Probabilistic and experimental method in Sum-Product Theory

Xiaobo Luo

April 19, 2021

1 Introduction

Let's start with some necessary definitions

Definition 1 (Sumset)**.** *Let $A, B \in G$ where $(G, +)$ is an Abelian Group, then the sumset of A, B is defined to be $A + B = \{a + b \mid a \in A, b \in B\}$.*

Definition 2 (Product Set)**.** *Let $A, B \in G$ where $(G, \cdot)$ is an Abelian Monoid, then the product set of A, B is defined to be $A \cdot B = \{a \cdot b \mid a \in A, b \in B\}$.*

Both definitions are defined in general setting and we are normally working with the construction of G being $\mathbb{Z}$ or $\mathbb{Z}_N$ (Multiplicative group of integers modulo n) or $\mathbb{R}$.

The sum set and product set is first investigated by Erdos and Szemeredi[1] in 1983. In their paper, they proved that for $A \subseteq \mathbb{Z}$ being a set of integers, then

$$\max(|A + A|, |A \cdot A|) \geq c|A|^{1+\epsilon}$$

for some small and positive ϵ where $|\cdot|$ denotes the size of the set. They further conjectured that it should be the case

$$\max(|A + A|, |A \cdot A|) \geq c|A|^{2-\delta}$$

for any positive δ.

This problem is further analyzed in the setting where $A \subseteq \mathbb{R}$ and after works by Nathanson[2], Ford[3] and Chang[4], Elekes[5] shows that $\epsilon \geq \frac{1}{4}$. This result is further extended to complex numbers by Toth[6] and Solymosi[7]. The best know bound is proven by Solymosi[8] which is

$$\max(|A + A|, |A \cdot A|) \geq \frac{c|A|^{\frac{14}{11}}}{\log^{\frac{3}{11}} |A|}$$

This is further analyzed in the setting of finite field but the situation becomes more complex as the key tool used in the analysis, Szemeredi-Trotter incidence theorem, doesn't hold in the same generality. It is first shown by Bourgain, Glibichuk and Konyagin [9] and Bourgain, Katz and Tao[10] that if q is a prime, than if $|A| \leq Cq^{1-\epsilon}$, for some $\epsilon > 0$, then there exists $\delta > 0$ such that

$$\max\{|A+A|, |A \cdot A|\} \geq c|A|^{1+\delta}$$

Hart, Iosevich and Solymosi [11] further improved this bound using incidence theorem and get

$$\max\{|A+A|, |A \cdot A|\} \geq c|A|^{\frac{8}{7}}$$

This gives us a better understanding of the size of $|A+A|$ and $|A \cdot A|$. In this paper, we will look at the setting where A is a random subset of $\mathbb{Z}_q$ where each element is chosen with probability p and investigate $E[|A+A|]$. Then we will extend to the more complex case where A is a random subset of $\{1, 2, \ldots, N\}$ and investigate $E[|A+A|]$. The same analysis is done on the product set $A \cdot A$ and we investigate its size in the setting of $\mathbb{Z}$ and $\mathbb{Z}_q$. We will approach this problem using both probabilistic method and experimental method which could aid us in gaining intuition and checking the result.

2 $E[|A+A|]$ in $\mathbb{Z}_p$

With this construction, Below is the first problem we are trying to solve

Question 1. *Let $A \in \mathbb{Z}_N$ be a randomly chosen subset where each element of $\mathbb{Z}_N$ is independently chosen to be in A with probability p, then what is the expected size of $|A+A|$?*

Before trying to solve this problem analytically, let's utilize the tool of simulation to see what the result would be as a function of p. For a fixed N and a fixed p, 100 samples of A are created and the corresponding $|A+A|$ is calculated. The mean of the 100 samples are than calculated and plotted. To make sure that the graph is on the same scale, we divide the mean by the corresponding N since this is the maximum size we could get.

The result is given in Figure 1 for the case $N = 20, 50, 100, 250$ and 500 which corresponds to different lines in the graph. We can see that as p goes to 1, the simulated expected size of $|A+A|$ converges to N and the bigger N is, the faster the convergence is.

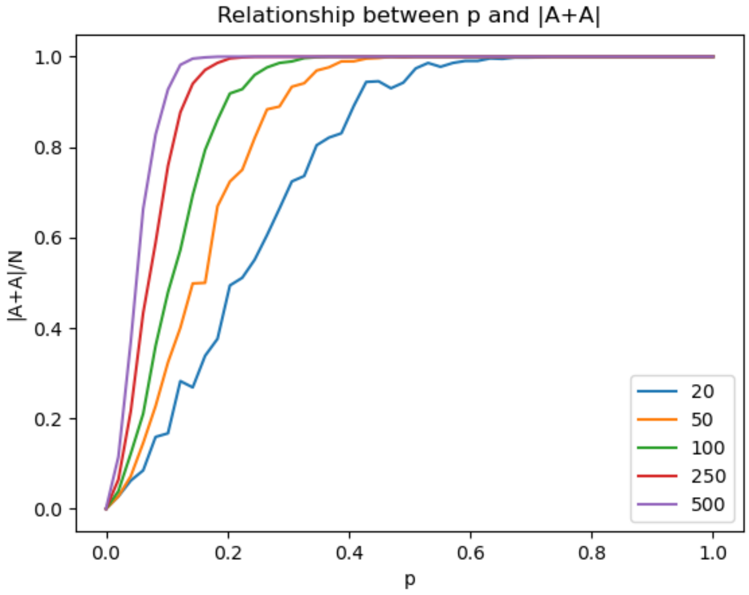

Figure 1: Relationship between p and $|A+A|$

Now we can analyze this problem analytically and check if this is consistent with what we have in our simulation.

Theorem 1. *Let $A \subseteq \mathbb{Z}_N$ be a randomly chosen subset where each element of $\mathbb{Z}_N$ is independently chosen to be in A with probability p. If N is odd, then*

$$E|A+A| = N\left(1-(1-p)(1-p^2)^{\frac{N-1}{2}}\right)$$

If N is even, then

$$E|A+A| = N\left(\frac{2-(1-p)^2(1-p^2)^{\frac{N}{2}-1}-(1-p^2)^{\frac{N}{2}}}{2}\right)$$

Proof.

For all N, we have

$$E[|A+A|] = E\left[\sum_{i=0}^{N-1} 1_{A+A}(i)\right] = \sum_{i=0}^{N-1} E\left[1_{A+A}(i)\right] = \sum_{i=0}^{N-1} P(i \in A+A)$$

Note that $i \in A+A$ if and only if there is some $u, v \in A$ where $u+v = i \mod N$

Therefore, if $i \notin A+A$, it has to be the case that for all (u,v) pair where $u+v = i \mod N$, at least one of them is not in A

If $u \neq v$, then the possibility of at least one of them not in A is $1-p^2$.

On the other hand, if $u = v$, then the possibility of at least one of them not in A is $1-p$

So the problem is essentially reduced to finding the number of such pairs

Suppose N is odd

I claim that for all $i \in \mathbb{Z}_N$, there is exactly one element $u \in \mathbb{Z}_N$ where $u+u = i \mod N$ and $\frac{N-1}{2}$ distinct pairs of $(u,v) \in \mathbb{Z}_N \times \mathbb{Z}_N$ s.t. $u + v = i \mod N$ (note that the distinct here is not the usual sense of distinct but rather distinct irrespective of order, so (u,v) and (v,u) are considered the same and all the distinct in this paper are defined in this way) and I denote it as lemma 1 and the proof is in the appendix

Using this fact, we have that the possibility of each i not being in $A + A$ is

$$P(i \notin A + A) = (1-p)(1-p^2)^{\frac{N-1}{2}}$$

Plugging this back into the expression, we have

$$E[|A+A|] = \sum_{i=0}^{N-1} P(i \in A + A) = \sum_{i=0}^{N-1} 1 - (1-p)(1-p^2)^{\frac{N-1}{2}} = N\left(1 - (1-p)(1-p^2)^{\frac{N-1}{2}}\right)$$

Suppose N is even

I claim that for all $i \in \mathbb{Z}_N$ which is odd, it has exactly $\frac{N}{2}$ distinct pairs of $(u,v) \in \mathbb{Z}_N \times \mathbb{Z}_N$ s.t. $u+v = i \mod N$ and there is no $u \in \mathbb{Z}_N$ where $u + u = i \mod N$.

