Some note on randomly choosing two integers to see if they are relatively prime

Stefanus[1]

[1] Samsung Semiconductor Inc
San Jose, CA 95134 USA

(Dated: December 25, 2018)

Abstract

Just for fun :)

On my longer than usual flight I had recently I encountered this problem on Shoup's ntb-v1 page 98 (number theory book) on the Chapter on probability. The question was

Question, show that the probability of choosing two random positive integers $\leq n$ is ≥ 0.25

Answer. To get that 0.25 we derive inspiration from apostol lattice derivation for the probability of two integers being co-prime when selected at random.

Consider n^2 lattice points with corners forming a square from $(1,1)$ (lower left) to (n,n) (upper right). Each lattice point is either visible from the origin or not, note that (a,b) is visible from the origin if and only if $\gcd(a,b) = 1$.

So we can think of the lattice points inside the square that are visible from the origin as generators of the rest of the lattice points inside the square because if a point is not visible from the origin then there must be a visible point that blocks it (there might be other points blocking it too for example $(3,3)$ is blocked by $(2,2)$ and $(1,1)$ but $(1,1)$ is the generator).

Another thing to note is that each point has a unique generator because if two generators generate the same point $(g,h) = r(a,b) = s(c,d)$ then $ra = sc, rb = sd \rightarrow \frac{a}{b} = \frac{c}{d}$ but since each generator is co-prime the fraction is already reduced and therefore $a = c, b = d$.

Now suppose that (a,b) is a generator, how many points does it generate inside the square? Assume without loss of generality that $b > a$ ($b = a$ is only possible if $b = a = 1$) then the points generated by (a,b) is $z(a,b)$ so as long as $zb \leq n$ we are good this means that there are $\left\lfloor \frac{n}{b} \right\rfloor$ points generated by (a,b) including itself.

So if we only have $0.25n^2$ generators how many points can we generate? from the argument above we must have the generator to be as small as possible, *i.e.* the larger coordinate of the generator to be as small as possible, to generate as many points as possible.

To get this maximization we need the generators to be at the lower left quadrant, *i.e.* the smaller square of size $\frac{n}{2}$ by $\frac{n}{2}$. This will guarantee that we have the smallest larger coordinate for each generator.

Note that in this hypothetical configuration we can have two generators generating the same point but for now we just want to see the upper bound of the points generated by these generators so it's OK.

So, including the generators themselves, how many points can we generate (at most)? the answer is

$$t = \sum_{k=1}^{n/2} (2k-1) \left\lfloor \frac{n}{k} \right\rfloor$$

where we denote the total as t, note that the upper limit of the sum might not make sense if n is odd but here we are only interested in the upper bound so it's OK. TO see why we get this formula note that there's only one generator whole larger coordinate is 1 *i.e.* $(1,1)$ and then there are 3 generators whose larger coordinate is 2, *i.e.* $(2,1), (2,2), (1,2)$ and so on (if you think about it visually they are just the outer layer of the $\frac{n}{2}$ by $\frac{n}{2}$ square).

We can now do what apostol did to estimate the floor

$$\lfloor x \rfloor = x - \{x\}$$

and so

$$\begin{aligned}
\sum_{k=1}^{n/2} (2k-1) \left\lfloor \frac{n}{k} \right\rfloor &= \sum_{k=1}^{n/2} (2k-1) \left[\frac{n}{k} - \left\{ \frac{n}{k} \right\} \right] \\
&= \sum_{k=1}^{n/2} (2k-1) \frac{n}{k} - \sum_{k=1}^{n/2} (2k-1) \left\{ \frac{n}{k} \right\} \\
&= \sum_{k=1}^{n/2} n - \sum_{k=1}^{n/2} \frac{n}{k} - \sum_{k=1}^{n/2} (2k-1) \left\{ \frac{n}{k} \right\}
\end{aligned}$$

because we want to maximize the LHS we can just ignore the last term now we need to minimize the second term we can do this by using an integral

$$\begin{aligned}
\sum_{k=1}^{n/2} \frac{1}{k} &> \int_1^{n/2+1} \frac{1}{k} dk \\
&= \log(\tfrac{n}{2} + 1)
\end{aligned}$$

Thus

$$t = \sum_{k=1}^{n/2} (2k-1) \left\lfloor \frac{n}{k} \right\rfloor < n^2 - n \log(\tfrac{n}{2} + 1)$$

and $\frac{t}{n^2} \to 1$ only when $n \to \infty$ thus we must have at least $0.25n^2$ generators which means that the probability of picking two co-prime integers at random from the set $\{1, 2, \ldots, n\}$

is at least 0.25. Of course if $n = 1$ the upper bound of t given above doesn't make sense but the argument itself doesn't make sense for $n = 1$.

