5AI25 – AIAI
(Algorithmic information and Artificial Intelligence)

Jean-Louis Dessalles

Evaluation – November 2025

Examples of answers

Q1. The objective of the Sudoku game is to fill a 9 × 9 grid with digits so that each column, each row, and each of the nine 3 × 3 subgrids that compose the grid contains all of the digits from 1 to 9. The puzzle setter provides a partially completed grid which has a single solution. Can you estimate the Kolmogorov complexity of the solution knowing the grid, if 36 out of the 81 cells are already filled.

It is equal to zero if the solution is unique, as the time required to find the solution is irrelevant.

Q2. The decimal expansion of a rational number is necessarily periodic. What can we conclude concerning the randomness of rational numbers?

Considered as an infinite sequence, the decimal expansion of a rational number, though infinite, can be compressed down to a finite amount. But rational numbers are integer fractions or, in other words, nothing more than a couple of integers. They correspond to a finite sequence and, as such, can be incompressible and therefore random.

Q3. A 20-bit long pattern is found to occur at locations 1670 and 2100 of a binary string. Do we spare information by encoding the repeat instead of merely leaving the bits intact?

Encoding the repeat requires coding for the pattern length and for the distance between its occurrences. This requires about $\lceil \log_2(1+20) \rceil + \lceil \log_2(1+2100 - 1679) \rceil$ = 5 bits + 9 bits. This is less than 20 (and there is still room for a few signal bits).

Q4. We want to describe the following ordered pairs of integers (x,y): ((1, 2), (2, 5), (3, 9) (4, 10), (5, 11), (6, 17)).

To do so, we hesitate between two regression functions: $y(x) = 2x$ or $y(x) = 3x$. The complexity of these functions is given by the complexity of their coefficient (so $C(2)$ or $C(3)$). The complexity of a (signed) integer n is given by $C(n) = 1 + \lceil \log_2(1+|n|) \rceil$ (we add one bit for the sign - codes are supposed to be delimited by spaces). Based on the Minimum Description Length principle, which of the two functions should we use to describe the data?

We evaluate the quantity $C(M) + \sum C(d_i \mid M)$, where M is the regression function. The term $\sum C(d_i \mid M)$ measures the corrections to perform.

For $y(x)=2x$:

$$C(D|M) = C(2\times 1 - 2) + C(2\times 2 - 5) + C(2\times 3 - 9) + C(2\times 4 - 10) + C(2\times 5 - 11) + C(2\times 6 - 17)$$

$$= 1 + 2 + 3 + 2 + 2 + 3 = 14 \text{ bits}$$

$$C(M) + C(D|M) = 3 + 14 = 17 \text{ bits}$$

For $y(x)$=3x:

$C(D|M)$ **= C(3×1 - 2) + C(3×2 - 5) + C(3×3 - 9) + C(3×4 - 10) + C(3×5 - 11) + C(3×6 - 17)**

= 2 + 2 + 0 + 2 + 3 + 2 = 11 bits

C(M) + C(D|M) = 3 + 11 = 14 bits

The model $y(x)$ = 3x is preferable.

Q5. You are designing a program that retrieves old events worth to tell. Event $E(t, a)$ occurred at time t in the past and involved an acquaintance a. Suppose $E(1, a)$ is considered worth telling by the program (t = 1 day), but that $E(60, a)$ is not. Now suppose that a has just been mentioned, independently from E. Being part of the context, a is therefore available for free when telling E. What is the condition on the description complexity $C(a)$ of a for $E(60, a)$ to become worth telling ("*Speaking of a, do you know what happened to her two months ago?…*")?

The complexity of events grows as $\log_2(t)$. So what we get from the availability of *a* should at least compensate for the older date: $C(a) > \log_2(60) = 6$ bits.