Устойчивость колмогоровских свойств при релятивизации

Андрей Мучник и Андрей Ромащенко

Аннотация

Предположим, что кортеж слов $\bar{x} = \langle x_1, \ldots, x_n \rangle$ имеет пренебрежимо малую взаимную информацию с некоторым словом y. Значит ли это, что свойства колмогоровской сложности набора слов $\bar{x}$ мало меняются при релятивизации относительно y ? Если аккуратно формализовать этот вопрос, то окажется, что получить на него полный ответ очень непросто. В данной работе мы изучим эту задачу для ограниченного класса свойств (для свойств, выразимых на языке $\exists$-формул). В частности, мы доказываем, что случайный относительно $\bar{x}$ оракул y не помогает выделять общую информацию из слов x_i.

1 Введение

Колмогоровской сложностью $K(x)$ двоичного слова x называется минимальная длина программы, порождающей x. Аналогично, относительной сложностью $K(x|y)$ (сложностью x при известном y) называется минимальная длина такой программы, которая получает на вход y и порождает x. При этом рассматриваются программы на некотором *оптимальном* языке программирования (см. детали определения и основные свойства колмогоровской сложности в [1, 4]).

Отметим, что можно также говорить о колмогоровской сложности (и относительной колмогоровской сложности) пар, троек, и вообще любых кортежей слов. Для этого мы фиксируем некоторую вычислимую нумерацию всех конечных кортежей двоичных слов. Сложностью кортежа называется колмогоровская сложность его *номера* в данной нумерации. Выбор конкретной нумерации не приницпиален: переход к другой вычислимой нумерации изменит колмогровскую сложность каждого кортежа лишь на ограниченную величину $\mathcal{O}(1)$. Изменение сложности на аддитивную константу не существенно, поскольку и колмогоровская сложность отдельных слов определена с точностью до аддитивного члена $\mathcal{O}(1)$ (зависящего от выбора способа программирования).

Основные определения и большинство результатов теории колмогоровской сложности легко релятивизуются — вместо обычных алгоритмов можно рассматривать алгоритмы с оракулом. Отметим, что если взять в качестве оракула конечный объект (слово) z, то нам не даже потребуется вводить новые обозначения для колмогоровской сложности с оракулом z. Релятивизация относительно z заключается в том, что мы добавляем z в условие всех сложностей; например, $K(x)$ заменяется на $K(x|z)$, а $K(x|y)$ на $K(x|y,z)$.

Информацией *о слове x, заключенной в слове y*, называется разница между простой колмогоровской сложностью x и относительной сложностью x при

известном y:

$$I(x:y) = K(x) - K(x|y)$$

Одним из фундаментальных свойств алгоритмической теории информации является теорема о симметрии взаимной информации:

Теорема 1 (Колмогоров–Левин, [1])

$$I(x:y) = I(y:x) + \mathcal{O}(\log N) = K(x) + K(y) - K(x,y) + \mathcal{O}(\log N),$$

где $N = K(x,y)$.

Таким образом, пренебрегая логарифмической поправкой, можно говорить о взаимной информации между x и y, не различая $I(x:y)$ и $I(y:x)$.

Если величина взаимной информации $I(x:y)$ пренебрежимо мала по сравнению с величинами $K(x)$, $K(y)$, $K(x,y)$ (скажем, если $I(x:y)$ имеет порядок логарифма от $N = K(x,y)$), то слова x и y часто называют *независимыми*. Этот жаргон весьма распространён в среди специалистов по алгоритмической теории информации. Интуитивно кажется, что если x и y *независимы*, то *разумные* алгоритмические свойства x (выразимые на языке колмогоровской сложности) не должны зависеть от релятивизации относительно y.

Гипотеза 1 (Основная) *Если взаимная информация между* $\langle x_1, \ldots, x_n \rangle$ *и* y *пренебрежимо мала, то релятивизация относительно* y *лишь незначительно меняет свойства* $x_1, \ldots, x_n$.

Приведём простейший пример, иллюстрирующий данное неформальное предположение. Возьмём в качестве $\bar{x}$ кортеж, состоящий из n слов: $\bar{x} = \langle x_1, x_2, \ldots, x_n \rangle$. Предположим, что взаимная информация между $\bar{x}$ и некоторым словом y пренебрежимо мала. Тогда нетрудно показать, что самые простые свойства колмогоровской сложности для $\bar{x}$ мало меняются при релятивизации относительно y:

$$K(x_i) \approx K(x_i|y),\ K(x_i,x_j) \approx K(x_i,x_j|y), \ldots,$$

для всех i, j, и т.д. Более, точно, имеет место следующее утверждение:

Утверждение 1 *Если* $\bar{x} = \langle x_1, x_2, \ldots, x_n \rangle$ *и* $\delta = K(\bar{x}) - K(\bar{x}|y)$, *то для всех наборов индексов* $i_1, \ldots, i_s \in \{1, \ldots, n\}$

$$|K(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) - K(x_{i_1}, x_{i_2}, \ldots, x_{i_s}|y)| \leq \delta + \mathcal{O}(\log N),$$

где $N = K(x_1, \ldots, x_n)$. *(Константа в* $\mathcal{O}$*-слагаемом зависит от* n, *но не от* N*).*

Доказательство этого утверждения тривиально; для полноты изложения мы приведём его в разделе 4.

Теперь рассмотрим более сложный пример, в котором свойства $\bar{x}$ уже не столь тривиальны. Обратимся к свойству *выделяемости общей информации*. Рассмотрим пару слов $\bar{x} = \langle x_1, x_2 \rangle$. Будем говорит, что у данной пары можно выделить α битов общей информации для уровня точности k, если

$$\exists z \text{ такое, что для } i = 1,2\ K(z|x_i) < k \text{ и } K(z) \geq \alpha$$

Легко показать, что для любого такого слова z выполнено неравенство

$$K(z) \le I(x_1 : x_2) + \mathcal{O}(k + \log K(x_1, x_2))$$

Это означает, что для достаточно малого уровня точности k невозможно извлечь из x_1, x_2 существенно больше, чем $I(x_1 : x_2)$ битов общей информации. (В огрублённом виде это наблюдение иногда формулируют так: общая информация не может быть больше взаимной информации).

Известно, что вопрос о выделяемости определённого числа битов общей информации у пары x_1, x_2 не сводится к вопросу о величинах $K(x_1)$, $K(x_2)$, $K(x_1, x_2)$. Например, зная, что $K(x_1) = K(x_2) = 2N$ и $K(x_1, x_2) = 3N$, мы не можем сказать ничего определённого о величине выделяемой общей информации пары. С одной стороны, существуют пары $\langle x_1, x_2\rangle$ с указанными сложностями, у которых можно выделить N битов общей информации для очень малого уровня точности $k = \mathcal{O}(1)$. С другой стороны, существуют пары слов с такими же сложностями, для которых даже для довольно больших значений уровня точности можно выделить лишь пренебрежимо малую общую информацию (порядка $\mathcal{O}(k+\log N)$). Подробное обсуждение вопроса о выделении общей информации можно найти в [2, 3, 6, 11].

Отметим, что вопрос о выделении общей информации можно ставить не только для пар слов, но и для кортежей произвольной длины $s \ge 2$. Но все интересные для нас нетривиальные свойства проявляются уже для $s = 2$. Чтобы не усложнять обозначения, мы в дальнейшем ограничимся свойством выделяемости общей информации только для пар.

Конкретизируем гипотезу 1 для свойства выделяемости общей информации:

Предположим, что взаимная информация между $\bar{x} = \langle x_1, x_2\rangle$ и y достаточно мала. Тогда α битов общей информации выделяются у слов x_1 и x_2 для уровня точности k, если и только если те же α битов общей информации можно выделить у этих слов, имея y в качестве оракула (возможно, для немного другого уровня точности).

Данное утверждение состоит из частей *если* и *только если*. Вторая часть (*только если*) тривиальна: если некоторая общая информация выделяется без оракула, то ту же самую информацию можно выделить и с оракулом. Интересна другая часть эквивалентности (*если*). Сформулируем её более строго. Нам представляется наиболее естественной формулировка с логарифмическими погрешностями:

Гипотеза 2 *Для любого $C_1 > 0$ существует число $C_2 > 0$ такое, что для всех $\bar{x} = \langle x_1, x_2\rangle$ и y, если $I(y : \bar{x}) \le C_1 \log N$ и*

$$\exists w : \ K(w|y) \ge \alpha, \ K(w|x_i, y) \le C_1 \log N \ (i = 1, 2),$$

где $N = K(\bar{x}, y)$, (т.е., α битов общей информации можно выделить у слов x_1, x_2 для уровня точности $C_1 \log N$, имея оракул y), то

$$\exists z : \ K(z) \ge \alpha, \ K(z|x_i) \le C_2 \log N \ (i = 1, 2),$$

т.е. те же α битов общей информации можно выделить у этих слов и без оракула (для уровня точности $C_2 \log N$).