I also claim that for all $i \in \mathbb{Z}$ which is even, it has exactly $\frac{N}{2} - 2$ distinct pairs of $(u,v) \in \mathbb{Z}_N \times \mathbb{Z}_N$ s.t. $u + v = i \mod N$ and there are 2 $u \in \mathbb{Z}_N$ where $u + u = i \mod N$. I denote these two claims as lemma 2 and the proof is in the appendix

Using this fact, we have that the possibility of each odd i not being in $A + A$ is

$$P(i \notin A + A) = (1-p^2)^{\frac{N}{2}}$$

and the possibility of each even i not being in $A + A$ is

$$P(i \notin A + A) = (1-p)^2(1-p^2)^{\frac{N}{2}-1}$$

Using the fact that there are exactly $\frac{N}{2}$ even numbers in $\mathbb{Z}_n$ and $\frac{N}{2}$ odd numbers in $\mathbb{Z}_n$, we can plug what we have above back into the expression and get

$$E[|A+A|] = \sum_{i=0}^{N-1} P(i \in A+A) = \frac{N}{2}\left(1 - (1-p^2)^{\frac{N}{2}}\right) + \frac{N}{2}\left(1 - (1-p)(1-p^2)^{\frac{N-2}{2}}\right) = N\left(\frac{2 - (1-p)^2(1-p^2)^{\frac{N}{2}-1} - (1-p^2)^{\frac{N}{2}}}{2}\right)$$

□

Now we have this analytical solution and we can plot the result to check if it agrees with the simulation. The result is given in Figure 2. Clearly, they match perfectly and we can confidently conclude that the result is correct!

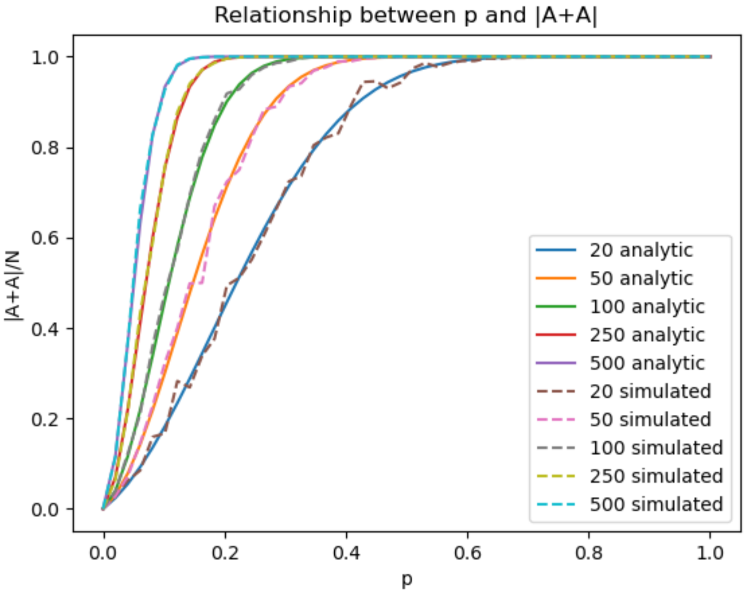

Figure 2: Comparison of simulation and analytic result

3 $E[|A+A|]$ in $\mathbb{Z}$

Now what if we extend our consideration to $\mathbb{Z}$ instead of $\mathbb{Z}_N$? There will definitely be more complication as we don't have the fact that the sum will always be in $\mathbb{Z}_N$. The new question is as follows

Question 2. *Let $A \in \mathbb{Z}$ be a randomly chosen subset where each element of $\{1,2,...,N\}$ is independently chosen to be in A with probability p, then what is the expected size of $|A+A|$?*

Again, let's begin by doing some simulation to have an intuitive understanding. For a fixed N and a fixed p, 100 samples of A are created and the corresponding $|A+A|$ is calculated. The mean of the 100 samples are than calculated and plotted. This time, the maximum size of $A+A$ becomes $2N-1$ (1 is never achieved in the sumset) so we divide the mean by $2N-1$ to make the graph on the same scale.

The result is given in Figure 3 for the case $N = 20, 50, 100, 250$ and 500 which corresponds to different lines in the graph. We can see that as p goes to 1, the simulated expected size of $|A+A|$ converges to $2N$ and the bigger N is, the faster the convergence is.

Compared to the previous case, we can see that the speed of convergence is slower which make sense since the sumset now lies in $[2,2N]$ which is bigger than the previous $\mathbb{Z}_N$ case.

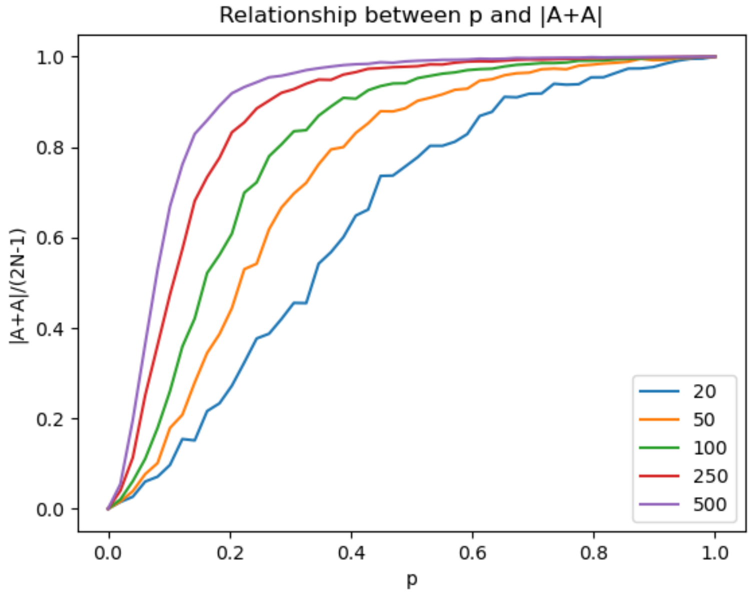

Figure 3: Relationship between p and $|A + A|$

Now we can analyze this problem analytically and check if this is consistent with what we have in our simulation. For simplicity, we will work with the case N being even so exactly half of the numbers in $\{1, 2, \ldots, N\}$ are even and half of them are odd

Theorem 2. *Let $A \subseteq \mathbb{Z}$ be a randomly chosen subset where each element of $\{1, 2, ..., N\}$ is independently chosen to be in A with probability p. If N is even, then*

$$E|A + A| = 2N - (4 - 2p - p^2)\frac{1 - (1 - p^2)^{\frac{N}{2}}}{p^2}$$

Proof. This time, we have

$$E[|A + A|] = E\left[\sum_{i \in \mathbb{Z}} 1_{A+A}(i)\right] = \sum_{i \in \mathbb{Z}} E\left[1_{A+A}(i)\right] = \sum_{i \in \mathbb{Z}} P(i \in A + A) = \sum_{i=1}^{2N} P(i \in A + A)$$

Note that we are summing from 1 simply for the simplicity of keeping the number of even and odd number the same. $P(1 \in A + A) = 0$ so it won't effect our analysis.

We can still reduce the problem to finding the number of pairs of $(u, v) \in [1, N] \times [1, N]$ s.t. $u + v = i$ and $u \neq v$ and number of $u \in [1, N]$ where $u + u = i$

There are a total of 4 cases for the value of i: whether it is $\leq N$ or $> N$ and whether it is odd or even

Case 1: $i \in [1, N]$ and i is odd

Case 2: $i \in [1, N]$ and i is even

Case 3: $i \in [N + 1, 2N]$ and i is odd

Case 4: $i \in [N + 1, 2N]$ and i is even

I claim that in each case, the number of (u, v) pairs and u we are looking for are as follows:

Case 1: There are $\frac{i-1}{2}$ distinct (u,v) pair and 0 u

Case 2: There are $\frac{i}{2}-1$ distinct (u,v) pair and 1 u

Case 3: There are $\frac{i-1}{2}-(i-N-1)$ distinct (u,v) pair and 0 u

Case 4: There are $\frac{i}{2}-1-(i-N-1)$ distinct (u,v) pair and 1 u

This claim is denoted as lemma 3 and is proved in the appendix

As before, the possibility of at least one element of the (u,v) pair not in A is $1-p^2$ and the possibility of u not in A is $1-p$

