Why Integer Factorising cannot be polynomial-time

Bhupinder Singh Anand

Update of June 29, 2016

Abstract. We show how the usual, linearly displayed, Eratosthenes sieve argument reveals the logical structure of divisibility when displayed as a 2-dimensional matrix representation of the residues $r_i(n)$, defined for all $n \geq 2$ and all $i \geq 2$ by $n + r_i(n) \equiv 0\ (mod\ i)$, where $i > r_i(n) \geq 0$, and the residues $r_i(n)$ can be viewed in two essentially different ways. (a) First as the values, for any given i, of a function $R_i(n)$ over the domain N of the natural numbers. Since we cannot define a probability function for the probability that a random n is prime over the probability space $(1, 2, 3, \dots,)$, this definition does not admit an argument which will allow us to conclude that the prime divisors of any integer n are independent. (b) Second as the values, for any given n, of the sequence $E(n) = \{r_i(n) : i \geq 1\}$. This now allows us to define a probability model from which we may conclude for any given $n > 1$, and any given prime $p > 1$, that the probability of the event $r_p(n) = 0$—whence p divides n—is $\frac{1}{p}$; and that the probability of the event $r_p(n) \neq 0$—whence p does not divide n—is $1 - \frac{1}{p}$. This allow us to conclude that, given $p, q > 1$ are two unequal primes, the compound probability that $r_p(n) = 0$ and $r_q(n) = 0$—whence both p and q divide n—is $\frac{1}{pq}$. We conclude that the prime divisors of any integer n are mutually independent. An immediate consequence is that integer factorising cannot be polynomial-time.

Keywords. complete system of incongruent residues, computational complexity, Eratosthenes sieve, integer factorising, mutually independent prime divisors, polynomial time algorithm, probability model, sample space.

2010 Mathematics Subject Classification. 03D15, 11A07, 11A51, 11Y05, 11Y11, 11Y16, 68Q15

Contents

1. Introduction: A 2-dimensional view of Eratosthenes sieve

We show how the usual, linearly displayed, Eratosthenes sieve argument reveals the logical structure of divisibility (ipso facto, of primality) when displayed as a 2-dimensional matrix representation of the residues $r_i(n)$, defined for all $n \geq 2$ and all $i \geq 2$ by:

$$n + r_i(n) \equiv 0\ (mod\ i), \text{ where } i > r_i(n) \geq 0.$$

'Density': For instance, the residues $r_i(n)$ can be defined for all $n \geq 1$ as the values of the non-terminating sequences $R_i(n) = \{i-1,\ i-2,\ \dots,\ 0,\ i-1,\ i-2,\ \dots,\ 0,\ \dots\}$, defined for all $i \geq 1$ (as illustrated below[1]).

Sequence:	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	R_{11}	. . .	R_n
$n = 1$	0	1	2	3	4	5	6	7	8	9	10	. . .	n-1
$n = 2$	0	0	1	2	3	4	5	6	7	8	9	. . .	n-2
$n = 3$	0	1	0	1	2	3	4	5	6	7	8	. . .	n-3
$n = 4$	0	0	2	0	1	2	3	4	5	6	7	. . .	n-4
$n = 5$	0	1	1	3	0	1	2	3	4	5	6	. . .	n-5
$n = 6$	0	0	0	2	4	0	1	2	3	4	5	. . .	n-6
$n = 7$	0	1	2	1	3	5	0	1	2	3	4	. . .	n-7
$n = 8$	0	0	1	0	2	4	6	0	1	2	3	. . .	n-8
$n = 9$	0	1	0	3	1	3	5	7	0	1	2	. . .	n-9

[1]For r_i read $r_i(n)$; for R_i read $R_i(n)$ in the following table.

$n = 10$	0	0	2	2	0	2	4	6	8	0	1	...	n-10
$n = 11$	0	1	1	1	4	1	3	5	7	9	0	...	n-11
n	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	...	0

We note that:

• For any $i \geq 2$, the non-terminating sequence $R_i(n)$ cycles through the values $(i-1,\ i-2,\ \ldots,\ 0)$ with period i;

• For any $i \geq 2$ the 'density'[2]—over the set of natural numbers—of the set $\{n\}$ of integers that are divisible by i is $\frac{1}{i}$; and the 'density' of integers that are not divisible by i is $\frac{i-1}{i}$.

