Unbounded Koopman operators and the distribution of prime numbers

Aidan Rocke
aidanrocke@gmail.com

February 25, 2021

Abstract

There exists a linear dynamical system that may be identified with an optimal prime-sieve. This dynamical system is interesting because its phase-space dimension is unbounded. From this we may deduce that the Kolmogorov Complexity of the Koopman operator is unbounded. This means that the shortest program in any language that could give the nth prime does not have finite length. More importantly, given that verifying the correctness of such a program would scale with program length, even if such a program existed we would not be able to verify its correctness in a finite number of steps.

1 An optimal prime-sieve

The goal of an optimal prime sieve is to find all prime numbers less than or equal to $N \in \mathbb{N}$. A relatively simple solution for finding the first n prime numbers $\mathbb{P}^n = \{p_i\}_{i=1}^n$ is given by the following program.

First, define the function f that takes as inputs $N \in \mathbb{N}$ and $\mathbb{P}^n = \{p_i\}_{i=1}^n$ and computes:

$$f(N, \mathbb{P}^n) = \prod_{i=1}^{n} (N \mod p_i) \tag{1}$$

If $f(N, \mathbb{P}^n) = 0$,

$$N := N + 1 \tag{2}$$

Otherwise, if $f(N, \mathbb{P}^n) = 1$, we execute the updates:

$$p_{n+1} = N \tag{3}$$

$$\mathbb{P}^{n+1} := p_{n+1} \cup \mathbb{P}^n \tag{4}$$

$$N := N + 1 \tag{5}$$

and we halt this process when the cardinality of $\mathbb{P}^n$ is as desired.

2 The Buckingham-Pi theorem

The Buckingham-Pi theorem states that if a dynamical system involves n physical variables and n also happens to be the number of physical dimensions involved then you need a system of scientific units whose cardinality is greater than or equal to n.

If you have a system of fundamental units $U = \{u_i\}_{i=1}^n$ and all the other physical units C are derived from U:

$$\forall c \in C, \exists \alpha_i \in \mathbb{Z}, c = \prod_{i=1}^{n} = u_i^{\alpha_i} \tag{6}$$

Furthermore, if we define the n-dimensional vector space:

$$\text{span}(\log U) = \big\{ \sum_{i=1}^{n} \alpha_i \log u_i | u_i \in U, \alpha_i \in \mathbb{Z} \big\} \tag{7}$$

the u_i are fundamental if they are dimensionally independent in the sense that $\forall u_i, u_{j\neq i} \in U, \log u_i \perp \log u_{j\neq i}$:

$$\exists \alpha_i \in \mathbb{Z}, \sum_{i=1}^{n} \alpha_i \log u_i \iff \alpha_i = 1 \land \alpha_{j\neq i} = 0 \tag{8}$$

Moreover, it can be shown that:

$$\log C = \text{span}(\log U) \tag{9}$$

as every element in C has a unique factorisation in terms of U.

3 The Koopman operator

In principle, any physical system may be modelled as a discrete dynamical system:

$$\forall k \in \mathbb{Z}, x_{k+1} = \Psi \circ x_k \tag{10}$$

where, without loss of generality, $x_k \in S \subset \mathbb{R}^n$, k is a discrete time index and $T : S \to S$ is a dynamic map. This representation is epistemologically sound as data collected from dynamical systems always comes in discrete-time samples.

Within this context, we may represent data as evaluations of functions of the state x_k, known as observables. In fact, if $g : S \to \mathbb{R}$ is an observable associated with the system then the collection of all such observables forms a vector space due to the Buckingham-Pi theorem.

Now, the Koopman operator Ψ is a linear transform on this vector space given by:

$$\Psi g(x) = g \circ \Psi(x) \tag{11}$$

which in a discrete setting implies:

$$\Psi g(x_k) = g \circ \Psi(x_k) = g(x_{k+1}) \tag{12}$$

where the linearity of the Koopman operator follows from the linearity of the composition operator:

$$\Psi \circ (g_1 + g_2)(x) = g_1 \circ \Psi(x) + g_2 \circ \Psi(x) \tag{13}$$

So we may think of the Koopman operator as lifting dynamics of the state space to the space of observables.

Furthermore, we may make the key observation that if Ψ is of rank n then Ψ describes the evolution of a dynamical system with an n-dimensional phase-space.

