A Random Model of the Primes

Frank Qiang, Aravinth Venkatesh Natarajan, Anthony Hong

Georgia Institute of Technology

April 25, 2023

Background

- Infinite number of primes (c. 300 BC, Euclid)
- $\pi(x) :=$ number of primes $\leq x$
- Prime number theorem: $\pi(x) \sim \frac{x}{\ln x}$ (1896, Hadamard and De la Vallée Poussin)

Prime Gaps

- Prime gap := difference between a pair of consecutive primes
- What does the sequence of prime gaps look like?

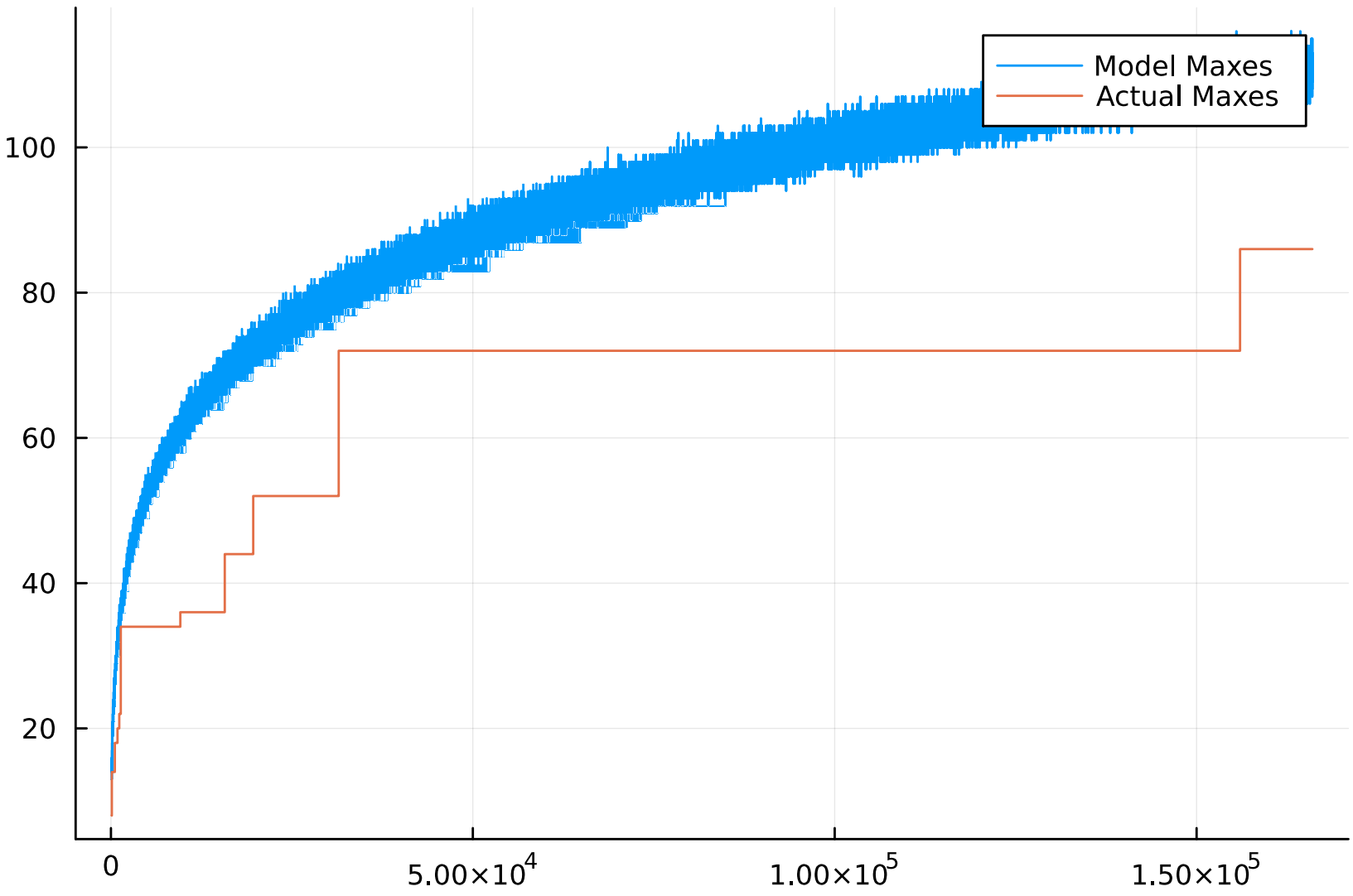
$$[1, 2, 2, 4, 2, 4, 2, 4, 6, 2]$$

  Notice that 1 is the only odd.
- How small/large can prime gaps get?