Primality: The residues $r_i(n)$ can be alternatively defined for all $i \geq 1$ as values of the non-terminating sequences, $E(n) = \{r_i(n) : i \geq 1\}$, defined for all $n \geq 1$ (as illustrated below[3]).

Sequence:	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	R_{11}	...	R_n
E(1):	*0*	1	2	3	4	5	6	7	8	9	10	...	n-1
E(2):	*0*	*0*	1	2	3	4	5	6	7	8	9	...	n-2
E(3):	*0*	**1**	*0*	1	2	3	4	5	6	7	8	...	n-3
E(4):	*0*	*0*	2	*0*	1	2	3	4	5	6	7	...	n-4
E(5):	*0*	**1**	**1**	**3**	*0*	1	2	3	4	5	6	...	n-5
E(6):	*0*	*0*	*0*	2	4	*0*	1	2	3	4	5	...	n-6
E(7):	*0*	**1**	**2**	**1**	**3**	**5**	*0*	1	2	3	4	...	n-7
E(8):	*0*	*0*	1	*0*	2	4	6	*0*	1	2	3	...	n-8
E(9):	*0*	1	*0*	3	1	3	5	7	*0*	1	2	...	n-9
E(10):	*0*	*0*	2	2	*0*	2	4	6	8	*0*	1	...	n-10
E(11):	*0*	**1**	**1**	**1**	**4**	**1**	**3**	**5**	**7**	**9**	*0*	...	n-11
...													
E(n):	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	...	*0*

We note that:

• The non-terminating sequences $E(n)$ highlighted in **bold** correspond to a prime[4] p (since $r_i(p) \neq 0$ for any $1 < i < p$) in the usual, linearly displayed, Eratosthenes sieve:

E(1), **E(2)**, **E(3)**, *E(4)*, **E(5)**, *E(6)*, **E(7)**, *E(8)*, *E(9)*, *E(10)*, **E(11)**, ...

• The non-terminating sequences $E(n)$ highlighted in *italics* identify a crossed out composite n (since $r_i(n) = 0$ for some $1 < i < n$) in the usual, linearly displayed, Eratosthenes sieve.

1.A. Introduction: The prime divisors of n are mutually independent

The residues $r_i(n)$ can thus be viewed in two essentially different ways.

> (a) First as the values, for any given i, of a function $R_i(n)$ over the domain N of the natural numbers. Classically, since we cannot define a probability function for the probability that a random n is prime over the probability space $(1, 2, 3, \ldots,)$, this definition does not admit an argument which will allow us to conclude that the prime divisors of any given integer n are independent.
>
> > The classical argument—that we cannot define a probability function for the probability that a random n is prime over the probability space $(1, 2, 3, \ldots,)$—is expressed informally[5] in a referee's critique of the author's original belief to the contrary:

[2] Strictly speaking, the probability defined by a discrete probability distribution; see [GS97], Chapter 5, pp.183-186; also [St02], Chapter 2, p.10.

[3] For r_i read $r_i(n)$; for R_i read $R_i(n)$ in the following table.

[4] Conventionally defined as integers that are not divisible by any smaller integer other than 1.

[5] See also [HL23], pp.36-37. A formal argument is given in [St02], Chapter 2, p.9, Theorem 2.1.

