#A65 INTEGERS 13 (2013)

INDEPENDENT DIVISIBILITY PAIRS ON THE SET OF INTEGERS FROM 1 TO N

Rosemary Sullivan
Dept. of Mathematics, West Chester University, West Chester, Pennsylvania
rsullivan@wcupa.edu

Neil Watling
Department of Mathematics, Widener University, Chester, Pennsylvania
nawatling@widener.edu

Received: 6/2/12, Revised: 8/6/13, Accepted: 8/30/13, Published: 9/26/13

Abstract

Let $\mathbb{S}_N$ denote the set of integers from 1 up to N and A_i be the event that a number selected from $\mathbb{S}_N$ is divisible by i. For the sample space $\mathbb{S}_N$, with the uniform probability measure, consider the question of the independence of the events A_i and A_j, $i \neq j$. We determine a characterization in terms of N, i and j. Using this we consider various situations and supplementary questions.

1. Introduction

Statistical independence is a fundamental concept in probability theory. Two events A and B are statistically independent if $P(A)P(B) = P(A \cap B)$; see Chung [3], Kac [5] or Ross [8] for example. Probabilists often illustrate statistical independence with examples from games of chance, whereas number theorists like to demonstrate the property using sets of positive integers. The natural density measure of a set A of natural numbers [2, 3, 5] is defined as,

$$D(A) = \lim_{n\to\infty} \frac{|A \cap \{1, 2, \ldots, n\}|}{n},$$

where $|\cdot|$ denotes the cardinality of a set. For example, if $A = \{3, 6, 9, \ldots\}$ then $D(A) = \frac{1}{3}$. If the "divisibility event", A_{q_1}, is the set of natural numbers divisible by q_1 then $D(A_{q_1})D(A_{q_2}) = D(A_{q_1} \cap A_{q_2})$ if and only if q_1 and q_2 are coprime. Here $D(A_{q_1}) = 1/q_1$, $D(A_{q_2}) = 1/q_2$ and $D(A_{q_1} \cap A_{q_2}) = D(A_{q_1 q_2}) = 1/q_1 q_2$. In this sense the "event of divisibility by q_1" and the "event of divisibility by q_2" are independent on the set of natural numbers if and only if q_1 and q_2 are coprime.

However, it should be noted that $D(A)$ is not a probability measure since it is not countably additive [2]; so this is not true statistical independence.

Suppose instead we restrict to a finite set $\mathbb{S}_N = \{1, \ldots, N\}$, where N is a natural number, with the uniform probability measure. We can analogously define A_{q_i} to be the event that a number selected from $\mathbb{S}_N$ is divisible by q_i. Now which events A_{q_1} and A_{q_2} are independent? This is more complicated to answer since we have the additional parameter N. As a simple illustration, consider the events A_2 and A_3. It is a straightforward computation to show these events are not independent on $\mathbb{S}_{10}$, even though 2 and 3 are coprime, but are independent on $\mathbb{S}_{12}$.

As another illustration if $N = 24$, $P(A_3) = 8/24$, $P(A_5) = 4/24$ and $P(A_3 \cap A_5) = P(A_{15}) = 1/24 \neq 8/24 \cdot 4/24$. So once again being coprime is not sufficient but it is, as we shall see, necessary. Clearly the cardinality of the set, N, plays a role. Perhaps we should require q_1 and q_2 to be coprime factors of N. This is in fact not necessary. For example, if we keep $N = 24$ but consider A_3 and A_7 then $P(A_3) = 8/24$, $P(A_7) = 3/24$ and $P(A_3 \cap A_7) = P(A_{21}) = 1/24 = 8/24 \cdot 3/24$ so these two events are independent.

Define A_{q_1} and A_{q_2} to be an *independent divisibility pair on* $\mathbb{S}_N$ if $P(A_{q_1})P(A_{q_2}) = P(A_{q_1} \cap A_{q_2})$ where $P(\cdot)$ is the uniform distribution on $\mathbb{S}_N$. Theorem 2.1 gives the necessary and sufficient conditions on N, q_1 and q_2 for A_{q_1} and A_{q_2} to be an independent divisibility pair on $\mathbb{S}_N$.

With the conditions for independence known, it is possible to consider various supplementary questions. If we fix q_1 and q_2, for which N are the events A_{q_1} and A_{q_2} an independent divisibility pair on $\mathbb{S}_N$? If instead we fix N, which pairs (A_{q_1}, A_{q_2}) are an independent divisibility pair on $\mathbb{S}_N$? These questions are examined in Sections 3 and 4 respectively. In Section 5, we analyze the question of characterizing those N which possess a specified number of independent divisibility pairs. In particular, we prove that if N is the product of a Sophie Germain prime and its corresponding safe prime then there is a unique independent divisibility pair given by these two primes. We conclude in Section 6 with some conjectures and generalizations.

2. The Necessary and Sufficient Conditions for A_{q_1} and A_{q_2} to be Independent on $\mathbb{S}_N$

Theorem 2.1. *Let N be a natural number and $\mathbb{S}_N = \{1, \ldots, N\}$. Let q_1 and q_2 be two natural numbers, with $1 < q_1 < q_2 < N$, and let A_{q_i} be the event that a number selected from $\mathbb{S}_N$ is divisible by q_i. The events A_{q_1} and A_{q_2} are independent on $\mathbb{S}_N$ if and only if q_1 and q_2 are coprime and*

$$N = tq_1q_2 + r_1q_1 = q_1(tq_2 + r_1), \tag{1}$$

with t a natural number and r_1 a non-negative integer such that $r_1q_1 < q_2$.