Using this fact and the result of lemma 3, we can plug in the expression like what we did before and get

$$\begin{aligned}
E[|A+A|] &= \sum_{i=1}^{2N} P(i \in A+A) \\
&= \sum_{i=1,odd}^{N} 1-(1-p^2)^{\frac{i-1}{2}} + \sum_{i=1,even}^{N} 1-(1-p^2)^{\frac{i}{2}-1}(1-p) \\
&\quad + \sum_{i=N+1,odd}^{2N} 1-(1-p^2)^{\frac{i-1}{2}-(i-N-1)} + \sum_{i=N+1,even}^{2N} 1-(1-p^2)^{\frac{i}{2}-1-(i-N-1)}(1-p) \\
&= I + II + III + IV
\end{aligned}$$

For part I, we have

$$\begin{aligned}
I &= \sum_{i=1,odd}^{N} 1-(1-p^2)^{\frac{i-1}{2}} \\
&= \sum_{k=1}^{\frac{N}{2}} 1-(1-p^2)^{\frac{2k-1-1}{2}} \\
&= \frac{N}{2} - \sum_{k=1}^{\frac{N}{2}} (1-p^2)^{k-1} \\
&= \frac{N}{2} - \frac{(1-p^2)^0-(1-p^2)^{\frac{N}{2}-1+1}}{1-(1-p^2)} \\
&= \frac{N}{2} - \frac{1-(1-p^2)^{\frac{N}{2}}}{p^2}
\end{aligned}$$

For part II, we have

$$\begin{aligned}
II &= \sum_{i=1,even}^{N} 1-(1-p^2)^{\frac{i}{2}-1}(1-p) \\
&= \sum_{k=1}^{\frac{N}{2}} 1-(1-p^2)^{\frac{2k}{2}-1}(1-p) \\
&= \frac{N}{2} - \sum_{k=1}^{\frac{N}{2}} (1-p^2)^{k-1}(1-p) \\
&= \frac{N}{2} - (1-p)\frac{1-(1-p^2)^{\frac{N}{2}}}{p}
\end{aligned}$$

For part III, we have

$$\begin{aligned}
III &= \sum_{i=N+1,odd}^{2N} 1-(1-p^2)^{\frac{i-1}{2}-(i-N-1)} \\
&= \sum_{i=1,odd}^{N} 1-(1-p^2)^{\frac{i+N-1}{2}-(i-1)} \\
&= \sum_{i=1,odd}^{N} 1-(1-p^2)^{\frac{N-i+1}{2}} \\
&= \sum_{k=1}^{\frac{N}{2}} 1-(1-p^2)^{\frac{N-(2k-1)+1}{2}} \\
&= \frac{N}{2}-(1-p^2)\sum_{k=1}^{\frac{N}{2}}(1-p^2)^{\frac{N}{2}-k} \\
&= \frac{N}{2}-(1-p^2)\sum_{k=0}^{\frac{N}{2}-1}(1-p^2)^{k} \\
&= \frac{N}{2}-(1-p^2)\frac{1-(1-p^2)^{\frac{N}{2}}}{p^2}
\end{aligned}$$

For part IV, we have

$$\begin{aligned}
IV &= \sum_{i=N+1,even}^{2N} 1-(1-p^2)^{\frac{i}{2}-1-(i-N-1)}(1-p) \\
&= \sum_{i=1,even}^{N} 1-(1-p^2)^{\frac{i+N}{2}-1-(i-1)}(1-p) \\
&= \sum_{i=1,even}^{N} 1-(1-p^2)^{\frac{N-i}{2}}(1-p) \\
&= \sum_{k=1}^{\frac{N}{2}} 1-(1-p^2)^{\frac{N-(2k)}{2}}(1-p) \\
&= \frac{N}{2}-(1-p)\sum_{k=1}^{\frac{N}{2}}(1-p^2)^{\frac{N}{2}-k} \\
&= \frac{N}{2}-(1-p)\sum_{k=0}^{\frac{N}{2}-1}(1-p^2)^{k} \\
&= \frac{N}{2}-(1-p)\frac{1-(1-p^2)^{\frac{N}{2}}}{p^2}
\end{aligned}$$

Putting everything together, we have

$$\begin{aligned}
E[|A+A|] &= I+II+III+IV \\
&= \frac{N}{2}-\frac{1-(1-p^2)^{\frac{N}{2}}}{p^2}+\frac{N}{2}-(1-p)\frac{1-(1-p^2)^{\frac{N}{2}}}{p}+\frac{N}{2}-(1-p^2)\frac{1-(1-p^2)^{\frac{N}{2}}}{p^2}+\frac{N}{2}-(1-p)\frac{1-(1-p^2)^{\frac{N}{2}}}{p^2} \\
&= 2N-(4-2p-p^2)\frac{1-(1-p^2)^{\frac{N}{2}}}{p^2}
\end{aligned}$$

□

A quick sanity check by comparing the analytical result with the simulation gives us Figure 4. Clearly, they match perfectly and we can confidently conclude that the result is correct!

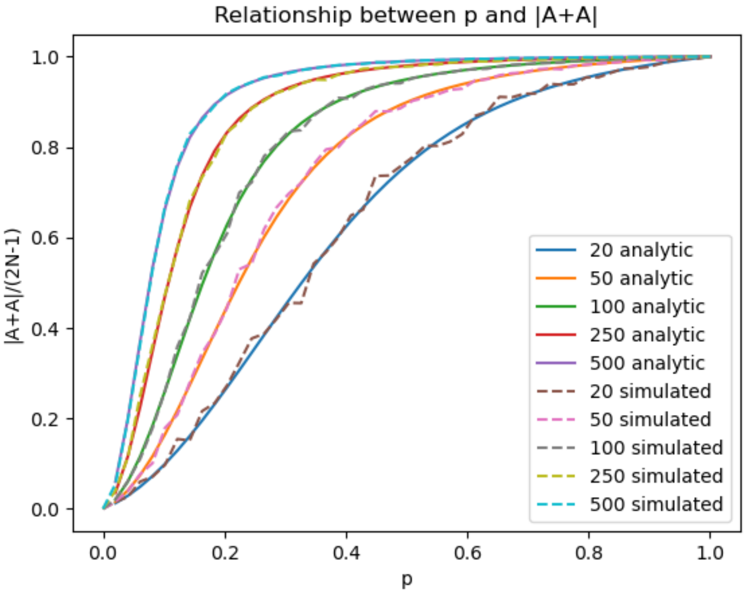

Figure 4: Comparison of simulation and analytic result

We could so similar calculation with the case N being odd and we will have

Theorem 3. *Let $A \subseteq \mathbb{Z}$ be a randomly chosen subset where each element of $\{1, 2, ..., N\}$ is independently chosen to be in A with probability p. If N is odd, then*

$$E|A+A| = 2N - (2-p)\frac{1-(1-p^2)^{\frac{N+1}{2}}}{p^2} - (2-p-p^2)\frac{1-(1-p^2)^{\frac{N-1}{2}}}{p^2}$$

The proof is almost identical to the N being even counter part except at handling some index more carefully so it is omitted.

4 $E[A \cdot A]$ in Z_p

Now we want to investigate the expected size of $A + A$ in Z_p and the problem we are trying to solve is as follows

Question 3. *Let $A \in \mathbb{Z}_N$ be a randomly chosen subset where each element of $\mathbb{Z}_N$ is independently chosen to be in A with probability p, then what is the expected size of $|A \cdot A|$?*

Again, let's begin by doing some simulation to have an intuitive understanding. The maximum size of $A \cdot A$ is N so we divide the mean by $2N - 1$ to make the graph on the same scale.