> "My objection is quite simply that I don't know what you mean by a randomly given positive integer n. If you want to make sense of it, then you need to assign to each positive integer n a probability $p(n)$. These probabilities must have two properties: that they are non-negative, and that their sum should be 1. If you do that, then you can talk about things like the probability that $m|n$. It will be $\sum_{d=1}^{\infty} p(dm)$.
>
> As an example, setting $p(n) = 2^{-n}$ for $n = 1, 2, 3, \ldots$ would satisfy the conditions for a probability distribution, though obviously this would be an unsuitable choice for your purposes. But the problem is that *every* possible way of choosing the $p(n)$ is unsuitable for your purposes. There does not exist a way of choosing the $p(n)$ such that for every m the equation $\sum_{d=1}^{\infty} p(dm) = 1/m$ holds.
>
> ... Consider first the probability of an unspecified integer n being divisible by an unspecified prime p. Given an arbitrary probability distribution on the positive integers, there will always be some prime p for which the above statement is false.
>
> To see this, suppose that the probability that n is chosen is not zero. Let's write this probability as $q(n)$. Now choose p so large that $1/p$ is less than $q(n)$. Then the probability that the remainder on division by p is n is at least $c(n)$ (since there is a probability $c(n)$ of choosing the integer n) and that is greater than $1/p$.
>
> ... A typical way that number theorists deal with a difficulty like this is to choose a random integer n in the range from N to $2N$ for some large integer N. But then you cannot say that the probability that n is a multiple of p is exactly $1/p$—it is only *approximately* $1/p$. And the various events are not exactly independent but only *approximately* independent. So there are error terms involved. And the entire difficulty of the subject is that these error terms accumulate and it becomes hard to say what the final answer is to any accuracy.
>
> ... Let me explain why what I did say is true. We pick an integer n uniformly at random from the set $\{N, N+1, N+2, , 2N\}$. What is the probability that n is even? If N is odd, then exactly half those integers are odd and half are even.
>
> If N is even, then we can write $N = 2M$, and in that case of the $N+1$ elements of the set, $M+1$ are even and M are odd, so the probability that n is even is $(M+1)/2M$. So that's already an example where the probability is only approximately equal to $1/p$ (which in this case is $1/2$). In general, the number of multiples of p in a set of R consecutive integers will be R/p if p happens to be a factor of R, and otherwise it will be one of the integers on either side of R/p.
>
> In the second case, which has to happen for several p (since R cannot be divisible by every prime less than R, or even than the square root of R), the best we can say is that the probability that an integer chosen uniformly at random from the R consecutive integers is a multiple of p is approximately equal to $1/p$.
>
> .... If you want to claim that you can make sense of the statement:
>
> 'The probability that an unspecified integer n is divisible by p is $1/p$',
>
> you will need to develop some kind of probability theory that allows you to do something that conventional probability theory (where you would need to specify a probability distribution on the positive integers) does not."

(b) Second as the values, for any given n, of the sequence $E(n) = \{r_i(n) : i \geq 1\}$. This now allows us to define a probability model from which we may conclude for any given $n > 1$, and any given prime $p > 1$, that the probability of the event $r_p(n) = 0$—whence p divides n—is $\frac{1}{p}$; and that the probability of the event $r_p(n) \neq 0$—whence p does not divide n—is $1 - \frac{1}{p}$. This allows us to argue (in §2.) that, given $p, q > 1$ are two unequal primes, the compound probability that $r_p(n) = 0$ and $r_q(n) = 0$—whence both p and q divide n—is $\frac{1}{pq}$. We conclude that the prime divisors of any given integer n are mutually independent. An immediate consequence (§4.) is that integer factorising cannot be polynomial-time.

2. The informal probability argument

(1) In this investigation—instead of seeking to define a probability function for the probability that a *random* n is prime over the probability space $(1, 2, 3, \ldots,)$—we shall address, instead, the questions:

(a) What is the probability for any *given* $n > i > 1$ and $i \geq 0$, where $i > u \geq 0$, that:

$n + u \equiv\ 0\ (mod\ i)$?

(b) What is the compound probability for any *given* $n > i,\ j > 1$, where $i \neq j$, $i > u \geq 0$, and $j > v \geq 0$, that:

$n + u \equiv\ 0\ (mod\ i)$, and $n + v \equiv\ 0\ (mod\ j)$?

(c) What is the compound probability for any *given* $n > i,\ j > 1$, where $i \neq j$, $i > u \geq 0$, and $j > v \geq 0$, that:

i divides $n + u$, and j divides $n + v$?

(2) We shall argue that:

(a) The probability that the roll of an i-sided cylindrical die will yield the value u is $\frac{1}{i}$ by the probability model for such an event as definable over the probability space $(0, 1, 2, \ldots, i - 1)$;

(b) The probability that the simultaneous roll of one i-sided cylindrical die and one j-sided cylindrical die will yield the values u and v, respectively, is $\frac{1}{i.j}$ by the probability model for such a simultaneous event as defined over the probability space $\{(u, v) : i > u \geq 0, j > v \geq 0\}$.