Удивительным образом, это естественное утверждение оказывается очень нетривиальным. В [7] данная гипотеза была доказана для $\alpha = I(x_1 : x_2)$. В данной статье мы докажем ослабленный вариант гипотезы, в которой логарифмические слагаемые заменены на o-малое:

Теорема 2 *Для любой функции $f(N)$, $f(N) = o(N)$ найдётся функция $g(N)$ (также $g(N) = o(N)$), что для всяких $\bar{x} = \langle x_1, x_2\rangle$ и y, если $I(y : \bar{x}) \le f(N)$ и*

$$\exists w : \ K(w|y) \ge \alpha, \ K(w|x_i, y) \le f(N) \ (i = 1, 2),$$

где $N = K(\bar{x}, y)$, то

$$\exists z : \ K(z) \ge \alpha, \ K(z|x_i) \le g(N) \ (i = 1, 2).$$

Мы докажем эту теорему в разделе 6

Отметим, что в алгоритмической теории информации редко встречаются естественные утверждения, которые выполнены с точностью $o(\cdot)$, но не с точностью $\mathcal{O}(\log N)$. Мы предполагаем, что гипотеза 2 верна и в *логарифмическом* варианте, однако известная нам техника не позволяет это доказать.

Мы рассмотрели наиболее естественные примеры утверждений, формализующих интуитивную гипотезу 1. В следующем разделе мы обсудим язык, на котором можно формализовать и доказать эту гипотезу в более общем виде.

2 Общая формулировка основной гипотезы

Мы интересуемся устойчивостью различных колмогоровских свойств кортежей слов при релятивизации. Прежде всего следует уточнить, какие свойства мы называем колмогоровскими. Говоря кратко и неформально, мы интересуемся свойствами, выразимыми в терминах колмогоровских сложностей заданных слов. При этом мы как правило будем интересоваться свойствами, выполняющимися с 'логарифмической точностью'. Прежде чем давать точное определение, приведём несколько примеров.

Пример: Для любых слов x_1, x_2 выполнены неравенства

$$K(x_1, x_2) \le K(x_1) + K(x_2) + \mathcal{O}(\log K(x_1, x_2)), \tag{1}$$

и

$$K(x_1, x_2) \ge K(x_1) - \mathcal{O}(1).$$

Кроме того, выполнено равенство

$$K(x_1, x_2) = K(x_1) + K(x_2|x_1) + \mathcal{O}(\log K(x_1, x_2)).$$

Это наиболее простые свойства колмогоровских сложностей пары слов, выразимые с помощью *линейных равенств и неравенств*. Поскольку нетривиальные соотношения для простой колмогоровской сложности выполнены лишь с точностью до аддитивного логарифмического члена, как правило ограничиваются рассмотрением подобных свойств именно с логарифмической точностью.

Отметим, что различные виды колмогоровской сложности (префиксная, монотонная сложности, сложность разрешения, априорная сложность, [14, 4]) отличаются от простой сложности на величины не более чем логарифмического

порядка. Поэтому свойства, выполненные с 'логарифмической точностью', одинаковы для всех разновидностей колмогоровской сложности. Это даёт дополнительные основания интересоваться свойствами сложности именно с логарифмической погрешностью.

Как можно описать класс такого рода свойств? Для набора слов $x_1, \ldots, x_n$ мы рассмотрим колмогоровские сложности всех кортежей $x_{i_1}, \ldots, x_{i_k}$, где $1 \leq i_1 < \ldots < i_k \leq n$. Таким образом, набору из n слов соответствует $(2^n - 1)$ сложностей. Упорядочив все поднаборы из данного набора слов некоторым каноническим образом (скажем, лексикографически), мы получаем *сложностной профиль* – вектор в $\mathbb{Z}_+^{2^n-1}$

$$\vec{K}(x_1, \ldots, x_n) = (K(x_1), K(x_1, x_2), \ldots, K(x_2), K(x_2, x_3), \ldots).$$

Аналогично определяется *относительны* сложностной профиль $\vec{K}(x_1, \ldots, x_n|y)$, состоящий из сложностей всевозможных наборов x_i относительно y (точные определения мы приведём в разделе 3).

Замечание. Не имеет смысла рассматривать кортежи, отличающиеся только порядком слов, а также кортежи, в которых некоторое слова встречается много раз. При перестановке или дублировании членов кортежа, его колмогоровская сложность меняется на величину $\mathcal{O}(1)$, зависящую только от числа слов в кортеже, но не от их сложностей. Поскольку мы собираемся пренебрегать логарифмической погрешностью, нет нужды различать между собой величины, отличающиеся друг от друга на $\mathcal{O}(1)$.

По тем же причинам нет необходимости принимать во внимание *относительные* колмогоровские сложности, поскольку с помощью теоремы Колмогорова – Левина их можно выразить как комбинации безусловных сложностей:

$$K(x_1, \ldots, x_n|y_1, \ldots, y_m) = K(x_1, \ldots, x_n, y_1, \ldots, y_m) - K(y_1, \ldots, y_m) + \mathcal{O}(\log N),$$

где $N = K(x_1, \ldots, x_n, y_1, \ldots, y_m)$.

Таким образом, мы можем говорить о свойствах сложностного профиля $\vec{K}(x_1, \ldots, x_n)$ как о простейших колмогоровских свойствах набора слов $x_1, \ldots, x_n$. Например, неравенство (1) соответствует утверждению о профиле пары x_1, x_2:

$$\vec{K}(x_1, x_2) = (K(x_1), K(x_1, x_2), K(x_2)) \in A = \{(u, v, w) \ : \ v \leq u + w + \mathcal{O}(\log w)\}.$$

Таким образом, простые утверждения о колмогоровских свойствах формулируются в виде

$$\vec{K}(x_1, \ldots, x_n) \in A$$

для всевозможных множеств A. В частности, в таком виде можно выразить утверждения об истинности линейных информационных неравенств (см. [8, 7, 9]).

Описанные выше простые утверждения, выражающие колмогоровские свойства, естественно рассматривать как бескванторные атомарные формулы. Разумеется, к такого вида формуле можно приписать кванторы всеобщности по всем переменным, получив (истинное или ложное) универсальное утверждение о колмогоровской сложности.

Далее мы рассмотрим более сложные свойства колмогоровской сложности. Чтобы выразить их, нам могут потребоваться формулы с переменами кванторов. Простейший пример – свойство выделяемости взаимной информации пары

слов, которые мы рассмотрели выше. Другие примеры колмогоровских свойств, требующих перемен кванторов в формулировке, можно найти, например, в [11]. Наиболее общий вид колмогоровского свойства для слов $x_1, \ldots, x_n$ может быть выражен формулой вида

$$\forall y_1 \exists y_2 \forall y_3 \ldots \vec{K}(x_1, \ldots, x_n, y_1, \ldots, y_m) \in A, \tag{2}$$

где $A \subset \mathbb{Z}^{2^{n+m}-1}$. Если для $\bar{x} = (x_1, \ldots, x_n)$, $\bar{x}' = (x'_1, \ldots, x'_n)$ и $C > 0$, каждому свойству кортежа $\bar{x}$ вида (2) соответствует свойство кортежа $\bar{x}'$ вида

$$\forall y_1 \exists y_2 \forall y_3 \ldots \vec{K}(x'_1, \ldots, x'_n, y_1, \ldots, y_m) \in A',$$

с C-близкими A и A' (то есть для всякой точки из A найдётся точка в A' на расстоянии меньше C, и наоборот), то мы можем говорить, что данные кортежи обладают одинаковыми свойствами для заданной степени точности C.

Итак, мы уточнили, какие свойства мы называем колмогоровскими. Теперь перейдём к вопросу об устойчивости этих свойств при релятивизации. Пусть задан некоторый оракул O (конечная или бесконечная двоичная последовательность). Мы можем рассматривать колмогоровскую сложность, релятивизованную относительно данного оракула. В случае конечного O релятивизованная сложность совпадает с обычной относительной сложностью.

Теперь мы готовы сформулировать основные результаты нашей работы. Для простоты мы ограничимся формулировками для конечного оракула O (слово, задающее оракул, мы будем обозначать z).