Proof. Assume q_1 and q_2 are coprime and satisfy (1). It is straightforward to compute the respective probabilities for A_{q_1}, A_{q_2} and the intersection:

$$P(A_{q_1}) = \frac{tq_2 + r_1}{N} = \frac{1}{q_1}, \qquad P(A_{q_2}) = \frac{tq_1}{N} = \frac{t}{tq_2 + r_1},$$

$$P(A_{q_1} \cap A_{q_2}) = P(A_{q_1 q_2}) = \frac{t}{N} = \frac{t}{q_1(tq_2 + r_1)},$$

confirming they are indeed independent.

Now assume $1 < q_1 < q_2 < N$, A_{q_1} and A_{q_2} are independent on $\mathbb{S}_N$ and $N \geq \operatorname{lcm}(q_1, q_2)$, the least common multiple of q_1 and q_2. (If $q_2 < N < \operatorname{lcm}(q_1, q_2)$ then the intersection event is empty. So $P(A_{q_1})P(A_{q_2}) \neq P(A_{q_1} \cap A_{q_2})$ since the left-hand side is nonzero whereas the right-hand side is zero. If $q_1 < N < q_2$ then A_{q_2} is empty, so the events A_{q_1} and A_{q_2} are degenerately independent but the conditions on q_1, q_2 and N have been violated.) By the definition of independence,

$$\left\lfloor \frac{N}{q_1} \right\rfloor \cdot \left\lfloor \frac{N}{q_2} \right\rfloor = N \cdot \left\lfloor \frac{N}{\operatorname{lcm}(q_1, q_2)} \right\rfloor. \tag{2}$$

If we suppose $2 \leq \gcd(q_1, q_2)$, the greatest common divisor of q_1 and q_2, then

$$\left\lfloor \frac{N}{q_1} \right\rfloor \cdot \left\lfloor \frac{N}{q_2} \right\rfloor \leq \frac{N}{q_1} \cdot \frac{N}{q_2} = N \cdot \frac{N}{q_1 q_2} < N \cdot \left\lfloor \frac{N \gcd(q_1, q_2)}{q_1 q_2} \right\rfloor = N \cdot \left\lfloor \frac{N}{\operatorname{lcm}(q_1, q_2)} \right\rfloor$$

which contradicts (2), so q_1 and q_2 must be coprime. Hence $\operatorname{lcm}(q_1, q_2) = q_1 q_2$ and (2) becomes

$$\left\lfloor \frac{N}{q_1} \right\rfloor \cdot \left\lfloor \frac{N}{q_2} \right\rfloor = N \cdot \left\lfloor \frac{N}{q_1 q_2} \right\rfloor. \tag{3}$$

We may assume $N = t(q_1 q_2) + r$ where $t \geq 1$ and $0 \leq r < q_1 q_2$, and also $r = r_1 q_1 + s_1$ or $r = r_2 q_2 + s_2$ with $r_1, r_2 \geq 0$, $0 \leq s_1 < q_1$, $0 \leq s_2 < q_2$. Then (3) gives

$$(tq_2 + r_1) \cdot (tq_1 + r_2) = (t(q_1 q_2) + r) \cdot t,$$

which can be simplified to

$$r_2(tq_2 + r_1) = ts_1. \tag{4}$$

Now if $r_2 \geq 1$, then $r_2(tq_2 + r_1) \geq tq_2 > tq_1 > ts_1$ giving a contradiction to (4), so $r_2 = 0$. Hence, from (4), $s_1 = 0$, and thus, $r = r_1 q_1 = s_2 < q_2$. So $N = t(q_1 q_2) + r_1 q_1$ where $t \geq 1$ and $0 \leq r_1 q_1 < q_2$ as required. □

It should be noted that Theorem 2.1 is a special case of Theorem 1 in [4]. Their requirement that N is composite is clearly true here but we expect a more restrictive condition since we have very specific events.

Observe (1) is satisfied if N is a multiple of two coprime numbers q_1 and q_2. However N need only be a multiple of the smaller number q_1, not the larger q_2. Conversely, if q_1 is not a divisor of N the events A_{q_1} and A_{q_2} are dependent on $\mathbb{S}_N$.

3. The Natural Density Measure of Independent A_{q_1} and A_{q_2}

Recall A_2 and A_3 are independent on $\mathbb{S}_N$ if $N = 12$ but not if $N = 10$. In fact, given Theorem 2.1, it is clear A_2 and A_3 are independent on $\mathbb{S}_N$ if and only if $N \equiv 0$ or $2 \mod 6$. That is, "one third of the time they are independent" and "two thirds of the time they are dependent." More precisely if we define $I_{q_1,q_2} = \{N : A_{q_1}, A_{q_2} \text{ are independent on } \mathbb{S}_N\}$ then,

$$D(I_{2,3}) = \lim_{n\to\infty} \frac{|I_{2,3} \cap \{1, 2, \ldots, n\}|}{n} = \frac{1}{3}.$$

Below are two more examples using Theorem 2.1.