(3) We shall trivially conclude that:

The compound probability of determining u and v correctly from the simultaneous roll of one i-sided cylindrical die and one j-sided cylindrical die, is the product of the probability of determining u correctly from the roll of an i-sided cylindrical die, and the probability of determining v correctly from the roll of a j-sided cylindrical die.

(4) We shall further conclude non-trivially that:

(a) If i and j are co-prime, the compound probability of correctly determining that i divides n and j divides n from the simultaneous roll of one i-sided cylindrical die and one j-sided cylindrical die, is the product of the probability of correctly determining that i divides n from the roll of an i-sided cylindrical die, and the probability of correctly determining that j divides n from the roll of a j-sided cylindrical die.

(b) The assumption that i and j be co-prime is also necessary, since 4(a) would not always be the case if i and j were not co-prime.

For instance, let $j = 2i$. The probability that an i-sided cylindrical die will then yield 0—and allow us to conclude that i divides n—is $\frac{1}{i}$, and the probability that a j-sided cylindrical die will then yield 0—and allow us to conclude that j divides n—is $\frac{1}{j}$; but the probability of determining both that i divides n, and that j divides n, from a simultaneous roll of the two cylindrical dice is $\frac{1}{j}$, and not $\frac{1}{i.j}$.

(5) We shall also conclude non-trivially that, if p and q are two unequal primes, the probability of algorithmically determining whether p divides n is independent of the probability of algorithmically determining whether q divides n.

2.A. The residues $r_i(n)$.

We begin by formally defining the residues $r_i(n)$ for all $n \geq 2$ and all $i \geq 2$ as below:

Definition 1. $n + r_i(n) \equiv 0 \ (mod\ i)$ *where* $i > r_i(n) \geq 0$.

Since each residue $r_i(n)$ cycles over the i values $(i-1, i-2, \ldots, 0)$, these values are all incongruent and form a complete system of residues[6] *mod* i.

It immediately follows that:

Lemma 2.1. $r_i(n) = 0$ *if, and only if,* i *is a divisor of* n. □

2.B. The probability model $\mathbb{M}_i$

By the standard definition of the probability $\mathbb{P}(e)$ of an event e[7], we have by Lemma 2.1 that:

Lemma 2.2. *For any* $n \geq 2,\ i \geq 2$ *and any given integer* $i > u \geq 0$*:*

- *the probability* $\mathbb{P}(r_i(n) = u)$ *that* $r_i(n) = u$ *is* $\frac{1}{i}$*;*
- $\sum_{u=0}^{u=i-1} \mathbb{P}(r_i(n) = u) = 1$*;*
- *and the probability* $\mathbb{P}(r_i(n) \neq u)$ *that* $r_i(n) \neq u$ *is* $1 - \frac{1}{i}$. □

By the standard definition of a probability model, we conclude that:

Theorem 2.3. *For any* $i \geq 2$, $\mathbb{M}_i = \{(0, 1, 2, \ldots, i-1),\ r_i(n), \frac{1}{i}\}$ *yields a probability model for each of the values of* $r_i(n)$. □

Corollary 2.4. *For any given* n, i *and* u *such that* $r_i(n) = u$, *the probability that the roll of an* i*-sided cylindrical die will yield the value* u *is* $\frac{1}{i}$ *by the probability model defined in Theorem 2.3 over the probability space* $(0, 1, 2, \ldots, i-1)$. □

Corollary 2.5. *For any* $n \geq 2$ *and any prime* $p \geq 2$, *the probability* $\mathbb{P}(r_p(n) = 0)$ *that* $r_p(n) = 0$, *and that* p *divides* n, *is* $\frac{1}{p}$*; and the probability* $\mathbb{P}(r_p(n) \neq 0)$ *that* $r_p(n) \neq 0$, *and that* p *does not divide* n, *is* $1 - \frac{1}{p}$. □

We also note the standard definition[8]:

Definition 2. *Two events* e_i *and* e_j *are mutually independent for* $i \neq j$ *if, and only if,* $\mathbb{P}(e_i \cap e_j) = \mathbb{P}(e_i).\mathbb{P}(e_j)$.