Далее мы будем использовать асимптотическое обозначение $\mathcal{O}(f(x_1, \ldots, x_n))$ для различных выражений $f(\cdot)$, в которые входит колмогоровская сложность. Мы имеем в виду, что мультипликативная константа в 'О-большом' зависит только от конкретизации функции колмогоровской сложности и от длины рассматриваемых кортежей. Более того, мы всегда считаем, что зависимость константы от длины кортежей эффективна.

Бесквантворные формулы. Нетрудно проверить, что для свойств, выражающихся бесквантворными формулами, малость взаимной информации между $\bar{x}$ и оракулом z является необходимым и достаточным условием для устойчивости элементарных колмогоровских свойств при релятивизации. Это есть переформулировка (тривиального) утверждения 1 из Введения:

Теорема 3 *Предположим, что для некоторых слов $\bar{x} = (x_1, \ldots, x_n)$ и слова z*

$$I(\bar{x} : z) \leq \delta.$$

Тогда соответствующие компоненты сложностных профилей $\vec{K}(\bar{x})$ и $\vec{K}(\bar{x}|z)$ отличаются не более чем на $\delta + \mathcal{O}(\log K(\bar{x}, z))$.

∃-формулы. Рассмотрим свойства, выражающиеся ∃-формулами (с параметрами). В этом случае наша общая гипотеза может быть сформулирована в виде следующей пары взаимно обратных утверждений (теоремы и гипотезы):

Теорема 4 *Предположим, что для некоторых слов $\bar{x}$, z*

$$I(\bar{x} : z) \leq \delta.$$

Тогда для любого $\bar{y} = \langle y_1, \ldots, y_m \rangle$ найдётся такой $\bar{y}' = \langle y'_1, \ldots, y'_m \rangle$, что соответствующие компоненты сложностных профилей $\vec{K}(\bar{x}, \bar{y})$ и $\vec{K}(\bar{x}, \bar{y}'|z)$ отличаются друг от друга не более чем на $\delta + \mathcal{O}(\log K(\bar{x}, \bar{y}, z))$.

Гипотеза 3 *Предположим, для некоторых слов* $\bar{x} = \langle x_1, \ldots, x_n \rangle$ *и* z

$$I(\bar{x} : z) \leq \delta.$$

Тогда для любого $\bar{y} = \langle y_1, \ldots, y_m \rangle$ *найдётся такой* $\bar{y}' = \langle y'_1, \ldots, y'_m \rangle$, *что соответствующие компоненты сложностных профилей* $\vec{K}(\bar{x}, \bar{y}|z)$ *и* $\vec{K}(\bar{x}, \bar{y}')$ *отличаются не более чем на* $\delta + \mathcal{O}(\log K(\bar{x}, \bar{y}, z))$.

Отметим, что гипотеза 2 является частным случаем гипотезы 3.

Нам удаётся доказать гипотезу 3 только для стохастических слов.

Определение 1 *Слово* x *называется* (α, β)*-стохастическим, если существует множество* A, *содержащее* x, *такое что*

- *сложность кортежа* $\hat{A}$, *состоящего из всех элементов* A *в лексикографическом порядке, не превосходит* α,
- $K(x|\hat{A}) \geq \log |A| - \beta$

(здесь и далее в нашем тексте все логарифмы берутся по основанию 2). В частности, всякое несжимаемое слово длины N (такое слово x длины N, что $K(x) \geq N$) является $(\log N + \mathcal{O}(1), \mathcal{O}(1))$-стохастическим.

Определение стохастичности очевидным образом переносится на *кортежи* слов. Для наших целей наибольший интерес будут представлять $(\mathcal{O}(\log N), \mathcal{O}(\log N))$-стохастические кортежи $\bar{x}$, где $N = K(\bar{x})$. Такие кортежи мы для краткости называем просто *стохастическими*.

Отметим, что сам по себе факт существования нестохастических слов является весьма нетривиальным [13]. В приложениях колмогоровской сложности как правило используются именно стохастические слова или кортежи слов. Поэтому изучение свойств стохастических наборов слов кажется заслуживающим внимания. Таким образом, следующая теорема отвечает на интересующий нас вопрос для наиболее типичной и наиболее важной для приложений ситуации:

Теорема 5 *Гипотеза 3 выполнена для стохастических* $\bar{x}$.

Для *нестахостических* кортежей гипотеза 3 остаётся недоказанной.

Существование кортежей, неэквивалентных стохастическим. Будем говорить, что слова a, b C-эквивалентны ($a \sim_C b$), если

$$K(a|b) \leq C \log K(a, b) \wedge K(b|a) \leq C \log K(a, b).$$

Далее, будем называть кортежи $\bar{a} = (a_1, \ldots, a_n)$ и $\bar{b} = (b_1, \ldots, b_n)$ C-эквивалентными (обозначая это $\bar{a} \sim_C \bar{b}$), если для каждого $i = 1, \ldots, n$ слово a_i C-эквивалентно b_i.

Поскольку мы изучаем колмогоровские свойства с логарифмической точностью, эквивалентные кортежи обладают одинаковыми свойствами. Может возникнуть искушение доказывать гипотезу 3, сводя произвольный кортеж $\bar{x}$ к эквивалентному ему стохастическому $\bar{x}'$ и применяя к $\bar{x}'$ теорему 5. Но можно ли для произвольного $\bar{x}$ найти эквивалентный ему стохастический $\bar{x}'$?

Прежде всего отметим, что для индивидуального слова x (для кортежа длины 1) можно найти C-эквивалентное ему $(C \log K(x), C \log K(x))$-стохастическое слово x'. (Разумеется, мы требуем, чтобы константа C была

достаточно большой, но независящей от x.) В самом деле, для любого слова x найдётся кратчайшее описание p для данного слова x, такое что

$$K(p|x) = \mathcal{O}(\log K(x)).$$

Это p и можно взять в качестве x'. Очевидно, данное слово $\mathcal{O}(\log K(x))$-эквивалентно слову x. В то же время $K(x') \geq |x'| - \mathcal{O}(1)$, так как оно является кратчайшим описанием x. Следовательно, x' стохастическое.

Верно ли аналогичное утверждение для пары $\langle x_1, x_2\rangle$? Если x_1, x_2 независимы, то мы можем отдельно заменить каждое из x_i на эквивалентное стохастическое x_i'. Нетрудно видеть, что пара $\langle x_1', x_2'\rangle$ является стохастической и эквивалентна $\langle x_1, x_2\rangle$. Другой простой пример: предположим, что $K(x_1|x_2) \leq \mathcal{O}(\log K(x_2))$. Тогда найдутся такие несжимаемые слова p и q, что x_1 эквивалентно p, а x_2 эквивалентно $\langle p, q\rangle$. При этом пара $(p, \langle p, q\rangle)$ является стохастической.

Однако в общем случае для пары слов нельзя найти эквивалентную ей стохастическую. Сформулируем это утверждение более точно.

Теорема 6 *Пусть даны рациональные числа $\alpha, \beta, \gamma, C > 0$ такие что $\alpha+\beta > \gamma$ и $\alpha, \beta < \gamma$. Тогда для достаточно больших n существует пара слов x_1, x_2, для которой*

- $K(x_1) = \alpha n + \mathcal{O}(\log n)$,
- $K(x_2) = \beta n + \mathcal{O}(\log n)$,
- $K(x_1, x_2) = \gamma n + \mathcal{O}(\log n)$,

и не существует $(C \log n, C \log n)$-стохастической пары x_1', x_2', которая была бы C-эквивалентна паре (x_1, x_2).

Замечание. Условие $\alpha + \beta > \gamma$ гарантирует, что x_1 и x_2 зависимы, а условия $\alpha, \beta < \gamma$ показывают, что ни одно из слов x_i не может быть слишком просто относительно другого.

Таким образом, уже для кортежа длины 2 в общем случае нельзя подобрать эквивалентный ему стохастический кортеж, и методы теоремы 5 не позволяют доказать гипотезу 3 в общем случае.

Как организована статья. В разделе 3 мы приводим основные определения и технические леммы (в основном, известные из других работ; для полноты изложения мы доказываем некоторые из этих технических леммы в Приложении). В разделе 4 мы доказываем основную гипотезу для бескванторных формул, а также для $\exists$-формул для стохастических кортежей слов. В разделе 5 мы объясняем, почему нестохастические кортежи (даже нестохастические пары) в общем случае нельзя заменить на эквивалентные им стохастические. В разделе 6 мы доказываем вариант нашей гипотезы для свойства выделения общей информации (в том числе и для нестахостических пар). В Заключении мы подводим итоги и комментируем вопросы, оставшиеся открытыми. Наконец, в Приложении мы приводим доказательства нескольких технических утверждений, известных из работ [8, 9, 12]. Отметим, что в каждом из разделов 4–6 мы используем разную технику доказательств, так что читатель может изучать эти разделы независимо и в произвольном порядке (после ознакомления с разделом 3).