Example 3.1. If $q_1 = 5$ and $q_2 = 53$ then the events A_5 and A_{53} are independent on $\mathbb{S}_N$ if and only if $N \equiv 0, 5, 10, \ldots, 50 \mod 265$. So $D(I_{5,53}) = 11/265 \approx 0.0415$.

Example 3.2. If $q_1 = 7$ and $q_2 = 510$ then the events A_7 and A_{510} are independent on $\mathbb{S}_N$ if and only if $N \equiv 0, 7, 14, \ldots, 504 \mod 3570$. So $D(I_{7,510}) = 73/3570 \approx 0.020448$.

In general, for q_1 and q_2 coprime, if $q_2 = pq_1 + r$, $1 \le r \le q_1 - 1$, then the events A_{q_1} and A_{q_2} are independent on $\mathbb{S}_N$ if and only if $N \equiv 0, q_1, 2q_1, \ldots, pq_1 \mod q_1q_2$. So $D(I_{q_1,q_2}) = (p+1)/(q_1q_2)$. Using the inequalities,

$$\begin{aligned} 0 < \frac{1}{q_1^2} < \frac{1}{q_1(q_1 - \frac{1}{p+1})} &= \frac{p+1}{q_1((p+1)q_1 - 1)} \le \frac{p+1}{q_1q_2} \\ &= \frac{p+1}{q_1(pq_1 + r)} \le \frac{p+1}{q_1(pq_1 + 1)} \le \frac{2}{q_1(q_1+1)} \le \frac{1}{3}, \end{aligned}$$

we can construct a table of intervals for possible values of $D(I_{q_1,q_2})$:

q_1	2	3	4	5	6	7	q_1
$D(I_{q_1,q_2})$	$(\frac{1}{4}, \frac{1}{3}]$	$(\frac{1}{9}, \frac{1}{6}]$	$(\frac{1}{16}, \frac{1}{10}]$	$(\frac{1}{25}, \frac{1}{15}]$	$(\frac{1}{36}, \frac{1}{21}]$	$(\frac{1}{49}, \frac{1}{28}]$	$(\frac{1}{q_1^2}, \frac{2}{q_1(q_1+1)}]$

If we fix q_1 and let q_2 get large, then p gets large, and we approach from above the left-hand end point of the interval, $1/q_1^2$.

Conversely, for $q_2 = q_1 + 1$ we have $p = 1$, $r = 1$ and the density becomes $\dfrac{2}{q_1(q_1+1)}$, with the maximum value of $\frac{1}{3}$ occurring when $q_1 = 2$. So consecutive pairs have the more explicit table of densities:

(q_1, q_2)	$(2,3)$	$(3,4)$	$(4,5)$	$(5,6)$	$(6,7)$	$(7,8)$	(q_1, q_1+1)
$D(I_{q_1,q_2})$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{10}$	$\frac{1}{15}$	$\frac{1}{21}$	$\frac{1}{28}$	$\frac{2}{q_1(q_1+1)}$

In both tables these densities are larger than one would intuitively expect. For the latter it appears to be double, and for the former the density is always larger than $1/q_1^2$.

4. The Independent Divisibility Pairs on $\mathbb{S}_N$ for a Particular N

If we choose a particular N, then which events A_{q_1} and A_{q_2} are an independent divisibility pair on $\mathbb{S}_N$? Equivalently, for which pairs of coprime natural numbers q_1, q_2 do N, q_1 and q_2 satisfy (1)?

Since $N = q_1(tq_2 + r)$ and $r \geq 0$, we have $q_2 \leq \dfrac{N}{q_1 t}$. Now $rq_1 < q_2$ so $N < q_1 t q_2 + q_2$, hence $q_2 > \dfrac{N}{q_1 t + 1}$. Thus, bounds for q_2 in terms of N, t and q_1 are:

$$\frac{N}{q_1 t + 1} < q_2 \leq \frac{N}{q_1 t}. \tag{5}$$

As $N = q_1(tq_2 + r)$ and $q_2 \geq q_1 + 1$, $t = \dfrac{1}{q_2}\left(\dfrac{N}{q_1} - r\right) \leq \dfrac{N}{q_1(q_1+1)}$, so bounds for t are:

$$1 \leq t \leq \frac{N}{q_1(q_1+1)}. \tag{6}$$

Also, $N \geq q_1(q_1+1)$ and consequently $q_1 \leq \frac{\sqrt{4N+1}-1}{2}$. Hence the bounds on q_1, a factor of N, are:

$$2 \leq q_1 \leq \frac{\sqrt{4N+1}-1}{2}. \tag{7}$$

Definition 4.1. Define (q_1, q_2) to be a valid pair on N if q_1, q_2 and N are natural numbers satisfying (1) and its conditions.

If (q_1, q_2) is a valid pair on N then the events A_{q_1} and A_{q_2} are an independent divisibility pair on $\mathbb{S}_N$.

Example 4.2. When $N = 24$ there are 8 valid pairs: $(2,3)$, $(2,5)$, $(2,9)$, $(2,11)$, $(3,4)$, $(3,7)$, $(3,8)$ and $(4,5)$.

