_M}\sum_{k=1}\sum_{j=2}^{M-k}|C_{(j-1,j+k)}|\mathbb{1}(\bigsqcup_{m=j+1}^{j+k}C_{(m-1,j+k]} \neq \emptyset)$$

Taking limit of both sides,

$$\begin{aligned}\lim_{M\to\infty}\frac{AC_M}{TN_M} &= \lim_{M\to\infty}\frac{1}{TN_M}\sum_{k=1}\sum_{j=2}^{M-k}|C_{(j-1,j+k)}|\mathbb{1}(\bigsqcup_{m=j+1}^{j+k}C_{(m-1,j+k]} \neq \emptyset)\\ &= \lim_{M\to\infty}\frac{1}{TN_M}\sum_{k=1}\sum_{j=2}^{M-k}[\lim_{j'\to\infty}|C_{(j'-1,j'+k)}|][\lim_{j'\to\infty}\mathbb{1}(\bigsqcup_{m=j'+1}^{j'+k}C_{(m-1,j'+k]} \neq \emptyset)]\end{aligned}$$

**Conjecture 2.** *$C_i$ are independent, and* $\frac{|C_{(a,b]}|}{|C_{(a,b-1]}|} = \frac{|C_{b-1}\cap C_b|}{|C_{b-1}|}$.

Under this assumption, we obtain

$$\begin{aligned}|C_{(j-1,j+k)}| &= |C_{j-1}^c \cap C_j|[\prod_{l=1}^{k-1}\frac{|C_{(j-1,j+l]}|}{|C_{(j-1,j+l-1]}|}]\frac{|C_{(j-1,j+k)}|}{|C_{(j-1,j+k-1]}|}\\ &= |C_{j-1}^c \cap C_j|[\prod_{l=1}^{k-1}\frac{|C_{j+l-1}\cap C_{j+l}|}{|C_{j+l-1}|}](1-\frac{|C_{j+k-1}\cap C_{j+k}|}{|C_{j+k-1}|})\end{aligned}$$

**Conjecture 3.** *For random $j \geq n$, $\mathbb{E}(|C_{j-1}^c \cap C_j|) \leq 1$ as $n \to \infty$.*

Taking limit of both sides,

$$\begin{aligned}\lim_{j\to\infty}|C_{(j-1,j+k)}| &= \lim_{j\to\infty}\{|C_{j-1}^c \cap C_j|[\prod_{l=1}^{k-1}\frac{|C_{j+l-1}\cap C_{j+l}|}{|C_{j+l-1}|}](1-\frac{|C_{j+k-1}\cap C_{j+k}|}{|C_{j+k-1}|})\}\\ &= \lim_{j\to\infty}|C_{j-1}^c \cap C_j|[\prod_{l=1}^{k-1}\lim_{j\to\infty}\frac{|C_{j+l-1}\cap C_{j+l}|}{|C_{j+l-1}|}]\lim_{j\to\infty}(1-\frac{|C_{j+k-1}\cap C_{j+k}|}{|C_{j+k-1}|})\\ &= \lim_{j\to\infty}\{|C_{j-1}^c \cap C_j|R^{k-1}(1-R)\}\end{aligned}$$

where $R = \lim_{n\to\infty}\frac{|C_{n-1}\cap C_n|}{|C_{n-1}|}$ and the limit is finite. We further note that

$$\begin{aligned}\mathbb{1}(\bigsqcup_{m=j+1}^{j+k}C_{(m-1,j+k]} \neq \emptyset) &\leq \sum_{m=j+1}^{j+k}|C_{(m-1,j+k]}|\\ &= \sum_{m=j+1}^{j+k}|C_{m-1}^c \cap C_m|[\prod_{l=m}^{j+k}\frac{|C_{(m-1,l]}|}{|C_{(m-1,l-1]}|}]\\ &= \sum_{m=j+1}^{j+k}|C_{m-1}^c \cap C_m|[\prod_{l=m}^{j+k}\frac{|C_{l-1}\cap C_l|}{|C_{l-1}|}]\end{aligned}$$