3 Определения и технические леммы

3.1 Сложностные профили

Обозначение 1 Пусть фиксирована n-ка слов $\bar{x} = x_1, \ldots, x_n$. Для любого множества $V = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ $(1 \leq i_1 < \ldots < i_k \leq n)$ будем обозначать $\bar{x}_V$ кортеж слов x_j с индексами $j \in V$:

$$\bar{x}_V = \langle x_{i_1}, \ldots, x_{i_k} \rangle.$$

Мы будем использовать данное соглашение для обозначения колмогоровской сложности соответствующего кортежа:

$$K(\bar{x}_V) := K(x_{i_1}, \ldots, x_{i_k}).$$

Если $V = \emptyset$, то полагаем $K(\bar{x}_V) := K(\lambda)$ (где λ – пустое слово).

Аналогичное обозначение будем использовать для условных сложностей: для любых подмножеств

$$V = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\} \text{ и } W = \{j_1, \ldots, j_l\} \subseteq \{1, \ldots, n\}$$

полагаем

$$K(\bar{x}_V | \bar{x}_W) := K(x_{i_1}, \ldots, x_{i_k} | x_{j_1}, \ldots, x_{j_l}).$$

При этом если W пусто, то считаем $K(\bar{x}_V | \bar{x}_W) := K(\bar{x}_V | \lambda)$.

Определение 2 Будем называть *сложностным профилем* n-ки слов $\bar{x} = \langle x_1, \ldots, x_n \rangle$ вектор, состоящий из $(2^n - 1)$ сложностей $K(\bar{x}_W)$ для всех непустых подмножеств $W \subseteq \{1, \ldots, n\}$ (считаем, что подмножества W располагаются в лексикографическом порядке):

$$\vec{K}(x_1, \ldots, x_n) = (K(x_1), K(x_1, x_2), \ldots, K(x_2), K(x_2, x_3), \ldots).$$

Будем называть *относительным сложностным профилем* n-ки $\bar{x}$ при условии y вектор, состоящий из $(2^n - 1)$ сложностей $K(\bar{x}_W | y)$ для всех непустых подмножеств $W \subseteq \{1, \ldots, n\}$ (снова полагаем, что все подмножества W располагаются в лексикографическом порядке):

$$\vec{K}(x_1, \ldots, x_n | y) = (K(x_1 | y), K(x_1, x_2 | y), \ldots, K(x_2 | y), K(x_2, x_3 | y), \ldots).$$

Определение 3 Будем называть *расширенным сложностным профилем* n-ки слов $x_1, \ldots, x_n$ вектор, состоящий из всех условных сложностей вида $K(\bar{x}_V | \bar{x}_W)$, где $V, W \subseteq \{1, \ldots, n\}$, причём $V \cap W = \emptyset$ и $V \neq \emptyset$. Заметим, что в случае $W = \emptyset$ мы получаем значения безусловной сложности: $K(\bar{x}_V | \bar{x}_\emptyset) = K(\bar{x}_V) + \mathcal{O}(1)$. Считаем, что все пары подмножеств (V, W) располагаются в лексикографическом порядке:

$$\vec{K}'(x_1, \ldots, x_n) = (K(x_1), K(x_1 | x_2), \ldots, K(x_2), K(x_2 | x_1), K(x_2 | x_3), \ldots).$$

Аналогично определим *расширенный относительъный сложностной профиль* $x_1, \ldots, x_n$ при условии y. Для этого рассмотрим вектор из всех сложностей вида $K(\bar{x}_V | \bar{x}_W, y)$:

$$\vec{K}'(x_1, \ldots, x_n | y) = (K(x_1 | y), K(x_1 | x_2, y), \ldots, K(x_2 | y), K(x_2 | x_1, y), K(x_2 | x_3, y), \ldots).$$

Нам потребуется сравнивать сложностные профили для разных наборов слов. Для этого мы введём обозначения для сравнения векторов в $\mathbb{R}^k$.

Обозначение 2 *Будем говорить что сложностной вектор $\bar{\alpha} \in \mathbb{R}^n$ не больше вектора $\bar{\beta} \in \mathbb{R}^n$ (обозначение: $\bar{\alpha} \leq \bar{\beta}$), если каждая компонента первого вектора не превосходит соответствующей компоненты второго вектора, т.е. $\alpha_i \leq \beta_i$ для $i = 1, \ldots, n$.*

Кроме того, будем использовать l_∞-норму для измерения расстояния между векторами:

$$\rho(\bar{\alpha}, \bar{\beta}) := \max_i \{|\alpha_i - \beta_i|\}.$$

В частности, будем говорить что сложностной профиль для $\bar{x} = (x_1, \ldots, x_n)$ не больше сложностного профиля для $\bar{y} = (y_1, \ldots, y_n)$, если каждая компонента первого профиля не превосходит соответствующей компоненты второго профиля, т.е. для каждого $V \subseteq \{1, \ldots, n\}$ выполнено $K(\bar{x}_V) \leq K(\bar{y}_V)$. Аналогично, мы будем говорить, что расстояние между сложностными профилями кортежей $\langle x_1, \ldots, x_n \rangle$ и $\langle y_1, \ldots, y_n \rangle$ не превосходит ε, если для каждого набора индексов V выполнено $|K(\bar{x}_V) - K(\bar{y}_V)| \leq \varepsilon$.

3.2 Типизация

В дальнейшем мы будем применять прием *типизации* (этот метод использовался в [8, 10, 9])

Определение 4 Пусть даны наборы слов $\bar{x} = \langle x_1, \ldots, x_n \rangle$ и $\bar{y} = \langle y_1, \ldots, y_m \rangle$. Будем называть типизацией $\bar{x}$ относительно $\bar{y}$ следующее множество n-ок слов:

$$T(\bar{x}|\bar{y}) := \{\bar{x}' = \langle x'_1, \ldots, x'_n \rangle \mid \vec{K}'(\bar{x}', \bar{y}) \leq \vec{K}'(\bar{x}, \bar{y})\}.$$

Далее, будем называть k-строгой типизацией $\bar{x}$ относительно $\bar{y}$ следующее множество:

$$ST_k(\bar{x}|\bar{y}) := \{\bar{x}' = \langle x'_1, \ldots, x'_n \rangle \mid \vec{K}'(\bar{x}', \bar{y}) \leq \vec{K}'(\bar{x}, \bar{y}) \text{ и } \rho(\vec{K}'(\bar{x}', \bar{y}), \vec{K}'(\bar{x}, \bar{y})) \leq k\}.$$

Отметим, что множество $T(\bar{x}|\bar{y})$ можно перечислять, зная $\bar{y}$ и все числа из расширенного профиля $\vec{K}'(\bar{x}, \bar{y})$. Множество $ST(\bar{x}|\bar{y})$ таким свойством не обладает.

Имеют место следующие леммы, доказанные в [8, 9].

Лемма 1 *Для любых наборов слов $\bar{x} = \langle x_1, \ldots, x_n \rangle$ и $\bar{y} = \langle y_1, \ldots, y_m \rangle$*

$$\log |T(\bar{x}|\vec{y})| = K(\bar{x}|\bar{y}) + \mathcal{O}(\log N),$$

где $N = K(\bar{x}, \bar{y})$.

Лемма 2 *Для любых натуральных n, m существует $C = C(n, m)$ такое, что для всякой n-ки $\bar{x} = \langle x_1, \ldots, x_n \rangle$ и всякой m-ки $\bar{y} = \langle y_1, \ldots, y_m \rangle$*

$$|ST_{C \log N}(\bar{x}|\bar{y})| > \frac{1}{2}|T(\bar{x}|\bar{y})|,$$

где $N = K(\bar{x}, \bar{y})$.

Обозначение 3 *Для краткости мы будем обозначать*

$$ST(\bar{x}|\bar{y}) = ST_{C\log N}(\bar{x}|\bar{y}),$$

где константа C такая же, как в лемме 2.

Элементы $ST(\bar{x}|\bar{y})$ мы будем также называть *клонами* $\bar{x}$ относительно $\bar{y}$.