---

[6] [HW60], p.49.
[7] See [Ko56], Chapter I, §1, Axiom III, pg.2.
[8] See [Ko56], Chapter VI, §1, Definition 1, pg.57 and §2, pg.58.

2.C. The prime divisors of any integer n are mutually independent

We note that:

Lemma 2.6. *If $n \geq 2$ and $n > i, j > 1$, where $i \neq j$, then:*

$$\mathbb{P}((r_i(n) = u) \cap (r_j(n) = v)) = \mathbb{P}(r_i(n) = u).\mathbb{P}(r_j(n) = v)$$

where $i > u \geq 0$ and $j > v \geq 0$.

Proof: (i) If $n \geq 2$ and $n > i, j > 1$, where $i \neq j$, then we can always determine a unique pair of residues $r_i(n) = u$ and $r_j(n) = v$, where $i > u \geq 0$, $j > v \geq 0$, i divides $n + u$, and j divides $n + v$.

(ii) There are $i.j$ pairs (u, v) such that $i > u \geq 0$ and $j > v \geq 0$.

(iii) The compound probability that the simultaneous roll of one i-sided cylindrical die and one j-sided cylindrical die will yield the values u and v, respectively, is thus $\frac{1}{i.j}$ by the probability model for such a simultaneous event as defined over the probability space $\{(u, v) : i > u \geq 0, j > v \geq 0\}$, where we note that:

- the probability $\mathbb{P}((r_i(n) = u) \cap (r_j(n) = v))$ that $r_i(n) = u$ and $r_j(n) = v$ is $\frac{1}{i.j}$;
- $\sum_{All\ (u,v):\ i>u\geq 0,\ j>v\geq 0} \mathbb{P}((r_i(n) = u) \cap (r_j(n) = v)) = 1$;

(iv) By Lemma 2.2, the product of the probability $\frac{1}{i}$ that the roll of an i-sided cylindrical die will yield the value u, and the probability $\frac{1}{j}$ that the roll of a j-sided cylindrical die will yield the value v, is $\frac{1}{i.j}$.[9]

(v) It follows that:

$$\mathbb{P}((r_i(n) = u) \cap (r_j(n) = v)) = \tfrac{1}{i.j}$$

$$\mathbb{P}(r_i(n) = u).\mathbb{P}(r_j(n) = v) = (\tfrac{1}{i})(\tfrac{1}{j}).$$

The lemma follows. □

Corollary 2.7. $\mathbb{P}((r_i(n) = 0) \cap (r_j(n) = 0)) = \mathbb{P}(r_i(n) = 0).\mathbb{P}(r_j(n) = 0)$. □

Since, by Lemma 2.1, $r_i(n) = 0$ if, and only if, i is a divisor of n, it follows from Corollary 2.7 that:

Theorem 2.8. *If i and j are co-prime and $i \neq j$, then whether, or not, i divides any given natural number n is independent of whether, or not, j divides n.* □

Proof: (i) By Corollary 2.6, we have that:

$$\mathbb{P}((r_i(n) = 0) \cap (r_j(n) = 0)) = \tfrac{1}{i.j}$$

$$\mathbb{P}(r_i(n) = 0).\mathbb{P}(r_j(n) = 0) = (\tfrac{1}{i})(\tfrac{1}{j}).$$

[9] In other words, the compound probability of determining u and v correctly from the simultaneous roll of one i-sided cylindrical die and one j-sided cylindrical die, is the product of the probability of determining u correctly from the roll of an i-sided cylindrical die, and the probability of determining v correctly from the roll of a j-sided cylindrical die.

(ii) Further, if i and j are co-prime, and $n+r_{i.j}(n) \equiv 0\ (mod\ i.j)$, then the $i.j$ integers $r_j(n).i+r_i(n).j$ are all incongruent and form a complete system of residues. It follows that $n = a.i$—whence i divides n—and also $n = b.j$—whence j divides n—if, and only if $r_i(n) = r_j(n) = r_{i.j}(n) = 0$.