Considering $C_{(m-1,j+k]}$ as random variables, we take limit of expectation of both sides,

$$\begin{aligned}
\lim_{j\to\infty}\mathbb{P}(\bigsqcup_{m=j+1}^{j+k}C_{(m-1,j+k]}\neq\emptyset) &\le \lim_{j\to\infty}\{\sum_{m=j+1}^{j+k}|C^c_{m-1}\cap C_m|[\prod_{l=m}^{j+k}\frac{|C_{l-1}\cap C_l|}{|C_{l-1}|}]\} \\
&\le \lim_{j\to\infty}\{\sum_{m=j+1}^{j+k}[\prod_{l=m}^{j+k}\frac{|C_{l-1}\cap C_l|}{|C_{l-1}|}]\} \\
&= \sum_{m=j+1}^{j+k}[\prod_{l=m}^{j+k}\lim_{m\to\infty}\frac{|C_{l-1}\cap C_l|}{|C_{l-1}|}] \\
&= R\frac{1-R^k}{1-R}
\end{aligned}$$

Therefore an upper bound for $\lim_{M\to\infty}\frac{AC_M}{TN_M}$ is obtained

$$\begin{aligned}
\lim_{M\to\infty}\frac{AC_M}{TN_M} &\le \lim_{M\to\infty}\sum_{k=1}\frac{1}{TN_M}\sum_{j=2}^{M-k}[\lim_{j'\to\infty}|C^c_{j'-1}\cap C_{j'}|R^{k-1}(1-R)](R\frac{1-R^k}{1-R}) \\
&= \sum_{k=1}R^{k-1}(1-R)(R\frac{1-R^k}{1-R})\lim_{M\to\infty}\{\sum_{k=1}\frac{1}{TN_M}\sum_{j=2}^{M-k}|C^c_{j-1}\cap C_j|\} \\
&= \sum_{k=1}R^{k-1}(1-R)(R\frac{1-R^k}{1-R}) \\
&= \frac{R}{1-R^2}
\end{aligned}$$

By Corollary 3 $R=\lim_{n\to\infty}\frac{|C_{n-1}\cap C_n|}{|C_{n-1}|}\le\frac{4}{9}$. Therefore $\lim_{M\to\infty}\frac{AC_M}{TN_M}<1$. □

### 4.1 Conjecture

By observing $G(n)$ for various $n$, we propose two conjectures of the tail $T=(a_i,b_i)$ of the longest length $l$ in $G(n)$.

**Conjecture 4.** *It is very likely that* $a_l+b_l>a_{l-1}+b_{l-1}$.

**Conjecture 5.** *The asymptotic behavior of* $a_l+b_l$ *is that* $a_l+b_l\sim 2n$.

# 5 Appendix

We obtain expressions for the number of product nodes of the form $p_1p_2...p_m$, $m=4,5$ and generalize to higher $m$. When $m=4$, a product node of the form $p_1p_2p_3p_4$ exists in $G(n)$ if and only if $p_2p_3\le n$ and $p_1p_4\le n$, or $p_1p_2p_3\le n$ and $p_4\le n$.

1. $p_2p_3\le n$ and $p_1p_4\le n$

$$\sum_{p_1=2}^{\lfloor\sqrt{n}\rfloor}\sum_{p_2=p_1}^{\lfloor\sqrt{n}\rfloor}\sum_{p_3=p_2}^{\lfloor\frac{n}{p_2}\rfloor}\sum_{p_4=p_3}^{\lfloor\frac{n}{p_3}\rfloor}1=\frac{1}{2}\sum_{p_1=2}^{\lfloor\sqrt{n}\rfloor}\sum_{p_2=p_1}^{\lfloor\sqrt{n}\rfloor}[(2\pi(\frac{n}{p_1})+2-\pi(p_2)-\pi(\frac{n}{p_2})][\pi(\frac{n}{p_2})-\pi(p_2)+1]$$

2. $p_1p_2p_3\le n$ and $p_4\le n$
   To avoid double-counting, we count the number of product nodes such that (2) is satisfied but (1) is not satisfied.