Нам потребуется следующий несложный технический факт:

Лемма 3 *Пусть даны наборы слов $\bar{x} = \langle x_1, \ldots, x_n\rangle$, $\bar{y} = \langle y_1, \ldots, y_m\rangle$. Имеют место следующие утверждения:*

(1) Для любого $\bar{x}' = (x'_1, \ldots, x'_n)$ если

$$\vec{K}'(\bar{x}', \bar{y}) \le \vec{K}'(\bar{x}, \bar{y}) + \delta_1 \cdot \vec{e},$$

и $K(\bar{x}', \bar{y}) \ge K(\bar{x}, \bar{y}) - \delta_2$, то

$$\vec{K}'(\bar{x}', \bar{y}) \ge \vec{K}'(\bar{x}, \bar{y}) - (2\delta_1 + \delta_2 + \mathcal{O}(\log N))\vec{e}$$

(здесь $N = K(\bar{x}, \bar{y})$, $\vec{e} = (1, \ldots, 1)$).

(2) Для любого z и для любого $\bar{x}' = (x'_1, \ldots, x'_n)$ такого, что

$$\vec{K}'(\bar{x}', \bar{y}) \le \vec{K}'(\bar{x}, \bar{y}) + \delta_1 \cdot \vec{e}$$

и $K(\bar{x}', \bar{y}|z) \ge K(\bar{x}, \bar{y}) - \delta_2$, имеем

$$\vec{K}'(\bar{x}', \bar{y}|z) \ge \vec{K}'(\bar{x}, \bar{y}) - (2\delta_1 + \delta_2 + \mathcal{O}(\log N))\vec{e},$$

где $\vec{e} = (1, \ldots, 1)$ и $N = K(\bar{x}, \bar{y})$.

3.3 Комбинаторная энтропия

Обозначение 4 *Пусть $A \subset X_1 \times \ldots \times X_n$ – некоторое множество n-ок (в качестве множеств X_i мы как правило будем брать конечные множества двоичных слов). Для произвольного набора индексов $I = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ будем обозначать $\pi_I(A)$ проекцию A на соответствующей оси координат:*

$$\pi_I(A) := \{\bar{x}_I \mid \bar{x} \in A\}.$$

В частности, для $I = \{1, \ldots, n\}$ получаем $\pi_I(A) = A$.

Далее, для всякого кортежа $\bar{x} = \langle x_1, \ldots, x_k\rangle$ будем обозначать $\sigma_I(A|\bar{x})$ сечение A, соответствующее значению $\bar{x}$ проекции на оси I:

$$\sigma_I(A|\bar{x}) := \{\bar{y} \mid \bar{y} \in A \text{ и } \bar{y}_I = \bar{x}\}.$$

Обозначение 5 *Пусть $X_1, \ldots, X_n$ – некоторые конечные множества, и $A \subset X_1 \times \ldots \times X_n$ – произвольное множество n-ок. Будем использовать следующие обозначения:*

- *$n_I(A)$ – число элементов в $\pi_I(A)$.*
- *$n_{I|J}(A|\bar{x})$ – число элементов в $\pi_I\sigma_J(A|\bar{x})$.*

- $n_{I|J}(A) = \max\limits_{\bar{x}\in\pi_J(A)} n_{I|J}(A|\bar{x})$. *В частности, если* $J = \emptyset$, *то* $n_{I|J}(A) = n_I(A)$.

Определение 5 Пусть $X_1, \ldots, X_n$ – некоторые конечные множества, и $A \subset X_1 \times \ldots \times X_n$ – произвольное множество n-ок. Для любых наборов индексов $I, J \subseteq \{1, \ldots, n\}$ полагаем

- $\mathrm{ent}_I(A) := \lceil \log n_I(A) \rceil$.
- $\mathrm{ent}_{I|J}(A) := \lceil \log n_{I|J}(A) \rceil$ (если $J = \emptyset$, то $\mathrm{ent}_{I|\emptyset}(A) = \mathrm{ent}_I(A)$).

Лемма 4 *Пусть* $A \subset \mathbb{B}^n$ *– некоторое конечное множество* n*-ок слов. Обозначим* $list(A)$ *список всех элементов* A *(в некоторой вычислимой кодировке). Тогда для любого* $\bar{x} \in A$, *для любых* $V, W \subset \{1, \ldots, n\}$

$$K(\bar{x}_V | \bar{x}_W, list(A)) \leq \mathrm{ent}_{V|W}(A) + \mathcal{O}(1).$$

Доказательство . Имея список всех элементов A и кортеж $\bar{x}_W$, можно найти список всех элементов в множестве

$$B = \pi_V \sigma_W(A|\bar{x}_W).$$

При этом $\bar{x}_V \in B$. Чтобы описать $\bar{x}_V$, остаётся указать номер кортежа $\bar{x}_V$ в списке элементов B, что требует не более $\lceil \log n_{V|W}(A) \rceil$ битов. ▲

3.4 Метод гроздей

В [12] было сформулировано следующее определение *грозди слов*:

Определение 6 Будем называть (α, β, γ)-гроздью множество слов X такое, что

1. $|X| = 2^\alpha$,
2. $K(x_1|x_2) < \beta$ для всех $x_1, x_2 \in X$,
3. $K(x) < \gamma$ для всех $x \in X$.

Лемма 5 ([12]) *Существует алгоритм, который для любой заданной тройки натуральных чисел* α, β, γ *перечисляет некоторый список* (α, β, γ)*-гроздей* $U_0, \ldots, U_q$ *со следующими свойствами:*

- *для любой* (α, β, γ)*-грозди* U *найдется номер* $i \leq q$ *такой, что* $|U \cap U_i| \geq 2^{\beta-\epsilon}$, *где* $\epsilon = 2(\beta - \alpha) + \mathcal{O}(1)$,
- $q < 2^{\beta+\gamma-2\alpha+\mathcal{O}(1)}$.

Нам потребуется несколько модифицировать определение грозди слов.

Определение 7 Будем называть (α, β, γ)-полугроздью множество слов X такое, что

1. $|X| = 2^\alpha$,

2. для любого $x_1 \in X$ для более чем половины $x_2 \in X$ выполнено $K(x_1|x_2) < \beta$

3. $K(x) < \gamma$ для всех $x \in X$.

Следующая лемма аналогична лемме 5 о свойствах гроздей.

Лемма 6 *Существет алгоритм, который для любой заданной тройки натуральных чисел α, β, γ перечисляет некоторый список (α, β, γ)-полугроздей $U_0, \ldots, U_q$ со следующими свойствами:*

- *для любой (α, β, γ)-полугрозди U найдется номер $i \leq q$ такой, что $|U \cap U_i| \geq 2^{\beta - \epsilon}$, где $\epsilon = 2(\alpha - \beta) + \mathcal{O}(1)$,*
- $q < 2^{\beta + \gamma - 2\alpha + \mathcal{O}(1)}$.

Данный алгоритм никогда не останавливается (он печатает лишь конечное число U_i, но мы не можем эффективно определить, когда именно алгоритм напечатал последнюю полугроздь).

Доказательство леммы 6 совершенно аналогично доказательству леммы 5. Полугрозди $U_0, \ldots, U_q$ из леммы 6 мы будем называть *стандартными* (для заданных α, β, γ). Очевидно, для любых α, β, γ и любого $i \leq q$ сложность списка элементов стандартной полугрозди U_i при известном i равна $\mathcal{O}(\log \gamma)$.

Для полноты изложения в Приложении мы доказываем леммы 1–3 и лемму 6.

4 Бесквантворные и экзистенциальные формулы

Доказательство теоремы 3 тривиально. С одной стороны, для любого набора индексов $V \subseteq \{1, \ldots, n\}$

$$K(\bar{x}_V|z) \leq K(\bar{x}_V) + \mathcal{O}(1).$$

С другой стороны, поскольку $I(\bar{x}_V : z) \leq I(\bar{x} : z) + \mathcal{O}(\log N)$, имеем

$$K(\bar{x}_V) - K(\bar{x}_V|z) = I(\bar{x}_V : z) \leq I(\bar{x} : z) + \mathcal{O}(\log N) \leq \delta + \mathcal{O}(\log N).$$

▲

Доказательство теоремы 4. Требуется доказать, что существует такой набор слов $\bar{y}'$, что расстояние между $\vec{K}(\bar{x}, \bar{y})$ и $\vec{K}(\bar{x}, \bar{y}'|z)$ не превосходит $\delta + \mathcal{O}(\log N)$. Мы докажем формально более сильный факт: расстояние между соответствующими *расширенными* сложностными профилями не превосходит $\delta + \mathcal{O}(\log N)$.

