Octobre 17th.

1 Kolmogorov complexity: the definition and basic properties

Definition 1.1. *Let $U : \{0,1\}^* \to \{0,1\}^*$ be a (partial) computable function. Then*

$$\mathrm{K}_U(x) := \begin{cases} \min\{|p| \ : \ U(p) = x\}, \\ \infty, \quad \textit{if there is no such } p. \end{cases}$$

Remark: In this context the function U is called a *decompressor*.

Definition 1.2. *A computable function U is a **better decompressor** than V if there exists a constant c such that for all x*

$$\mathrm{K}_U(x) \leq \mathrm{K}_V(x) + c.$$

Theorem 1.1. *There exists a partial computable function U that is a better decompressor than any other V. (Such a function U is called an **optimal decompressor**.)*

Remark: There exist infinitely many optimal decompressors. If U_1 and U_2 are both optimal, then there exists a constant c such that for all binary strings x

$$|\mathrm{K}_{U_1}(x) - \mathrm{K}_{U_2}(x)| \leq c.$$

We fix once and for all some optimal decompressor U_0. In what follows we omit the subscript and denote $\mathrm{K}(x) := \mathrm{K}_{U_0}(x)$. The value of $\mathrm{K}(x)$ is called *Kolmogorov complexity* of x.

Some simple properties of Kolmogorov complexity:

- $\mathrm{K}(x) < \infty$ for all x;
- $\mathrm{K}(x) \leq |x| + O(1)$;
- $\mathrm{K}(xx) \leq \mathrm{K}(x) + O(1)$;
- $\mathrm{K}(f(x)) \leq \mathrm{K}(x) + O(1)$ for every computable f;
- for every integer $n > 0$ there exists a binary string x of length n such that $\mathrm{K}(x) \geq n$ (an *incompressible* string);
- there exists a constant c such that for every integer $n > 0$ at least 99% of strings x of length n

$$n - c \leq \mathrm{K}(x) \leq n + c.$$

Proposition 1.1. *There exists no algorithm $\mathcal{A} : n \mapsto x_n$ that returns for every input n a string x_n such that $\mathrm{K}(x_n) > n$.*

Corollary 1. The mapping $x \mapsto \mathrm{K}(x)$ is not computable.

Corollary 2. If U is an optimal decompressor, then U is not a total function.

Theorem 1.2 (a version of Gödel's incompleteness theorem)**.** [1] *There exists a number N such that* ZFC *cannot prove any statement of the form* "$\mathrm{K}(x) > N$."

Proposition 1.2. *If a set $A \subset \{0,1\}^*$ is* decidable *(or at least* enumerable*) and $x \in A \cap \{0,1\}^n$, then $\mathrm{K}(x) \leq \log |A \cap \{0,1\}^n| + O(\log n)$.*

[1] Optional, not necessary for the final exam.

Theorem 1.3. *(a) Let x be a string of length n, with pn ones and $(1-p)n$ zeros. Then*

$$\mathrm{K}(x) \le \left(p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \right) n + O(\log n).$$

(b) Let p be a real number such that $0 < p < 1$. Then for every n there exists a binary string x of length n with $\lceil pn \rceil$ zeros and $\lfloor (1-p)n \rfloor$ ones such that

$$\mathrm{K}(x) = \left(p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \right) n + O(\log n).$$

[The proof of Theorem 1.3 is postponed until next lecture.]

Exercises 1.1 (a simplified version of Stirling's formula)**.** *(a) Prove that $N! = \tau(N) \cdot \left(\frac{N}{e}\right)^N$, where $\tau(N)$ is a function such that*

$$1 \le N! \le \mathrm{poly}(N)$$

for some polynomial $\mathrm{poly}(N)$.

(b) Prove that $N! = \lambda(N) \cdot \sqrt{N} \left(\frac{N}{e}\right)^N$, where $\lambda(N)$ is a function such that

$$c_1 \le \lambda(n) \le c_2$$

for some reals c_1, c_2.

Hint: *Estimate the difference between* $\ln(N!) = \sum_{k=1}^{N} \ln k$ *and* $\int_1^N \ln x \, dx = (x \ln x - x)\Big|_1^N$.

Exercises 1.2. *Does there exist an optimal decompressor U such that for every x the value of $\mathrm{K}_U(x)$ is*
(a) an even number*?*
(b) a power of 2*?*
(c) a prime number?

Exercises 1.3. *Prove that the relation « U* is a better decompressor than *V » is a partial order (there exist computable functions U and V that are incomparable as decompressors).*

Exercises 1.4. *Prove that the set of pairs $\{(x, n) \mid \mathrm{K}(x) < n\}$ is recursively enumerable but not decidable.*

Exercises 1.5. *If x is a palindrome, then $\mathrm{K}(x) \le |x|/2 + O(1)$.*

Exercises 1.6. *Denote by xy the concatenation of words x and y. Prove that $\mathrm{K}(x) \le \mathrm{K}(xy) + O(\log |x|)$ and $\mathrm{K}(xy) \le \mathrm{K}(x) + \mathrm{K}(y) + O(\log |x|)$.*

References

[1] Ming Li and Paul Vitanyi, *An Introduction to Kolmogorov Complexity and Its Applications*. Springer Verlag, 3rd Edition, 2008.

[2] Bruno Durand, Alexandr Zvonkin, *Complexité de Kolmogorov*.
https://www.labri.fr/perso/zvonkin/Research/kolmogorov.pdf

[3] Laurent Bienvenu, Mathieu Hoyrup, *Une brève introduction á la théorie effective de l'aléatoire*.
http://www.liafa.jussieu.fr/~lbienven/docs/publications/Une_breve_introduction_a_la_theorie_effective_de_l_aleatoire-SMF.pdf

[4] Alexander Shen, *Algorithmic Information Theory and Kolmogorov Complexity*.
http://www.lirmm.fr/~ashen/uppsala-notes.pdf

[5] Lance Fortnow, *Kolmogorov complexity*.
http://people.cs.uchicago.edu/~fortnow/papers/kaikoura.pdf

[6] A. Shen, V.A. Uspensky, N.K. Vereshchagin, *Kolmogorov complexity and algorithmic randomness*, 2014.
http://www.lirmm.fr/~ashen/kolmbook-eng.pdf