Kolmogorov complexity in quantum information

Jovan Odavić

June 7, 2023

1 Introduction

The value of literary art (and any art) is subjective and depends on the reader's individual experience and taste. One way I learned to appreciate novels is thought their informational content, which refers to the minimum amount of information needed to reproduce the entire novel. Interestingly, the repetition of words and word roots means that the novel is not a completely random sequence of words and can be reconstructed using an algorithm developed by Lempel and Ziv in 1976 (LEMPEL; ZIV, 1976). This algorithm allows for the creation of a minimal dictionary from which the entire novel can be reconstructed. Notably, this dictionary is smaller than the total language dictionary, as some words from the latter do not appear in the novel itself. To illustrate the algorithm, let's consider the following simple example.

2 Informational complexity using simple examples

Lets imagine an initially empty dictionary of words. Suppose now that a novel starts with the following word with 16 letters

A	A	A	B	B	A	B	A	A	B	A	A	A	B	A	B
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

This example has been taken from Ref. (LEMPEL; ZIV, 1976). First row is the example string. Second row represents integer indexing of each letter in the example. We shall use these indices in the following. The construction of the dictionary work in the following manner.

We begin the construction and filling of the dictionary as follows. We start with the shortest possible word, which currently does not exist in the dictionary and consists of only one letter, starting from the right. In this case, that letter is A. This allows us to write the previous word as

$$A \cdot AABBABAABAAABAB$$

and currently our dictionary consists of only one element, which we can write as $\{A\}$. Next, the second letter of the word is again A, but A already exists in the dictionary, therefore we move on to the third letter in combination with the second, and we have AA. Within AA, there is already a root word that exists in the dictionary, which is A. Therefore, AA does not represent a new word. Furthermore, by combining the second, third, and fourth letters, we obtain AAB. This represents a new word because we cannot form it exclusively from the words in the current dictionary. Thus, we have

$$A \cdot AAB \cdot BABAABAAABAB,$$

and the dictionary now consists of the set $\{A, AAB\}$. The fifth letter is the letter B, which already exists in the dictionary as an isolated pattern, so it does not enter the dictionary as such.

Therefore, we must consider the fifth and sixth letters together, i.e. BA. This combination does not exist as a pattern in the current dictionary, and we obtain

$$A \cdot AAB \cdot BA \cdot BAABAAABAB$$

Repeating this process, we arrive at the final separation of this word

$$A \cdot AAB \cdot BA \cdot BAA \cdot BAAA \cdot BAB$$

with the corresponding dictionary $\{A, AAB, BA, BAA, BAA, BAB\}$. We have managed to reduce the given word to 6 parts, and therefore we assign the following value as $c = 6$.

It is possible to work with zeros and ones instead of letter. Some examples are

$$1011010 \to 1 \cdot 0 \cdot 11 \cdot 010 \quad \text{where} \quad c = 4$$
$$0010 \to 0 \cdot 01 \cdot 0 \quad \text{where} \quad c = 3, \quad \text{Ref.} \quad [2]$$
$$000000... \to 0 \cdot 0000... \quad \text{where} \quad c = 2, \quad \text{Ref.} \quad [2]$$
$$01010101... \to 0 \cdot 1 \cdot 010101... \quad \text{where} \quad c = 3, \quad \text{Ref.} \quad [2]$$

Kolmogorov complexity (KC) can now be computed using c if we apply a proper normalization. Typically we normalize (divide) the obtained partition number with the maximally possible complex partition that a word can have. The best possible example of such complex pattern would be digits of an irrational number. If we map the digits into zeros and ones we can expect the dictionary to be infinitely long and that no repetitions of periodic patterns occur. In the original paper Ref. (LEMPEL; ZIV, 1976) it was shown

$$\text{norm} = \frac{N}{\log_2 N}, \tag{1}$$

where N is the length of the string. To practically compute the complexity we evaluate

$$\text{KC} = \frac{c}{\text{norm}} = \frac{c}{N} \frac{\log_{10} N}{\log_{10} 2}. \tag{2}$$

One can also consider the following examples of the computation of the KC

- AAABBABAABAAABAB

$$\text{KC} = \frac{6}{16} \frac{\log_{10} 16}{\log_{10} 2} = 1.5, \tag{3}$$

- 1011010

$$\text{KC} = \frac{4}{7} \frac{\log_{10} 7}{\log_{10} 2} = 1.6042, \tag{4}$$

- 0010

$$\text{KC} = 1.5, \tag{5}$$

- 000000000000000 (x15 zeros)

