Probability to Compute Divisor of a Hidden Integer

Xingbo Wang\textsuperscript{1,2,3}, Jianhui Li\textsuperscript{2}, Zhikui Duan\textsuperscript{1,3} & Wen Wan\textsuperscript{3}

\textsuperscript{1} Department of Mechatronic Engineering, Foshan University, China
\textsuperscript{2} State Key Laboratory of Mathematical Engineering and Advanced Computing, China
\textsuperscript{3} National Supercomputer Center in Guangzhou, China

Correspondence: Xingbo Wang, Department of Mechatronic Engineering, Foshan University, No.18, Jiangwanyi Road, Foshan, China. E-mail: xbwang@fusu.edu.cn & 153668@qq.com

Received: August 7, 2017    Accepted: October 30, 2017    Online Published: November 14, 2017

doi:10.5539/jmr.v10n1p1    URL: https://doi.org/10.5539/jmr.v10n1p1

Abstract

The article makes an investigation on the probability of finding the greatest common divisor between a given integer and a hidden integer that lies in an integer interval. It shows that, adding the integers that are picked randomly in the interval results in a much bigger probability than subtracting the picked integers one with another. Propositions and theorems are proved and formulas to calculate the probabilities are presented in detail. The research is helpful in developing probabilistic algorithm of integer factorization.

Keywords: probability, greatest Common Divisor, integer factorization, probabilistic algorithm

1. Introduction

Given an odd integer \( N \) that has the greatest common divisor (GCD) \( d \) with another odd integer \( e \) in a large odd interval that consists in consecutive odd numbers, what is the probability to find out \( d \) ? This question is close to the problem of integer factorization, as investigated in (Xingbo Wang, 2017(1)), (Jianhui Li, 2017), (Dongbo Fu, 2017) and (Xingbo Wang, 2017(2)). As stated in (Xingbo Wang, 2017(3)) the answer to the question relies on a detail study on the properties of odd integers on an odd interval. This article makes an investigation on the probability of finding the integer \( e \) in a large odd interval. The research shows that, by adding two or more terms contained in the interval, it has a very big probability to find out \( d \).

2. Preliminaries

2.1 Symbols and Notations

In this whole article, an odd interval \([a, b]\) is a set of consecutive odd numbers that take \( a \) as lower bound and \( b \) as upper bound, for example, \([3, 11]\) = \([3, 5, 7, 9, 11]\). Symbol \([x]\) is the floor function of real number \( x \) that satisfies \([x] \leq x < [x] + 1\) . Symbol \( P(n, o) \) or \( P(S, o) \) is to express the total probability to find successfully an objective number \( o \) in a set \( S \) consisting in \( n \) terms and symbol \( P_k(S, o) \) is the probability to find successfully \( o \) by picking randomly \( k \) terms in set \( S \). Symbol \( \left( \begin{array}{c} n \\end{array} \right) \) is the binomial coefficient defined by

\[
\left( \begin{array}{c} n \\end{array} \right) = \frac{n!}{k!(n-k)!}
\]

2.2 Lemmas

Lemma 1(See in Xingbo Wang, 2016 & 2017(3))  
Let \( p \) be a positive odd integer; then among \( p \) consecutive positive odd integers there exists one and only one that can be divisible by \( p \). Let \( q \) be a positive odd number and \( S \) be a finite set that is composed of consecutive odd integers; then \( S \) needs at least \((n-1)q+1\) elements to have \( n \) multiples of \( q \). If \( s_0 \in S \) is a multiple of \( o \), then so is \( s_{o+1} \). If \( S \) consists in \( n \) consecutive odd integers and odd number \( o \) has one and only one multiple in \( S \), then \( o \geq \left[ \frac{n}{2} \right] \).

Lemma 2(See in Xingbo Wang, 2017(3))  
Let \( n \) be a positive integer, \( p \) be a odd integer and \( S = \{s_1, s_2, \cdots, s_n\} \) be a sequence that consists in \( n \) consecutive odd integers in which there is one and only one term divisible by \( p \); if \( n > p \) and \( S^* = \{s_i - s_j | 1 \leq i < j \leq n\} \); then there are \( v(p, n) = an - \frac{a(a+1)}{2}p \) \( p \)'s multiples in \( S^* \), where \( a = \left\lfloor \frac{n}{p} \right\rfloor \).

