THE PRIME NUMBER THEOREM, OR THE INCOMPRESSIBILITY OF THE PRIMES

Aidan Rocke
aidanrocke@gmail.com

February 16, 2021

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 expect:

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

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.

There is another useful interpretation of (4). If we break $\sum_{k=1}^{N} \frac{1}{k}$ into $\pi(N)$ disjoint blocks of size $[p_k, p_{k+1}]$ where $p_k, p_{k+1} \in \mathbb{P}$ and $p_1 = 1$:

$$\sum_{k=1}^{N} \frac{1}{k} \approx \sum_{k=1}^{\pi(N)} \sum_{b=p_k}^{p_{k+1}} \frac{1}{b} = \sum_{k=1}^{\pi(N)} (p_{k+1} - p_k) \cdot P(p_k) \approx \ln(N) \tag{5}$$

where $P(p_k) = \frac{1}{(p_{k+1}-p_k)} \sum_{b=p_k}^{p_{k+1}} \frac{1}{b}$. So we see that (4) also approximates the expected number of observations per prime where $P(p_k)$ may be interpreted as the probability of a successful observation in a frequentist sense. This is consistent with John Wheeler's *it from bit* interpretation of entropy [5], where the entropy measures the average number of bits(i.e. yes/no questions) per prime number.

Now, given (3),(4) and (5) this implies that the average number of bits per prime number is given by:

$$\frac{N}{\pi(N)} \sim \ln(N) \tag{6}$$

which is in complete agreement with the predictions of the prime number theorem.

2 The Shannon source coding theorem, and the compressibility of the prime numbers

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 the primes are incompressible. In fact, given that the expected Kolmogorov Complexity equals the Shannon entropy for computable probability distributions for a prime encoding of length N we must asymptotically obtain:

$$\mathbb{E}[K(X_N)] \sim \pi(N) \cdot \ln(N) \sim N \tag{7}$$

where $K(\cdot)$ is the Kolmogorov Complexity.

We shall now proceed by contradiction. If there is an algorithmic method which may be used to prove that the Riemann Hypothesis is true then we may construct a program of finite length $textitzeta$ which takes as input a strictly increasing integer sequence of length N, U_N, and outputs a prime encoding of length N, X_N, by correctly deciding whether each element in that sequence is prime or not.

By the hypothesis on *zeta* and U_N, an application of the Minimum Description Length principle yields:

$$\mathbb{E}[K(\ \textit{zeta} \circ U_N)] \leq -\ln(P(X_N|\textit{zeta} \circ U_N)) + \text{Cst} = \text{Cst} \tag{8}$$

since $P(X_N|\textit{zeta} \circ U_N) = 1$, as there exists a prime encoding $X_N \in \{0,1\}^N$ such that $\textit{zeta} \circ U_N = X_N$. So we must also have:

$$\mathbb{E}[K(\ \textit{zeta} \circ U_N)] = \mathbb{E}[K(X_N)] \tag{9}$$

However, X_N is known to be incompressible due to (5) so we have:

$$\lim_{N\to\infty} \frac{\mathbb{E}[K(\ \textit{zeta} \circ U_N)]}{\mathbb{E}[K(X_N)]} \sim \lim_{N\to\infty} \frac{\text{Cst}}{N} = 0 \tag{10}$$

which is a contradiction.

From this analysis we may conclude 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.

3 Testable predictions for machine learning researchers:

I have colleagues in the machine learning community that believe their powerful machine learning methods may be used to predict the distribution of primes with much greater accuracy than any of the statistical algorithms developed so far. However, my model implies that independently of the amount of data and computational resources at their disposal, if the best machine learning model predicts the next N primes to be at $\{a_i\}_{i=1}^{N} \in \mathbb{N}$ then for large N their model's statistical performance will converge to an accuracy that is no better than:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{a_i} \tag{11}$$

and the the scenario where they predict all N primes accurately occurs with a frequency of:

$$\prod_{i=1}^{N}\frac{1}{a_i} \tag{12}$$

4 Discussion

Concerning the MaxEnt model I propose it is the simplest model that correctly predicts the average number of bits per prime number. In this sense, it is also unique in satisfying Occam's razor or what has been formalised as the Minimum Description Length principle [3]. Furthermore, I can also explain why this model works. It turns out that there is a general connection between maximum entropy distributions and incompressible signals.

The primes are incompressible for two fundamental reasons. We may first note that they are dimensionally independent which implies that every integer has a unique prime factorisation. Since any mathematical system that is sufficiently powerful to formulate a general theory of computation must contain arithmetic and the integers have a unique representation in terms of the primes, it follows that all the other types in such a system are derived types relative to the prime numbers.

For a deeper insight into why the prime numbers appear to behave in an algorithmically random manner, one may consider the fact that if the human brain may be simulated by a Turing machine then algorithmic randomness is not observer independent [6]. This is a reasonable proposition if you consider the all-or-nothing operation of neurons as well as the observation that all organisms have finite sensory and behavioural state spaces.

References

[1] Dániel Schumayer and David A. W. Hutchinson. Physics of the Riemann Hypothesis. Arxiv. 2011.

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

[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] Peter Shor. Shannon's noiseless coding theorem. lecture notes. 2010.

[6] Aidan Rocke (https://cstheory.stackexchange.com/users/47594/aidan-rocke), On the equivalence of incompressibility and incompleteness in machine learning, URL (version: 2021-02-21): https://cstheory.stackexchange.com/q/48443