Alternative solution #1 The following is the solution when $n \to \infty$, taken from

https://www.cut-the-knot.org/m/Probability/TwoCoprime.shtml but I'll modify it to suit finite n.

Let q denote the sought probability of two random integers being co-prime. Pick an integer k and two random numbers a and b. The probability that k divides a is $\frac{1}{k}$, and the same holds for b.

But that only works if $n \to \infty$ if n is finite say for example $n = 10$ and $k = 3$ then the probability of getting a multiple of 3 is $\frac{3}{10}$ so for a finite n the probability that k divides a is

$$\frac{1}{K} = \left\lfloor \frac{n}{k} \right\rfloor \frac{1}{n}$$

Therefore, the probability that both a and b are divisible by k equals $\frac{1}{K^2}$. The probability that a and b have no other factors, i.e., that $\gcd\left(\frac{a}{k}, \frac{b}{k}\right) = 1$ equals q, by our initial assumption. But $\gcd\left(\frac{a}{k}, \frac{b}{k}\right) = 1$ is equivalent to $gcd(a, b) = k$. Assuming independence of the events, it follows that the probability that $gcd(a, b) = k$ equals $\frac{1}{K^2} \cdot q = \frac{q}{K^2}$.

Now, k was just one possibility for the greatest common divisor of two random numbers. Any number could be the $gcd(a, b)$. Furthermore, since the events $gcd(a, b)$ are mutually exclusive (the gcd of two numbers is unique) and the total probability of having a gcd at all is 1 leads to:

$$1 = \sum_{k=1}^{n} \frac{q}{K^2},$$

implying that

$$q = \left[\sum_{k=1}^{n} \frac{1}{K^2} \right]^{-1}$$

We now need to expand K

$$\sum_{k=1}^{n} \frac{1}{K^2} = \frac{1}{n^2} \sum_{k=1}^{n} \left\lfloor \frac{n}{k} \right\rfloor^2$$

we now use what apostol did to estimate the error of floors

$$\lfloor x \rfloor = x - \{x\}$$

where $0 \leq \{x\} < 1$ so that

$$\begin{aligned}\frac{1}{n^2}\sum_{k=1}^{n}\left\lfloor\frac{n}{k}\right\rfloor^2 &= \frac{1}{n^2}\sum_{k=1}^{n}\left(\frac{n}{k}\right)^2 - 2\left\{\frac{n}{k}\right\}\left(\frac{n}{k}\right) + \left\{\frac{n}{k}\right\}^2 \\ &= \sum_{k=1}^{n}\frac{1}{k^2} - 2\frac{1}{n}\sum_{k=1}^{n}\left\{\frac{n}{k}\right\}\frac{1}{k} + \frac{1}{n^2}\sum_{k=1}^{n}\left\{\frac{n}{k}\right\}^2\end{aligned}$$

Since the probability is given by $\left[\frac{1}{n^2}\sum_{k=1}^{n}\left\lfloor\frac{n}{k}\right\rfloor^2\right]^{-1}$ we want to maximize the above so for the second term because of the negative sign in front of it we can set $\left\{\frac{n}{k}\right\} = 0$ and the third term we just take $\left\{\frac{n}{k}\right\} = 1$ therefore

$$\sum_{k=1}^{n}\frac{1}{k^2} - 2\frac{1}{n}\sum_{k=1}^{n}\left\{\frac{n}{k}\right\}\frac{1}{k} + \frac{1}{n^2}\sum_{k=1}^{n}\left\{\frac{n}{k}\right\}^2 < \sum_{k=1}^{\infty}\frac{1}{k^2} - 0 + \frac{1}{n}$$

and the max for $\frac{1}{n}$ is when $n = 1$ and for the first term the max is when $n \to \infty$ which means we get $\frac{\pi^2}{6}$ so

$$q = \left[\sum_{k=1}^{n}\frac{1}{K^2}\right]^{-1} > \frac{1}{\frac{\pi^2}{6}+1} > 0.37 > 0.25$$

For $n = \infty$ we just replace $\frac{1}{K}$ with $\frac{1}{k}$ and we get everything else like normal.

Graveyard of dead Ideas and Stuffs

1. My first idea involving lattice points is to show that for every 4 lattice points at least one of them is visible from the origin but a short python program shows that this is not the case from a grid of 1000 by 1000 lattice points one finds 2136 consecutive lattice points where all of them have gcd bigger than 1.