4 The prime numbers satisfy the criteria of Buckingham-Pi

It may be shown that $\mathbb{P}$ forms a fundamental system of units in the sense of Buckingham-Pi since the elements of $\log \mathbb{P}$ are dimensionally independent $\forall p_j, p_{i \neq j} \in \mathbb{P}, \log p_{i \neq j} \perp \log p_j$ in the sense that:

$$\exists \alpha_i \in \mathbb{Z}, \sum_{i=1}^{\infty} \alpha_i \log p_i = \log p_j \iff \alpha_j = 1 \wedge \alpha_{i \neq j} = 0 \tag{14}$$

Furthermore, if we define the infinite-dimensional vector space:

$$\text{span}(\log \mathbb{P}) = \Big\{ \sum_{i=1}^{\infty} \alpha_i \log p_i < \infty | p_i \in \mathbb{P}, \alpha_i \in \mathbb{Z} \Big\} \tag{15}$$

and if we define the first n primes, $\mathbb{P}^n = \{p_i\}_{i=1}^n$ then by definition:

$$\text{span}(\log \mathbb{P}^n) \subset \text{span}(\log \mathbb{P}) \tag{16}$$

Moreover, due to the unique factorisation of the integers, it may be shown that:

$$\text{span}(\log \mathbb{Q}_+) = \text{span}(\log \mathbb{P}) \tag{17}$$

5 A linear dynamical system associated with optimal prime sieving

Given the optimal prime-sieve that was introduced earlier, we may identify the first n prime numbers $\mathbb{P}^n$ with the n-dimensional vector space:

$$\text{span}(\log \mathbb{P}^n) = \Big\{ \sum_{i=1}^{n} \alpha_i \log p_i \cdot \vec{e_i} | p_i \in \mathbb{P}, \alpha_i \in \mathbb{N} \Big\} \tag{18}$$

where $\vec{e_i}$ is the usual n-dimensional unit vector with components δ_{ij}.

Now, given that the dynamics on the integers(our states) may be described by the linear model:

$$\forall n \in \mathbb{N}, f \circ n = n + 1 \tag{19}$$

we may construct a bijective mapping G which lifts dynamics from the state-space $\mathbb{N}$ to the space of observables $\text{span}(\log \mathbb{P}^n)$ such that:

$$\forall n \in \mathbb{N} \exists! X_n \in \text{span}(\log \mathbb{P}^n), X_n = G \circ n \tag{20}$$

where $X = \{X_i\}_{i=1}^n, X_i = \alpha_i \cdot \log p_i$ assuming that:

$$\exists \alpha_i \in \mathbb{N}, \log n = \sum_{i=1}^{n} \alpha_i \cdot \log p_i \tag{21}$$

Now, let's suppose that there exists a linear map $\Psi \in \mathbb{R}^{n \times n}$ such that:

$$X_{n+1} = \Psi_n \circ X_n \tag{22}$$

Given the infinitude of primes, the phase-space associated with the evolution of X_n is unbounded. Furthermore, the Expected Kolmogorov Complexity $K(\cdot)$ of Ψ at the nth iteration will scale with:

$$\mathbb{E}[K(\Psi_n)] \gtrsim \pi(n) \cdot \ln(n) \sim n \tag{23}$$

due to the Prime Number Theorem and the Shannon source coding theorem. (Check the Appendix.)

6 Analysis

From this analysis we may deduce that the shortest program in any programming language that could give the nth prime for all $n \in \mathbb{N}$ would have unbounded length. Furthermore, given that verifying the correctness of this program would scale with program length, $\sim n$, we may deduce that even if such a program existed we would not be able to verify its correctness in a finite number of steps.

As a corollary, it is only possible for mathematicians to define:

$$\zeta^n(s) = \prod_{i=1}^{n} \frac{1}{1 - p_i^{-s}} \tag{24}$$

where $\zeta^n(s) \neq \zeta(s)$ and n is bounded due to resource-constraints. This means that while a single counter-example may be used to prove that the Riemann Hypothesis is false, we can't prove that the Riemann Hypothesis is true.

References

[1] Bernard Koopman. Hamiltonian systems and Transformations in Hilbert Space.

[2] Steven L. Brunton. Notes on Koopman operator theory. 2019.

[3] Peter D. Grünwald. The Minimum Description Length Principle . MIT Press. 2007.

[4] M. Li and P. Vitányi. An Introduction to Kolmogorov Complexity and Its Applications. Graduate Texts in Computer Science. Springer. 1997.

[5] Fine Rosenberger. Number Theory: An Introduction Via the Distribution of Primes. 2007. [6] Don Zagier. Newman's short proof of the Prime Number Theorem. The American Mathematical Monthly, Vol. 104, No. 8 (Oct., 1997), pp. 705-708

7 Appendix

7.1 An information-theoretic derivation of the prime number theorem

If we know nothing about the primes in the worst case we may assume that each prime number less than or equal to N is drawn uniformly from $[1, N]$. So our source of primes is:

$$X \sim U([1, N]) \tag{1}$$

where $H(X) = \ln(N)$ is the Shannon entropy of the uniform distribution.

Now, given a strictly increasing integer sequence of length N, $U_N = \{u_i\}_{i=1}^N$, where $u_i = i$ we may define the *prime encoding* of U_N as the binary sequence $X_N = \{x_i\}_{i=1}^N$ where $x_i = 1$ if u_i is prime and $x_i = 0$ otherwise. With no prior knowledge, given that each integer is either prime or not prime, we have 2^N possible prime encodings(i.e. arrangements of the primes) in $[1, N] \subset \mathbb{N}$.