Рассмотрим множество $T(\bar{y}|\bar{x})$. Согласно лемме 1 в данном множестве содержится $2^{K(\bar{y}|\bar{x}) + \mathcal{O}(\log N)}$ m-ок слов. Следовательно, мы можем выбрать $\bar{y}' \in T(\bar{y}|\bar{x})$ такое, что

$$K(\bar{y}'|\bar{x}, z) \geq K(\bar{y}|\bar{x}) - \mathcal{O}(\log N).$$

Для выбранного набора слов $\bar{y}'$

$$\begin{aligned} K(\bar{x}, \bar{y}'|z) = K(\bar{x}|z) + K(\bar{y}'|\bar{x}, z) &\geq K(\bar{x}) - \delta + K(\bar{y}|\bar{x}) - \mathcal{O}(\log N) = \\ &= K(\bar{x}, \bar{y}) - \delta - \mathcal{O}(\log N). \end{aligned}$$

Следовательно, можно применить лемму 3, т.е.

$$\rho(\vec{K}'(\bar{x},\bar{y}),\vec{K}'(\bar{x},\bar{y}'|z)) \le \delta + \mathcal{O}(\log N).$$

▲

Мы предполагаем, что верна гипотеза 3, являющаяся обращением теоремы 4. Нам известно доказательство этой гипотезы лишь для *стохастических* $\langle x_1,\dots,x_n\rangle$:

Доказательство теоремы 5.

Шаг 1. Пусть $N = K(\bar{x},\bar{y},z)$. Заменим каждое из слов $y_1,\dots,y_m$ на соответствующую ему кратчайшую программу (в оптимальном языке программирования). При этом величины сложностей в расширенном профиле $\vec{K}'(\bar{x},\bar{y}|z)$ изменятся не более чем на $\mathcal{O}(\log N)$. Так что далее без ограничения общности мы будем предполагать, что $\bar{y} \in (\mathbb{B}^N)^m$. По условию n-ка $\bar{x}$ является стохастической, то есть лежит в некотором простом множестве $S \subset (\mathbb{B}^*)^n$:

$$K(S) = \mathcal{O}(\log N),$$

$$\log|S| = K(\bar{x}) + \mathcal{O}(\log N).$$

Таким образом, $\langle\bar{x},\bar{y}\rangle \in S \times (\mathbb{B}^N)^m$.

Шаг 2. Рассмотрим множество $A_0 = T(\bar{x},\bar{y}|z) \cap \big(S \times (\mathbb{B}^N)^m\big)$. Размеры проекций и сечений A_0 не превосходят экспоненты от соответствующих значений расширенного сложностного профиля $\vec{K}'(\bar{x},\bar{y}|z)$. Будем называть множество $A \subset S \times (\mathbb{B}^N)^m$ *правильным*, если все его комбинаторные энтропии $\mathrm{ent}_{I|J}(A)$ не превосходят соответствующих комбинаторных энтропий A_0. Другими словами, множество A называется правильным, если для любых $I,J \subseteq \{1,\dots,(n+m)\}$

$$\log n_{I|J}(A) \le \mathrm{ent}_{I|J}(A_0).$$

В частности, само множество A_0 является правильным. Отметим, что из правильности A следует $n_{I|J}(A) < 2n_{I|J}(A_0)$.

Зная все координаты расширенного сложностного профиля $\vec{K}'(\bar{x},\bar{y}|z)$ и все элементы множества S, можно алгоритмически найти список *всех* правильных множеств (этот список чрезвычайно велик, но конечен!). Поскольку список всех элементов множества S имеет лишь логарифмическую сложность, список всех правильных множеств также может быть получен со сложностью $\mathcal{O}(\log N)$. Обозначим $A_1, A_2, \dots$ лексикографически упорядоченный список всех правильных множеств.

Согласно определению, размер каждого сечения правильного множества не более чем в двое превосходит размер соответствующего сечения A_0. При этом, однако, некоторые сечения могут быть значительно меньше указанной границы.

Назовём *сильной проекцией* множества A_i на первые n координат (т.е. на S) множество B_i, состоящее из точек проекции множества A_i на S, которым соответствуют достаточно большие сечения:

$$B_i = \{\bar{x}' \in \pi_{1,\dots,n}(A_i) \mid \log|\sigma_{1,\dots,n}(A_i|\bar{x}')| \ge \mathrm{ent}_{n+1,\dots,n+m|1,\dots,n}(A_0) - C_1\log N\}$$

(константу C_1 мы выберем ниже). В частности, сильную проекцию множества A_0 будем называть B_0. Для дальнейшего зафиксируем два алгоритма, которые по слову z, компонентам расширенного сложностного профиля $\vec{K}'(\bar{x},\bar{y})$ и константе C_1 перечисляют A_0 и B_0 соответственно.

Заметим, что зная z, x и расширенный профиль $\vec{K}'(\bar{x},\bar{y}|z)$, мы можем перечислять элементы сечения A_0, соответствующего значению $\bar{x}$ (в наших обозначениях это $\sigma_{1,\ldots,n}(A_0|\bar{x})$). Следовательно,

$$\mathrm{ent}_{n+1,\ldots,n+m|1,\ldots,n}(A_0) \le K(\bar{y}|\bar{x},z) \le \log|\sigma_{1,\ldots,n}(A_0|\bar{x})| + \mathcal{O}(\log N).$$

Таким образом, мы можем выбрать такое значение C_1, чтобы $\bar{x}$ принадлежало B_0.

Шаг 3. Теперь выберем из последовательности правильных множеств специальную подпоследовательность по следующем правилу. Пусть правильные множества $A_1,\ldots,A_{s-1}$ уже рассмотрены, причем $A_{i_1},\ldots,A_{i_k}$ включены в подпоследовательность. Очередное по списку правильное множество A_s включается в подпоследовательность, если разность

$$B_s \setminus \left(\bigcup_{r\le k} B_{i_r}\right)$$

имеет мощность не меньше $2^{K(\bar{x}|z)-C_2\log N}$ (константа C_2 будет выбрана позже).

Отметим, что в данную подпоследовательность будет включено не более $\frac{|S|}{2^{K(\bar{x}|z)-C_2\log N}} = 2^{I(\bar{x}:z)+C_2\log N+\mathcal{O}(\log N)}$ правильных множеств. Обозначим $\hat{A}$ объединение всех правильных множеств из выбранной подпоследовательности $A_{i_1}, A_{i_2},\ldots$ Проекцию $\hat{A}$ на первые n координат обозначим $\hat{B}$.

Очевидно, $K(\hat{A}) = \mathcal{O}(\log N + \log C_2)$ и $K(\hat{B}) = \mathcal{O}(\log N + \log C_2)$, т.к. список элементов этих множеств может быть найден алгоритмически, если известны расширенный сложностной профиль $\vec{K}'(\bar{x},\bar{y}|z)$, множество S и константа C_2.

Замечание. Очевидно, что

$$\log n_{I|J}(\hat{A}) \le \mathrm{ent}_{I|J}(A_0) + I(\bar{x}:z) + C_2\log N + \mathcal{O}(\log N)$$

для любых $I,J\subset\{1,\ldots,n\}$. При этом список всех элементов $\hat{A}$ может быть получен со сложностью $\mathcal{O}(\log N + \log C_2)$. Применяя лемму 4, находим, что для любого $\bar{u}\in\hat{A}$

$$K(\bar{u}_I|\bar{u}_J) \le \mathrm{ent}_{I|J}(A_0) + I(\bar{x}:z) + C_2\log N + \mathcal{O}(\log N + \log C_2)$$

для любых $I,J\subset\{1,\ldots,n\}$.

Лемма 7 *$\bar{x}\in\hat{B}$ (при подходящем выборе $C_2 = C_2(n,m)$).*

Доказательство леммы: Предположим противное: $\bar{x}$ не принадлежит сильной проекции построенного множества $\hat{A}$. В этом случае множество A_0 (которое входит в список всех правильных множеств) не было включено в выбранную подпоследовательность. Но это значит, что мощность разности $B_0\setminus\hat{B}$ заведомо меньше, чем $2^{K(\bar{x}|z)-C_2\log N}$. Следовательно, чтобы задать n-ку $\bar{x}$ при известных z и $\vec{K}'(\bar{x},\bar{y})$, достаточно указать список всех элементов $\hat{B}$, и номер кортежа $\bar{x}$ в списке всех элементов множества $B_0\setminus\hat{B}$ в порядке перечисления. Таким образом,

$$K(\bar{x}|z) \le \log|(B_0\setminus\hat{B})| + \mathcal{O}(\log N + \log C_2) \le K(\bar{x}|z) - C_2\log N + \mathcal{O}(\log N + \log C_2).$$