This is equivalent to the conditions $p_1p_2p_3 \le n$ and $\frac{n}{p_1} < p_4 \le n$. Since $p_3 \le \frac{n}{p_1}$, for a fixed $p_1$ the choices of $(p_1, p_2, p_3)$ and $p_4$ are independent. Therefore we have

$$\sum_{p_1=2}^{\lfloor \sqrt[3]{n} \rfloor}[(\sum_{p_2=p_1}^{\lfloor \sqrt{\frac{n}{p_1}} \rfloor}\sum_{p_3=p_2}^{\lfloor \frac{n}{p_1p_2} \rfloor}1)(\sum_{p_4=\frac{n}{p_1}+1}^{\lfloor n \rfloor}1)] = \sum_{p_1=2}^{\lfloor \sqrt[3]{n} \rfloor}[(\sum_{p_2=p_1}^{\lfloor \sqrt{\frac{n}{p_1}} \rfloor}\sum_{p_3=p_2}^{\lfloor \frac{n}{p_1p_2} \rfloor}1)(\pi(n)-\pi(\frac{n}{p_1}+1))]$$

When $m = 5$, a product node of form $p_1p_2p_3p_4p_5$ can be written as the product of two numbers $(a, b)$, $a, b \le n$ in $\binom{5}{1} + \binom{5}{2} = 15$ ways.

$$(p_1p_2p_3, p_4p_5), (p_1p_2p_4, p_3p_5), (p_1p_2p_5, p_3p_4), (p_1p_3p_4, p_2p_5), (p_2p_3p_4, p_1p_5)$$
$$(p_1p_3p_5, p_2p_4), (p_1p_4p_5, p_2p_3), (p_2p_3p_5, p_1p_4), (p_2p_4p_5, p_1p_3), (p_3p_4p_5, p_1p_2)$$
$$(p_2p_3p_4p_5, p_1), (p_1p_3p_4p_5, p_2), (p_1p_2p_4p_5, p_3), (p_1p_2p_3p_5, p_4), (p_1p_2p_3p_4, p_5)$$

We note that every pair in the second lines implies a pair in the first line. For example, $p_1p_4p_5 \le n$, $p_2p_3 \le n$ implies $p_1p_2p_3 \le n$, $p_4p_5 \le n$. In addition, every pair in the last line implies the last pair in the last line.

1. $(p_1p_2p_3, p_4p_5)$

$$\sum_{p_1=2}^{\lfloor \sqrt[3]{n} \rfloor}\sum_{p_2=p_1}^{\lfloor \sqrt{\frac{n}{p_1}} \rfloor}\sum_{p_3=p_2}^{\lfloor \frac{n}{p_1p_2} \rfloor}\sum_{p_4=p_3}^{\lfloor \sqrt{n} \rfloor}\sum_{p_5=p_4}^{\lfloor \frac{n}{p_4} \rfloor}1$$

2. $(p_1p_2p_4, p_3p_5)$
   $p_1p_2p_4 \le n$, $\frac{n}{p_4} < p_5 \le \frac{n}{p_3}$

$$\sum_{p_1=2}^{\lfloor \sqrt[3]{n} \rfloor}\sum_{p_2=p_1}^{\lfloor \sqrt{\frac{n}{p_1}} \rfloor}\sum_{p_3=p_2}^{\lfloor \sqrt{n} \rfloor}\sum_{p_4=p_3}^{\lfloor \frac{n}{p_1p_2} \rfloor}\sum_{p_5=\max(p_4\frac{n}{p_4}+1)}^{\lfloor \frac{n}{p_3} \rfloor}1$$

3. $(p_1p_2p_5, p_3p_4)$
   $p_3p_4 \le n$, $\frac{n}{p_3} < p_5 \le \frac{n}{p_1p_2}$

$$\sum_{p_1=2}^{\lfloor \sqrt[3]{n} \rfloor}\sum_{p_2=p_1}^{\lfloor \sqrt{\frac{n}{p_1}} \rfloor}\sum_{p_3=p_2}^{\lfloor \sqrt{n} \rfloor}\sum_{p_4=p_3}^{\lfloor \frac{n}{p_3} \rfloor}\sum_{p_5=\max(p_4,\frac{n}{p_3}+1)}^{\lfloor \frac{n}{p_1p_2} \rfloor}1$$

4. $(p_1p_3p_4, p_2p_5)$
   $p_1p_3p_4 \le n$, $\frac{n}{p_3} < p_5 \le \frac{n}{p_2}$, $n < p_1p_2p_5$.