If there are $\pi(N)$ primes less than or equal to N then the average number of bits per arrangement gives us the average amount of information gained from correctly identifying each prime number in U_N as:

$$S_c = \frac{\log_2(2^N)}{\pi(N)} = \frac{N}{\pi(N)} \tag{2}$$

Furthermore, if we assume a maximum entropy distribution over the primes then we would expect that each prime is drawn from a uniform distribution as in (1) so we would have:

$$S_c = \frac{N}{\pi(N)} \sim \ln(N) \tag{3}$$

As for why the natural logarithm appears in (3), we may first note that the base of the logarithm in the Shannon Entropy may be freely chosen without changing its properties. Moreover, given the assumptions if we define $(k, k+1] \subset [1, N]$ the average distance between consecutive primes is given by the sum of weighted distances l:

$$\sum_{k=1}^{N-1} \frac{1}{k} |(k, k+1]| = \sum_{k=1}^{N-1} \frac{1}{k} \approx \sum_{l=1}^{\lambda} l \cdot P_l \approx \ln(N) \tag{4}$$

where $P_l = \frac{1}{l} \cdot \sum_{k=\frac{l\cdot(l-1)}{2}}^{\frac{l\cdot(l-1)}{2}+l-1} \frac{1}{k+1}$ and $\lambda = \frac{\sqrt{1+8(N+1)}-1}{2}$.

This is consistent with the maximum entropy assumption in (1) as there are k distinct ways to sample uniformly from $[1, k]$ and a frequency of $\frac{1}{k}$ associated with the event that a prime lies in $(k-1, k]$. The computation (4) is also consistent with Boltzmann's notion of entropy as a measure of possible arrangements.

Now, we note that given (3), and (4) we have:

$$\pi(N) \sim \frac{N}{\ln(N)} \tag{5}$$

which happens to be equivalent to the prime number theorem.

By the Shannon source coding theorem, we may also infer that $\pi(N)$ primes can't be compressed into fewer than $\pi(N) \cdot \ln(N)$ bits so this result tells us something about the incompressibility of the primes. Why might they be incompressible? By definition, all integers have non-trivial prime factorisations except for the prime numbers.

In fact, there is a strong connection between maximum entropy distributions and incompressible signals. This connection may be clarified by the following insight:

$$\mathbb{E}[K(X_N)] \sim \pi(N) \cdot \ln(N) \sim N \tag{6}$$

where the implicit assumption here is that for any recursive probability distribution, the expected value of the Kolmogorov Complexity equals the Shannon entropy. The distribution of the prime numbers is such a distribution as it is computable.

7.2 Buckingham-Pi

If we know the appropriate physical physical units for the observable Ω then we have made significant progress in understanding its behaviour. The Buckingham-Pi theorem tells us how many physical units we need and what we can do with free parameters if we happen to have more physical measurements than what is necessary to model the behaviour of Ω.

Now, let's suppose we have succeeded in identifying an equation that describes the evolution of Ω as a function of N physical units. Then we have:

$$\exists \alpha_i \in \mathbb{Z}, \Omega = \prod_{i=1}^{N} \omega_i^{\alpha_i} \tag{1}$$

where ω_i are our physical units(ex. Joules).

Now, we also know that each unit of Ω may be expressed in terms of the fundamental units $U = \{u_i\}_{i=1}^{k}$ so we have:

$$\exists \beta_i \in \mathbb{Z}, \Omega = \prod_{i=1}^{k} u_i^{\beta_i} \tag{2}$$

and for each physical unit,

$$\exists \lambda_{i,j} \in \mathbb{Z}, \omega_j^{\alpha_j} = \prod_{i=1}^{k} u_i^{\lambda_{i,j} \cdot \alpha_j} \tag{3}$$

This allows a representation in terms of the system of equations:

$$\sum_{j=1}^{N} \lambda_{i,j} \cdot \alpha_j = \beta_i \tag{4}$$

so we have

$$\Lambda \cdot \vec{\alpha} = \vec{\beta} \tag{5}$$

Now, in order to translate between different civilisations which might have different metrology institutes, it makes sense to model the problem using dimensionless parameters. This amounts to finding $\vec{\Delta\alpha}$ such that:

$$\Lambda \cdot (\vec{\alpha} + \vec{\Delta\alpha}) = \Lambda \cdot \vec{\alpha} = \vec{\beta} \tag{6}$$

where the set of all possible $\vec{\Delta\alpha}$ defines the null-space of Λ.

From the rank-nullity theorem we may deduce that:

$$\dim(\text{Null}(\Lambda)) = N - k \tag{7}$$

is the number of dimensionless parameters with which we may describe our physical system.

A useful interpretation of the null-space is that it measure the number of physical units that are not dimensionally independent. This number is exactly $N - k$ and we may use these free parameters to model the physical problem in a dimensionless manner. In other words, an over-parametrised problem is better than one that is under-parametrised.