   But the other way round is obvious, *i.e.* there are no 4 consecutive lattice points where all of them are co-prime because at least one of them is even.

   But one interesting thing I notice is

   Question: For a 4 consecutive lattice points (a, b), $(a+m, b)$, $(a, b+m)$, $(a+m, b+m)$ if m is odd then there's no consecutive points where all of them are co-prime and if

m is even there will be consecutive points where all of them are co-prime. This is very easy to prove

Answer. For m odd it is obvious because one of them will be even even but m even we can easily construct such consecutive points, start with the upper right corner $(m+1, m+2)$ this is definitely co-prime the lower left would then be $(1,2)$ also co-prime, the lower right is $(m+1,2)$ which is co-prime because m is even and so $m+1$ is odd and the upper left will be $(1, m+2)$ which is co-prime.

Question:. Another easy question is, given any co-prime point (a,b) is there an m such that all four consecutive points $(a,b),(a+m,b),(a,b+m),(a+m,b+m)$ are co-prime?

Answer. The answer is no, and it's very easy to construct a counter example. Suppose there is such an m then we can pick a point $(1, m(m+1))$ which is obviously co-prime the next point to the right will be $(1+m, m(m+1))$ which is not co-prime.

Another way would be $(m+1, m(m+m+1))$ (a co-prime point) and the next point will be $(m+m+1, m(m+m+1))$ which is not co-prime and there are many other ways.

Question:. The next question will be is there a geometric shape (rather than the square $(a,b),(a+m,b),(a,b+m),(a+m,b+m)$) such that if whenever a point in the shape is co-prime every other point is also co-prime

Answer. We know a straight line (vertical or horizontal) won't do because of the above this includes a diagonal line and this is because of $(1,1)$. For lines with other slopes they won't do either and it is very simple to see. Say we want to make lines with gradient (m,n) so for a point (a,b) (co-prime) we get $(a,b),(a+m,b+n)$.

Suppose that there is m,n such that for every co-prime (a,b) we also have $(a+m,b+n)$ co-prime. Suppose (m,n) itself is co-prime then $(m+m,n+n)$ is definitely not. Now suppose that $d = \gcd(m,n)$ then $(m/d, n/d)$ will be a co-prime point and if we make the line $(m/d,n/d),(m/d+m,n/d+n)$ we will have $(m/d+m,n/d+n)$ not co-prime for obvious reason :)

Thus there is no such shape such that for any co-prime point the rest of the points in the shape will be co-prime as well. But of course there will be certain points where all points in the shape will be co-prime, however, you cannot just choose a random co-prime point, put any point of your shape on that point and expect all other points in the shape will be co-prime as well.

Question:. Next, how about the largest square area where each lattice points (not just the corner points) are all co-prime (or all not co-prime), is there a limit or given any n you can find such area?

Answer. For all co-prime it's a simple answer there's no such thing because we will hit even even lattice points thus no such area exists for any n.

For all non co-prime, from playing around with Python we need quite a big start for $n = 3$ to get all 9 points to be non co-prime for $n = 4$ I haven't seen any but this might be because my starting point is not big enough. Because we need to have numbers who have each others' factors.

Then I was reminded by a theorem in apostol on chapter 5 about congruences, and there it was, it says that given any k there's always a square (of lattice points) with sides k all of which is non co-prime. The prove uses Chinese Reminder Theorem in a very smart way, please see page 118.

I will not repeat the proof here but it gives a new way of generating consecutive composite numbers besides the $(N+1)! + 2 \to (N+1)! + N + 1$ trick. This time the difference is we can choose what factors the consecutive numbers have.

First, set up a Chinese Remainder system with

$$x \equiv 0 - i + 1 \pmod{p_i^2}$$

where i takes the values $1, 2, \ldots, N$ and p_i is the i^{th} prime number, it's p^2 instead of p to make sure that it's a composite number (although we might get away with just using 2^2 and just p for the rest), for example for $N = 3$ we have

