THE MONTE CARLO HYPOTHESIS AND ITS IMPLICATIONS

Aidan Rocke
aidanrocke@gmail.com

June 17, 2021

ABSTRACT

The Monte Carlo hypothesis implies a non-trivial interpretation of the prime number theorem. The Prime Number Theorem says how the prime numbers are distributed but not why. On the other hand, an information-theoretic analysis of the Prime Number Theorem indicates that they are distributed in this way because prime encodings are algorithmically random and the prime numbers have a maximum entropy distribution.

1 An information-theoretic derivation of the density of primes

Let's define the indicator variable X_p where $X_p = 1$ if x is divisible by p and $X_p = 0$ otherwise. From a frequentist perspective,

$$\mathbb{E}[X_p] = 1 \cdot \frac{1}{p} + 0 \cdot (1 - \frac{1}{p}) = \frac{1}{p} \tag{1}$$

as the numbers divisible by p occur with frequency $\frac{1}{p}$.

Given this frequentist perspective, we may define the frequency P_N with which an integer $x \in [1, N]$ is not divisible by any prime in $[1, N]$ as:

$$\forall x \sim U([1, N]), P(x \in \mathbb{P}) = \prod_{p \leq N} (1 - \frac{1}{p}) \tag{2}$$

and due to Mertens' second theorem we have:

$$-\ln P_N \approx \sum_{p \leq N} \frac{1}{p} \approx \ln \ln N + \mathcal{O}(1) \tag{3}$$

so for large N, the density of primes is on the order of:

$$P_N \propto \frac{1}{\ln N} \tag{4}$$

Now, in order to get an exact asymptotic for (4) let's suppose there exists a function $\lambda_N = f(N)$ such that:

$$\ln \lambda_N = \mathcal{O}(\lim_{N \to \infty} -\ln P_N) \tag{5}$$

$$\frac{1}{\lambda_N} = \mathcal{O}(\lim_{N \to \infty} \frac{\partial}{\partial \lambda_N} - \ln P_N) \tag{6}$$

where (5) and (6) denote Occam approximations, clarified in the Appendix.

It follows that for large N, $-\ln P_N$ behaves like the entropy of a random variable distributed uniformly in $[1, N]$ with frequency $\frac{1}{\lambda_N}$. In fact, the asymptotic rate of change in the entropy equals the asymptotic frequency with which a prime is observed in $[1, N]$.

Due to Mertens' third theorem, we may set $\lambda_N := \ln N$ in (5) so we have:

$$\frac{1}{\ln N} = \mathcal{O}(\lim_{N\to\infty} \frac{\partial}{\partial \lambda_N} - \ln P_N) \tag{7}$$

As a result, we have:

$$P_N \sim \frac{1}{\ln N} \tag{8}$$

and if $\pi(N)$ counts the number of primes less than N,

$$\frac{\pi(N)}{N} \sim \frac{1}{\ln N} \tag{9}$$

Finally, given that the density of the primes in $[1, N]$ is on the order of $\frac{1}{\ln N}$ we may also verify that:

$$\pi(N) \sim \int_2^N \frac{1}{\ln x} dx \sim \frac{N}{\ln N} \tag{10}$$

which tells us how the prime numbers are distributed, but not why.

2 The Monte Carlo Hypothesis, or the algorithmic randomness of prime encodings

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

$$X \sim U([1, N]) \tag{11}$$

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

Now, we may define the prime encoding of $[1, N]$ as the binary sequence $X_N = \{x_n\}_{n=1}^{N}$ where $x_n = 1$ if n is prime and $x_n = 0$ otherwise. With no prior knowledge, given that each integer is either prime or not prime, we have 2^N possible prime encodings 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 in $[1, N]$ as:

$$S_c = \frac{\log_2(2^N)}{\pi(N)} = \frac{N}{\pi(N)} \tag{12}$$

and 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 (11). So we would have:

$$S_c = \frac{N}{\pi(N)} \sim \ln N \tag{13}$$

In fact, given the assumptions the expected information gained from observing each prime in $[1, N]$ is on the order of:

$$\sum_{k=1}^{N-1} \frac{1}{k} \cdot |(k, k+1]| = \sum_{k=1}^{N-1} \frac{1}{k} \approx \ln N \tag{14}$$

as there are k distinct ways to sample uniformly from $[1, k]$ and a frequency of $\frac{1}{k}$ associated with the event that $k \in \mathbb{P}$.

