Algorithms and Data Structures **Fall 2019**

Lecture Notes: Primality Testing

Lecturer: Kavitha T *Scribe: Aparna Shankar*

We will now look at another example of a randomised algorithm - for primality testing. Primality testing is a fundamental problem and has applications in many fields like cryptography.

1 Problem Statement

We want an algorithm that carries out the following:

Input: An odd integer $n \geq 3$.
Output: "prime" if n is prime, and "composite" otherwise.

Further, we want the algorithm to run in time $O(\text{poly}(\log n))$. This is because the number n, when stored in binary, takes $\log n$ bits, and hence the input size is $\log n$.

The naive algorithm that checks if n is divisible by any number from 2 to $n-1$ (or $\sqrt{n}$) does not run in this time.

We also want the success probability of the algorithm to be at least $\frac{3}{4}$.

Remark. *When we discuss success probability of an algorithm, it is over the locally random choices made by the algorithm for a given input, not over the input distribution. In particular, for this problem, if our algorithm returned "composite" on every input, we cannot say its error probability is bounded by the density of primes (some $O(\frac{1}{\log n})$).*

2 First Attempt

Our first idea is to guess a possible divisor of n.

Algorithm 1 Tests if n is prime

procedure IS-PRIME(n) ▷ n is odd and $n \geq 3$
  Choose a uniformly at random from $\{1, 2, \ldots, n-1\}$ ▷ We can delete 1 since it always divides n, but this does not affect the analysis much
  if $\gcd(a, n) \neq 1$ **then return** "composite"
  end if
end procedure

If $\gcd(a, n) = 1$, then a and n are coprime. It is too soon, of course, to conclude that n is prime just because we chose a number coprime to it. Before doing further tests let us look at some relevant theorems.

3 Some Theorems

Theorem 1 (Fermat's Little Theorem)**.** *If n is prime, then for every $a \in \mathbb{Z}_n \setminus \{0\}$,*

$$a^{n-1} \equiv 1 \mod n$$

First, consider the set $\mathbb{Z}_n^* = \{a \in \mathbb{Z}_n \mid \gcd(a, n) = 1\}$. This is nothing but the set of residues (remainders) of numbers relatively prime to n.

Claim 2. *$(\mathbb{Z}_n^*, *_n)$ forms a group where $*_n$ is multiplication modulo n.*

Proof. We need to show that the operation is associative - this can be seen from the fact that ordinary multiplication of integers is associative. We also need to show the existence of an identity and inverses for each element. 1 is clearly the identity. To see that every element a has an inverse, note that the gcd algorithm implies that there are integers α and β such that $\alpha a + \beta n = 1$ (ordinary integer addition). Looking at this equality modulo n, $\alpha' a \equiv 1$ where α' is the residue of α modulo n.

□

The size of this group $|\mathbb{Z}_n^*|$ is denoted $\phi(n)$.

Now, consider $\langle a \rangle = \{a, a^2, a^3, \dots\}$ for $a \in \mathbb{Z}_n^*$. This is a subgroup of $\mathbb{Z}_n^*$:

Claim 3. *$\langle a \rangle$ contains 1.*

Proof. Since $\mathbb{Z}_n^*$ is finite, some element must eventually appear twice i.e. $a^i = a^j$ for some i and j. Then $a^{i-j} = 1$.

□

Theorem 4 (Lagrange's Theorem)**.** *Let G be any finite group and H a subgroup of G. Then $|H|$ divides $|G|$.*

Proof. If $H = G$, we are done. Otherwise, there is some element $a \in G \setminus H$. If $H = \{h_1, h_2, \dots, h_t\}$, define $aH = \{ah_1, ah_2, \dots ah_t\}$. H and aH must be disjoint since if they were not, $h_i = ah_j \implies a = h_i h_j^{-1} \in H$, which is a contradiction. Further, $|H| = |aH|$, since $ah_i = ah_j \implies h_i = h_j$.

Now, if H and aH together cover G, we are done, and $|G| = 2|H|$. Otherwise, we have $b \in G \setminus (H \cup aH)$. Repeat the process to get a set bH, and so on until there are no remaining elements of G.

□

Proof of Theorem 1. Suppose $|\langle a \rangle| = k$. k must divide $n-1$; say $n-1 = kl$. Then, $a^{n-1} = (a^k)^l = 1^l = 1$. □

This can be used to improve our first attempt.