$$x \equiv 0 \pmod{2^2}$$
$$x \equiv -1 \pmod{3^2}$$
$$x \equiv -2 \pmod{5^2}$$

therefore $x + 1$ is a multiple of 3^2 and $x + 2$ is a multiple of 5^2 thus we have three consecutive numbers each divisible by $2^2, 3^2, 5^2$ respectively and this way we can generate N consecutive composite numbers. This generates interestingly a much larger number compared to the $(N + 1)!$ method. Here's a python code to calculate Chinese Remainder system

```
from fractions import gcd
from math import factorial

def extended_gcd(a,b):
    if (b == 0):
# if b =0, then return g =a, m=1, n=0
        #print (a,"\t",b,"\t",a,"\t",1,"\t",0)
        return (a,1,0)
    else:
# if b is not 0, then recursively call the function to get the value of
# g,m,n at each step. The code also displays the value of a,b at each step
        (g,m,n) = extended_gcd(b,a%b)
        #print a,"\t",b,"\t",g,"\t",n,"\t",m-(a//b)*n
        return(g,n,m-(a//b)*n)

def Chinese(vals, mods):
    if not mods or not vals or len(mods) != len(vals):
        return
    m = mods[0]
    v = vals[0]

    for r in range(1,len(mods)):

        if gcd(m, mods[r]) > 1:
            return

        (g,i,j) = extended_gcd(m, mods[r])
        t = i*(vals[r] - v) % mods[r]
        x = (v + t*m) % (m * mods[r])

        m = m * mods[r]
        v = x

    return x

pp = [2**2,3**2,5**2,7**2,11**2,13**2,17**2,19**2,23**2,29**2,31**2,37**2,41**2,43

def Compare(N):
    for i in range(2,N+1):
        print(factorial(i+1), Chinese(range(0,-i,-1),pp[:i]))
```

This also gives an idea on how to generate all prime numbers albeit inefficiently :) We can set up a Chinese Remainder system with the first n primes like so $x \equiv a_i \pmod{p_i}$ where $a_i \not\equiv 0, \not\equiv 1 \pmod{p_i}$. This will generate a bunch of new numbers not divisible by any of hte p_i.

The smallest of these numbers (mod $\prod_{i=1}^{n} p_i$) must be a prime number and it must be the next prime number after p_n. This is because if it's composite it must be divisible by one of the p_i or it can be a power of a new prime number.

But any number in the range $[1, \prod_{i=1}^{n} p_i]$ can be mapped into a Chinese Remainder map $\pmod{p_i}$ bijectively. Therefore if a power of a new prime number is in the range then the prime number itself must be in the range therefore the smallest number $\not\equiv 0 \pmod{p_i}$ is a new prime number.

We then repeat the process with our latest recruit. In a way this is just the sieve method in disguise :)

The question now is whether we can find the smallest number $\not\equiv 0 \pmod{p_i}$ in an efficient way.

Another idea is to use the fact that at least $0.25n^2$ are visible and then to proceed as follows. Assume we have four consecutive lattice points and all of them are not co-prime then from Bezout we get

$$\begin{aligned} am_1 + bn_1 &= d_1 \\ (a+1)m_2 + bn_2 &= d_2 \\ am_3 + (b+1)n_3 &= d_3 \\ (a+1)m_4 + (b+1)n_4 &= d_4 \end{aligned}$$

But any of the two d's are co-prime because otherwise they divide both a and $(a+1)$ (or b and $(b+1)$) and obviously $d < a, b$ therefore we generate 12 lattice points that are co-prime so every 4 non co-prime consecutive points generate 12 co-prime lattice points.

So my strategy was to show that if we have $0.25n^2$ co-prime lattice points then the non co-prime ones will generate more than $0.25n^2$ co-prime lattice points but this was quite hard to do.

2. My first idea was to use the proof for $\frac{6}{\pi^2}$. First, the number of pairs of positive integers $\leq n$ with a *common factor* a prime p is $\left\lfloor \frac{n}{p} \right\rfloor$. Note that we are talking about common factor here **not** greatest common factor. I want then to establish a one to one correspondence between these pairs are co-prime pairs. However, we don't usually get one to one mapping, *e.g.* $(1, 2), (2, 4), (3, 6)$ all map to $(1, 2)$.

So I devised a plan to generate a one to one mapping, from Bezout we know that $\gcd(a,b)=1$ means that $ax+by=1$ and there are other x, y, $x'=x-c\frac{b}{d}, y'=y+c\frac{a}{d}$, so we can always decrease y until it is less than a. So we can map a redundant one to (a,r) where r is $y \mod a$.

The problem is that we might not have enough r. We can also map it to (r',b) where $r'=x \mod b$ because we can also decrease x but then again we still might not have enough slots to copy the redundant ones.

The thing is that if we think in terms lattice points again, in any row say $y=a$ the number of lattice points that are co-prime is $\varphi(a)$. But the number of elements in the row, *i.e.* n, might not be a multiple of $\varphi(a)$ so it's not guaranteed that we can map redundant ones to empty slots (a,r) or (r',b).