$$\text{KC} = 0.52091, \tag{6}$$

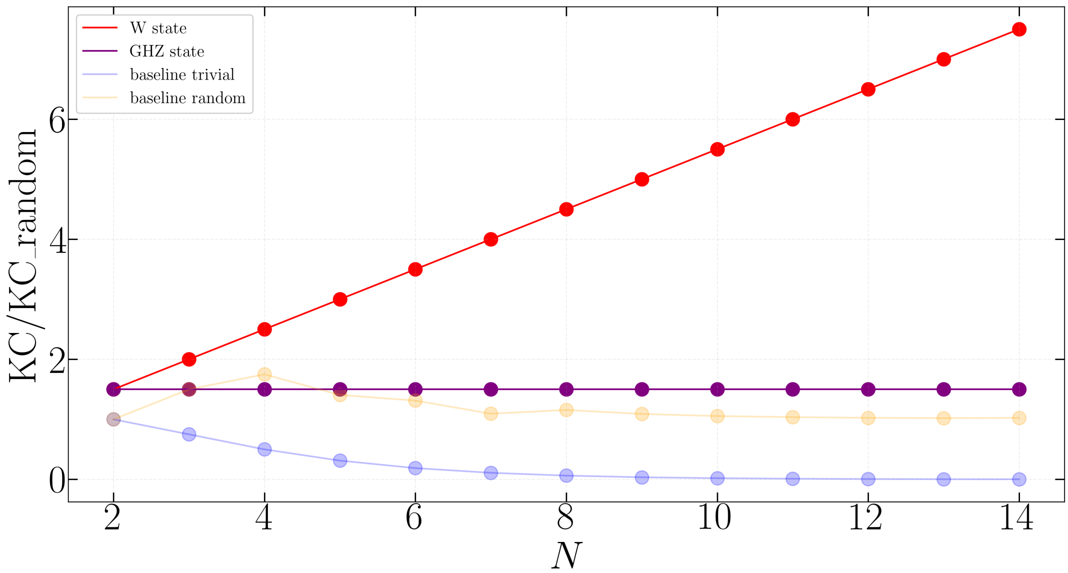

Figure 1: Comparison between KC of the two different multipartite entangled states of the GHZ (red) and W (purple). In yellow we display the KC value of a random string of zeros of one of size 2^N while in light blue if the string is all zeros of size 2^N. We have normalized the GHZ and W state values with the KC value of the random string. The label is wrong therefore for the two baseline examples.

- 000000000000000000000000000000 (x30 zeros)

$$\mathrm{KC} = 0.3271. \tag{7}$$

It should be emphasized that if a sequence is completely random (e.g. decimals of irrational numbers) then KC $\to 1$. If a series is completely trivial (for example, only a series of zeros and ones as in the given example), KC $\to 0$ tends to zero, i.e. in the case of infinite series, the complexity value ranges between 0 and 1. We see in fact, if the string is actually of finite length, then those values are not in between those values. Certainly this algorithm enables the comparison of strings of different lengths and even normalization to infinity as we have presented so far.

In practice, when we observe some experimental response or output from a simulation, it is necessary to convert the original sequence into zeros and ones (maybe also into A, B) by some rule. This rule can be, for example, to compare individual entries with the mean value of the total sequence. If the individual input is less than the mean value, then we assign zero, while if it is greater, then one, and in this way we can reduce each string and than evaluate its complexity.

3 Information complexity in quantum information

Lets consider two of the hallmark multipartite entangled states in quantum information

- the Greenberger-Horne-Zeilinger (GHZ) state (GREENBERGER; HORNE; ZEILINGER, 1989)

$$|\mathrm{GHZ}\rangle = |000...0\rangle + |111...1\rangle \tag{8}$$

- the W state (DÜR; VIDAL; CIRAC, 2000)

$$|\mathrm{W}\rangle = |100..0\rangle + |010..0\rangle + ... + |000...1\rangle \tag{9}$$

In writing the states we have ignored the normalization constants. In this way the state vector is represented with zeros and ones. The number of qubits is N while the state vector is of size 2^N.

Using Kolmogorov Complexity we analyse the two different class of states and show the results in Fig. 1. This analysis does not say anything about quantum complexity of these states but just reflects on the traditional informational complexity of the states. Obviously, the extensive number of states of the generalized W-state yields a linear increase in Kolmogorov complexity as compared to the globally entangle GHZ state with a fixed complexity and irrespective of the number of qubits considered.

References

DÜR, W.; VIDAL, G.; CIRAC, J. I. Three qubits can be entangled in two inequivalent ways. **Physical Review A**, American Physical Society, v. 62, n. 6, p. 062314, Nov. 2000.

GREENBERGER, Daniel M.; HORNE, Michael A.; ZEILINGER, Anton. Going Beyond Bell's Theorem. In: **Bell's Theorem, Quantum Theory and Conceptions of the Universe**. Ed. by Menas Kafatos. Dordrecht: Springer Netherlands, 1989. P. 69–72. (Fundamental Theories of Physics).

LEMPEL, A.; ZIV, J. On the Complexity of Finite Sequences. **IEEE Transactions on Information Theory**, v. 22, n. 1, p. 75–81, Jan. 1976.