The bounds (5), (6), and (7) enable us to write an algorithm that can be used to generate all possible valid pairs for a particular N. Start at $q_1 = 2$ and proceed through all possible values for q_1 given by (7). For each q_1 check if it is a factor of N. If so, use (6) to compute the range of values for t and (5) to compute the range of values for potential q_2. Now just increment through all these potential q_2's and at each stage check if the potential q_2 is coprime to q_1. If so then this value for q_2 would make (q_1, q_2) a valid pair on N.

Figure 1 has some annotated lines of pseudocode showing this algorithm.

Definition 4.3. Define the function $v(N)$ to be the number of valid pairs on N.

The algorithm in Figure 1 can be used to determine the values of $v(N)$. Figure 2 has a table listing N and $v(N)$ for $N \leq 100$.

```
Input N
count=0;
for q1=2 to (√(4 * N + 1) − 1)/2,                                  % from (7)
  if N mod q1 == 0,                              % check q1 is a factor of N
   for t=1 to N/(q1*(q1+1)),                                       % from (6)
    for potentialq2 = ceiling(N/(t*q1+1)) to N/(t*q1),             % from (5)
     if gcd(q1,potentialq2) == 1,                    % check q1, q2 coprime
      print(q1,potentialq2), count++                % if so then valid pair
return(count)                                       % number of valid pairs
```

Figure 1: Pseudocode to generate all valid pairs on N.

5. Bounds on $v(N)$ and the Characterization of those N with $v(N) < 5$

If we allow q_1, q_2 and N to vary, is there a characterization for those N with a particular number of valid pairs? Can we characterize those N for which $v(N) = 0$ or 1 for example? To begin we will give some simple bounds for $v(N)$.

Theorem 5.1. *Let $N = p_1^{n_1} p_2^{n_2} \dots p_r^{n_r}$ be the prime decomposition of N and $v(N)$ be the number of valid pairs on N. Then upper and lower bounds for $v(N)$ are respectively,*

$$v(N) < \frac{N \ln N}{2}, \tag{8}$$

and

$$v(N) \geq \frac{[(2n_1+1)\dots(2n_r+1)] - 2[(n_1+1)\dots(n_r+1)] + 1}{2}. \tag{9}$$

Proof. From (1) we have $q_2 = \frac{(N/q_1)-r_1}{t}$. But $t \geq 1$ and $r_1 \geq 0$, so $q_1 < q_2 \leq N/q_1$ and hence, for a given factor q_1 of N, the maximum number of valid q_2's is $(N/q_1) - q_1$. Since q_1 is a factor of N between 2 and $\lfloor\sqrt{N}\rfloor$, we have

$$v(N) < \sum_{q_1=2}^{\lfloor\sqrt{N}\rfloor} \left(\frac{N}{q_1} - q_1\right) < \sum_{q_1=2}^{\lfloor\sqrt{N}\rfloor} \frac{N}{q_1} < N \ln(\sqrt{N}) = \frac{N \ln N}{2}.$$

If we consider (1) and fix $r_1 = 0$, the number of choices for coprime factors q_1 and q_2, with $1 < q_1 < q_2$, will clearly give a lower bound for $v(N)$.

We can view this counting problem in terms of boxes and colored balls instead of factors and primes as follows: the two coprime values q_1 and q_2 correspond to two different boxes and the r distinct primes correspond to r different colors with n_i balls of color i, $i = 1, \dots, r$. We now employ the inclusion-exclusion principle.

The q_i are coprime factors of N so we do not need to use all the balls, but we cannot have the same colored ball in both boxes. For each color, i, there are

N	$v(N)$	N	$v(N)$	N	$v(N)$	N	$v(N)$	N	$v(N)$
1	0	21	1	41	0	61	0	81	5
2	0	22	3	42	13	62	12	82	11
3	0	23	0	43	0	63	10	83	0
4	0	24	8	44	12	64	11	84	32
5	0	25	0	45	6	65	4	85	3
6	1	26	4	46	6	66	17	86	13
7	0	27	3	47	0	67	0	87	10
8	1	28	4	48	16	68	12	88	17
9	0	29	0	49	0	69	7	89	0
10	1	30	10	50	9	70	17	90	35
11	0	31	0	51	6	71	0	91	2
12	3	32	7	52	9	72	25	92	20
13	0	33	3	53	0	73	0	93	8
14	3	34	5	54	14	74	14	94	15
15	2	35	2	55	1	75	12	95	6
16	1	36	9	56	12	76	14	96	29
17	0	37	0	57	3	77	2	97	0
18	4	38	6	58	8	78	21	98	17
19	0	39	4	59	0	79	0	99	13
20	5	40	9	60	25	80	21	100	21

Figure 2: Values of $v(N)$ for $N \leq 100$

$(2n_i+1)$ ways to place some or none of the i colored balls into the two boxes. Hence for all r colors we have $(2n_1+1)\ldots(2n_r+1)$ total possibilities.

However, we cannot have an empty box, since neither q_1 nor q_2 can be 1; so we need to remove these possibilities from the above total. These situations correspond to putting the balls into only one box rather than two. Consequently, each color has (n_i+1) choices, and for all colors there are $(n_1+1)\ldots(n_r+1)$ possibilities. Now observe that either the q_1 box or the q_2 box could be empty, so multiply this latter number by 2. However, the case where both boxes are simultaneously empty has now been counted once but excluded twice, hence correct this by adding back one possibility.

