

Information-theoretic measures predict the human judgment of rhythm complexity

Remi de Fleurian, Tim Blackwell, Oded Ben-Tal, Daniel Müllensiefen

Supplementary material A

Generative algorithms and associated complexity values

A very large number of generative sequences have been defined and examined in the mathematical and scientific literature (the Online Encyclopedia of Integer Sequences,¹ or OEIS, has over 220000 entries). We devised four broad groupings and collected, as far as possible, representative sequences which have known H , h , E , and T . We computed our own upper bound estimates to the Kolmogorov complexity by compressing sequences of length 10^7 with the Lempel-Ziv 1978 algorithm and counting the number of codewords per symbol (Li & Sleep, 2004). We computed h , E and T wherever values were not available by analyzing generative sequences of 10^7 symbols. The sequences were all chosen for ease of generation; the algorithms in each case amounted to only a few lines of Java code.

The groupings reflect the means of generation and overall statistical properties of the sequences. We have deliberately avoided any sequence of musical origin; our sequences are purely abstract entities.

1. Deterministic-periodic

These sequences are repetitions of finite length substrings. They are generated by deterministic algorithms. We chose one sequence of period 5 (P5c: (10101)*), three sequences of period 12 (P12.165: (001010010101)*, P12.88: (000010111101)*, P12.116: (000100100101)*) and a single sequence of period 16 (P16: (1010111011101110)*). The sequences of period 12 were chosen to have entropy close to 1 (approximately equal numbers of 0's and 1's) and to have T values that spanned the possible range. The complexities of all inequivalent sequences of period 12 were computed; two sequences of period p are equivalent if they cannot be decomposed into sequences of smaller period, if they are identical under interchange of 0 and 1 or if one is a cyclic permutation of the other. The sequence complexities are given in Table A1. The entropy rates are zero as expected, and the K -complexities are low (0.01 - 0.0178) since these sequences are easily compressed.

2. Stochastic

This group contains sequences where each symbol is drawn independently from a probability distribution. Two sequences were produced using a pseudorandom number generator to produce 0's and 1's with probability 0.5 of either (Bernoulli-0.5 sequence), and with probability 0.7 of one symbol and 0.3 of the other (Bernoulli-0.7 sequences). Table A1 shows the complexity values. In this case, E and T are zero because the sequences are produced from sources without memory (symbols do not depend on the past) and because there is nothing to synchronize to. The K -complexities are high (0.0514 - 0.0574) because the sequences, ideally, are incompressible.

¹ Accessible at: <http://oeis.org/>

3. Deterministic-stochastic

These processes mix a deterministic rule with a stochastic one. The ‘golden mean’ process (GM) produces strings with no consecutive 0’s. A 0 or a 1 is generated with probability 1/2. The next symbol is certainly a 1 if a 0 was generated; otherwise a 0 or 1 is again generated with equal probabilities. The RRXOR process proceeds as follows: two Bernoulli-0.5 generations are immediately followed by their XOR i.e. sub-sequences 000, 011, 101, 110 occur with equal probability. The even process generates strings with even numbers of 1’s bounded by any number of 0’s. A 0 or a pair of 1’s is generated with even probability followed by a Bernoulli-0.5 trial. A full account of these sequences can be found in Crutchfield and Feldman (2003). Table A1 tabulates the complexities. None of the measures is zero, indicating the mix of stochastic and deterministic elements. The sequences are fairly incompressible with high K -complexities (0.0415 - 0.0469).

4. Deterministic-aperiodic

These sequences are deterministic but are not strictly periodic. The Thue-Morse process (Crutchfield & Feldman, 2003) is produced as follows. The sequence starts with 01. The complement of 01 is then added to produce 01 10. This is again followed by the complement to produce 0110 1001. The procedure is iterated.

The Fibonacci Word sequence (sequence A003849 of OEIS), or FW, is produced by adding the sequences at iteration n and at iteration $n + 1$ to produce the sequence at iteration $n + 2$. It begins $S(0) = 0, S(1) = 01, S(2) = S(0)S(1) = 0 + 01 = 001$ and continues $S(3) = 01 + 001 = 01001$.