The lemma follows. □

We thus conclude that:

Corollary 2.9. *The prime divisors of any integer n are mutually independent.* □

3. Approximating $\pi(n)$ by the 'density' of integers that are not divisible by the first $\pi(\sqrt{n})$ primes

It also follows from Definition 1 and Lemma 2.1 that, if $\pi(k)$ denotes the primes $\leq k$, the $p_1.p_2.\ldots.p_{\pi(\sqrt{n})}$ numbers $u_{p_1}.p_1 + u_{p_2}.p_2 + \ldots + u_{p_{\pi(\sqrt{n})}}.p_{\pi(\sqrt{n})}$, where $p_i > u_{p_i} \geq 0$, are all incongruent and form a complete system of residues[10] *mod* $(p_1.p_2.\ldots.p_{\pi(\sqrt{n})})$.

We thus have by Corollary 2.9 that:

Lemma 3.1. *Given any integers $k \geq 2$, $n \geq 2$, the first $\pi(\sqrt{n})$ primes $p_1, p_2, \ldots, p_{\pi(\sqrt{n})}$, and the set of $\pi(\sqrt{n})$-tuples $\{(u_{p_1}, u_{p_2}, \ldots, u_{p_{\pi(\sqrt{n})}})\}$ for all $0 \leq u_{p_i} < p_i$:*

- *the probability $P(\cap_{i=1}^{\pi(\sqrt{n})} r_{p_i}(k) = u_{p_i})$ that:*

  $$r_{p_1}(k) = u_{p_1}\ \&\ r_{p_2}(k) = u_{p_2} \ldots \&\ r_{\pi(\sqrt{n})}(k) = u_{\pi(\sqrt{n})}$$

  is $\frac{1}{p_1.p_2.\ldots.p_{\pi(\sqrt{n})}}$;

- $\sum_{(u_{p_1}, u_{p_2}, \ldots, u_{p_{\pi(\sqrt{n})}})} P(\cap_{i=1}^{\pi(\sqrt{n})} r_{p_i}(k) = u_{p_i}) = 1$;

- *and the probability $P(\cup_{i=1}^{\pi(\sqrt{n})} r_{p_i}(k) \neq u_{p_i})$ that some $r_{p_i}(k) \neq u_{p_i}$ is $1 - \frac{1}{p_1.p_2.\ldots.p_{\pi(\sqrt{n})}}$.* □

It further follows that:

Corollary 3.2. *Given any integers $k \geq 2$, $n \geq 2$, the first $\pi(\sqrt{n})$ primes $p_1, p_2, \ldots, p_{\pi(\sqrt{n})}$, and the set of $\pi(\sqrt{n})$-tuples $\{(u_{p_1}, u_{p_2}, \ldots, u_{p_{\pi(\sqrt{n})}})\}$ for all $0 \leq u_{p_i} < p_i$:*

- *the probability $P(\cap_{i=1}^{\pi(\sqrt{n})} r_{p_i}(k) \neq 0)$ that:*

  $$r_{p_1}(k) \neq 0\ \&\ r_{p_2}(k) \neq 0 \ \ldots\ \&\ r_{\pi(\sqrt{n})}(k) \neq 0$$

  is $\prod_{i=1}^{\pi(\sqrt{n})}(1 - \frac{1}{p_i})$;

- $\sum_{(u_{p_1}, u_{p_2}, \ldots, u_{p_{\pi(\sqrt{n})}})} P(\cap_{i=1}^{\pi(\sqrt{n})} r_{p_i}(k) = u_{p_i}) = 1$;

- *and the probability $P(\cup_{i=1}^{\pi(\sqrt{n})} r_{p_i}(k) = 0)$ that some $r_{p_i}(k) = 0$ is $1 - \prod_{i=1}^{\pi(\sqrt{n})}(1 - \frac{1}{p_i})$.* □

[10] [HW60], p.52, Theorem 59.