Finally, only half the remaining pairs satisfy the requirement $q_1 < q_2$, so we divide the expression by two. □

The next two theorems give relatively simple characterizations for those N with exactly zero, one, two, three or four valid pairs.

Theorem 5.2. *N has no valid pairs, that is $v(N) = 0$, if and only if N is a prime or the square of a prime.*

Proof. It is clear that if N is prime then it is impossible to decompose N in the form specified by (1). (In fact it is known that the set $\mathbb{S}_N$ is a dependent set, so it has no independent events whatsoever [4].) If N has at least two distinct prime factors, then choose these as q_1 and q_2. With $r_1 = 0$ and $t = N/(q_1q_2)$ equation (1) is satisfied demonstrating there is at least one valid pair. Consequently the only remaining cases are powers of a single prime. If $N = p^n$, $n \geq 3$, choose $q_1 = p$ and $q_2 = p^{(n-1)} - 1$, then q_1 and q_2 are coprime with $q_1 < q_2$ and $N = q_1q_2 + q_1$. However, if $N = p^2$, where p is prime, we must have $q_1 = p$ in which case it is impossible to choose q_2 satisfying $q_2 > q_1$ and $q_1q_2 \leq N$ simultaneously. □

Once again this is more restrictive than the general situation. It is known that if N is prime then $\mathbb{S}_N$ is a dependent set and any two events are dependent. However, if N is the square of a prime, it is possible to have independent events. For example, if $N = 9$ there are no independent sets of our divisibility type, but $A = \{1, 3, 5\}$ and $B = \{5, 7, 9\}$ are clearly independent.

Theorem 5.3. *Let $v(N)$ denote the number of valid pairs on N. Then*

(a) $v(N) = 1$ if and only if $N = 6 = 2 \cdot 3$, $N = 8 = 2^3$, $N = 16 = 2^4$ or $N = p_1p_2$, with p_1, p_2 prime and $p_2 = 2p_1 + 1$. The unique valid pair is $(2, 3)$, $(2, 3)$, $(2, 7)$ and $(p_1, 2p_1 + 1)$ respectively.

(b) $v(N) = 2$ if and only if $N = p_1p_2$ with p_1, p_2 prime, $2 < p_1$, and $p_1 + 2 \leq p_2 \leq 2p_1 - 1$.

(c) $v(N) = 3$ if and only if $N = 12 = 2^2 \cdot 3$, $N = 14 = 2 \cdot 7$, $N = 22 = 2 \cdot 11$, $N = 27 = 3^3$, $N = 57 = 3 \cdot 19$, $N = p_1p_2$ with p_1, p_2 prime and $p_2 = 3p_1 + 2$, or $N = p_1p_2$ with p_1, p_2 prime and $p_2 = 4p_1 + 1$, $p_1 \geq 5$.

(d) $v(N) = 4$ if and only if $N = 18 = 2 \cdot 3^2$, $N = 26 = 2 \cdot 13$, $N = 28 = 2^2 \cdot 7$, $N = 39 = 3 \cdot 13$, $N = p_1p_2$ with p_1, p_2 prime, $p_1 \geq 5$, $2p_1 + 3 \leq p_2 \leq 3p_1 - 2$, or $N = p_1p_2$ with p_1, p_2 prime, $p_1 \geq 5$, $p_2 = 4p_1 + 3$, with $2p_1 + 1$ not divisible by 3.

Proof. If N is the product of at least three distinct prime factors, then (9) gives $v(N) \geq 6$. Hence, we can restrict to N having one or two distinct prime factors.

If $N = p^n$, from Theorem 5.2, $n \geq 3$ in which case $(p, p^{n-1} - c)$, $1 \leq c \leq p - 1$ are valid pairs. So we need only consider $p \leq 5$. If $p = 5$, then $(5, 5^{n-1} - 1)$, $(5, 5^{n-1} - 2)$, $(5, 5^{n-1} - 3)$, $(5, 5^{n-1} - 4)$, and $(5, \frac{5^{n-1}-1}{2})$ are five valid pairs. If $p = 3$, then $(3, 3^{n-1} - 1)$, $(3, 3^{n-1} - 2)$, $(3, 3^{n-1} - 4)$, $(3, 3^{n-1} - 5)$, and $(3, 3^{n-1} - 7)$ are five valid pairs if $n \geq 5$. For $N = 3^3 = 27$ and $N = 3^4 = 81$ there are three and five valid pairs respectively (see Figure 2). If $p = 2$, then $(2, 2^{n-1} - 1)$, $(2, 2^{n-1} - 3)$, $(2, 2^{n-1} - 5)$, $(2, 2^{n-1} - 7)$, and $(2, 2^{n-1} - 9)$ are five valid pairs if $n \geq 6$. It is easy to check that $(2, 3)$ and $(2, 7)$ are the unique valid pair for $N = 2^3 = 8$ and $N = 2^4 = 16$ respectively. The case $N = 2^5 = 32$ has seven valid pairs (see Figure 2).