The quasi-periodic linear sequence (Tao, 2006), or QP-linear, is formed by the linearly structured set

$$L = \{n: [\alpha n] < \delta\} \quad (5)$$

where $[x]$ denotes the fractional part of x and, for this experiment, $\alpha = \sqrt{7}$ and $\delta = 0.5$. The n ’th term of the sequence is 1 if n is in L and 0 otherwise. The resulting sequence is almost periodic since symbols at n and $n + L$ are correlated by virtue of the identity

$$[\alpha(n + L)] - [\alpha n] = [\alpha L]. \quad (6)$$

The quasi-periodic quadratic sequence (QP-quad) is formed in a similar way to QP-linear. The quadratically structured set is

$$Q = \{n: [\alpha n^2] < \delta\} \quad (7)$$

and the n ’th term is 1 if n is in Q . The sequence is more random than QP-linear but has some vestige of periodicity.

The quasi-periodic random sequence (QP-rand) has even more stochasticity; it is formed in a similar way to QP-quad except values of n in Q are further subject to a test. They are included in R with probability δ' . The definition is

$$Q = \{n: [\alpha n^2] < \delta \text{ AND } U(0, 1) < \delta'\} \quad (8)$$

where $U(0,1)$ is the uniform distribution on $[0,1]$ and, for this experiment, $\delta' = 0.5$.

The increasing randomness of the quasi-periodic sequences is reflected by the increasing h and K values. The very high T complexities indicate the unpredictable nature of these sequences.

Finally, sequence W is designed to be particularly challenging for the LZ78 algorithm (Shor, 2005). It contains all possible codewords at each length. The sequence starts 0|1, and continues 0|1|00|01|10|11. Its empirical K -complexity as computed by LZ78 compressibility (Table A1) is even higher than the incompressible Bernoulli-generated sequences. This result is presumably due to the finite length of the sequences and the imperfections of the pseudorandom number generator.

Table A1
Complexity Values of Generative Algorithms used in the Trials

Group	Algorithm	H	h	E	T	K
1	P5c	.971	0 ^a	2.32 ^a	4.87 ^a	.00100
	P12.165	.980	0	3.59	14.8	.00154
	P12.88	1	0	3.59	8.42	.00155
	P12.116	.918	0	3.59	11.7	.00154
	P16	.896	0	4	16.6	.00178
2	B-0.5	1	1	0	0	.0574
	B-0.7	.881	.881	0	0	.0514
3	GM	.918	.667 ^a	0.252 ^a	0.252 ^a	.0415
	RRXOR	1	.667 ^a	2 ^a	9.43 ^a	.0469
	EVEN	.918	.667 ^a	0.902 ^a	3.03 ^a	.0432
4	TM	1	.083 ^a	4.168	16.4	.00674
	QP-linear	1	.019	5.14	33.2	.00488
	QP-quadratic	.970	.023	3.7	26.2	.0121
	QP-random	.877	.650	1.52	10.2	.0417
	W	1	.693	5.59	53.84	.0582
	FW	.960	.087	2.63	9.04	.00474

^athese values of h , E and T were taken from Crutchfield and Feldman (2003).

References

- Crutchfield, J. P., & Feldman, D. P. (2003). Regularities unseen, randomness observed: Levels of entropy convergence. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 13(1), 25-54.
- Li, M., & Sleep, R. (2004). Melody classification using a similarity metric based on Kolmogorov complexity. In *Proceedings of the Sound and Music Computing Conference*. Paris, France.
- Shor, P. (2005). *Lempel-Ziv* [PDF document]. Retrieved from lecture notes online website: http://www-math.mit.edu/~shor/PAM/lempel_ziv_notes.pdf
- Tao T. (2006). The dichotomy between structure and randomness, arithmetic progression, and the primes. In M. Sanz-Solé, J. Soria, J. L. Varona, & J. Verdera (Eds.), *Proceedings of the International Congress of Mathematicians* (pp. 581-608). Madrid, Spain: EMS Publishing House.

Supplementary material B

Sub-sequences and answer key

We extracted three sub-sequences from each of the generative algorithms we used. We then removed the last symbol from each of these sub-sequences, in order to create the answer key for the prediction task. The full set of rhythmic sequences and the answer key are given below for the deterministic-periodic algorithms (Table B1), the stochastic algorithms (Table B2), the deterministic-stochastic algorithms (Table B3), and the deterministic-aperiodic algorithms (Table B4).