Lemma 3(See in Klambauer G,1979, p52)  
Suppose \( m, n \) are positive integers with \( m < n \); then

\[
\sum_{k=1}^{m} \frac{m!(n-k)!}{n!(n-k)!} = \frac{m}{n-m+1}
\]
3. Main Results and Proofs

Proposition 1 Let \( p \) be an odd number and \( n \) be a positive integer with \( n \leq p \); suppose \( s_1, s_2, \ldots, s_n \) are \( n \) consecutive odd numbers and \( S = \{s_j - s_i \mid 1 \leq i < j \leq n\} \); then there is not a \( p \)'s multiple in \( S \).

Proof. Referring to the proof of Theorem 2 in (Xingbo Wang, 2017(3)) immediately yields the conclusion.

Proposition 2 Suppose \( N \) is a composite odd integer and \( n \) is a positive integer; let \( S = \{s_1, s_2, \ldots, s_n\} \) be a set that consists in \( n \) consecutive odd integers that satisfy, for an \( m \) with \( 1 \leq m \leq n \), \( \text{GCD}(N, s_i) = \{d > 1, i = m ; 1, i \neq m\} \); then \( P_1(S, d) = \frac{1}{n} \).

Proof. (Omitted)

Proposition 3 Let \( N \) be an odd number and \( n \) be a positive integer. Suppose \( S = \{s_1, s_2, \ldots, s_n\} \) are \( n \) consecutive odd integers that satisfy, for an \( m \) with \( 1 \leq m \leq n \), \( \text{GCD}(N, s_i) = \{d > 1, i = m ; 1, i \neq m\} \); then \( \alpha = \left[ \frac{n}{d} \right] \) ways to compute \( d \) by adding two or more terms in \( S \).

Proof. By Lemma 1 it knows \( n \leq d \) because \( s_1 \) is the unique term that is a multiple of \( d \). Let \( s_1 = ds \) with an odd integer \( s \); then \( S \) can be rewritten by

\[
S = \{ds, ds + 2, \ldots, ds + 2(n-1)\}, n \leq d
\]

Considering \( ds + 2i + ds + 2(d - i) = 2d(s + 1) \) is a multiple of \( d \), \( d \leq n \) and adding three or more terms may also lead to \( d \)'s multiples, one knows that there are more than \( \left[ \frac{n}{d} \right] \) + 1 ways to compute \( d \).

Example 1 Let \( N = 15 \) and \( S = \{25, 27, 29, 31, 33\} \); then \( 27 + 33 = 60, 29 + 31 = 60, 25 + 29 + 31 = 85, 25 + 27 + 33 = 85, 27 + 29 + 31 = 120 \) and \( 25 + 27 + 29 + 31 + 33 = 145 \) are all multiples of 5.

Theorem 2 Suppose \( N \) is a composite odd integer and \( n \) is a positive integer; let \( S = \{s_1, s_2, \ldots, s_n\} \) be a set that consists in \( n \) consecutive odd integers that satisfy, for an \( m \) with \( 1 < m < n \), \( \text{GCD}(N, s_i) = \{d > 1, i = m ; 1, i \neq m\} \); then there are totally \( 2^{n-m} \) ways to compute \( d \) if \( m < n < 2m - 1 \), and there are totally \( 2^{m-1} \) ways to compute \( d \) if \( n = 2m - 1 \).

Proof. When \( 1 < m < n \), there always are terms on the left and the right of \( s_m \). For convenience, the terms on the left are called left-terms and the terms on the right are called right-terms. Since \( S \) consists in \( n \) consecutive odd integers, when \( n \geq 2m - 1 \) it knows that, there are \( m - 1 \) left-terms and there are at least \( m - 1 \) right-terms. This time, it yields the following facts:

1. There are \( \binom{m-1}{1} = m - 1 \) ways to obtain \( 2s_m \) by choosing \( s_m-j \) and \( s_m+j \) that fit \( s_m-j + s_m+j = 2s_m \) with \( j = 1, 2, \ldots, m-1 \).
2. There are \( \binom{m-1}{2} = \frac{(m-1)(m-2)}{2} \) ways to obtain \( 4s_m \). In fact, choosing two consecutive terms, \( s_m-j \) and \( s_m-j-1 \), and another two consecutive terms, \( s_m+j \) and \( s_m+j+1 \), will result in \( 4s_m \) with \( j = 1, 2, \ldots, m-2 \). Likewise, arbitrary symmetrically-distributed four terms, say \( s_{m-k}, s_{m-l}, s_{m+k} \) and \( s_{m+l} \), lead to \( 4s_m \). Hence there are \( \binom{m-1}{2} = \frac{(m-1)(m-2)}{2} \) ways to obtain \( 4s_m \).
3. Similarly, there are \( \binom{m-1}{k} \) ways to obtain \( 2ks_m \). For example, choosing \( k \) terms \( s_{m-j}, s_{m-j+1}, \ldots, s_{m-j+k} \) and their respective symmetric terms \( s_{m+j}, s_{m+j+1}, \ldots, s_{m+j+k} \) yields \( s_{m-j} + \ldots + s_{m-j+k} + s_{m-j+k} + s_{m+j+k} = 2ks_m \) with \( j = 1, \ldots, m-k-1 \). Actually, adding arbitrary \( k \) left-terms and their respectively symmetric right-terms leads to \( 2ks_m \).

4. There is one way to obtain \( s_m \), that is, to choose \( s_m \) itself.

Consequently, the total ways of picking one to \( k \) left-terms, denoted by \( \Lambda_k \), that can produce multiples of \( s_m \) are calculated by

\[
\Lambda_k = \binom{m-1}{1} + \binom{m-1}{2} + \cdots + \binom{m-1}{k} + 1 \tag{4}
\]

Note that

\[
\Lambda_k = \binom{m-1}{0} + \binom{m-1}{1} + \binom{m-1}{2} + \cdots + \binom{m-1}{k} \tag{5}
\]

it yields when \( k = m-1 \)

\[
\Lambda_{m-1} = \binom{m-1}{0} + \binom{m-1}{1} + \binom{m-1}{2} + \cdots + \binom{m-1}{k} = 2^{m-1}
\]

Now consider the case \( m \leq n < 2m-1 \). There are \( m-1 \) left-terms but there are merely \( n-m \) right-terms, as shown in figure 1.

![Figure 1. Distribution of terms around \( s_m \)](http://jmr.ccsenet.org.

It knows that, this time there are at most \( n-m \) consecutive right-terms and hence it yields

\[
\Lambda_{n-m} = \binom{n-m}{1} + \binom{n-m}{2} + \cdots + \binom{n-m}{n-m} + 1 \tag{6}
\]

which results in

\[
\Lambda_{n-m} = 2^{n-m} \tag{7}
\]

**Proposition 5** Suppose \( N \) is a composite odd integer and \( n \) is a positive integer; let \( S = \{s_1, s_2, \ldots, s_n\} \) be a set that consists in \( n \) consecutive odd integers that satisfy \( \gcd(N, s_i) = \{ d > 1, i = n \ \text{or} \ \gcd(N, s_i) = \{ d > 1, i = 1, i \neq n \} \) \) then by picking terms in \( S \) and adding the picked terms together the probability \( P(S, d) \) is bigger than \( \frac{1}{n-1} \).

**Proof.** Referring to Theorem 1 and its proof, it holds

\[
P(S, d) > \left(\frac{n}{2}\right) + 1 > \frac{n}{2(n-1)} = \frac{1}{n-1}
\]

**Proposition 6** Suppose \( N \) is a composite odd integer and \( n \) is a positive integer; let \( S = \{s_1, s_2, \ldots, s_n\} \) be a set that consists in \( n \) consecutive odd integers that satisfy , for an \( m \) with \( 1 < m < n \), \( \gcd(N, s_i) = \{ d > 1, i = m \ \text{or} \ \gcd(N, s_i) = \{ d > 1, i \neq m \} \) \) then by picking terms in \( S \) and adding the picked terms together the probability \( P(S, d) \) is \( \frac{1}{n} + \frac{m}{nm+1} \) when \( m < n < 2m-1 \); when \( n = 2m-1, P(2m-1, d) \geq \frac{2}{3} \).