Выбирая достаточно большую константу C_2, получаем противоречие. ▲

Шаг 4. Таким образом, $\bar{x} \in \hat{B}$. Обозначим Q сечение множества $\hat{A}$, соответствующее $\bar{x}$:

$$Q = \{\bar{y} | \langle \bar{x}, \bar{y} \rangle \in \hat{A}\}.$$

Из построения $\hat{A}$ следует, что число элементов в Q не может быть слишком мало. Точнее,

$$\log |Q| \geq K(\bar{y}|\bar{x}, z) - \mathcal{O}(\log N).$$

Остается выбрать из Q m-ку $\bar{y}'$, имеющую максимальную возможную сложность относительно $\bar{x}$. А именно, существует $\bar{y}' \in \hat{S}$, для которой

$$K(\bar{y}'|\bar{x}) \geq \log |Q| \geq K(\bar{y}|\bar{x}, z) - \mathcal{O}(\log N).$$

Учитывая

$$K(\bar{x}) \geq K(\bar{x}|z) + I(\bar{x} : z) - \mathcal{O}(\log N),$$

мы получаем

$$K(\bar{x}, \bar{y}') \geq K(\bar{y}'|\bar{x}) + K(\bar{x}) \geq K(\bar{x}, \bar{y}|z) + I(\bar{x} : z) - \mathcal{O}(\log N).$$

С другой стороны, из построения $\hat{A}$ следует (см. замечание выше), что

$$\vec{K}'(\bar{y}'|\bar{x}) \leq \vec{K}'(\bar{y}|\bar{x}) + (I(\bar{x} : z) - \mathcal{O}(\log N)) \cdot \vec{e}.$$

Полагая в лемме 3 $\delta_1 = -\delta_2 = I(\bar{x} : z)$ имеем

$$\rho(\vec{K}'(\bar{x}, \bar{y}'), \vec{K}'(\bar{x}, \bar{y}|z)) \leq I(\bar{x} : z) + \mathcal{O}(\log N),$$

что заканчивает доказательство теоремы. ▲

5 Не для всякой пары существует эквивалентная ей стохастическая

Доказательство теоремы 6.

Для краткости мы будем называть $(C \log n, C \log n)$-стохастические пары просто стохастическими. Зафиксируем достаточно большое n. Обозначим S_1 множество слов длины αn и S_2 множество слов длины βn. Мы построим некоторое перечислимое множество $A \subset S_1 \times S_2$ размера $2^{\gamma n - \mathcal{O}(\log n)}$. При этом мы покажем, что некоторый элемент A является искомым – он имеет требуемый сложностной профиль, и для него нет C-эквивалентной стохастической пары.

Нам будет удобно представлять A как множество рёбер в двудольном графе с левой долей S_1 и правой долей S_2.

Всякое слово x_1' C-эквивалентное некоторому слову из S_1 имеет сложность не более $\alpha n + 2C \log n$. Обозначим множество всех слов с такими сложностями L_1. Аналогично, всякое слово x_2', эквивалентное некоторому слову из S_2, имеет сложность не более $\beta n + 2C \log n$. Множество слов с таким сложностями мы обозначим L_2.

Нас будут интересовать множества $R \subset L_1 \times L_2$ такие, что

- $|R| \leq 2^{\gamma n + 3C \log n}$,
- $K(R) \leq C \log n$, то есть список элементов R можно получить с помощью алгоритма сложности не более $C \log n$.

Очевидно, если пара слов (x_1, x_2) из множества A C-эквивалентна некоторой стохастической паре (x'_1, x'_2), то данная пара (x'_1, x'_2) должна принадлежать одному из множеств R указанного вида. Число всех таких множеств R не превосходит $2^{C \log n}$. Зная их точное число (чтобы сообщить это число, нужно логарифмическое число битов), мы можем найти все множества R. Обозначим $\hat{R}$ их объединение. Множество $\hat{R}$ также удобно рассматривать как двудольный граф; правую и левую доли вершин составляют L_1 и L_2 соответственно.

Таким образом, с помощью программы длины $\mathcal{O}(\log n)$ мы можем получить двудольный граф $\hat{R}$. Остаётся построить граф A, у которого будет достаточно много рёбер неэквивалентных ни одному из рёбер $\hat{A}$.

Некоторая трудность состоит в том, что отношение C-эквивалентности не является разрешимым. Однако мы можем описать относительно небольшой (и вычислимый) класс отношений, который будет заведомо содержать интересующее нас отношение C-эквивалентности. А именно, назовём *отношением сходства* всякое

$$D \subset S_1 \times L_1 \cup S_2 \times L_2,$$

удовлетворяющее следующим условиям:

- для любого $x \in S_i$ имеется не более $2^{C \log n + 1}$ элементов $y \in L_i$, для которых $(x, y) \in D$;
- для любого $y \in L_i$ имеется не более $2^{C \log n + 1}$ элементов $x \in S_i$, для которых $(x, y) \in D$,

$i = 1, 2$. Очевидно, отношение

$$D_0 = \{(x, y) \in S_1 \times L_1 \cup S_2 \times L_2 \ : \ x \sim_C y\}$$

удовлетворяет данным условиям. Заметим, что общее число отношений сходства не превосходит

$$(|L_1|^{\mathrm{poly}(n)})^{|S_1|} \cdot |L_2|^{\mathrm{poly}(n)})^{|S_2|}.$$

Без ограничения общности можно считать, что $\alpha \geq \beta$. Тогда число различных отношений сходства равно

$$2^{2^{\alpha n + \mathcal{O}(\log n)}}.$$

Будем говорить, что ребро $(x_1, x_2) \in A$ D-сходно с ребром $(x'_1, x'_2) \in \hat{R}$, если $(x_i, x'_i) \in D$ для $i = 1, 2$.

Теперь мы готовы перейти к построению графа A. Будем называть ребро $(x_1, x_2) \in A$ *правильным*, если степени вершин x_1 и x_2 не превосходят $2^{(\gamma-\alpha)n + C_1 \log n}$ и $2^{(\gamma-\beta)n + C_1 \log n}$ соответственно. Будем требовать, чтобы

$$|A| = 2^{\gamma n - C_1 \log n},$$

а также чтобы выполнялось следующее условие:

$$\begin{gathered} \textit{Для любого отношения сходства } D \textit{ найдется не менее} \\ 2^{\gamma n - C_2 \log n} \textit{ правильных рёбер } (x_1, x_2) \in A, \textit{ не являющих-} \\ \textit{ся } D\textit{-сходными ни с одним ребром из } \hat{R} \end{gathered} \tag{3}$$

(константы C_1, C_2 будут выбраны позднее).

Мы построим такое множество A эффективно, то есть его сложность будет равна $\mathcal{O}(\log n)$. При этом не менее $2^{\gamma n - C_2 \log n}$ рёбер из A не будет иметь эквивалентных стохастических пар. Остаётся лишь выбрать среди данных рёбер одно, имеющее сложность не менее $\gamma n - C_2 \log n$. Для выбранной пары (x_1, x_2) будут выполнены условия

$$K(x_1) \leq \alpha n, K(x_2) \leq \beta n,$$

а также

$$K(x_2|x_1) \leq (\gamma - \alpha)n + C_1 \log n$$

и

$$K(x_1|x_2) \leq (\gamma - \beta)n + C_1 \log n.$$

Кроме того, при достаточно большой константе C_1

$$K(x_1, x_2) < \gamma n,$$

и тем самым теорема будет доказана.

Отметим, что свойство (3) можно проверить алгоритмически. Мы покажем, что для множества, состоящего из $2^{\gamma n - C_1 \log n}$ случайно выбранных рёбер из $S_1 \times S_2$ с положительной вероятностью выполнено условие (3). Тем самым будет доказано, что множества с нужными нам свойствами существуют, и мы сможем найти одно из них (скажем, лексикографически первое) перебором.

Зафиксируем одно из отношений сходства D. Назовём ребро $(x_1, x_2) \in S_1 \times S_2$ *плохим*, если для него найдется D-сходное ребро в $\hat{R}$. Подсчитаем вероятность оказаться плохим для ребра случайно выбранного среди всех пар $S_1 \times S_2$. Граф $\hat{R}$ содержит $2^{\gamma n + \mathcal{O}(\log n)}$ рёбер. Для каждого из них имеется не более $\mathrm{poly}(n)$ D-сходных пар в $S_1 \times S_2$. Следовательно,

$$\mathrm{Prob}[(x_1, x_2) \text{ плохое }] \leq \frac{2^{\gamma n + \mathcal{O}(\log n)}}{2^{(\alpha+\beta)n}} \ll 1/2.$$

Здесь мы использовали условие $\alpha + \beta > \gamma$.