This implies that the average number of bits per prime number is given by $\frac{N}{\pi(N)} \sim \ln N$. Rearranging, we find:

$$\frac{\pi(N)}{N} \sim \frac{1}{\ln N} \tag{15}$$

which is in complete agreement with the Prime Number Theorem.

2.1 The Shannon source coding theorem and the algorithmic randomness of prime encodings

By the Shannon source coding theorem, we may infer that $\pi(N)$ primes can't be compressed into fewer than $\pi(N) \cdot \ln N$ bits. Furthermore, as the expected Kolmogorov Complexity equals the Shannon entropy for computable probability distributions:

$$\mathbb{E}[K(X_N)] \sim \pi(N) \cdot \ln N \sim N \tag{16}$$

where the identification of $\mathbb{E}[K(X_N)]$ with $\pi(N) \cdot \ln N$ is a direct consequence of the fact that $K(X_N)$ measures the information gained from observing X_N and $\pi(N) \cdot \ln N$ measures the expected information gained from observing X_N.

However, as it is not possible to prove that a particular object is incompressible within algorithmic information theory we may experimentally verify the hypothesis (16) that prime-encodings are algorithmically random and prime numbers have a maximum entropy distribution using machine learning methods.

3 The Monte Carlo Hypothesis

3.1 Challenges in experimental mathematics

The maximum entropy distribution of the prime numbers and the algorithmic randomness of prime encodings are naturally divided into two separate challenges.

3.2 Probabilistic Primality testing

The hypothesis that the prime numbers have a maximum entropy distribution implies that the true positive rate of a primality test with bounded algorithmic information content converges to zero.

The appropriate data for such an experimental verification consists of binary encodings of the integers which for all prime numbers less than N requires $\log_2 N$ bits for representation.

3.3 Weighted Next-bit test

The hypothesis that prime encodings are algorithmically random implies that a sequence of prime encodings pass the weighted next-bit test. Given a sequence of N consecutive prime encodings, the probability that a machine learning model M accurately predicts the next term converges to $p \in [0, \frac{1}{2}]$.

Together, these hypotheses succinctly expressed as:

$$\mathbb{E}[K(X_N)] \sim \pi(N) \cdot \ln N \sim N \tag{17}$$

are known as the Monte Carlo Hypothesis. So far both hypotheses have been tested for all prime numbers under 10^9. Feel free to check https://github.com/AidanRocke/Monte-Carlo-Hypothesis for a getting-started guide.

3.4 Discussion

In order to develop a better intuition for (17), we may derive it from the assumptions on prime encodings. We shall note that all the information in X_N is contained in the location of the primes $\pi(N)$, and since it is assumed that each prime occurs independently of the others we may decompose $\mathbb{E}[K(X_N)]$ into:

$$\mathbb{E}[K(X_N)] \sim \pi(N) \cdot I_{\pi(N)} \tag{18}$$

where $I_{\pi(N)}$ is the expected information gained per prime.

Given the assumptions:

$$I_{\pi(N)} = \sum_{n=1}^{N} \frac{1}{n} \sim \ln N \tag{19}$$

and since $I_{\pi(N)} = \frac{N}{\pi(N)} \sim \ln N$, we may deduce that:

$$\mathbb{E}[K(X_N)] \sim \pi(N) \cdot \ln N \sim N \tag{20}$$

4 Implications for Cramér's conjecture

Given the consecutive primes $p_n, p_{n+1} \in \mathbb{P}$ we may define the prime encoding:

$$X_{p_n} = \{x_k\}_{k=p_n}^{p_{n+1}} \tag{21}$$

where $x_{p_n} = x_{p_{n+1}} = 1$ and $x_k = 0$ for $k \in [p_n + 1, p_{n+1} - 1]$.
Now, if we recall that:

$$\forall x \sim U([p_n, p_{n+1}]), P(x \in \mathbb{P}) \sim \frac{1}{\ln p_n} \tag{22}$$

in the worst case the occurrences of p_n and p_{n+1} are independent so we have:

$$P(x_{p_n} \wedge x_{p_{n+1}}) \geq \frac{1}{(\ln p_n)^2} \tag{23}$$

where $\ln p_n < \ln p_{n+1} < \ln p_n + 1$ due to Bertrand's postulate.
The independence of consecutive primes is certainly possible if X_{p_n} is incompressible or algorithmically random:

$$\mathbb{E}[K(X_{p_n})] \sim p_{n+1} - p_n \tag{24}$$

which is consistent with the Monte Carlo Hypothesis.
Moreover, with (20) we may define the probability that X_{p_n} halts at $x_{x>p_n}$ as:

$$P(\text{halts at } x_k) \geq \frac{1}{(\ln p_n)^2} \tag{25}$$

so we have:

$$P(\text{halts at } x_{k \leq p_{n+1}}) = \sum_{k=p_n+1}^{p_{n+1}} P(\text{halts at } x_k) = 1 \tag{26}$$

and the combination of (22) and (23) implies:

$$p_{n+1} - p_n = \mathcal{O}((\ln p_n)^2) \tag{27}$$

as conjectured by Harald Cramér.

5 Implications for the Riemann Hypothesis

Experimental verification of the Monte Carlo Hypothesis would imply that knowledge of more informative prime number distributions is inaccessible to civilisations within the Turing limit.

References

[1] Aidan Rocke (https://mathoverflow.net/users/56328/aidan-rocke), information-theoretic derivation of the prime number theorem, URL (version: 2021-02-20): https://mathoverflow.net/q/384109

[2] Terence Tao. Structure and randomness in the prime numbers: A small selection of results in number theory. Slides. 2007.

[3] A. N. Kolmogorov Three approaches to the quantitative definition of information. Problems of Information and Transmission, 1(1):1–7, 1965

[4] G. J. Chaitin On the length of programs for computing finite binary sequences: Statistical considerations. Journal of the ACM, 16(1):145–159, 1969.

[5] R. J. Solomonoff A formal theory of inductive inference: Parts 1 and 2. Information and Control, 7:1–22 and 224–254, 1964.

[6] Schnorr, C. P. (1971). "A unified approach to the definition of a random sequence". Mathematical Systems Theory.

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

[8] David Deutsch. Quantum theory, the Church–Turing principle and the universal quantum computer. 1985.

[9] Don Zagier. Newman's short proof of the Prime Number Theorem. The American Mathematical Monthly, Vol. 104, No. 8 (Oct., 1997), pp. 705-708

[10] John A. Wheeler, 1990, "Information, physics, quantum: The search for links" in W. Zurek (ed.) Complexity, Entropy, and the Physics of Information. Redwood City, CA: Addison-Wesley.

[11] Guillermo Valle Pérez, Chico Camargo, Ard Louis. Deep Learning generalizes because the parameter-function map is biased towards simple functions. 2019.

[12] Aidan Rocke (https://cstheory.stackexchange.com/users/47594/aidan-rocke), Understanding the Physical Church-Turing thesis and its implications, URL (version: 2021-02-22): https://cstheory.stackexchange.com/q/48450

[13] Igor V. Volovich. Number Theory as the Ultimate Physical Theory. Pleiades Publishing. 2010.

[14] Cramér, H. "On the Order of Magnitude of the Difference Between Consecutive Prime Numbers." Acta Arith. 2, 23-46, 1936.

6 Appendix

6.1 Occam's razor and non-linear approximations

Let's suppose we have a smooth function $g : \mathbb{R} \to \mathbb{R}$ that describes recently-discovered physical phenomena whose behaviour is non-linear in the neighbourhood of $x = a$. In order to describe the behaviour of g near $x = a$, we may define the function space F of Taylor polynomials:

$$\lim_{x \to a} \frac{g(x)}{f(x)} = 1 \implies f \in F \tag{28}$$

where F contains an *Occam approximation* of g in the neighbourhood of $x = a$ if it satisfies two properties:

1. Non-linearity in the neighbourhood of $x = a$ implies that the lowest-order Taylor polynomial in F must be greater than or equal to two. This requirement implies that F contains descriptions of the behaviour of g.

2. F is constrained to satisfy the *prefix property*. Intuitively, this means that $f' \in F$ if and only if it appears as a prefix of some $f \in F$ where $f \neq f'$. A more precise definition of this prefix property shall be given in the next paragraph in terms of computation graphs.