But this time, we need to be careful on the choice of N because Z_N will be a monoid or a group depending on whether N is a prime so when doing experiment, it is better to put in some prime numbers and see if there is any difference

The result is given in Figure 5 for the case $N = 19, 20, 47, 50, 97, 100, 241, 250, 499$ and 500 which corresponds to different lines in the graph. We can see that as p goes to 1, the simulated expected size of $|A + A|$ converges to N in all cases. However, we can see that when N is prime, we are having faster convergence rate. For example, the blue line corresponds to the case $n = 19$ and the orange line corresponds to the case $n = 20$. Though $19 < 20$, the convergence rate is faster for the $n = 19$ case. Similar thing is happening for the pair $(47, 50)$, $(97, 100)$, $(241, 250)$ and $(499, 500)$. So instead of having a pure monotonic relationship between N and the convergence rate, we are having two cases, one is N being prime and one is N being composite. In each case, the convergence rate increases as N increases but this is not the case across the two groups. This gives us a sanity check criteria for our later result which means that the formula should be different for the case N being prime or not

Compared to the sum set $A + A$, we can see that the speed of convergence is slower. The only intuitive explanation is that some numbers in Z_N are unlikely to be contained in $A \cdot A$ so the convergence is happening more slowly. A quick analysis will tell us that the probability each elements will be in $A \cdot A$ varies based on the number of divisors and unlike the sumset case where the number of pairs that sum up to a specific number is almost the same across $\mathbb{Z}_N$, they are not uniformly distributed in the product set and this is a big problem we will face when we analyze the problem analytically

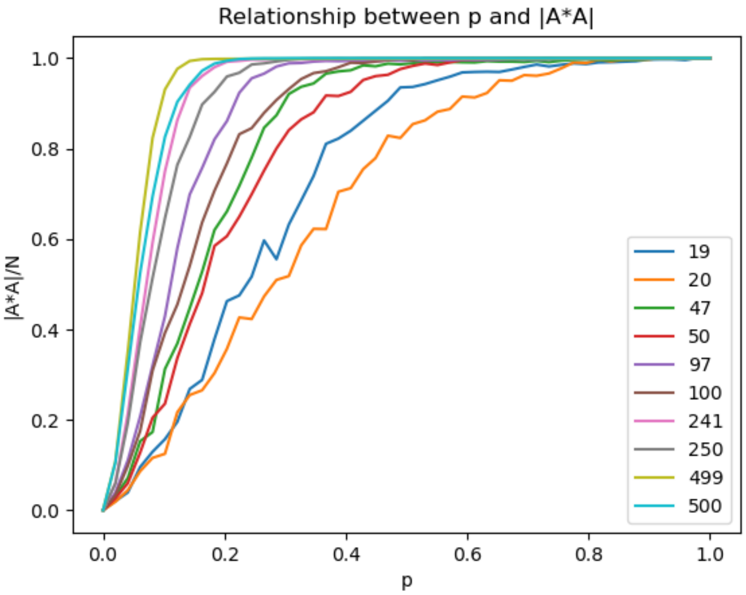

Figure 5: Relationship between p and $|A + A|$

To approach this problem analytically, we define the following notations

Definition 3 (Divisor Counting Function in $\mathbb{Z}_N$)**.** *Let $x \in \mathbb{Z}_N$, define $d(x)$ to be the number of (α, β) pair where $\alpha, \beta, \in \mathbb{Z}_N$, $\alpha \leq \beta$ in the canonical representation, i.e. $0 \leq 1 \leq 2 \leq \ldots \leq N - 1$ and $\alpha \cdot \beta = x$.*

This is essentially the number of pairs of divisors of x disregarding the order

Definition 4 (Square Root Counting Function in $\mathbb{Z}_N$)**.** *Let $x \in \mathbb{Z}_N$, define $e(x)$ to be the number of $\alpha \in \mathbb{Z}_N$ where $\alpha \cdot \alpha = \alpha^2 = x$. This is essentially the number of "square root" of x*

Theorem 4. *Let $A \subseteq \mathbb{Z}_N$ be a randomly chosen subset where each element of $\mathbb{Z}_N$ is independently chosen to be in A with probability p, then*

$$E[|A \cdot A|] = \sum_{i=0}^{N-1} 1 - (1-p)^{e(i)}(1-p^2)^{d(i)-e(i)}$$

Proof.

For all N, we have

$$E[|A \cdot A|] = E\left[\sum_{i=0}^{N-1} 1_{A \cdot A}(i)\right] = \sum_{I=0}^{N-1} E\left[1_{A \cdot A}(i)\right] = \sum_{i=0}^{N-1} P(i \in A \cdot A)$$

Similar to previous problems, $i \in A \cdot A$ if and only if there is some $u, v \in A$ where $u \cdot v = i \mod N$

If $u \neq v$, then the possibility of at least one of them not in A is $1 - p^2$.

On the other hand, if $u = v$, then the possibility of at least one of them not in A is $1 - p$

By definition 3 and 4, we can directly conclude that there are $d(i) - e(i)$ pairs for the first case and $e(i)$ pairs for the second case

Plugging back into the expression gives us

$$E[|A \cdot A|] = \sum_{i=0}^{N-1} 1 - (1-p)^{e(i)}(1-p^2)^{d(i)-e(i)}$$

However, it is worth noticing that we may run into the problem of 0^0

This may only happen if $p = 1$ and it could be the case that $e(i) = 0$, i.e. i is not a "perfect square" or $d(i) = e(i)$, i.e. i can only be expressed as perfect squares.

In these cases, we can simply define $0^0 = 1$ and we will be fine with the calculation still being valid

□

This formula is not a closed form formula because we need a closed form expression of $d(i)$ and $e(i)$ to get this. When N is an odd prime, then we can show that $e(i)$ is either 0 or 2 and this is proven in appendix as lemma 4. Stangl[12] has shown that the number of perfect squares in Z_N is $\frac{N+1}{2}$ if N is an odd prime but other than this, we don't have more explicit information on each $d(i), e(i)$

But we could get a lower bound which is of cleaner form. For simplicity, we are only dealing with the case where $p \in (0,1)$ as the result of the two extreme cases are clear: when $p = 0$, we have $E[A \cdot A] = 0$ and when $p = 1$, we have $E[A \cdot A] = N$. We will also focus our attention for N being an odd prime so there is more nice property of $\mathbb{Z}_N$

Theorem 5. *Let $A \subseteq \mathbb{Z}_N$ be a randomly chosen subset where each element of $\mathbb{Z}_N$ is independently chosen to be in A with probability $p \in (0,1)$. If N is an odd prime then*

$$E[|A \cdot A|] > \frac{p^2}{2}N(N-1) - \frac{p^4}{8}(N^3 - 7N + 2)$$

Proof.

Note that $p \in (0,1)$ so we have

$$p > p^2$$
$$1 - p < 1 - p^2$$
$$(1-p)^{e(i)}(1-p^2)^{d(i)-e(i)} \le (1-p^2)^{e(i)}(1-p^2)^{d(i)-e(i)}$$
$$1-(1-p)^{e(i)}(1-p^2)^{d(i)-e(i)} \ge 1-(1-p^2)^{e(i)}(1-p^2)^{d(i)-e(i)}$$
$$1-(1-p)^{e(i)}(1-p^2)^{d(i)-e(i)} \ge 1-(1-p^2)^{d(i)}$$

This is true for all i so we can plug this into the conclusion of the previous theorem and get

$$E[|A \cdot A|] \ge \sum_{i=1}^{N} 1 - (1-p^2)^{d(i)}$$

Now we have successfully removed $e(i)$ from our expression and we want to further remove this $d(i)$

I claim that $\sum_{i=0}^{N-1} d(i) = \frac{N(N+1)}{2}$ and this is proven in appendix as lemma 5

To utilize this equation, we can use Taylor Theorem to expand $(1-p^2)^{d(i)}$ as a function of p^2

The underlying function will be $f(x) = (1-x)^{d(i)}$ where $d(i)$ is some fixed constant for now that satisfied the property $d(i) \ge 0$ and $d(i) \in \mathbb{Z}$

Then we have $f'(x) = -d(i)(1-x)^{d(i)-1}$ and $f''(x) = d(i)(d(i)-1)(1-x)^{d(i)-2}$

So we can expand f at 0 and by Taylor Theorem, we have

$$f(x) = 1 - d(i)x + \frac{d(i)(d(i)-1)(1-\xi)^2}{2}x^2$$

for some $\xi \in (0,x)$

So if we plug in $x = p^2$, we have

$$(1-p^2)^{d(i)} = 1 - d(i)p^2 + \frac{d(i)(d(i)-1)(1-\xi)^2}{2}p^4$$

for some $\xi \in (0,p^2)$

Plugging this into the bound we have for $E[|A \cdot A|]$ gives us

$$\begin{aligned}
E[|A \cdot A|] &\ge \sum_{i=0}^{N-1} 1 - 1 + d(i)p^2 - \frac{d(i)(d(i)-1)(1-\xi)^2}{2}p^4 \\
&= \sum_{i=0}^{N-1} d(i)p^2 - \frac{d(i)(d(i)-1)(1-\xi)^2}{2}p^4 \\
&\ge \sum_{i=0}^{N-1} d(i)p^2 - \frac{d(i)(d(i)-1)}{2}p^4 \\
&= p^2\frac{N(N+1)}{2} - \frac{p^4}{2}\sum_{i=0}^{N-1} d(i)(d(i)-1)
\end{aligned}$$

At this point, the second term involves the sum of $d(i)^2$ which we don't have an estimate on. What we could do is to pull out one of them as $\max d(i)$ and get a bound on this term.

