

---

# ASYMPTOTIC INCOMPRESSIBILITY AND ITS APPLICATION TO RARE-EVENT MODELLING

---

Aidan Rocke  
aidanrocke@gmail.com

April 8, 2021

## ABSTRACT

The objective of this article is to present a theory of algorithmic information that allows us to characterise rare events that are recurrent, of variable frequency, and unpredictable. Within this theory, it may be shown that such data-generating processes are asymptotically incompressible and therefore identifiable via an invariance theorem for algorithmically random data. While such phenomena define the worst case in rare-event modelling, a robust approach for distinguishing such cases which are intractable from cases which are algorithmically tractable has so far been lacking.

## 1 Unpredictable phenomena are asymptotically incompressible

Given a rare-event process  $X$ , a scientist may only collect a finite number of observations  $X_N = \{x_i\}_{i=1}^N$  from that process. Moreover, let's suppose  $x_i = 1$  when the event of interest is observed and  $x_i = 0$  otherwise. Then the location of the rare events, i.e.  $i$  where  $x_i = 1$ , are necessary and sufficient to define the binary encoding  $X_N$ .

This process is deterministic if there exists a computable function  $f$  such that:

$$x_{n+1} = f \circ x_n \quad (1)$$

and if there exists an asymptotic formula  $\pi(N)$  such that for large  $N$ ,

$$\pi(N) \sim \sum_{i=1}^N x_i \quad (2)$$

then we say that  $X$  is an unpredictable process if the average amount of information gained from the occurrence of each rare-event is given by the combinatorial entropy:

$$S_c = \frac{\log_2(2^N)}{\pi(N)} = \frac{N}{\pi(N)} \sim \ln(N) \quad (3)$$

where  $2^N$  is exactly the number of possible binary encodings of length  $N$  and  $\sim \ln(N)$  implies that the rate of entropy production is maximal. (3) also implies that  $X_N$  may not be compressed into fewer than  $\pi(N) \cdot \ln(N)$  bits without being certain that information will be lost.

In fact,  $X_N$  is asymptotically incompressible in the sense that:

$$\mathbb{E}[K(X_N)] \sim \pi(N) \cdot \ln(N) \sim N \quad (4)$$

as (3) states that on average each observation  $x_i$  is a surprising event.

Now, a direct implication of (4) is that any approximation  $\hat{f} \in F_\theta$  of  $f$  discovered using machine learning or another scientific method such that the Kronecker delta satisfies:

$$\forall n \in [1, N], \delta_{\hat{f}(x_n), x_{n+1}} = 1 \quad (5)$$

has an algorithmic complexity that scales as follows:

$$K(\hat{f}) \sim N \quad (6)$$

as such high levels of accuracy are due to memorisation and not discovering regularities in  $X_N$ . Given (4), (5), and (6) we may state that the process  $X$  is algorithmically random with respect to  $F_\theta$ .

## 2 A derivation of the maximum entropy distribution

Let's define the sequence  $\mathbb{P} = \{p_k\}_{k=1}^\infty \subset \mathbb{N}$  such that  $p_k \in \mathbb{P}$  if and only if  $x_{p_k} = 1$ . Then the  $\ln(N)$  term in (4) implies that the elements of  $\mathbb{P}$  are in some sense uniformly distributed in  $X_N$ .

Given that there are  $k$  distinct ways to sample uniformly from  $[1, k]$  and a frequency of  $\frac{1}{k}$  associated with the event  $(k-1, k] \cap \mathbb{P} \neq \emptyset$  the average entropy rate has a natural interpretation.

By breaking  $\sum_{k=1}^N \frac{1}{k}$  into  $\pi(N)$  disjoint blocks of size  $[p_k, p_{k+1}]$  where  $p_k, p_{k+1} \in \mathbb{P}$ :

$$\sum_{k=1}^N \frac{1}{k} \approx \sum_{k=1}^{\pi(N)} \sum_{n=p_k}^{p_{k+1}} \frac{1}{n} = \sum_{k=1}^{\pi(N)} (p_{k+1} - p_k) \cdot P(p_k) \approx \ln(N) \quad (7)$$

where  $P(p_k) = \frac{1}{p_{k+1} - p_k} \sum_{n=p_k}^{p_{k+1}} \frac{1}{n}$ .

