UM. Autumn 2020.
Supplementary exercise to the course «Information theory».

Problem 1. We say that a binary string x *contains only isolated zeros* if x contains no factors 00 (two consecutive zeros). For example, the string 111010101 contains only isolated zeros, while 0010001 contains factors 00. Prove that there exist real numbers $\lambda_1 < 1$ and $\lambda_2 > 0$ such that for all binary strings x with only isolated zeros

$$C(x) \le \lambda_1 |x| + \lambda_2$$

(since $\lambda_1 < 1$, this mean that all long enough x with only isolated zeros are compressible, i.e., $C(x) < |x|$).

Problem 2. We say that a binary string x is incompressible if $C(x) \ge |x|$.
(i) Prove that there exist infinitely many incompressible strings x.
(ii) Prove that all long enough incompressible strings x contain the factors "000", "0101", and "01011001."

Problem 3. Prove that there exist infinitely many binary strings x such that

$$0.49|x| < C(x) < 0.51|x|.$$

Problem 4. Let $\bar{x}$ be a string obtained from x by inverting all bits (e.g., $\overline{00101} = 11010$).
(i) Prove that there is a constant d_1 such that for all strings x

$$|C(x) - C(\bar{x})| \le d_1.$$

(ii) Prove that there is a constant d_2 such that for all strings x

$$|C(x\bar{x}) - C(x)| \le d_2$$

($x\bar{x}$ is a concatenation of x and $\bar{x}$).

Problem 5. Prove that there exist infinitely many binary strings x with extremely small Kolmogorov complexity :

$$C(x) \le \underbrace{\log\log\log\ldots\log}_{100} |x|.$$

Problem 6. Prove that there exist constants d_1, d_2 such that for all binary strings x

$$C(x) - d_1 \le C(xx) \le C(x) + d_2$$

(here xx is the string x repeated twice).

Problem 7. Prove that there exist constants d_1, d_2 such that for every binary strings x with *exactly* $n/2$ zeros and $n/2$ ones we have

$$C(x) \le n - d_1 \log n + d_2.$$

(This means that the strings x with exactly balanced number of zeros and ones are compressible. So in a "truly random" strings the fractions of ones and zeros should slightly deviate from 50%.)
Hint : count the number of "balanced" strings of length n and use Stirling's approximation for the factorial.

Problem 8. Prove that there exist constants d_1, d_2, d_3 such that for all binary strings x, y

(i) $I(x : y) \leq C(x) + d_1 \log C(x) + d_2$,
(ii) $I(x : x) \geq C(x) - d_3$.

Problem 9. Prove that there exist real numbers d_1, d_2 such that all binary strings x, y

$$I(xxx : yyy) \leq I(x : y) + d_1 \log(C(x) + C(y)) + d_2$$

(here xxx is the string x repeated three times).

Problem 10. Prove that the sum of $2^{-C(x)}$ over all binary strings x diverges, i.e.,

$$\sum_{x \in \{0,1\}^*} 2^{-C(x)} = \infty.$$