Since n is a prime if, and only if, n is not divisible by any prime $p \leq \sqrt{n}$, and it follows from Lemma 2.1 and Corollary 3.2 that the 'density' of integers which are not divisible by the first $\pi(\sqrt{n})$ primes is $\prod_{i=1}^{\pi(\sqrt{n})}(1-\frac{1}{p_i})$, the estimated number of such integers in the interval $(1,n)$ is thus a non-heuristic approximation[11] of $\pi(n)$, given by[12]:

Corollary 3.3. $\pi(n) \approx n.\prod_{i=1}^{\pi(\sqrt{n})}(1-\frac{1}{p_i}) \approx 2e^{-\lambda}\frac{n}{log_e(n)}$[13]. □

4. Integer Factorising cannot be polynomial-time

In a seminal paper[14], Agarwal et al have shown that deciding whether a given integer n is a prime or not can be done in polynomial time.

Given that n is composite, Theorem 2.8 now immediately yields the computational complexity consequence[15] that no deterministic algorithm[16] can further compute a factor of n in polynomial time[17].

We note the standard definition[18]:

Definition 3. *A deterministic algorithm computes a number-theoretical function $f(n)$ in polynomial-time if there exists k such that, for all inputs n, the algorithm computes $f(n)$ in $\leq (log_e\ n)^k + k$ steps.*

It then follows from Theorem 2.8 that:

Corollary 4.1. *Any deterministic algorithm that always computes a prime factor of n cannot be polynomial-time.*

Proof: Any computational process that successfully identifies a prime divisor of n must necessarily appeal to at least one logical operation for identifying such a factor.

Since any given integer n is a prime if, and only if, it is not divisible by any prime $p \leq \sqrt{n}$, and since n may be the square of a prime, it follows from Theorem 2.8 that we necessarily require at least one logical operation for each prime $p \leq \sqrt{n}$ in order to logically determine whether p is a prime divisor of n.

Since, by Corollary 3.3, the number of such primes is of the order $O(\frac{\sqrt{n}}{log_e\ n})$, the number of computations required by any deterministic algorithm that always computes a prime factor of n cannot be polynomial-time—i.e. of order $O((log_e\ n)^c)$ for any c—in the length of the input n. The corollary follows. □

[11] Though not necessarily the best approximation.

[12] Compare Chebychev's Theorem $\pi(x) \asymp \frac{x}{log_e x}$; [HW60], p.9, Theorem 7 and p.345, §22.4.

[13] By Mertens' Theorem, where $2.e^{-\gamma} \approx 1.12292$.

[14] We have refrained from citing the Agarwal et al paper or related work, as the principal conclusion of this paper (a trivial corollary in an independent investigation) did not emerge out of a study of computational complexity, nor out of pursuing the PvNP problem, and it would have been misleading to include any reference to, or survey of, any work on the subject that might suggest otherwise.

[15] cf. [Cook].

[16] A deterministic algorithm computes a mathematical function which has a unique value for any input in its domain, and the algorithm is a process that produces this particular value as output.

[17] cf. [Cook], p.1; also [Br00], p.1, fn.1.

[18] cf. [Cook], p.1; also [Br00], p.1, fn.1: "For a polynomial-time algorithm the expected running time should be a polynomial in the length of the input, i.e. $O((logN)^c)$ for some constant c".

Appendix I: Definitions of some terms and concepts of Probability Theory

Probability model[19]: *A probability model* is a mathematical representation of a random phenomenon. It is defined by its sample space, events within the sample space, and probabilities associated with each event.

- *The sample space* S for a probability model is the set of all possible outcomes.
- *An event* A is a subset of the sample space S.
- *A probability* is a numerical value assigned to a given event A.

Distribution Function[20]: Let X be a random variable which denotes the value of the outcome of a certain experiment, and assume that this experiment has only finitely many possible outcomes. Let Ω be the sample space of the experiment (i.e., the set of all possible values of X, or equivalently, the set of all possible outcomes of the experiment). A *distribution function* for X is a real-valued function m whose domain is Ω and which satisfies:

1. $m(\omega) \geq 0$, for all $\omega \in n$, and
2. $\sum_{\omega \in \Omega} m(\omega) = 1$.

For any subset E of Ω, we define the *probability* of E to be the number $P(E)$ given by $P(E) = \sum_{\omega \in E} m(\omega)$.

Some notations[21]: Let A and B be two sets. Then the union of A and B is the set $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$. The intersection of A and B is the set $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$.