Table B1

Set of the three Sub-Sequences extracted from each Deterministic-Periodic Algorithm for the Prediction Task, with the associated Answer Key. Each vertical line represents a drum hit and each dot represents a rest.

Table B2

Set of the three Sub-Sequences extracted from each Stochastic Algorithm for the Prediction Task, with the associated Answer Key. Each vertical line represents a drum hit and each dot represents a rest.

Algorithm	Sub-sequence	Correct Answer
B-0.5	. .	.

	
B-0.7	

Table B3

Set of the three Sub-Sequences extracted from each Deterministic-Stochastic Algorithm for the Prediction Task, with the associated Answer Key. Each vertical line represents a drum hit and each dot represents a rest.

Table B4

Set of the three Sub-Sequences extracted from each Deterministic-Aperiodic Algorithm for the Prediction Task, with the associated Answer Key. Each vertical line represents a drum hit and each dot represents a rest.

Supplementary material C

Example of calculations of complexity values

The following outlines complexity measure calculations for the period-5 sequence, P5c, a periodic repetition of the length 5 block 10101, i.e. ...101011010110101... and B-0.5, a sequence of independent 0 and 1s, each occurring with probability 1/2. The principal references for this supplementary material are Cover and Thomas (2006) and Crutchfield and Feldman (2003).

1. Shannon entropy, H

The Shannon entropy,

$$H \equiv - \sum_i p_i \log(p_i) \quad (9)$$

is easily calculated once p_i , the probability of the i th symbol, is known or has been estimated. The logarithm is usually taken to base 2. There are just two symbols, 0 and 1, in the binary sequences employed here.

The entropy of B-0.5 is therefore

$$-\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1. \quad (10)$$

This is the maximum possible entropy and expresses the complete unpredictability of each successive term on the sequence.

Since P5c is an infinite number of repetitions of the 10101, $p_0 = 2/5$ and $p_1 = 3/5$. The Shannon entropy,

$$-\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} \approx 0.971 \quad (11)$$

is slightly less than 1 due to the unequal weighting of the symbols. Although random, the sequence has some predictability in the sense that the best guess at the next unseen (or unheard) symbol is 1.

2. Entropy rate, h

The entropy rate,

$$h \equiv \lim_{L \rightarrow \infty} \left(\frac{H(L)}{L} \right) \quad (12)$$

expresses the uncertainty per symbol given complete knowledge of the underlying probability distribution. $H(L)$, the Shannon entropy of blocks (subsequences of consecutive symbols) of length L , is calculated from the Shannon entropy formula. The p_i 's, in this case, are the probabilities of each distinct block i of length L . If the block probabilities are not known, an empirical estimate can be made by determining the relative frequency of each distinct block for $L = 1, 2, \dots, L_{max}$ in a finite string of length N . $L_{max} < N$ is chosen to ensure a meaningful estimation (Lesne, Blanc, & Pezard, 2009; Speidel, Titchener, & Yang, 2006).

The entropy rate estimation can be made by plotting $H(L)$ against L . The plot is concave and non-decreasing. h is the gradient of the linear asymptote to $H(L)$ at large L .

The B-0.5 sequence entropy rate can be calculated theoretically. There are 2^L different blocks of length L , each with probability of 2^{-L} and therefore

$$H(L) = -2^L \times 2^{-L} \log_2(2^{-L}) = L. \quad (13)$$

Hence

$$h = \lim \left(\frac{L}{L} \right) = 1. \quad (14)$$

This is the largest attainable value for h and expresses the maximal randomness of fair coin tosses.

Even though h is known to be exactly 0 for all periodic sequences, a detailed calculation proceeds as follows. $H(1)$ has already been calculated: it is exactly the same as H . There are three different $L = 2$ blocks, 10, 01, and 11, and these occur with probabilities, $2/5$, $2/5$, and $1/5$ respectively. Hence

$$H(2) = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{2}{5} \log_2 \frac{2}{5} - \frac{1}{5} \log_2 \frac{1}{5} \approx 1.522. \quad (15)$$