  - Small: twin prime conjecture; we have a constant bound by Yitang Zhang (2013), since improved to gap ≤ 246 infinitely often
  - Large: can find gaps of arbitary size (think $n!$); so slightly different goal, want to see how large of a gap we can expect in a given interval

Cramer's Random Model

- We know prime number distribution isn't exactly random, but it seems to behave that way: try to model as if it were
- In the interval $[2, n]$, we expect roughly $\frac{n}{\ln n}$ primes by prime number theorem. There are also roughly n total numbers $[2, n]$. So n has probability $\frac{1}{\ln n}$ of being prime?
- Naively assign each $x \in \mathbb{N}_{>2}$ a probability of $\frac{1}{\ln x}$ of being prime

First Iteration



Plot of estimated largest gap $\leq n$ vs. n, with 100 trials per n.

Why?

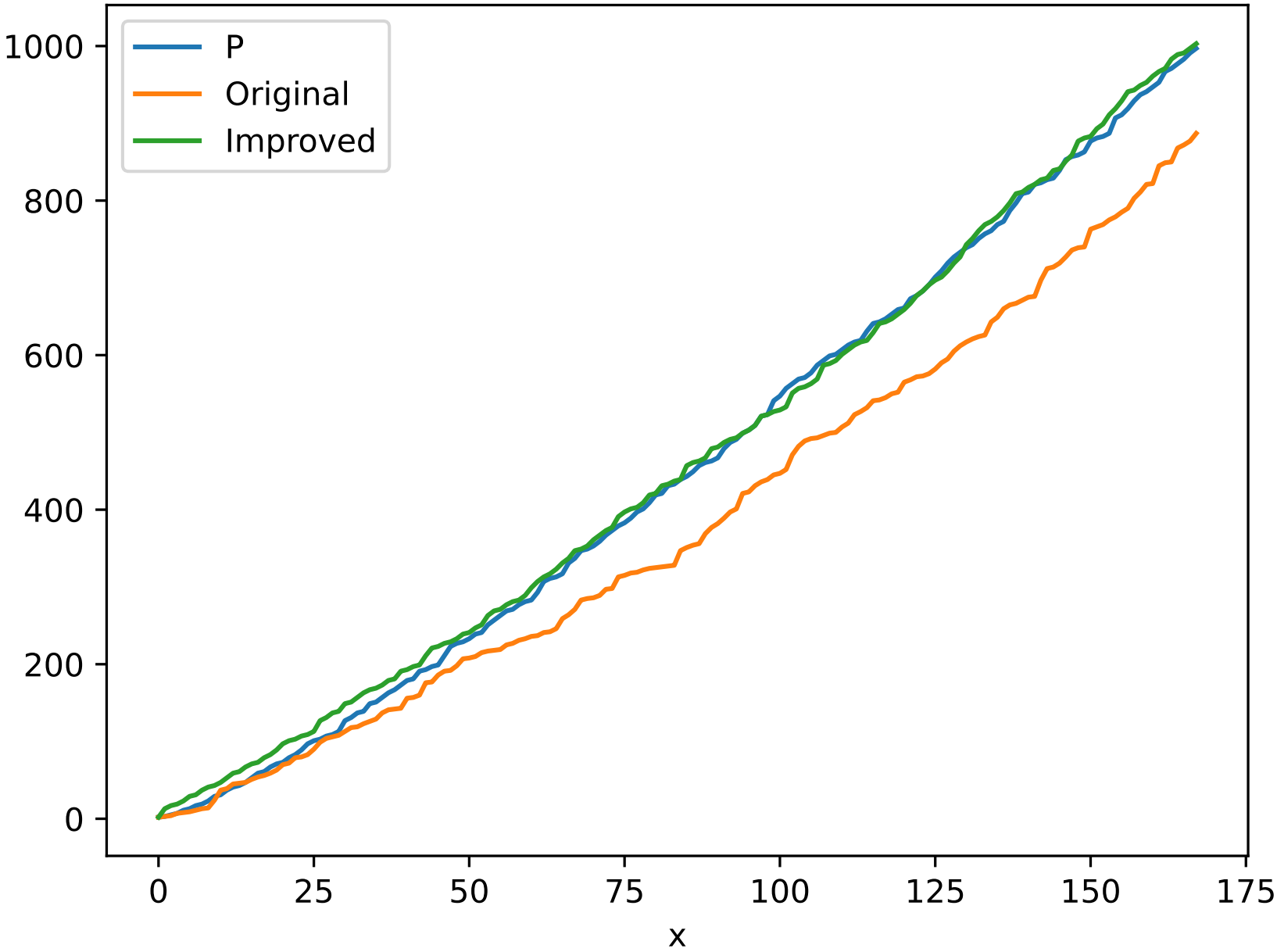
- We can study this random distribution using ideas from probability and statistics, can give extra insight
- Conducting many trials comes at relatively low cost (can be parallelized, will scale with current trend towards multithreaded computers)
- We can consider disjoint intervals completely independently (not true of traditional sieve methods)

