Колмогоровская сложность

Будем рассматривать функции над множеством *двоичных слов* (слов из нулей и единиц) $\{0,1\}^*$. Длину слова $x \in \{0,1\}^*$ обозначим $|x|$.

Определение. Пусть фиксирована (частичная) вычислимая функция $D\colon \{0,1\}^* \to \{0,1\}^*$, называемая *способом описания. Колмогоровской сложностью* слова $x \in \{0,1\}^*$ относительно способа описания D называется число

$$KS_D(x) = \min\{|y| \mid D(y) = x\}.$$

Если же такого y, что $D(y) = x$, не существует, то $KS_D(x) = +\infty$.

Задача 1. Докажите, что для любых двух способов описания D_1 и D_2 существует такой способ описания D_0, что $KS_{D_0}(x) \leqslant KS_{D_1}(x) + c$ и $KS_{D_0}(x) \leqslant KS_{D_2}(x) + c$, где c — константа, не зависящая от x.

Теорема 1 (Теорема Соломонова–Колмогорова (об оптимальном способе описания))**.** *Существует такой способ описания D, что для любого другого способа описания D' существует константа c, такая что для всех $x \in \{0,1\}^*$*

$$KS_D(x) \leqslant KS_{D'}(x) + c.$$

В дальнейшем будем считать, что фиксирован некоторый оптимальный способ описания D, и колмогоровскую сложность будем обозначать просто $KS(x)$ (эта величина определена с точностью до аддитивной константы, зависящей от выбора оптимального способа описания).

Задача 2. Докажите, что существует константа c, такая что для любого слова x верно $KS(x) \leqslant |x| + c$.

Задача 3. Докажите, что если f вычислима, то для некоторой константы c неравенство $KS(f(x)) \leqslant KS(x) + c$ верно для всех x.

Задача 4. **(а)** Докажите, что существует константа c, такая что для любых двух слов x и y верно

$$KS(xy) \leqslant KS(x) + KS(y) + 2\log_2 KS(x) + c,$$

где xy — слово, получаемое приписыванием слова y к слову x.

(б) Докажите более сильное неравенство

$$KS(xy) \leqslant KS(x) + KS(y) + \log_2 KS(x) + 2\log_2\log_2 KS(x) + c.$$

Задача 5. Покажите, что если в качестве описаний использовать слова в четырёхбуквенном алфавите (скажем, конечные последовательности цифр 0, 1, 2, 3), то сложность (измеряемая как длина кратчайшего описания) будет равна половине обычной (с точностью до аддитивной константы).

Задача 6. Докажите, что для любого n существует «несжимаемое» слово x длины n, такое что $KS(x) \geqslant n$.

Задача 7. Покажите, что число слов сложности меньше n заключено между 2^{n-c} и 2^n (при некотором c и всех n).

Задача 8. Чему асимптотически равно среднее арифметическое сложностей всех слов длины n?

Задача 9. **(а)** Докажите, что множество $S = \{\langle n, x\rangle \mid KS(x) < n\}$ является перечислимым, причём $|S_n| = |\{x \mid \langle n, x\rangle \in S\}| < 2^n$ при всех n. **(б)** Докажите, что если $V = \{\langle n, x\rangle\}$ является перечислимым, причём $|V_n| = |\{x \mid \langle n, x\rangle \in V\}| < 2^n$ при всех n, то найдётся такое c, что $KS(x) < n + c$ для любого n и для любого $x \in V_n$.

Интуитивный смысл только что доказанного можно объяснить так: утверждения «объектов определённого вида мало» (меньше 2^i) и «объекты этого вида просты» (имеют сложность меньше i) равносильны, если мы рассматриваем перечислимые семейства и измеряем сложность с точностью до аддитивной константы (а число элементов — с точностью до мультипликативной).

Задача 10. **(а)** Докажите, что если у слова изменить последний бит, то его сложность изменится не более чем на константу. **(б)** Докажите, что при изменении одного бита слова его сложность может увеличиться на величину не более чем порядка логарифма его длины. **(в)** Докажите, что оценка в предыдущем пункте точная: бывает, что при изменении одного бита сложность слова увеличивается на величину порядка логарифма его длины.

Задача 11. Пусть D — оптимальный способ описания. Докажите, что D не является всюду определённой функцией.

Задача 12. Существует ли вычислимая функция $f\colon \mathbb{N} \to \{0,1\}^*$, такая что $KS(f(n)) \geqslant n$ для всех n?

Задача 13. **(а)** Докажите, что функция $KS : x \mapsto KS(x)$ невычислима. **(б)** Докажите, что у функции KS не существует нетривиальных вычислимых нижних оценок: если вычислимая функция $k\colon \{0,1\}^* \to \mathbb{N}$ такова, что для всех x из области определения k верно $k(x) \leqslant KS(x)$, то k ограничена сверху константой.