If $N = p_1^{n_1}p_2^{n_2}$, $n_1, n_2 \geq 1$, from (9) $v(N) \geq n_1n_2$. In particular there are the n_1n_2 valid pairs given by (p_1^a, p_2^b) (possibly in reverse order) where $1 \leq a \leq n_1$,

$1 \leq b \leq n_2$. So we need only consider when $n_1 n_2 \leq 4$.

In the cases $(n_1, n_2) = (2, 2)$, $(n_1, n_2) = (4, 1)$, and $(n_1, n_2) = (1, 4)$ a fifth valid pair is given by $(p_1, p_1 p_2^2 - 1)$, $(p_1, p_1^3 p_2 - 1)$, and $(p_2, p_1 p_2^3 - 1)$ respectively.

If $(n_1, n_2) = (3, 1)$, then $(p_1, p_1^2 p_2 - 1)$ and $(p_1, p_1 p_2 - 1)$ give two additional valid pairs whereas for $(n_1, n_2) = (1, 3)$ they are $(p_2, p_1 p_2^2 - 1)$ and $(p_2, p_2^2 - 1)$.

If $(n_1, n_2) = (2, 1)$, then $(p_1, p_1 p_2 - c)$, with $1 \leq c \leq p_1 - 1$ are valid pairs. So there are at least four additional pairs if $p_1 \geq 5$. If $p_1 = 3$, then $(3, 3p_2 - 1)$, $(3, 3p_2 - 2)$, and $(3, \frac{3p_2 - 1}{2})$ are three additional valid pairs. If $p_1 = 2$, then $(2, 2p_2 - 1)$, $(2, 2p_2 - 3)$, and $(2, 2p_2 - 5)$ are three additional valid pairs if $p_2 \geq 11$. This leaves $N = 2^2 \cdot 3 = 12$, $N = 2^2 \cdot 5 = 20$ and $N = 2^2 \cdot 7 = 28$ which have 3, 5, and 4 valid pairs respectively (see Figure 2).

If $(n_1, n_2) = (1, 2)$, then $(p_2, p_1 p_2 - c)$, $1 \leq c \leq p_1 - 1$ are valid pairs. Hence there are at least four additional pairs if $p_1 \geq 5$. If $p_1 = 3$, then $(p_2, 3p_2 - 1)$, $(p_2, 3p_2 - 2)$, and $(p_2, \frac{3p_2 - 1}{2})$ are three additional valid pairs. If $p_1 = 2$, then $(2, p_2^2 - 2)$, $(2, p_2^2 - 4)$, and $(2, p_2^2 - 6)$ are three additional valid pairs if $p_2 \geq 5$. For $N = 2 \cdot 3^2 = 18$ there are 4 valid pairs (see Figure 2).

Hence we are reduced to the case where N is the product of two primes p_1 and p_2. First suppose that $p_1 = 2$ so $N = 2p_2$. Here $(2, p_2)$, $(2, p_2 - 2)$, $(2, p_2 - 4)$, $(2, p_2 - 6)$, and $(2, p_2 - 8)$ are valid pairs if $p_2 > 24$. So we are left with $p_2 = 3$, 5, 7, 11, 13, 17, 19, and 23. These have 1, 1, 3, 3, 4, 5, 6, and 6 valid pairs respectively (see Figure 2). Now suppose $p_1 = 3$, so $N = 3p_2$. In this case $(3, p_2)$ together with four of $(3, p_2 - 1)$, $(3, p_2 - 2)$, $(3, p_2 - 3)$, $(3, p_2 - 4)$, $(3, p_2 - 5)$ and $(3, p_2 - 6)$ are valid pairs provided $p_2 > 24$. This leaves $p_2 = 5$, 7, 11, 13, 17, 19 and 23. These have 2, 1, 3, 4, 6, 3 and 7 valid pairs respectively (see Figure 2).

Finally, assume $p_1 \geq 5$. So p_1 and p_2 are odd primes, $p_2 \neq c(p_1 + 1)$, $p_2 \neq cp_1$ and at most one of $p_2 - 1$, $p_2 - 2$, $p_2 - 3$, $p_2 - 4$ and $p_2 - 5$ is a multiple of p_1. Now suppose that (p_1, q_3) is a valid pair on $N = p_1 p_2$ and $(p_1, q_3) \neq (p_1, p_2)$. Then we must have $p_2 = tq_3 + r$ with $r \geq 1$, $t \geq 1$, q_3 coprime to p_1, and $q_3 > rp_1$. Hence $q_3 = \dfrac{p_2 - r}{t}$ and $p_2 > trp_1 + r$. So restrictions on the value of p_2 will generate restrictions on the possible values of r and t, which in turn give restrictions on the possible forms for q_3, and hence the possible valid pairs. Consequently, we have the following:

If $p_2 > 5(p_1 + 1)$, at least five of (p_1, p_2), $(p_1, p_2 - 1)$, $(p_1, (p_2 - 1)/2)$, $(p_1, p_2 - 2)$, $(p_1, p_2 - 3)$, $(p_1, p_2 - 4)$, and $(p_1, p_2 - 5)$ are valid pairs.

If $4(p_1 + 1) < p_2 < 5(p_1 + 1)$, then $p_2 - 1$ is coprime to p_1, so at least five of (p_1, p_2), $(p_1, p_2 - 1)$, $(p_1, (p_2 - 1)/2)$, $(p_1, p_2 - 2)$, $(p_1, p_2 - 3)$, and $(p_1, p_2 - 4)$ are valid pairs.