Improvement to the Original Model

- Something obvious, the only even prime should be 2
- However, Cramer's original scheme gives all natural numbers $x > 2$ chance $\frac{1}{\ln x}$
- Idea: For every even number e, give its chance to $e + 1$
- Even number, $p(x) = 0$, odd number, $p(x) \approx \frac{2}{\ln x}$

Improvement to the Original Model

- Extrapolate idea: Instead of just 2, consider a small prefix of primes
- Only consider numbers coprime to those primes to be a part of the model
- Maintains randomness, but incorporates some structure from that prefix


P
Original
Improved
1000
800
600
400
200
0
0
25
50
75
100
125
150
175
x

Extension of Cramer's Model

- Cramer's model can be applied to other seemingly random sequences with known density
- e.g. Ulam numbers
- $U(1) = 1, U(2) = 2$
- $U(x)$ is least number which is a unique sum of two distinct earlier terms

Primes Themselves

- Fermat's Little Theorem, $a^{n-1} \equiv 1 \mod n \; \forall a$ coprime to n
- But, how accurate is this?
- Rewrite $n - 1 = 2^s d$
- $a^d \equiv 1 \pmod{n}$
- $a^{2^r d} \equiv -1 \pmod{n}, \quad 0 \leq r < s.$

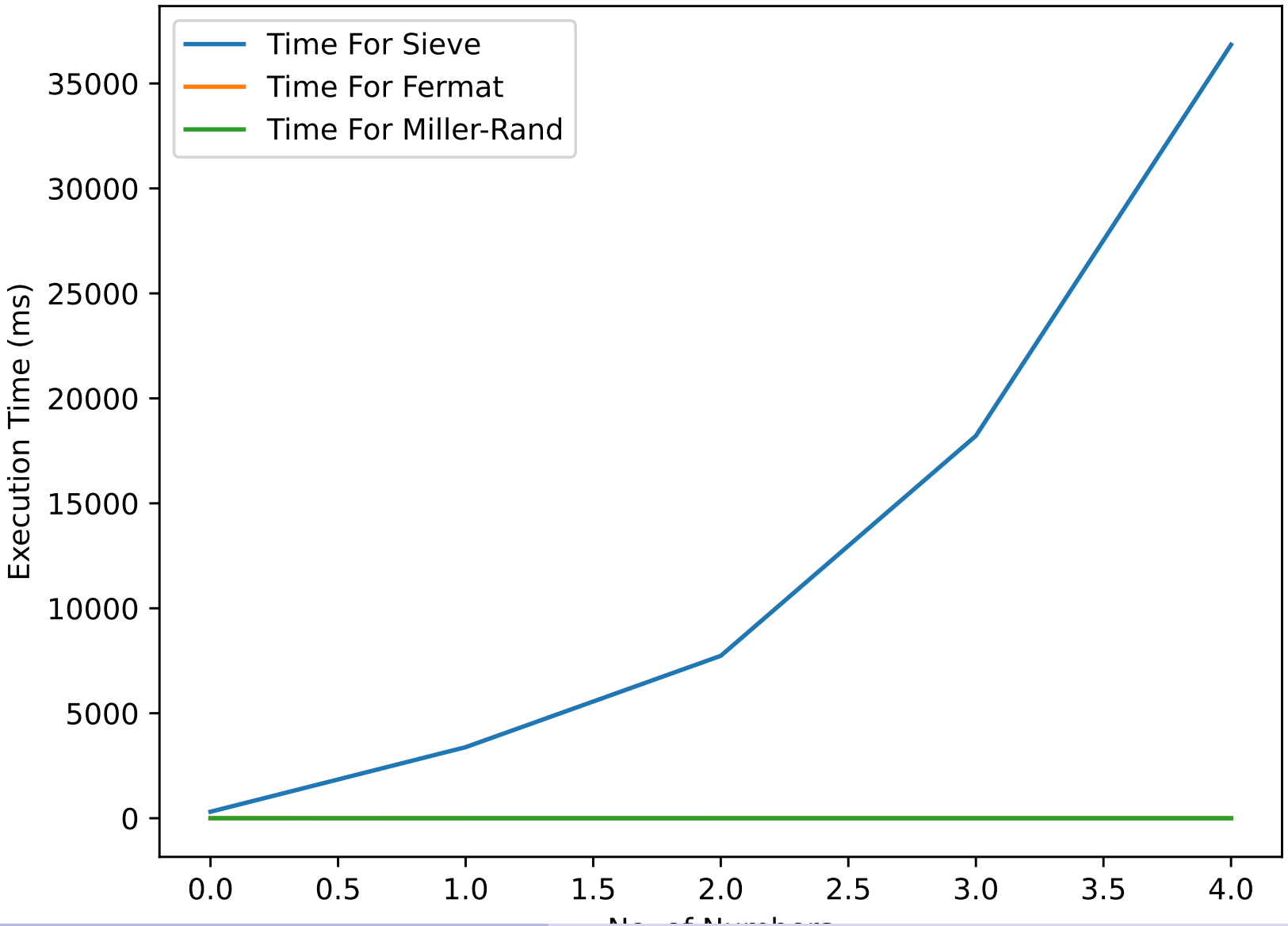
Miller Rabin

- No non-prime, odd, number exists that can pass for every a!!
- If the a chosen reveals that n is composite, it is called a witness. If it turns out it told us that a composite was prime, we call it a liar
- We don't know any way to deterministically find witnesses for a number
- So... Randomize!!
- It is shown that for $2 < a < n - 1$, there are at most $(1/4) * n$ liars