I claim that $\max_{i \in \mathbb{Z}_N, i \ne 0} d(i) = \frac{N+1}{2}$ and $d(0) = N$ and this is proven in appendix as lemma 6

We can see that $d(0)$ is like the outlier that would mess up with our estimation on $\max d(i)$ by a factor of roughly 2. To have a tighter bound, we split the sum into two parts

$$\begin{aligned}
[|A\cdot A|] &\geq p^2\frac{N(N+1)}{2}-\frac{p^4}{2}d(0)(d(0)-1)-\frac{p^4}{2}\sum_{i=1}^{N-1}d(i)(d(i)-1)\\
&\geq p^2\frac{N(N+1)}{2}-\frac{p^4}{2}N(N-1)-\frac{p^4}{2}\max_{i\in\mathbb{Z}_N,i\neq 0}(d(i))\sum_{i=1}^{N-1}d(i)-1\\
&= p^2\frac{N(N+1)}{2}-\frac{p^4}{2}N(N-1)-\frac{p^4}{2}\frac{N+1}{2}\left(\frac{N(N+1)}{2}-N-(N-1)\right)\\
&= p^2\frac{N(N+1)}{2}-\frac{p^4}{2}N(N-1)-\frac{p^4}{8}(N+1)(N^2-3N+2)\\
&= p^2\frac{N(N+1)}{2}-\frac{p^4}{8}(N-1)(N^2+3N-2)
\end{aligned}$$

□

We could interpret this result as $p^2\frac{N(N+1)}{2}$ is like the main term we are interested in and the $-\frac{p^4}{8}(N-1)(N^2+3N-2)$ is the remainder. We can solve for the range of p where the main term is dominant and the range is $p^2 \leq \frac{4N(N+1)}{(N-1)(N^2+3N-2)}$ which is roughly on the order of $\frac{1}{N}$. This is not a coincidence and is a direct result of our estimation on $\max d(i)$.

Now let's utilize the simulation tool again and see how this lower bound is performing. Since we are only investigating the case when N is a prime, we only check the case when $N = 19, 47, 97, 241, 499$ and compare the simulated result with the lower bound. The result is given in Figure 6 where the right one is a zoomed in graph to see its performance in the low p range more clearly. The lower bounds are all plotted only in the rang $0 \leq p^2 \leq \frac{4N(N+1)}{(N-1)(N^2+3N-2)}$.

We can see that the sharpness of the bound increases as N increases. If we take the same $0 \leq p^2 < \frac{4N(N+1)}{(N-1)(N^2+3N-2)}$ as the range of p, then asymptotically, the remainder will be dominated and the meaning is that every pair of $(\alpha,\beta) \in A\cdot A$ is given us different $\alpha\cdot\beta$. But it is also worth noticing that the upper bound of the range of p also goes to 0 asymptotically. So only when we are selecting the elements for A "sparse" and random enough, can we obtain this uniqueness.

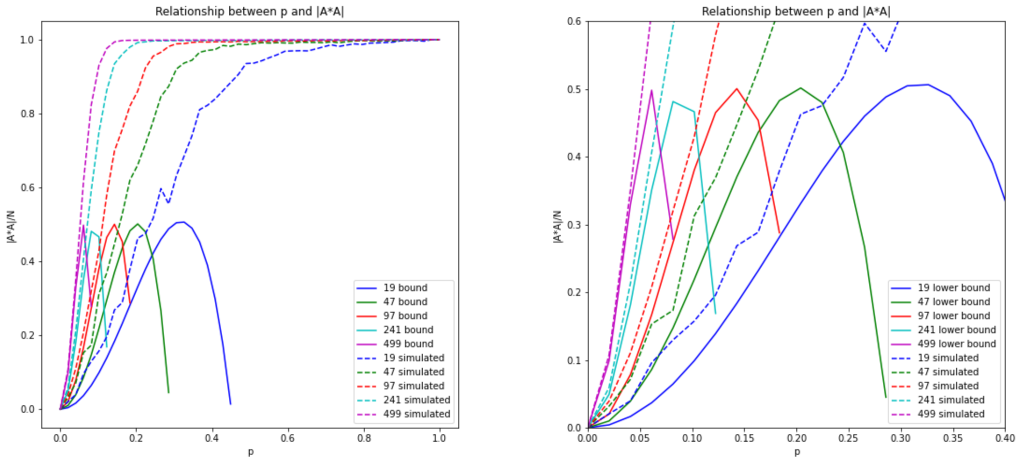

Figure 6: Comparison of simulation and analytical lower bound

5 $E|A \cdot A|$ in $\mathbb{Z}$

The extension of the problem from $\mathbb{Z}_N$ to $\mathbb{Z}$is very natural and the question we have now is

Question 4. *Let $A \in \{1, 2, \dots, N\}$ be a randomly chosen subset where each element of $\{1, 2, \dots, N\}$ is independently chosen to be in A with probability p, then what is the expected size of $|A \cdot A|$?*

Generally, there are many similarities with the above section where we worked in $\mathbb{Z}_N$. Before we dig into the detail, let's first do simulation as always and see what we will get. The range of the product always lies in the range $\{1, 2, \dots, N^2\}$ so let's just divide the size of simulated $A \cdot A$ by N^2 to normalize the graph as it is unclear from first glance what the maximum size will be. The simulation is done for $N = 20, 50, 100, 250, 500, 1000$ and the result is given in Figure 7.

First, we could see that the ratio is not going to 1. This should be expected as even if all the (α, β) pair gives us distinct product irrespective of order, we should only expect the final size to be $\frac{N(N+1)}{2}$. But even as p approaches 1, this ratio is still lower than 0.5 which suggests that there are some number in the range of $[1, N^2]$ that could never appear in $A \cdot A$. A quick thought will reveal that the prime numbers in the range $(N, N^2]$ will not be in the range. Also, numbers in the same range with a prime divisor that is bigger than N will also not appear in $A \cdot A$. An interesting fact is that the higher N is, the lower the proportion is.

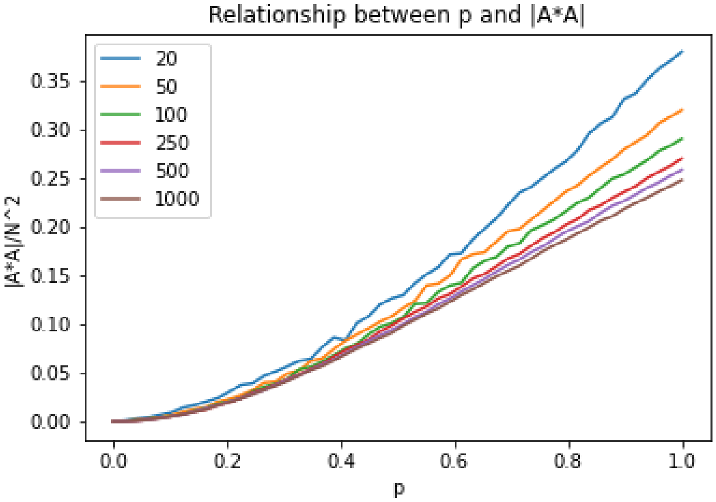

Figure 7: Relationship between p and $|A \cdot A|$

Now we can approach this problem analytically, for simplicity, we again define the corresponding divisor counting function

Definition 5 (Divisor Counting Function in $\mathbb{Z}$)**.** *Let $x \in \{1, \ldots, N^2\}$, define $D(x)$ to be the number of (α, β) pair where $\alpha, \beta, \in \{1, \ldots, N\} \times \{1, \ldots, N\}$, $\alpha \leq \beta$ and $\alpha \cdot \beta = x$. This is essentially the number of pairs of divisors of x disregarding the order*

Note that there is no need to define the square root counting function anymore since in $\mathbb{Z}$, a number will either not be a perfect square or it has a unique positive square root, i.e. the square root counting formula is either 0 and 1.