$$\sum_{p_1=2}^{\lfloor \sqrt[3]{n} \rfloor}\sum_{p_2=p_1}^{\lfloor \sqrt{n} \rfloor}\sum_{p_3=p_2}^{\lfloor \sqrt{\frac{n}{p_1}} \rfloor}\sum_{p_4=p_3}^{\lfloor \frac{n}{p_1p_3} \rfloor}\sum_{p_5=\max(p_4,\frac{n}{p_3}+1,\frac{n}{p_1p_1}+1)}^{\lfloor \frac{n}{p_2} \rfloor}1$$

5. $(p_2p_3p_4, p_1p_5)$
   $p_2p_3p_4 \le n$, $\frac{n}{p_2} < p_5 \le \frac{n}{p_1}$

$$\sum_{p_1=2}^{\lfloor \sqrt{n} \rfloor}\sum_{p_2=p_1}^{\lfloor \sqrt[3]{n} \rfloor}\sum_{p_3=p_2}^{\lfloor \sqrt{\frac{n}{p_2}} \rfloor}\sum_{p_4=p_3}^{\lfloor \frac{n}{p_2p_3} \rfloor}\sum_{p_5=\max(p_4,\frac{n}{p_2}+1)}^{\lfloor \frac{n}{p_1} \rfloor}1$$

6. $(p_1p_2p_3p_4, p_5)$
   The case occurs, without any of the above cases occurring, if and only if $p_1p_5 > n$

$$\sum_{p_1=2}^{\lfloor \sqrt[4]{n} \rfloor} \sum_{p_2=p_1}^{\lfloor \sqrt[3]{\frac{n}{p_1}} \rfloor} \sum_{p_3=p_2}^{\lfloor \sqrt{\frac{n}{p_1p_2}} \rfloor} \sum_{p_4=p_3}^{\lfloor \frac{n}{p_1p_2p_3} \rfloor} \sum_{p_5=\max(p_4, \frac{n}{p_1}+1)}^{\lfloor n \rfloor} 1$$

## 5.1 Generalization to Higher $m$

Let the product be $p_1p_2...p_m$ in $G(n)$.

1. The number of ways $p_1p_2...p_m$ can be written as the product of two numbers $(a, b)$, $a, b \leq n$ is $2^{m-1} - 1$.

2. Some cases may include other cases. For example, $(\prod_{i=1}^{m-1} p_i, p_m)$ includes $(\prod_{\substack{i=1 \\ i \neq j}}^{m} p_i, p_j)$.

3. Choose one case $(p_{r_1}p_{r_2}...p_{r_a}, p_{s_1}p_{s_2}...p_{s_b})$, $r_{i-1} < r_i$, $s_{i-1} < s_i$. Without loss of generality suppose $r_a < s_b$. The number of product nodes of this form is

$$\sum_{p_1=2}^{n} \sum_{p_2=p_1}^{n} ... \sum_{p_i=p_{i-1}}^{n} ... \sum_{p_{r_a}=p_{r_a-1}}^{\frac{n}{p_{r_1}p_{r_2}\cdots p_{r_{a-1}}}} ... \sum_{p_i=p_{i-1}}^{n} ... \sum_{p_{s_b}=p_{s_b-1}}^{\frac{n}{p_{r_1}p_{r_2}\cdots p_{r_{b-1}}}}$$

4. Suppose we calculated the number of product nodes of $M$ forms and seek to calculate the additional number of product nodes of the $(M+1)$th form. Inequalities of the $(M+1)$th form are satisfied and at least one of the two conditions of each of the previous $M$ forms is not satisfied. Therefore instead of summing $p_i$ from $p_{i-1}$, we sum $p_i$ from $\max(p_{i-1}, A_1, ..., A_j)$, where $A$ represent inequalities not satisfied.

5. We seek an improved upper bound for summation of each of $p_{r_i}$, $p_{s_i}$. Consider $p_{r_1}p_{r_2}...p_{r_a} \leq n$, then $p_{r_i} \leq \sqrt[a+1-i]{\frac{n}{p_{r_2}\cdots p_{r_{i-1}}}}$. Similarly $p_{s_i} \leq \sqrt[b+1-i]{\frac{n}{p_{s_2}\cdots p_{s_{i-1}}}}$

6. To find the maximum for $m$, consider $2^m \leq n \iff m \leq \log_2(n)$