**Proof.** Referring to the proofs of Theorem 1 and Theorem 2 and considering the case \( 1 < n - m \leq m - 1 \), one can see that, picking an arbitrary right-term or one more arbitrary right-terms among \( s_{m+1}, s_{m+2}, \ldots, s_n \) can always result in one or one more multiples of \( s_m \) by adding the picked terms with their symmetric left-terms. Hence the whole events can be considered as follows.
1. Choose randomly 1 term in \( S \), the probability \( P_1(S, d) \) that a right-term \( s_{m+j} \) with \( j = 1, 2, ..., n-m \) is picked is calculated by

\[
P_1(S, d) = \binom{n-m}{1} \frac{1}{n} = \frac{(n-1)(n-m)!}{n!(n-m-1)!}, 1 \leq n-m
\]

(8)

2. Choose randomly 2 terms in \( S \), the probability \( P_2(S, d) \) that two right-terms are picked is calculated by

\[
P_2(S, d) = \binom{n-m}{2} \frac{1}{binomn2} = \frac{(n-2)(n-m)!}{n!(n-m-2)!}, 2 \leq n-m
\]

(9)

3. Choose randomly \( k \) terms in \( S \), the probability that \( k \) right-terms are picked is calculated by

\[
P_k(S, d) = \binom{n-m}{k} \frac{1}{n} = \frac{(n-k)(n-m)!}{n!(n-m-k)!}, k \leq n-m
\]

(10)

4. Choosing \( s_m \) to obtain \( d \) leads to a probability \( P_1(S, d) = \frac{1}{n} \).

Consequently, the total probability \( P(S, d) \) is given by

\[
P(S, d) = \frac{1}{n} + \sum_{1 \leq k \leq n-m} P_k(S, d) = \frac{1}{n} + \sum_{1 \leq k \leq n-m} \frac{(n-k)(n-m)!}{n!(n-m-k)!}
\]

(11)

By Lemma 3, it yields

\[
P(S, d) = \frac{1}{n} + \sum_{1 \leq k \leq n-m} \frac{(n-k)(n-m)!}{n!(n-m-k)!} = \frac{1}{n} + \frac{n-m}{m+1}
\]

(12)

When \( n = 2m-1 \), it yields

\[
P(S, d) = \frac{1}{n} + \frac{n-1}{n+3} = 1 - \frac{3}{n} + \frac{12}{n(n+3)}
\]

(13)

Since \( m \geq 1 \), \( n \)'s minimal value is 3; hence

\[
\min(P(S, d)) = \frac{2}{3}
\]

(14)

In fact, let \( f(x) = \frac{1}{3} - \frac{12}{(x+3)(x+4)} \); then \( f'(x) = \frac{3(x-3)(x+1)}{2(x+3)^2} \) and \( f''(x) = \frac{6(x^3-3x^2-9x-9)}{x^2(x+3)^3} \). Since \( f'(3) = 0 \) and \( f''(3) = -\frac{47}{27} \), \( f(x) \) reaches its mmaximal value \( \frac{1}{3} \) when \( x = 3 \), and thus \( 1 - f(x) \) reaches its minimal value \( \frac{2}{3} \).

4. Conclusions and Future Work

Finding the GCD between a given integer and a hidden integer in the sequence of integers plays an important role in solving the problems of factoring integers. There was successful approach like famous Pollards rho algorithm (Eric Bach, 1991) that applies integer minus another one to increase the probability of computing the GCD. Pollards rho method being applied on an infinite integer interval, its probability does increase to a relative big quantity. However, when the interval that contains the objective integer is limited to a finite range, as analyzed in (Xingbo Wang, 2017(1)), the subtraction of two terms in the interval cannot increase the probability as expected, as proved in Proposition 2. By inequality (1) in Proposition 2, when the GCD \( d \) is bigger enough, \( \alpha = \frac{3}{2} \) tends to be zero and the probability also tends to be zero. On the contrary, adding the terms in the interval, as Theorem 1 and Theorem 2 say, can increase a lot of probability. According to Proposition 6, if the objective number lies near the middle of the interval, the probability increase with \( n \)'s increasing. These results provide a different way from the classical method of subtraction. This is the future work that is required to do: developing new probabilistic algorithm. Hope to see better achievements.

Acknowledgements

The research work is supported by the State Key Laboratory of Mathematical Engineering and Advanced Computing under Open Project Program No.2017A01, Department of Guangdong Science and Technology under projects 2015A030401105 and 2015A010104011, Foshan Bureau of Science and Technology under projects 2016AG100311, Projects 2014SFKC30 and 2014QTLXXM42 from Guangdong Education Department, Project gg040981 from Foshan University. The authors sincerely present thanks to them all.
References


Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).