Theorem 6. *Let $A \subseteq \{1, \ldots, N^2\}$ be a randomly chosen subset where each element of $\{1, \ldots, N^2\}$ is independently chosen to be in A with probability p, then*

$$E[|A \cdot A|] = \sum_{i=1, i \text{ not square}}^{N^2} 1 - (1-p^2)^{D(i)} + \sum_{i=1, i \text{ is square}}^{N^2} 1 - (1-p)(1-p^2)^{D(i)-1}$$

Proof.
The proof is essentially the same as the proof of Theorem 4 so I will follow the same logic and be concise

Change the range we are summing, we have

$$E[|A \cdot A|] = E\left[\sum_{i=1}^{N^2} 1_{A \cdot A}(i)\right] = \sum_{i=1}^{N^2} E\left[1_{A \cdot A}(i)\right] = \sum_{i=1}^{N^2} P(i \in A \cdot A)$$

Again, $i \in A \cdot A$ if and only if there is some $u, v \in A$ where $u \cdot v = i$

If $u \neq v$, then the possibility of at least one of them not in A is $1 - p^2$.

On the other hand, if $u = v$, then the possibility of at least one of them not in A is $1 - p$

If i is not a perfect square, then its probability of being in $A \cdot A$ is $1 - (1 - p^2)^{D(i)}$ and if it is a perfect square, then its probability of being in $A \cdot A$ is $1 - (1 - p)(1 - p^2)^{D(i)-1}$

Therefore, splitting into two cases, we have

$$E[|A \cdot A|] = \sum_{i=1, i \text{ not square}}^{N^2} 1 - (1 - p^2)^{D(i)} + \sum_{i=1, i \text{ is square}}^{N^2} 1 - (1 - p)(1 - p^2)^{D(i)-1}$$

Again, we may face the problem of 0^0 when $p = 1$ and as before, we define it to be 1

□

We could approach similarly as in the above section and obtain a cleaner lower bound

Theorem 7. *Let $A \subseteq \{1, \ldots, N^2\}$ be a randomly chosen subset where each element of $\{1, \ldots, N^2\}$ is independently chosen to be in A with probability $p \in (0, 1)$, then*

$$E[|A \cdot A|] \geq p^2 \frac{N(N+1)}{2} - p^4 \frac{o(N^\epsilon) - 1}{4} N(N+1)$$

where $o(N^\epsilon)$ means that in little o notation, this quantity is $o(N^\epsilon)$ for all $\epsilon > 0$, i.e. it is "sub-polynomial"

Proof.

Again, the proof will be similar to that of theorem 5.

Since $p \in (0, 1)$, we have

$$p > p^2$$

$$1 - p < 1 - p^2$$

$$(1 - p)(1 - p^2)^{D(i)-1} \leq (1 - p^2)(1 - p^2)^{D(i)-1}$$

$$1 - (1 - p)(1 - p^2)^{D(i)-1} \geq 1 - (1 - p^2)(1 - p^2)^{D(i)-1}$$

$$1 - (1 - p)(1 - p^2)^{D(i)-1} \geq 1 - (1 - p^2)^{D(i)}$$

So we can merge the two sums in Theorem 6 and get

$$E[|A \cdot A|] \geq \sum_{i=1, i \text{ not square}}^{N^2} 1 - (1 - p^2)^{D(i)} + \sum_{i=1, i \text{ is square}}^{N^2} 1 - (1 - p^2)^{D(i)} = \sum_{i=1}^{N^2} 1 - (1 - p^2)^{D(i)}$$

I claim that in this setting, we still have $\sum_{i=1}^{N^2} D(i) = \frac{N(N+1)}{2}$ and this is proved in appendix as lemma 7

Then we again use Taylor theorem on $f(x) = (1 - x)^{D(i)}$ and get

$$\begin{aligned}
E[|A \cdot A|] &\geq \sum_{i=1}^{N^2} 1 - 1 + D(i)p^2 - \frac{D(i)(D(i) - 1)(1 - \xi)^2}{2} p^4 \\
&= \sum_{i=1}^{N^2} D(i)p^2 - \frac{D(i)(D(i) - 1)(1 - \xi)^2}{2} p^4 \\
&\geq \sum_{i=1}^{N^2} D(i)p^2 - \frac{D(i)(D(i) - 1)}{2} p^4 \\
&= p^2 \frac{N(N+1)}{2} - \frac{p^4}{2} \sum_{i=1}^{N^2} D(i)(D(i) - 1)
\end{aligned}$$

We still want an estimate of $\max_{i\in\{1,\dots,N\}} D(i)$ to pull one of the $D(i)$ out.
It is shown in Apostol's textbook [13] that the number of divisor $\sigma(n)$ is $o(N^{\epsilon})$ so we should investigate the relationship between the number of divisors $\sigma(n)$ and our $D(i)$
First note that $D(i)$ is counting pairs of divisors regardless of order so it will be at most be $\frac{\sigma(n)}{2}$
Moreover, we are restricting our attention to divisor that is smaller than N so this will further reduce $D(i)$ by some unknown quantity so it is safe to use $\frac{\sigma(n)}{2}$ as our upper bound for $D(i)$
Since little o notation doesn't care constant, we could directly write the upper bound of $D(i)$ as $o(N^{\epsilon})$
It is worth noticing that when plugging in the bound, we can't pull out the $D(i)$ as we did before since $\sum_{i=1}^{N^2} D(i) - 1$ will then be a negative number
So we pull out the $D(i) - 1$ instead and get

$$\begin{aligned}
[|A\cdot A|] &\geq p^2\frac{N(N+1)}{2} - \frac{p^4}{2}\sum_{i=1}^{N^2} D(i)(D(i)-1) \\
&\geq p^2\frac{N(N+1)}{2} - \frac{p^4}{2}\max_{i\in\{1,\dots,N^2\}}(D(i)-1)\sum_{i=1}^{N^2} D(i) \\
&= p^2\frac{N(N+1)}{2} - \frac{p^4}{2}(o(N^{\epsilon})-1)\frac{N(N+1)}{2}
\end{aligned}$$

□

It will be harder to visualize the result due to the fact that the little o notation is not a very precise notation. The bound on $\sigma(n)$ is more precisely calculated by Wigert[14] to be $\limsup_{n\to\infty}\frac{\log\sigma(n)}{\log n/\log\log n} = \log 2$. This means that the maximum of $\sigma(n)$ is on the order of $e^{\frac{\log(2)\log(n)}{\log(\log(n))}}$. As argued above, our upper bound for $D(n)$ is $\frac{\sigma(n)}{2}$ so I will substitute $\frac{1}{2}e^{\frac{\log(2)\log(n)}{\log(\log(n))}}$ for $o(N^{\epsilon})$ to plot the lower bound for visualization. Again, we solve for the range of p where the main term will be dominant and it turns out to be $p^2 \in \left(0, \frac{2}{e^{\frac{\log(2)\log(n)}{\log(\log(n))}}-1}\right)$. The result is given in Figure 8 where the right one is a zoomed in graph to see its performance in the low p range more clearly. The lower bounds are all plotted only in the rang $0 \leq p^2 \leq \frac{2}{e^{\frac{\log(2)\log(n)}{\log(\log(n))}}-1}$.

We can see that at the lower end and as N increases, this bound is pretty sharp and the range of p where the main term dominates is much larger compared to the case of $A\cdot A$ in Z_N setting.

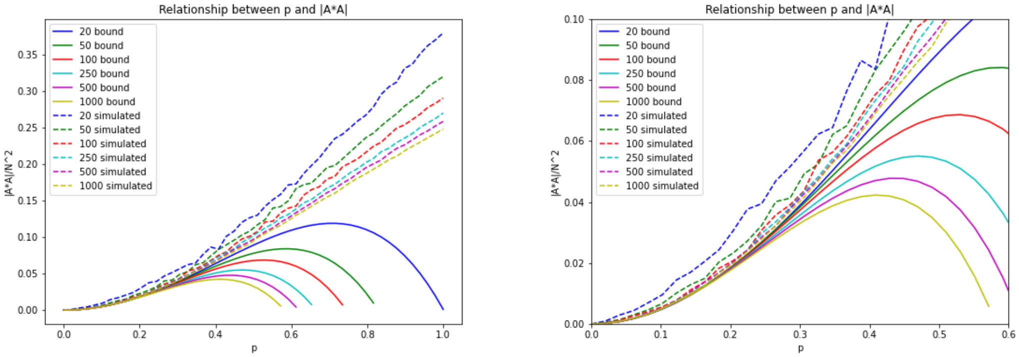

Figure 8: Comparison of simulation and analytical lower bound

6 Appendix

6.1 Proof of Lemmas

Lemma 1. *Suppose N is odd, then for all $i \in \mathbb{Z}_N$, there is exactly one element $u \in \mathbb{Z}_N$ where $u + u = i \mod N$ and $\frac{N-1}{2}$ distinct pairs of $(u, v) \in \mathbb{Z}_N \times \mathbb{Z}_N$ s.t. $u + v = i \mod N$*