If $p_2 = 4p_1 + 3$, then (p_1, p_2), $(p_1, p_2 - 1)$, $(p_1, (p_2 - 1)/2)$, and $(p_1, p_2 - 2)$ are valid pairs. If $2p_1 + 1$ is divisible by 3, then $(p_1, (p_2 - 1)/3)$ is a fifth valid pair.

If $p_2 = 4p_1 + 1$, then (p_1, p_2), $(p_1, p_2 - 2)$, and $(p_1, p_2 - 3)$ are the only valid pairs.

If $3(p_1+1) < p_2 \leq 4p_1-1$, then (p_1, p_2), (p_1, p_2-1), $(p_1, (p_2-1)/2)$, (p_1, p_2-2), and (p_1, p_2-3) are valid pairs. If $2p_1-1$ is divisible by 3, then $(p_1, (p_2-1)/3)$ is a sixth valid pair.

If $p_2 = 3p_1+2$, then (p_1, p_2), (p_1, p_2-1), and $(p_1, (p_2-1)/2)$ are the only valid pairs.

If $2p_1+3 \leq p_2 \leq 3p_1-2$, then (p_1, p_2), (p_1, p_2-1), $(p_1, (p_2-1)/2)$, and (p_1, p_2-2) are the only valid pairs.

If $p_2 = 2p_1+1$, then (p_1, p_2) is the only valid pair.

If $p_1+2 \leq p_2 \leq 2p_1-1$, then (p_1, p_2) and (p_1, p_2-1) are the only valid pairs. □

6. Remarks, Conjectures, and Generalizations

Remark 6.1. Primes p such that $2p+1$ is also prime are called Sophie Germain primes with $2p+1$ called a safe prime. The numbers are listed in sequence A005384 in [6]. For more information see [7], [9], and [1]. It is interesting to note that the largest known Sophie Germain prime, found in April 2012, is $18543637900515 \cdot 2^{666667} - 1$ which has 200,701 digits. It is conjectured there are infinitely many such primes.

Remark 6.2. Primes p such that $3p+2$ or $4p+1$ are also prime do not appear to be named, but the numbers are listed in sequences A023208 and A023212 in [6]. (Primes of the form $(4n+1)$ can be called Pythagorean, [6, A002144], since they correspond to the length of the hypotenuse of a right triangle with integer sides.) Figure 3 contains a table of all N values less than 50000 with exactly three valid pairs, together with the corresponding p_1 and p_2 values.

N	p_1	p_2	N	p_1	p_2	N	p_1	p_2
12	2	3	689	13	53	18023	67	269
14	2	7	901	17	53	18881	79	239
22	2	11	1121	19	59	20833	83	251
27	3	-	1633	23	71	21389	73	293
33	3	11	2581	29	89	23941	89	269
57	3	19	4181	37	113	25043	79	317
85	5	17	5513	37	149	28421	97	293
161	7	23	5633	43	131	32033	103	311
203	7	29	7439	43	173	37733	97	389
533	13	41	10561	59	179	48641	127	383

Figure 3: Values of $N < 50000$ with $v(N) = 3$

Given the lower bound, (9), it is clear that N exist for which $v(N) > M$ for any non-negative integer M.

Conjecture 6.3. N exist such that $v(N) = M$ for any non-negative integer M.

We have used Mathematica to confirm values of N for M up to 20,000.

Definition 6.4. Define the function $r(N)$ to be the ratio of the number of valid pairs on N to the value of N, that is, $r(N) = v(N)/N$.

Conjecture 6.5. For all N, the ratio of the number of valid pairs on N to the value of N itself is less than one half. That is, $r(N) < 0.5$.

Note that the conjecture is equivalent to $v(N) < N/2$, which is a better upper bound for $v(N)$ than (8). We can confirm this is true for $N \leq 1,000,000$ with the two largest values of $r(N)$ being 0.46317 and 0.465922 when $N = 360360$ and $N = 720720$ respectively. Figure 4 contains a table with some larger values of N.

N	Prime Factorization	$v(N)$	$r(N)$
360,360	$2^3.3^2.5.7.11.13$	166,908	0.46317
720,720	$2^4.3^2.5.7.11.13$	335,799	0.465922
12,252,240	$2^4.3^2.5.7.11.13.17$	5,782,591	0.471962
232,792,560	$2^4.3^2.5.7.11.13.17.19$	111,124,547	0.477354
5,354,228,880	$2^4.3^2.5.7.11.13.17.19.23$	2,576,614,559	0.48123
26,771,144,400	$2^4.3^2.5^2.7.11.13.17.19.23$	12,955,084,847	0.48392
80,313,433,200	$2^4.3^3.5^2.7.11.13.17.19.23$	39,012,925,759	0.485758

Figure 4: Values of $r(N)$ for select N

These N values have the property that they are the smallest number containing all factors from 2 to m, where m =15, 16, 18, 22, 24, 26 and 28 respectively. The next number in this sequence would be $N = 2^4.3^3.5^2.7.11.13.17.19.23.29$, containing all factors from 2 to 30. (These are the distinct entries in the integer sequence, A003418 [6], where a_n is the least common multiple of 1 up to n.)