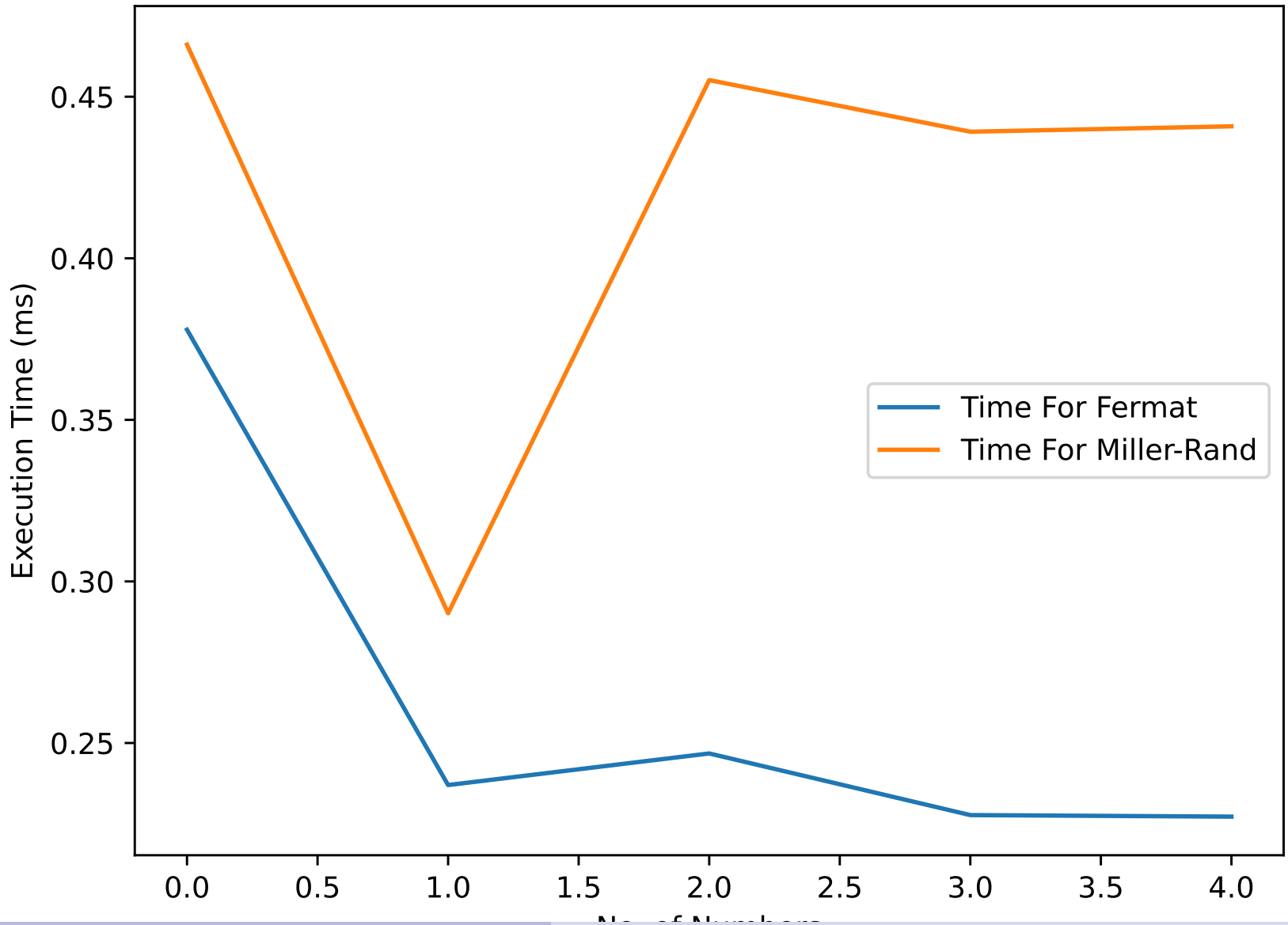
Miller-Rabin

- Doing k iterations of the algorithm, we have a confidence of 0.25^k that the number is prime
- In reality, the number of liars is much higher for most numbers
- This produces something known as an 'industry-grade' prime
- The running time is incredibly fast. The next slide will show Miller-Rabin in comparison with other algorithms

Running Time



Frame Title


0.45
0.40
0.35
0.30
0.25
Execution Time (ms)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Time For Fermat
Time For Miller-Rand

Other Primality tests

- The **The Baillie-PSW** primality test is one that is possibly deterministic
- It is a mix between a Miller-Rabin test with a $= 2$(chosen arbitrarily) and another class of numbers called Lucas primes
- There is no known crossover between the two sets of numbers upto 2^{64}