3. We can also try to modify the proof for $\frac{6}{\pi^2}$ as follows. The number of positive integers $\leq n$ that is divisible by p is $\left\lfloor\frac{n}{p}\right\rfloor$ so the number of pairs of integers divisible by p is $\left\lfloor\frac{n}{p}\right\rfloor^2$. Note that p is just a common divisor not the greatest common divisor.

   We now need to use the principle of cross classification to count the number of co-prime pairs

$$n^2-\sum_{p}\left\lfloor\frac{n}{p}\right\rfloor^2+\sum_{p,q}\left\lfloor\frac{n}{pq}\right\rfloor^2-\sum_{p,q,r}\left\lfloor\frac{n}{pqr}\right\rfloor^2+\dots$$

   So first we remove all pairs that have one prime as a common factor (note again that this is just a common factor not the greatest common factor) but then we remove too many because we removed the multiple of pq twice and so on.

For now let's forget about the floor thing and just write everything as

$$\begin{aligned}
&n^2 - \sum_p \left(\frac{n}{p}\right)^2 + \sum_{p,q} \left(\frac{n}{pq}\right)^2 - \sum_{p,q,r} \left(\frac{n}{pqr}\right)^2 + \dots \\
&= n^2 \left[1 - \sum_p \left(\frac{1}{p}\right)^2 + \sum_{p,q} \left(\frac{1}{pq}\right)^2 - \sum_{p,q,r} \left(\frac{1}{pqr}\right)^2 + \dots\right] \\
&= n^2 \prod_p \left(1 - \frac{1}{p^2}\right) \\
&= n^2 \, \frac{1}{\prod_p \left(1 - p^{-2}\right)^{-1}}
\end{aligned}$$

But $\frac{1}{\prod_p (1-p^{-2})^{-1}} = \frac{1}{\sum_k \frac{1}{k^2}} = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}$ so that's how we get $\frac{6}{\pi^2}$ to be the probability of getting two random integers to be co-prime.

Now going back to the original formula with floors and stuff, if n is finite then each sum terminates and the whole thing is bounded.

When taking the limit $n \to \infty$ we can do what apostol does for floors

$$\lfloor x \rfloor = x - \{x\}$$

so each sum gives us

$$\sum_p \left(\frac{n}{p}\right)^2 - 2\left\{\frac{n}{p}\right\}\left(\frac{n}{p}\right) + \left\{\frac{n}{p}\right\}^2$$

now $\sum_{p \le x} \frac{1}{p} = O(\log\log x)$ and $\sum_{p \le x} \{x/p\}^2 = O(x/\log x)$ so what we have is

$$\begin{aligned}
\frac{1}{n^2} \sum_p -2\left\{\frac{n}{p}\right\}\left(\frac{n}{p}\right) + \left\{\frac{n}{p}\right\}^2 &= \frac{1}{n^2} O(\log\log n) + \frac{1}{n^2} O\left(\frac{n}{\log n}\right) \\
&= \frac{1}{n^2} O\left(\frac{n}{\log n}\right)
\end{aligned}$$

The problem now is that we have infinite of these errors. One from each of the infinite sums $-\sum_p + \sum_{p,q} - \sum_{p,q,r} + \dots$. So even though each error tends to zero an infinite of them might not be zero.

But I think we might guesstimate the whole error as such

$$O\left(\frac{n \cdot n}{n^2 \log n}\right) = O\left(\frac{1}{\log n}\right)$$

This is because sum of the sums $-\sum_p + \sum_{p,q} - \sum_{p,q,r} + \dots$ were initially zero, *e.g.* if $n = 5$ the second sum is already zero because pq starts with $2 \cdot 3 = 6$. So the number of sums increases as we increase n and it actually increases slower than n because we are considering the product of distinct primes thus the approximation above is sufficient.

Now that we have dealt with the error we can see that $\frac{6}{\pi^2}$ is the lower bound because for finite n we have some of the prime missing from $\prod_p (1 - p^{-1})^{-2}$ which makes $\zeta(2)$ smaller but $\zeta(2)$ is in the denominator so for finite n the probability is higher than $\frac{6}{\pi^2} > 0.25$.

But as you can see there are problems in terms of estimating the errors and getting back from $n \to \infty$ to n finite because we had to add back the error. And also you can see that in this case it is easier to use $\sum_k 1/k^2$ rather than $\prod(1 - 1/p^2)$ just like what was done in Alternative Solution 1 above.

More on the errors, going from $n \to n + 1$ what might happen is that a new prime $p | n + 1$ but $p \not| n$