The $L = 3$ blocks 101, 010, 011, 110 occur with probabilities $2/5$, $1/5$, $1/5$, $1/5$ and $H(3) \approx 1.922$. At $L = 4$, each possible block, 1010, 0101, 1011, 0110, 1101, is equally probable and therefore

$$H(4) = -5 \times \left(\frac{1}{5} \log_2 \frac{1}{5} \right) = \log_2 5. \quad (16)$$

Similarly, there are five equally probable blocks, 10101, 01011, 10110, 01101, 11010, of length 5 and indeed just five equally probable blocks at any bigger L ; hence

$$H(L \geq 4) = \log_2 5 \approx 2.322. \quad (17)$$

The sequence

$$\left(\frac{H(1)}{1}, \frac{H(2)}{2}, \frac{H(3)}{3}, \dots \right) \quad (18)$$

evidently converges to 0 and therefore $h = 0$. The sequence is predictable and lacks uncertainty.

3. Excess entropy, E

The excess entropy,

$$E \equiv \lim_{L \rightarrow \infty} (H(L) - hL) \quad (19)$$

can be expressed in terms of the finite- L entropy rate

$$\frac{dH(L)}{dL} = \frac{H(L) - H(L-1)}{L - (L-1)} = H(L) - H(L-1) \equiv \Delta H(L). \quad (20)$$

Therefore

$$\begin{aligned} E &\equiv \lim_{\substack{L \rightarrow \infty \\ \infty}} (H(L) - hL) \\ &= \sum_{\substack{L=1 \\ \infty}} H(L) - H(L-1) - (Lh - (L-1)h) \\ &= \sum_{L=1} (\Delta H(L) - h) \end{aligned} \quad (21)$$

where we have defined $H(0) = 0$. If there are no correlations between symbols, i.e. if each symbol is generated independently, the finite- L rate will equal the actual entropy rate and in consequence, $E = 0$. Otherwise, blocks of scale L and below will fail to capture structure on scales larger than L and will hence appear more random per symbol. $\Delta H(L)$ will therefore overestimate the actual entropy rate; the surplus is

$$\Delta H(L) - h > 0. \quad (22)$$

E is equal to the sum total of all these surpluses. E therefore quantifies the amount of structure in the sequence (and by implication, the complexity of the sequence). The sums and limits in the above formulae may even diverge for sequences with non-periodic structure on all block scales.

If E is finite then the asymptote to $H(L)$ at a large L has equation

$$H^\infty = E + hL. \quad (23)$$

The excess entropy can be estimated by finding the intercept of the asymptote with the line $L = 0$. Alternatively, E can be approximated by terminating the sums or limit at L_{max} . If h is known then

$$E \approx H(L_{max}) - L_{max}h. \quad (24)$$

For B-0.5, since $H(L) = L$ and $h = 1$, $E = 0$.

For P5c, the asymptote to $H(L)$ is a horizontal line ($h = 0$). Since $H(L)$ is constant along this line and has the value $\log_2 5$, $E = \log_2 5$.

4. Transient information, T

The complete list of quantities $H(L)$ for $L = 1, 2, \dots$ provides information on correlation and structure. Although the list may be infinite, in many cases $H(L)$ approaches the asymptote H^∞ , and we can characterize the full set $H(L)$ with a small number of parameters. One of these, the excess entropy, is the $L = 0$ intercept, i.e. $H^\infty(0)$. Another parameter is the gradient h of the asymptote. However two parameters are not always enough to capture all structural aspects of a sequence.

The gradient and intercept of the asymptote are two obvious aspects of the asymptote. A third is the area enclosed between the asymptote and $H(L)$; this is the transient information

$$T \equiv \sum_L (E + hL - H(L)). \quad (25)$$

Periodic sequences of the same period will share the same h and E but will in general differ in T . In fact T vanishes if E is zero; otherwise finite T may be able to distinguish sequences of identical E .

The transient information can also be written by converting the entropy rate surplus $\Delta H(L) - h$ at size L (the excess entropy rate is the sum of all surpluses) into an entropy by multiplying by L (Crutchfield & Feldman, 2003):

$$T = \sum_{L=1}^{\infty} L(\Delta H(L) - h). \quad (26)$$

The sum truncates at the smallest block length for which $\Delta H(L) - h = 0$. Denoting this length as L' , then

$$\begin{aligned} T &= \sum_{L=1}^{L'-1} L(\Delta H(L) - h) \\ &= (L' - 1)H(L' - 1) - \sum_{L=0}^{L'-2} H(L) - h \sum_{L=1}^{L'-1} L \\ &= (L' - 1)H(L' - 1) - \sum_{L=1}^{L'-2} H(L) - h \frac{(L' - 1)L'}{2} \end{aligned} \quad (27)$$

where the identity $H(0) = 0$ has been used in the last line.