4 Improved First Attempt

Remark. *a^{n-1} can be computed in time $O(\text{poly}(\log n))$ by repeatedly squaring a.*

Algorithm 2 Tests if n is prime

```
procedure IS-PRIME(n)                                   ▷ n is odd and n ≥ 3
    Choose a uniformly at random from {1, 2, ..., n − 1}
    if gcd(a, n) ≠ 1 then return "composite"
    else if a^{n−1} ≠ 1 mod n then return "composite"
    end if
return "prime"
end procedure
```

Note that no prime number is mistakenly labelled as composite by this algorithm, however a composite number might "slip through the cracks" and be labelled as prime. In other words, if a number n is said to be composite by the algorithm, then it is definitely composite; however if the algorithm says n is prime, then n may be prime or composite.

5 Carmichael Numbers

Suppose the converse of Fermat's Little Theorem held (i.e. if n is composite, then there is some $a \in \mathbb{Z}_n^*$ such that $a^{n-1} \neq 1 \mod n$). Then we can bound the error probability of the above algorithm.

Claim 5. *If the converse of Fermat's Little Theorem is true, then at least half the elements $a \in \mathbb{Z}_n^*$ have $a^{n-1} \neq 1 \mod n$.*

Proof. Consider the set of elements that do *not* satisfy this, i.e $\{a \in \mathbb{Z}_n^* \mid a^{n-1} = 1 \mod n\}$. We claim that these form a subgroup. Then, by Lagrange's theorem, the size of this set must divide $|\mathbb{Z}_n^*|$. Since this set is not the whole group (by the assumption), the size of the set must be at most half of $|\mathbb{Z}_n^*|$.

To show this is a subgroup, we need to show closure, identity and inverses. To see closure, observe that if a and b are elements, $(ab)^{n-1} = a^{n-1}b^{n-1} = 1$. Clearly $1^{n-1} = 1$, and if $a^{n-1} = 1, (a^{-1})^{n-1} = 1$ by multiplying both sides by a^{n-1}.

□

Unfortunately, the converse of Fermat's Little Theorem is not true! The counterexamples are known as Carmichael numbers: composite numbers n such that $a^{n-1} = 1$ for every $a \in \mathbb{Z}_n^*$. There are infinitely many Carmichael numbers and the smallest is 561.

Notice that at this point, $a^{n-1} = 1$ and $n - 1$ is even. We claim that if $a^{\frac{n-1}{2}} \notin \{\pm 1\}$, then a is composite.

Claim 6. *If there is a nontrivial square root of 1 in $\mathbb{Z}_n^*$, then n is composite.*

Proof. Suppose $x^2 - 1 = 0 \mod n$ and x is not 1 or -1. Then $(x+1)(x-1) = 0 \mod n$. Then n divides either $x + 1$ or $x - 1$, which are not both zero since we assumed $x \notin \{\pm 1\}$. □

Now, since $n - 1$ is even, we can write $n - 1 = 2^k t$, for some odd number t. Then, the idea of the following algorithm is: compute $a^t, a^{2t}, a^{4t}, \ldots, a^{2^k t}$. Look at the number just before the first occurrence of 1, and if it is not 1 or -1, return "composite". 1 must surely occur in the sequence since $a^{n-1} = 1$.

6 Second Attempt - Miller-Rabin Algorithm

The following algorithm is due to Miller and Rabin.

Algorithm 3 Tests if n is prime

procedure IS-PRIME(n) ▷ n is odd and $n \geq 3$

if $n = a^b$ for some integers $a, b \geq 2$ **then return** "composite" ▷ This is so we can apply the Chinese remainder theorem
end if
Choose a uniformly at random from $\{1, 2, \ldots, n-1\}$
if $\gcd(a, n) \neq 1$ **then return** "composite"
else if $a^{n-1} \neq 1 \mod n$ **then return** "composite"
else
Write $n = 2^k t$, t odd
Compute $a^t, a^{2t}, \ldots$ until a 1 is seen
if the previous number is not 1 or $-1 \mod n$ **then return** "composite"
end if
end if
return "prime"
end procedure

Remark. *When checking if $n = a^b$, b can be at most $\log n$. For a fixed b, a can be found using binary search, hence the whole step takes time $O(\text{poly}(\log n))$.*

Remark. *As it stands, the success probability is at least $\frac{1}{2}$, but this can be amplified to $\frac{3}{4}$ by repeating the experiment.*

Again, no prime number is mistakenly labelled as composite. If n is not Carmichael, as earlier, the success probability is at least $\frac{1}{2}$. We formalise these claims.

Claim 7. *If n is prime, then the algorithm always returns "prime".*

Claim 8. *If n is composite and not Carmichael, then the algorithm returns "composite" with probability at least $\frac{1}{2}$.*

Now, we only need to consider the case when n is Carmichael.