Computation graphs of Taylor polynomials:

In order to define Occam's approximation of g near $x = a$, we may represent each polynomial $f \in F$ in terms of its computation graph $G \circ f$:

$$f = \sum_{n=1}^{N} a_n(x) \implies G \circ f = \{a_n\}_{n=1}^{N} \tag{29}$$

where $\forall n \in [1, N], a_n(x) = \alpha_n \cdot x^{n-1}$ for some constant $\alpha_n \in \mathbb{R}$. Given $G \circ f$, we may also readily recover f:

$$f = G^{-1} \circ (G \circ f) \tag{30}$$

Using (2), we may represent the computation of a Taylor polynomial f by a directed Hamiltonian path from the node a_1 to the node a_n where the edge between consecutive nodes is the addition operation. In this formalism, f is a sub-computation of $f' \in F$ if $G \circ f$ appears as a sub-graph of $G \circ f'$:

$$G \circ f \subset G \circ f' \tag{31}$$

In fact, (4) captures the prefix property of F in the sense that:

$$f' \in F \implies \exists f \in F, f' \neq f \wedge G \circ f \subset G \circ f' \tag{32}$$

and given (5), Occam's approximation of g near $x = a$ is none other than:

$$f = G^{-1} \circ \bigcap_{f' \in F} G \circ f' \tag{33}$$

and for the sake of brevity:

$$f = \mathcal{O}\big(\lim_{x \to a} g(x)\big) \tag{34}$$

The case where the lowest-order Taylor polynomial is infinite:

If the lowest-order Taylor polynomial in F is infinite, we may use a space of special functions H instead in order to guarantee that all computation graphs have a finite number of nodes.

By requiring that H satisfy the prefix property, we may guarantee that Occam's approximation of a special function g near $x = a$ with respect to H is well-defined.

6.2 A complexity bound on Gödel numbers

By assigning a Gödel number to each logical proposition formulated with respect to a formal system F, a variation of Chaitin's incompressibility argument leads us to a complexity-bound on the largest Gödel number whose truth-value is decidable. In other words, if a sufficiently complex mathematical proposition is true then it is not provable.

Assigning Gödel numbers to logical propositions:

Given a formal system F with N symbols, each symbol may be assigned a unique integer from $A \subset \mathbb{N}$ where $|A| = N$. In particular, to encode any logical proposition such as:

$$a = (a_1, a_2, ..., a_n) \in A^n \tag{35}$$

we may use the prime factorisation:

$$G(a) = \prod_{i=1}^{n} p_i^{a_i} = 2^{a_1} \cdot 3^{a_2} ... \cdot p_n^{a_n} \tag{36}$$

and by the fundamental theorem of arithmetic, i.e. unique prime factorisation, it is possible to recover the original logical proposition from its Gödel number.

Chaitin's incompressibility argument:

For concreteness, let's suppose F corresponds to Peano Arithmetic(PA). Given that a universal theory of computation may be modelled with PA and any formal system necessarily has a finite description, F may be described in a finite number of bits:

$$\exists f \in \mathbb{N}, K(F) = f \tag{37}$$

with respect to prefix-free Kolmogorov Complexity.

Moreover, for any Gödel number $x \in \mathbb{N}^*$ decidable in F we may express:

$$K(x) \geq \log_2(x) \tag{38}$$

where $\log_2(x)$ is the length of the binary encoding of x.

Now, let's define a large Gödel number $N >> f$ and perform a stochastic search until we find a Gödel number $x \in \mathbb{N}$ such that:

$$K(x) \geq \log_2(x) \geq N + \text{Cst} \tag{39}$$

Given that almost all integers are incompressible the search algorithm has converged to (3) almost surely. However, given that the search algorithm $\mathcal{A}$ is assumed to be consistent with F it can't contain more independent axioms than F so we have:

$$K(\mathcal{A}) \leq f \tag{40}$$

and since x is defined in terms of N with respect to $\mathcal{A}$ it may be encoded using at most $\log_2(N)$ bits so we have:

$$K(x) \leq \log_2(N) + f \tag{41}$$

which leads us to a contradiction almost surely.

References:

1. Whitehead, Alfred North and Bertrand Russell (1963). Principia Mathematica. Cambridge: Cambridge University Press

2. Smoryński, C., 1977, "The incompleteness theorems," in Handbook of Mathematical Logic, J. Barwise (ed.), Amsterdam: North-Holland

3. G. J. Chaitin, Information-theoretic limitations of formal systems, J. Assoc. Comput. Mach. 21 (1974); 403–424 (Reprinted in: [19], 113–128).