Proof.
First, I want to show the existence of u
If i is even, then clearly $\frac{i}{2} + \frac{i}{2} = i$
If i is odd, then $\frac{i+N}{2} + \frac{i+N}{2} = i + N = i \mod N$ and $\frac{i+N}{2} \in \mathbb{Z}_N$ since i, N are both odd

Then, I want to show the uniqueness of u
Note that if $u + u = i \mod N$, then $u + u = i + kN$ for some $k \in \mathbb{Z}$
Since $0 \le u \le N - 1$, we have $0 \le 2u \le 2N - 2$
So $k = 0$ or 1
Now if $2u = i \mod N$ and $2u' = i \mod N$ for $0 \le u, u' \le N - 1$ but $u \ne u'$, then it has to be the case that one of them has $k = 0$ and the other has $k = 1$
WLOG, we assume $2u = i$ and $2u' = i + N$
Since N is odd, we know that i and $i + N$ are of different parity
However, $2u$ and $2u'$ are both even so there is a contradiction
So it has to be the case that $u = u'$ and the uniqueness part is proven

Now I want to show that there are $\frac{N-1}{2}$ distinct pairs of (u, v) s.t. $u + v = i \mod N$
Note that for a give u, there is a unique v s.t. $u + v = i \mod N$ since $\mathbb{Z}_N$ is a group under addition and we just showed that there is exactly one pair where $v = u$
So there will be $N - 1$ pairs of (u, v) that satisfy the condition $u + v = i \mod N$ but $u \ne v$
But our uniqueness definition allows switching the order of u, v so we have to divide this number by 2 and we have that there are $\frac{N-1}{2}$ distinct pairs

□

Lemma 2. *Suppose N is even, then for all $i \in \mathbb{Z}_N$ which is odd, it has exactly $\frac{N}{2}$ distinct pairs of $(u, v) \in \mathbb{Z}_N \times \mathbb{Z}_N$ s.t. $u + v = i \mod N$ and there is no $u \in \mathbb{Z}_N$ where $u + u = i \mod N$. For all $i \in \mathbb{Z}$ which is even, it has exactly $\frac{N}{2} - 1$ distinct pairs of $(u, v) \in \mathbb{Z}_N \times \mathbb{Z}_N$ s.t. $u + v = i \mod N$ and there are two $u \in \mathbb{Z}_N$ where $u + u = i \mod N$.*

Proof.
First, we consider the case i is odd
Then like the proof of lemma 1, we have $u + u = i \mod N$ if $2u = i$ or $2u = i + N$

However, since i is odd and N is even, the RHS is always odd but the LHS is even so neither could happen
So there are no u s.t. $u + u = i \mod N$

Now I want to show that there are $\frac{N}{2}$ distinct pairs of (u, v) s.t. $u + v = i \mod N$
Note that for a give u, there is a unique v s.t. $u + v = i \mod N$ since $\mathbb{Z}_N$ is a group under addition and we just showed that there is no pair where $v = u$
So there will be N pairs of (u, v) that satisfy the condition $u + v = i \mod N$ but $u \neq v$
But our uniqueness definition allows switching the order of u, v so we have to divide this number by 2 and we have that there are $\frac{N}{2}$ distinct pairs

Then, we consider the case i is even
Our previous proof have shown that there are a maximum of 2 elements where $u + u = i \mod N$ since either $2u = i$ or $2u = i + N$
In the current case, both $\frac{i}{2}$ and $\frac{i+N}{2}$ are in $\mathbb{Z}_N$ so we are done showing that there are exactly 2 u where $u + u = i \mod N$

Finally, I want to show that there are $\frac{N}{2} - 1$ distinct pairs of (u, v) s.t. $u + v = i \mod N$
Note that for a give u, there is a unique v s.t. $u + v = i \mod N$ since $\mathbb{Z}_N$ is a group under addition and we just showed that there are 2 pair where $v = u$
So there will be $N - 2$ pairs of (u, v) that satisfy the condition $u + v = i \mod N$ but $u \neq v$
But our uniqueness definition allows switching the order of u, v so we have to divide this number by 2 and we have that there are $\frac{N}{2} - 1$ distinct pairs

□

Lemma 3. *If N is even, then*

1. *If $i \in [1, N]$ and i is odd, then there are exactly $\frac{i-1}{2}$ distinct (u, v) pairs s.t. $(u, v) \in [1, N] \times [1, N]$, $u + v = i$ and $u \neq v$ and no $u \in [1, N]$ s.t. $u + u = i$*

2. *If $i \in [1, N]$ and i is even, then there are exactly $\frac{i}{2} - 1$ distinct (u, v) pairs s.t. $(u, v) \in [1, N] \times [1, N]$, $u + v = i$ and $u \neq v$ and 1 $u \in [1, N]$ s.t. $u + u = i$*

3. *If $i \in [N + 1, 2N]$ and i is odd, then there are exactly $\frac{i-1}{2} - (i - N - 1)$ distinct (u, v) pairs s.t. $(u, v) \in [1, N] \times [1, N]$, $u + v = i$ and $u \neq v$ and no $u \in [1, N]$ s.t. $u + u = i$*

4. *If $i \in [N + 1, 2N]$ and i is even, then there are exactly $\frac{i}{2} - 1 - (i - N - 1)$ distinct (u, v) pairs s.t. $(u, v) \in [1, N] \times [1, N]$, $u + v = i$ and $u \neq v$ and 1 $u \in [1, N]$ s.t. $u + u = i$*

Proof.

This is a simple book keeping exercise

1. If $i \in [1, N]$ and i is odd, then clearly there is no u where $u + u = i$ since $2u$ has to be even

   Since we are considering (u, v) the same as (v, u), it is enough the count the number of pairs s.t. $u < v$

   We have $1 + (i - 1) = i$, $2 + (i - 2) = i$ all the way to $\frac{i-1}{2} + \frac{i+1}{2} = i$

   So there are a total of $\frac{i-1}{2}$ pairs

2. If $i \in [1, N]$ and i is even, then clearly there is one u where $u + u = i$ which is $u = \frac{i}{2}$

   Since we are considering (u, v) the same as (v, u), it is enough the count the number of pairs s.t. $u < v$

   We have $1 + (i - 1) = i$, $2 + (i - 2) = i$ all the way to $\frac{i-1}{2} + \frac{i+1}{2} = i$

   So there are a total of $\frac{i}{2} - 1$ pairs

3. If $i \in [N + 1, 2N]$ and i is odd, then clearly there is no u where $u + u = i$ since $2u$ has to be even

   Since we are considering (u, v) the same as (v, u), it is enough the count the number of pairs s.t. $u < v$

   We still have $1 + (i - 1) = i$, $2 + (i - 2) = i$ all the way to $\frac{i-1}{2} + \frac{i+1}{2} = i$ but the first few cases may have $v > N$

   So we can only start from $(i - N) + N = i$ and the first $i - N - 1$ pairs are dropped

   So there are a total of $\frac{i-1}{2} - (i - N - 1)$ pairs

4. If $i \in [N + 1, 2N]$ and i is even, then clearly there is one u where $u + u = i$ which is $u = \frac{i}{2}$

   Since we are considering (u, v) the same as (v, u), it is enough the count the number of pairs s.t. $u < v$

   We still have $1 + (i - 1) = i$, $2 + (i - 2) = i$ all the way to $\frac{i}{2} - 1 + \frac{i}{2} + 1 = i$ but the first few cases may have $v > N$

   So we can only start from $(i - N) + N = i$ and the first $i - N - 1$ pairs are dropped

   So there are a total of $\frac{i}{2} - 1 - (i - N - 1)$ pairs

□

Lemma 4. *If N is an odd prime, then $e(i) = 0$ or 2 where $e(i)$ is the square root counting formula*

Proof.