Consider a table with rows indexed by the elements of $\mathbb{Z}_n^*$ and columns indexed by $t, 2t, 4t, \ldots, 2^k t = n-1$. Each entry is the row index raised to the power of the column index modulo n, that is, row a_i has entries $a_i^t, a_i^{2t}, \ldots, a_i^{n-1}$.

	t	$2t$	$\cdots$	$2^h t$	$2^{h+1}t$	$\cdots$	$n-1$
a_1	1	1		1	1		1
a_2	$*$	$*$		$*$	1		1
$\vdots$							
$a_{\phi(n)}$	-1	1		-1	1		1

Since n is Carmichael, the last column consists of all 1's. Now, the first column is not all 1's, since t is odd and $(-1)^t$ must be -1.

This implies that there is some column $2^h t$ which is not all $1's$, followed by a column $2^{h+1}t$ which is all 1's. We will later show that at least half the entries in column $2^h t$ are neither 1 nor -1.

Theorem 9 (Chinese Remainder Theorem)**.** *Let n be a composite number that is not a prime power. That is, n can be written as $n = rs$ where $\gcd(r, s) = 1$. Then, there is a bijection (in fact, an isomorphism) $f : \mathbb{Z}_n \to \mathbb{Z}_r \times \mathbb{Z}_s$ where $f(a) = (a \mod r, a \mod s)$.*

Proof. The two sets $\mathbb{Z}_n$ and $\mathbb{Z}_r \times \mathbb{Z}_s$ have the same size, hence it is enough to prove that the function is surjective. Let (r', s') be in the codomain, we need to show it has a pre-image. Since r and s are relatively prime, there are integers x and y such that $xr + ys = 1$.

Then,

$$xr = 1 \mod s$$
$$ys = 1 \mod r$$

$$s'xr = s' \mod s$$
$$r'ys = r' \mod r$$

$$(s'xr + r'ys) = s' \mod s$$
$$(s'xr + r'ys) = r' \mod r$$

Then $(s'xr + r'ys)$ is the required pre-image. □

Now consider the map when restricted to $\mathbb{Z}_n^*$. We claim that $f : \mathbb{Z}_n^* \to \mathbb{Z}_r^* \times \mathbb{Z}_s^*$ is also a bijection and in fact an isomorphism.

The reason it is well-defined and a bijection is that if a is relatively prime to n, $a \mod r$ is relatively prime to r and similarly for s.

To show it is an isomorphism, we need to show $f(ab) = f(a)f(b)$ where on the right side, multiplication is coordinate-wise. Suppose

$$a = kr + \alpha, b = k'r + \beta$$

Then $f(ab) = ab \mod r = (kr + \alpha)(k'r + \beta) \mod r = \alpha\beta$, and similarly for s.

Example. *For $n = 15$, $15 = 3 \times 5$. Then $1 \mapsto (1, 1), 2 \mapsto (2, 2), 4 \mapsto (1, -1)$ and so on.*

Note that the trivial square roots of 1 are 1 and 14, and they map to $(1, 1)$ and $(-1, -1)$ respectively. The nontrivial square roots, 4 and 7, map to $(1, -1)$ and $(-1, 1)$.

We are now ready to prove the claim indicated earlier.

Claim 10. *If the column indexed by $2^h t$ has at least one entry that is neither 1 nor -1, then at least half the entries are neither 1 nor -1.*

Proof. Let $H = \{a \in \mathbb{Z}_n^* \mid a^{2^h t} = \pm 1 \mod n\}$. H is a subgroup, the proof is very similar to the one done earlier. By the assumption, H is not the entire set, hence it must have size at most half the size of the whole set. □

Claim 11. *At least one entry is neither* 1 *nor* -1.

Proof. We prove this by contradiction. Suppose all the entries were either 1 or -1. We use this to construct an entry that is neither 1 nor -1.

By choice of h, the column has at least one -1, say $b^{2^h t} = -1$. Also, there is some a (for example, 1) such that $a^{2^h t} = 1$. We will show that $c^{2^h t} \notin \{\pm 1\}$ for some c.

Recall that n is a Carmichael number and not a prime power. Then we can write n as $r \times s$ and apply the Chinese Remainder Theorem. Then,

$$f(b^{2^h t}) = f(-1) = (-1, -1)$$

But since f is an isomorphism,

$$(-1, -1) = f(b^{2^h t}) = [f(b)]^{2^h t}$$

Say $f(b) = (b_1, b_2)$. Then $(b_1^{2^h t}, b_2^{2^h t}) = (-1, -1)$.

Now consider the pre-image of $(b_1, 1)$ (which exists, by the theorem) and call it c. Then, $f(c^{2^h t}) = (b_1^{2^h t}, 1) = (-1, 1)$.

This is the image of neither 1 nor -1, hence $c^{2^h t} \notin \{\pm 1\}$.

□