So we see that (4) approximates the expected number of observations per rare event 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 where entropy measures the average number of bits(i.e. yes/no questions) per rare event.

Interestingly, (7) may also be interpreted as the expected distance or waiting time between consecutive rare events as we have:

$$\mathbb{E}[|p_{n+1} - p_n|] = \sum_{k=1}^{\pi(N)} (p_{k+1} - p_k) \cdot P(p_k) \approx \ln(N) \quad (8)$$

Having clarified the rate of entropy production of  $X$  we may consider a practical method for identifying data-generating processes that are asymptotically incompressible.

### 3 An invariance theorem for algorithmically random data

Let's suppose we have a natural signal described by the process  $X$ :

$$x_n \in \{0, 1\}, x_{n+1} = \varphi \circ x_n \quad (9)$$

If we should use machine learning to approximate  $\varphi$  given the datasets  $X_N^{train} = \{x_i\}_{i=1}^N$ ,  $X_N^{test} = \{x_i\}_{i=N+1}^{2N}$  such that for any  $\hat{f} \in F_\theta$ :

$$x_{n+1} = \hat{f} \circ x_n \Rightarrow \delta_{\hat{f}(x_n), x_{n+1}} = 1 \quad (10)$$

then  $X_N$  is asymptotically incompressible if for large  $N$  any solution to the empirical risk minimisation problem:

$$\hat{f} = \max_{f \in F_\theta} \frac{1}{N} \sum_{i=1}^N \delta_{f(x_n), x_{n+1}} \quad (11)$$

has an expected performance:

$$\frac{1}{N} \sum_{n=N+1}^{2N-1} \delta_{f(x_n), x_{n+1}} \leq \frac{1}{2} \quad (12)$$

Furthermore, if the dataset is imbalanced i.e.  $\frac{1}{N} \sum_{i=1}^N x_i \neq \frac{1}{2}$  then we may generalise this result by introducing the auxiliary definitions:

$$y_n = x_{n+1} \quad (13)$$

$$\hat{y}_n = \hat{f} \circ x_n \quad (14)$$

$$\beta_n = \delta_{y_n, \hat{y}_n} \quad (15)$$

$$N_1 = \sum_{n=N+1}^{2N} \delta_{y_n, 1} \quad (16)$$

$$N_0 = \sum_{n=N+1}^{2N} \delta_{y_n, 0} \quad (17)$$

and so for large  $N$ , we have:

$$\frac{1}{2} \sum_{n=N+1}^{2N-1} \left( \frac{\delta_{y_n, 1} \cdot \beta_n}{N_1} + \frac{\delta_{y_n, 0} \cdot \beta_n}{N_0} \right) \leq \frac{1}{2} \quad (18)$$

which implies that the true-positive rate is an unbiased estimator of (18):

$$\frac{1}{N_1} \sum_{n=N+1}^{2N-1} \delta_{y_n, 1} \cdot \beta_n \leq \frac{1}{2} \quad (19)$$

Finally, as these results are invariant to transformations that preserve the phase-space dimension of  $X$ , this theorem may be used as an overfitting test for algorithmically-random data.

## 4 Discussion

Given that the prime numbers are the only known rare-event process with a distribution satisfying (4):

$$\pi(N) \sim \frac{N}{\ln(N)} \quad (20)$$

the overfitting test (19) may be applied to the scientific problem of identifying cosmic signals communicated by civilisations within the Turing limit that are sufficiently advanced to do number theory.

## References

- [1] 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.
- [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] M. Li and P. Vitányi. An Introduction to Kolmogorov Complexity and Its Applications. Graduate Texts in Computer Science. Springer. 1997.
- [4] Peter Shor. Shannon's noiseless coding theorem. lecture notes. 2010.
- [5] Jerome H. Friedman, Robert Tibshirani, and Trevor Hastie. The Elements of Statistical Learning. Springer. 2001.
- [6] Andrew Chi-Chih Yao. Theory and applications of trapdoor functions. In Proceedings of the 23rd IEEE Symposium on Foundations of Computer Science, 1982.
- [7] Sullivan, Walter (September 29, 1968). "The Universe Is Not Ours Alone; From the edge of the galaxy". The New York Times.
- [8] Alexander L. Zaitsev. Rationale for METI. Arxiv. 2011.