Again, we assume that every number is written in canonical form

It is enough to show that if $e(i) \neq 0$, then $e(i) = 2$

Suppose $x^2 = i \mod N$, then $(-x)^2 = i$

Note that $x \neq -x$ since otherwise, we would have $-x + N = x$ and thus $N = 2x$ which contradicts with

N being odd prime

This shows that $e(i) \geq 2$

Now suppose y is another solution, i.e. $y^2 = i \mod N$

Then $x^2 - y^2 = 0 \mod N$

Since $\mathbb{Z}_N$ is abelian, we have $(x-y)(x+y) = 0 \mod N$

Since N is odd prime, we know that there is no zero divisors so either $x - y = 0 \mod N$ or $x + y = 0 \mod N$

In the first case, we have $x = y$ and in the second case we have $x = -y$ so y is already counted

So there will be only 2 solutions which is x and $-x$

□

Lemma 5. *Consider the divisor counting Function $d(x)$ in the setting of $\mathbb{Z}_N$, then $\sum_{i=0}^{N-1} d(i) = \frac{N(N+1)}{2}$*

Proof.

Note that $\sum_{i=0}^{N-1} d(i)$ is essentially summing up the number of divisors of all the elements in $\mathbb{Z}_N$

All pair (α, β) where $\alpha, \beta \in \mathbb{Z}_N$ and $\alpha \leq \beta$, we will have $\alpha \cdot \beta \in \mathbb{Z}_N$ so it will be counted exactly once when we sum up all the divisor pairs

So $\sum_{i=0}^{N-1} d(i)$ is essentially the number of (α, β) pairs where $\alpha, \beta \in \mathbb{Z}_N$ and $\alpha \leq \beta$

This is a simple combinatoric problem and the answer is $\frac{N(N+1)}{2}$ so we are done □

Lemma 6. *In the setting of $\mathbb{Z}_N$ where N is an odd prime and $d(x)$ is the divisor counting function, then $d(0) = N$ and $\max_{i \in \mathbb{Z}_N, i \neq 0} d(i) = \frac{N-1}{2}$*

Proof.

First I want to show that $d(0) = N$

It is know that when N is an odd prime, $\mathbb{Z}_N$ has no zero divisors, i.e. there is no $\alpha, \beta \in \mathbb{Z}_N$ where $\alpha, \beta \neq 0$ but $\alpha \cdot \beta = 0 \mod N$

It is also know that $\alpha \cdot 0 = 0 \mod N$ for all α and this will be the only possible case

There are a total of N choices of α so there are N divisors of 0

Now I want to show that $\max_{i \in \mathbb{Z}_N, i \neq 0} d(i) = \frac{N-1}{2}$

Since N is an odd prime, we know that $\mathbb{Z}_N$ is a group under multiplication

So if we fix i, then for all $\alpha \in \mathbb{Z}_N \backslash \{0\}$ there is a unique β s.t. $\alpha\beta = i$

By lemma 4, we know that the number of "square root" of i is either 0 or 2

There are a total of $N - 1$ numbers in $\mathbb{Z}_N$ so if there is no square root, then there will be $\frac{N-1}{2}$ distinct pairs of (α, β) whose product if i disregarding the order

If there are 2 square root, then there will be $1 + 1 + \frac{N-3}{2} = \frac{N+1}{2}$ distinct pairs of (α, β) whose product if i disregarding the order

So the maximum we could have for $d(i)$ is $\max(\frac{N-1}{2}, \frac{N+1}{2}) = \frac{N+1}{2}$ □

Lemma 7. *Consider the divisor counting Function $D(x)$ in the setting of $\mathbb{Z}$, then* $\sum_{i=1}^{N^2} D(i) = \frac{N(N+1)}{2}$

Proof.

Similar to lemma 5, $\sum_{i=1}^{N^2} D(i)$ is essentially summing up the number of pairs of (α, β) where $\alpha, \beta \in \{1, \dots, N\}$, $\alpha \leq \beta$ and $\alpha\beta \leq N^2$

This is again a simple combinatoric problem and the answer is $\frac{N(N+1)}{2}$ so we are done □

6.2 Simulation Code

All the simulation done in this thesis is written in Python. In all four cases, the code has high similarity besides from the operator used (plus or times) and the setting (whether need to calculate mod N) so I only included the $A \cdot A$ in $\mathbb{Z}_N$ as an example. This is a paralleled version that speeds up the simulation by running the program across cpu cores simultaneously.

```
# -*- coding: utf-8 -*-
"""
Created on Sat Apr  3 14:50:59 2021

@author: danny
"""


import numpy as np
from scipy.spatial.distance import pdist
import pandas as pd
from matplotlib import pyplot as plt
from multiprocessing import cpu_count
from multiprocessing import Pool
import os


class Function:
    def __init__(self, N):
        self.N = N


    def get_size(self, p):
        result = []
        for i in range(100):
            chosen_index = np.random.binomial(1, p, self.N)
            A = np.unique(chosen_index * range(1, self.N+1)) - 1
            A = A[A>=0]
            AplusA = np.unique(pdist(A.reshape((-1,1)),metric = lambda x,y: x*y)) % self.N
            AplusA = np.append(AplusA, A**2 % self.N)
            AplusA = np.unique(AplusA)
            result.append(len(AplusA))
        print(p)
        return np.mean(result)
def main():
    N_list = [19,20,47,50,97,100,241,250,499,500]
    plt.figure()
    for N in N_list:
        p_list = np.linspace(0,1,50)
        if not os.path.isfile("Data/AtimesA_"+str(N)+"_result.csv"):
            function = Function(N)
            pool = Pool(cpu_count())
            results = pool.starmap(function.get_size, p_list.reshape(-1,1))
            pool.close()
            pool.join()
            plt.plot(p_list, np.array(results)/N, label = str(N))
            results = pd.DataFrame(np.array(results)/N, columns = ["portion"])
            results.to_csv("Data/AtimesA_"+str(N)+"_result.csv")
        else:
            results = pd.read_csv("Data/AtimesA_"+str(N)+"_result.csv", index_col = 0)
            plt.plot(p_list, results.values, label = str(N))
    plt.xlabel("p")
    plt.ylabel("|A*A|/N")
    plt.legend()
    plt.title("Relationship between p and |A*A|")
    plt.savefig("Figure/AtimesA_modN.png")

if __name__ == "__main__":
    main()
```

References

[1] P. Erdős and E. Szemerédi, "On sums and products of integers," in *Studies in pure mathematics.* Springer, 1983, pp. 213–218.

[2] M. Nathanson, "On sums and products of integers," *Proceedings of the American Mathematical Society*, vol. 125, no. 1, pp. 9–16, 1997.

[3] K. Ford, "Sums and products from a finite set of real numbers," *The Ramanujan Journal*, vol. 2, no. 1, pp. 59–66, 1998.

[4] M.-C. Chang, "A sum-product estimate in algebraic division algebras," *Israel Journal of Mathematics*, vol. 150, no. 1, pp. 369–380, 2005.

[5] G. Elekes, "On the number of sums and products," *Acta Arithmetica*, vol. 81, no. 4, pp. 365–367, 1997.

[6] C. D. Tóth, "The szemerédi-trotter theorem in the complex plane," *Combinatorica*, vol. 35, no. 1, pp. 95–126, 2015.

[7] J. Solymosi, "On sum-sets and product-sets of complex numbers," *Journal de théorie des nombres de Bordeaux*, vol. 17, no. 3, pp. 921–924, 2005.

[8] ——, "On the number of sums and products," *Bulletin of the London Mathematical Society*, vol. 37, no. 4, pp. 491–494, 2005.

[9] J. Bourgain, A. A. Glibichuk, and S. V. Konyagin, "Estimates for the number of sums and products and for exponential sums in fields of prime order," *Journal of the London Mathematical Society*, vol. 73, no. 2, pp. 380–398, 2006.

[10] J. Bourgain, N. Katz, and T. Tao, "A sum-product estimate in finite fields, and applications," *Geometric & Functional Analysis GAFA*, vol. 14, no. 1, pp. 27–57, 2004.

[11] D. Hart, A. Iosevich, and J. Solymosi, "Sum-product estimates in finite fields via kloosterman sums," *International Mathematics Research Notices*, vol. 2007, no. 9, pp. rnm007–rnm007, 2007.

[12] W. D. Stangl, "Counting squares in zn," *Mathematics Magazine*, vol. 69, no. 4, pp. 285–289, 1996.

[13] T. M. Apostol, *Introduction to analytic number theory.* Springer Science & Business Media, 1998.

[14] G. H. Hardy, E. M. Wright *et al.*, *An introduction to the theory of numbers.* Oxford university press, 1979.