Mutual Independence[22]: A set of events $\{A_1,\ A_2,\ \ldots,\ A_n\}$ is said to be *mutually independent* if for any subset $\{A_i,\ A_j,\ \ldots,\ A_m\}$ of these events we have

$$P(A_i \cap A_j \cap \ldots \cap A_m) = P(A_i)P(A_j)\ldots P(A_m),$$

or equivalently, if for any sequence $\overline{A_1},\ \overline{A_2},\ \ldots,\ \overline{A_n}$ with $\overline{A_j} = A_j$ or $\overline{A_j}$,

$$P(\overline{A_i} \cap \overline{A_j} \cap \ldots \cap \overline{A_m}) = P(\overline{A_i})P(\overline{A_j})\ldots P(\overline{A_m}).$$

Acknowledgements: I am indebted to my erstwhile classmate, Professor Chaitanya Kumar Harilan Mehta, for his unqualified encouragement and support for my scholarly pursuits over the years, without which this extension of a 1964 investigation into the nature of divisibility and the structure of the primes—begun whilst yet classmates—would have vanished into some black hole of the informal universe of seemingly self-evident truths. I am also grateful for the extra-ordinary patience and indulgent persistence of Professor William Timothy Gowers in impressing upon me that my original argument conflated the concept of the probability of a number being a prime with that of the 'density' of primes.

References

[Br00] Richard P. Brent. 2000. *Recent Progress and Prospects for Integer Factorisation Algorithms.* In *Computing and Combinatorics*, Lecture Notes in Computer Science, Volume 1858, 2000, pp.3-22, Springer, New York/Heidelberg.

[Cook] Stephen Cook. 2000. *The P versus NP Problem*. Official description provided for the Clay Mathematical Institute, Cambridge, Massachusetts.

[El79a] P. D.T. A. Elliott. 1979. *Probabilistic Number Theory I.* Springer-Verlag, New York.

[El79b] P. D.T. A. Elliott. 1979. *Probabilistic Number Theory II.* Springer-Verlag, New York.

[GS97] Charles M. Grinstead and J. Laurie Snell. 2003. *Introduction to Probability*, *The CHANCE Project* Version dated 4 July 2006 of the Second Revised Edition, 1997, American Mathematical Society, Rhode Island, USA.

[HL23] G.H. Hardy and J.E. Littlewood. 1923. *Some problems of 'partitio numerorum:' III: On the expression of a number as a sum of primes,* Acta Mathematica, December 1923, Volume 44, pp.1-70.

[HW60] G. H. Hardy and E. M. Wright. 1960. *An Introduction to the Theory of Numbers.* 4th edition. Clarendon Press, Oxford.

[Ka59] Mark Kac. 1959. *Statistical Independence in Probability, Analysis and Number Theory.* 1959. *The Carus Mathematical Monographs: Number Twelve* The Mathematical Association of America, Second Impression, 1964.

[Ko56] A. N. Kolmogorov. 1933. *Foundations of the Theory of Probability.* Second English Edition. Translation edited by Nathan Morrison. 1956. Chelsea Publishing Company, New Yourk (*sic*).

[19] cf. http://www.stat.yale.edu/Courses/1997-98/101/probint.htm.
[20] Excerpted from [GS97], Chapter 1, §1.2, p.19.
[21] Excerpted from [GS97], Chapter 1, §1.2, p.21.
[22] Excerpted from [GS97], Chapter 4, §4.1, Definition 4.2, p.141.

[St02] Jörn Steuding. 2002. *Probabilistic Number Theory.* The Pennsylvania State University CiteSeerX Archives, doi=10.1.1.118.4755.

[An14a] Bhupinder Singh Anand. 2014. *A probability-based proof that any deterministic algorithm which always computes a prime factor of n cannot be polynomial-time.* Submitted on 01/01/2015 to the *Calcutta Statistical Association Bulletin*, Kolkata, India.

Address: Bhupinder Singh Anand, #1003 B Wing, Lady Ratan Tower, Dainik Shivner Marg, Gandhinagar, Worli, Mumbai - 400 018, Maharashtra, India.
Email: bhup.anand@gmail.com