Conjecture 6.6. The only values of N for which $r(N) > 0.45$ are multiples of 420, that is, have factors 2 through 7 inclusive.

Once again we can confirm this for $N \leq 1,000,000$.

Definition 6.7. Define the function $a(N)$ to be the average number of valid pairs from 1 up to N, that is, $a(N) = \dfrac{\sum_{i=1}^{N} v(i)}{N}$.

Figure 5 shows a plot of the values of $\frac{a(N)}{N}$ for $N \leq 1,000,000$.

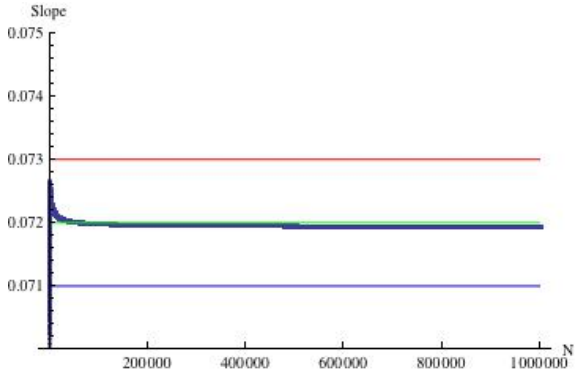

Figure 5: Plot of $\frac{a(N)}{N}$

Conjecture 6.8. $a(N) \approx 0.072N$ or $\sum_{i=1}^{N} v(i) \approx 0.072N^2$.

Definition 6.9. Define the function $\rho(N)$ to be the average ratio from 1 up to N, that is, $\rho(N) = \dfrac{\sum_{i=1}^{N} r(i)}{N}$.

Figure 6 shows a plot of the values of $\rho(N)$ for $N \leq 1,000,000$. The black curve corresponds to a model of the form $a\left(\dfrac{\ln(N)}{N}\right) + b$. The green line on the graph corresponds to 0.143878 the value of b given by the model.

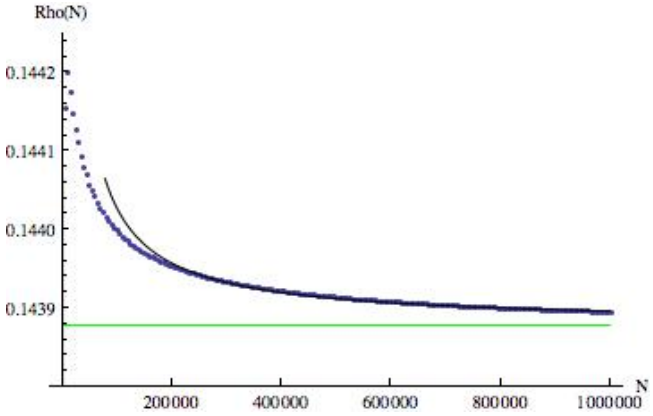

Figure 6: Plot of $\rho(N)$ and model $a\left(\dfrac{\ln(N)}{N}\right) + b$

Conjecture 6.10. The limiting value of $\rho(N)$ is between 0.14387 and 0.14389.

Remark 6.11. Statistical independence is also defined for more than two events, [3, 8], so it is possible to generalize these ideas for more than two divisibility events. In the case of three divisibility events we have the analog of Theorem 2.1:

Theorem 6.12. *Let N be a natural number and $\mathbb{S}_N = \{1, \ldots, N\}$. Let q_1, q_2, and q_3 be three natural numbers, with $1 < q_1 < q_2 < q_3 < N$, and let A_{q_i} be the event that a number selected from $\mathbb{S}_N$ is divisible by q_i. The events A_{q_1}, A_{q_2}, and A_{q_3} are independent on $\mathbb{S}_N$ if and only if q_1, q_2, and q_3 are coprime and*

$$N = tq_1q_2q_3 + r_1q_1q_2 = q_1q_2(tq_3 + r_1),$$

with t a natural number and r_1 a non-negative integer such that $r_1q_1q_2 < q_3$.

We can then consider the same questions as for pairs. The results, not surprisingly, are more complicated in this situation and are not presented here.

References

[1] Chris Caldwell, *The Prime Pages*, http://primes.utm.edu/

[2] Yung-Pin Chen, *On Primes, Density Measures, and Statistical Independence*, College Math. J. **36** (2005), 284–288.

[3] Kai Lai Chung, *Elementary Probability Theory with Stochastic Processes*, 3rd ed., Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1979.

[4] Bennett Eisenberg and B. K. Ghosh, *Independent Events in a Discrete Uniform Probability Space*, Amer. Statist. **41** (1987), 52–56.

[5] Mark Kac, *Statistical Independence in Probability, Analysis, and Number Theory*, The Carus Mathematical Monographs, No. 12 (1959), The Mathematical Association of America.

[6] The On-Line Encyclopedia of Integer Sequences, published electronically at http://oeis.org, 2010.

[7] James E. Pommersheim, Tim K. Marks and Erica L. Flapan, *Number Theory*, John Wiley & Sons, 2010.

[8] Sheldon Ross, *A First Course in Probability*, 5th ed., Prentice Hall, 1998.

[9] Eric W. Weisstein, "Sophie Germain Prime." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/SophieGermainPrime.html