Since $E = 0$ for the fair coin, $T = 0$.

For P5c, the sum in the defining expression for T terminates at $L' = 5$ since $H(4) = H(5) = H(6) \dots$ The previous calculations provide the entropies

$$[H(L)]_{L=1}^5 = [0.971, 1.522, 1.922, 2.322, 2.322] \quad (28)$$

and since, in addition, $h = 0$, T can be calculated:

$$T_{P5c} = -(0.971 + 1.522 + 1.922) + 4 \times 2.322 = 4.873. \quad (29)$$

5. Kolmogorov complexity, K

The Kolmogorov complexity, K , of a string s , is the length of the shortest program that outputs s . K is defined for finite strings, and makes no assumption of stationarity. It is non-computable, but can be upper-bounded by sequence compressibility with respect to a particular compressor. A popular choice is the Lempel-Ziv 1978 (LZ78) compressor.

LZ compression functions by the construction of a dictionary of different codewords (Li & Sleep, 2004; Shor, 2005). A repeated substring can be referred to by a compact reference to a dictionary codeword. The number of codewords in the dictionary, K_{LZ78} , provides an estimation of the Kolmogorov complexity.

K_{LZ78} is expected to be closer to K for very random, incompressible strings. It is impossible, in general, to find the maximal compression of a structured and significantly

compressible string. Since LZ78 is one of many possible compressors, it might overestimate K by some way.

The dictionary is built by scanning the sequence from left to right. Any unseen subsequences are added as codewords. For example, consider the compression of the first 45 digits (nine repetitions of 10101) of P5c:

The first unseen subsequence is the first symbol, 1:

The next codeword is 0:

The next symbol, 1, is already in the dictionary, so the compressor scans ahead. 10 has not been seen and is added to the dictionary:

The dictionary eventually contains the codewords:

$$1 \ 0 \ 10 \ 11 \ 01 \ 011 \ 010 \ 110 \ 101 \ 1010 \ 1101 \ 0110 \ 1011 \ 0101 \ 10101. \quad (34)$$

We notice that, even in this short sequence, four and five symbol codewords have been added. The appearance of longer codewords and the omission of shorter ones (for example, 111 is not a codeword of P5c) reflect the regularity of the sequence. Coding of B-0.5 will proceed in exactly the same way, but there will be many more short codewords due to the sequence's inherent randomness; it will therefore have a larger dictionary for a string of the same size. Since the K_{LZ78} complexity is the number of codewords in the dictionary, B-0.5, in comparison to P5c, will have a larger K_{LZ78} per-symbol complexity. Indeed, B-0.5 should have one of the highest K_{LZ78} complexities. Table A1 shows that only sequence W has a higher K_{LZ78} complexity. W was deliberately engineered to be the least LZ compressible (Shor, 2005).

References

- Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Crutchfield, J. P., & Feldman, D. P. (2003). Regularities unseen, randomness observed: Levels of entropy convergence. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 13(1), 25-54.

Lesne, A., Blanc, J. L., & Pezard, L. (2009). Entropy estimation of very short symbolic sequences. *Physical Review E*, 79, 046208.

Li, M., & Sleep, R. (2004). Melody classification using a similarity metric based on Kolmogorov complexity. In *Proceedings of the Sound and Music Computing Conference*. Paris, France.

Shor, P. (2005). *Lempel-Ziv* [PDF document]. Retrieved from lecture notes online website: http://www-math.mit.edu/~shor/PAM/lempel_ziv_notes.pdf

Speidel, U., Titchener, M., & Yang, J. (2006). How well do practical information measures estimate the Shannon entropy? In M. D. Logothetis, & Z. Ghassemlooy (Eds.), *Proceedings of the 5th International Symposium on Communication Systems, Networks, and Digital Signal Processing* (pp. 861-865). Patras, Greece: Northumbria University.