Пусть $k = 2^{\gamma n - C_1 \log n}$ и $l \leq 2^{\gamma n - C_2 \log n + 1}$. Тогда вероятность того, что среди k случайно выбранных рёбер $(k - l)$ оказались плохими, не превосходит

$$C_k^l \cdot (1/2)^{k-l} \leq k^l (1/2)^{k-l} \leq 2^{2^{\gamma n - C_2 \log n + \mathcal{O}(\log n)}} \cdot (1/2)^{2^{\gamma n - C_1 \log n + \mathcal{O}(\log n)}},$$

что равно $\frac{1}{2^{2^{\gamma n - \mathcal{O}(\log n)}}}$ при достаточно большой разнице между C_1 и C_2. Теперь мы должны просуммировать данную вероятность по всем $l = 0, \ldots, 2^{\gamma n - C_2 \log n + 1}$. Ясно, что умножение полученной вероятности на число различных l (то есть на $2^{\gamma n}$) не изменит существенно асимптотики убывания.

Итак, мы оценили вероятность того, что случайно выбранное A содержит много плохих рёбер для фиксированного отношения сходства D. Остаётся просуммировать эту вероятность по всем отношениям сходства. Это даст оценку вероятности того, что в случайно выбранном множестве A для любого отношения сходства D (в том числе и для собственно отношения C-эквивалентности) не менее $2^{\gamma n - C_2 \log n + 1}$ рёбер не эквивалентны ни одному из рёбер $\hat{R}$. А именно, данная вероятность не меньше

$$\frac{2^{2^{\alpha + \mathcal{O}(\log n)}}}{2^{2^{\gamma n - \mathcal{O}(\log n)}}} < 1.$$

Здесь мы использовали условие $\alpha < \gamma$.

Мы доказали, что с положительной вероятностью p_0 случайно выбранное множество A содержит не менее $2^{\gamma n - C_2 \log n + 1}$ рёбер, не эквивалентных ни одному из пар в $\hat{R}$. Однако некоторые из этих рёбер могут оказаться *неправильными* (одна из вершин имеет слишком большую степень). Покажем, что с вероятностью $p_1 > (1 - p_0)$ число неправильных рёбер в A не превосходит $2^{\gamma n - C_2 \log n}$. Таким образом, мы докажем, что с положительной вероятностью множество A содержит не менее $2^{\gamma n - C_2 \log n}$ рёбер, которые одновременно являются и *хорошими*, и *правильными*, то есть удовлетворяют условию (3).

Рассмотрим произвольную вершину $x \in S_1$. Среднее число рёбер, инцидентных x в случайно выбранным A, равно

$$\frac{2^{\gamma n - C_1 \log n}}{2^{\alpha n}} = 2^{(\gamma - \alpha)n - C_1 \log n}.$$

Из неравенства Чебышёва следует, что вероятность того, что вершина x инцидентна более чем $2^{(\gamma-\alpha)n + C_1 \log n}$ рёбрам, не превосходит $1/n^{2C_1}$. Следовательно, математическое ожидание числа вершин из S_1, имеющих аномально большую степень (более $2^{(\gamma-\alpha)n + C_1 \log n}$) не превосходит $2^{\alpha n}/n^{2C_1}$. Ещё раз воспользуемся неравенством Чебышёва: вероятность того, что число вершин в S_1 с аномально большой степенью оказалось больше $2^{(\gamma-\alpha)n - C_1 \log n}$, не может быть больше $1/n^{C_1}$. Аналогично, вероятность того, что число вершин в S_2 с аномально большой (более $2^{(\gamma-\beta)n + C_1 \log n}$) степенью превысит $2^{\beta n - C_1 \log n}$, также не может быть больше $1/n^{C_1}$. При больших n

$$p_1 \geq 1 - 2/n^{C_1} \gg 1 - p_0.$$

Таким образом, существование множества A, удовлетворяющего (3), доказано. ▲

6 Свойство выделяемости общей информации

Доказательство теоремы 2. Сначала введём обозначения и сделаем некоторые предположения. Без ограничения общности мы можем считать, что $f(N) > \log N$, и функция $f(N)$ нигде не убывает ($f(N+1) \geq f(N)$ для всех N).

Далее мы выбираем $g(N)$ и $\delta(N)$ которые растут не слишком быстро и не слишком медленно (с тем, чтобы работала конструкция в нашем доказательстве). В выборе этих функций имеется значительный произвол. Для определённости положим $\delta(N) = N/\sqrt{\log \frac{N}{f(N)}}$ и

$$g(N) = C(3^{D\sqrt{\log \frac{N}{f(N)}}} \cdot f(N) + \delta(N))$$

(константы C и D будут зафиксированы ниже). Для краткости мы будем писать просто δ, если значение аргумента N ясно из контекста.

Неформальная идея: Основной трюк в доказательстве состоит в типизации y и w относительно $\bar{x}$. Мы берём множество 'клонов' пары $\langle y, w \rangle$, которые имеют примерно одинаковые сложностные профили (относительно $\bar{x}$). Возможны два случая:

Простой случай: Пусть множество 'клонов' оказалось плотно консолидированным: большинство клонов имеют достаточно большую взаимную информацию. В этом случае мы применяем лемму 6 и выделяем из множества клонов общее ядро z. Слово z содержит примерно α битов информации; при этом данное слово просто относительно каждого из x_i. Таким образом, мы выделили у слов x_i примерно α битов общей информации без оракула, и теорема доказана.

Сложный случай: Пусть множество 'клонов' не является плотно консолидированным. Тогда найдётся пара клонов, у которых взаимная информация довольно мала. На данном этапе мы не можем выделить из слов x_i нужную общую информацию. Зато мы можем заменить слово y на такое слово y_1, что относительно y_1 из заданных слов x_1, x_2 можно выделить α_1 битов общей информации (и α_1 больше α). Таким образом, мы сводим исходную задачу о выделении информации у слов x_1, x_2 с оракулом y и параметром α к аналогичной задаче для слов x_1, x_2, оракула y_1 и параметра α_1. За увеличение параметра α мы платим ухудшением параметра точности: вместо уровня точности $f(N)$, в новой задаче мы вынуждены использовать несколько более слабое ограничение $f_1(N)$.

Остаётся объяснить, как построить нужный нам оракул y_1. Напомним, что множество 'клонов' не является плотно консолидированным. Случайно выберем два клона; обозначим их $\langle y', w'\rangle$ и $\langle y'', w''\rangle$. Тогда пара $\langle y', y''\rangle$ может играть роль y_1. Действительно, с новым оракулом мы можем извлечь из каждого из x_i как w', так и w''. Вместе w' и w'' дают α_1 битов общей информации ($\alpha_1 > \alpha$; более точно, мы получим оценку $\alpha_1 \geq \alpha + \delta/2$).

Мы итерируем описанный трюк, пока на некотором этапе не будет получено плотно консолидированное множество клонов.

Далее мы изложим формальное доказательство, следующее описанному выше плану.

Строгое рассуждение: По условию теоремы имеется такое слово w, что выполнено $K(w|x_i, y) \leq f(N)$ (для $i = 1, 2$). Без ограничения общности можно считать, что $\alpha = K(w|y)$ (если $K(w|y) > \alpha$, мы увеличим значение α; это сделает доказываемое утверждение только сильнее). Обозначим $m = K(y)$. Нашей целью является построение такого слова z, что $K(z|x_i) \leq g(n)$ и $K(z) \geq \alpha - g(N)$.

Рассмотрим строгую типизацию пары $\langle y, w\rangle$ относительно x: положим $A = ST(y, w|\bar{x})$. По лемме 1 получаем $|A| = 2^{K(y,w|\bar{x}) - \mathcal{O}(f(N))}$. Далее, мы имеем соотношения

$$K(y, w|\bar{x}) = K(y|\bar{x}) + K(w|y, \bar{x}) + O(\log N),$$

$K(y|\bar{x}) \geq K(y) - f(N)$ (взаимная информация между y и $\bar{x}$ пренебрежимо мала) и $K(w|y, \bar{x}) \leq f(N)$ (слово w можно легко извлечь из каждого из x_i, имея y в качестве оракула). Следовательно, $|A| = 2^{m - \mathcal{O}(f(N))}$. Отметим, что для всех $\langle y', w'\rangle \in A$ выполняется

$$K(y', w') = K(y') + K(w'|y) + \mathcal{O}(\log N) = m + \alpha + \mathcal{O}(f(N)).$$

Возможны два случая:

Случай 1^0: Для каждого $\langle y', w'\rangle \in A$ для большинства $\langle y'', w''\rangle \in A$

$$I(y'w' : y''w'') \geq \alpha - \delta.$$

Данное неравенство означает, что

$$K(y'w'|y''w'') = K(y', w') - I(y'w' : y''w'') \leq m + \delta + \mathcal{O}(f(N)).$$

Таким образом, множество A является полугроздью с параметрами

$$(m - \mathcal{O}(f(N)), m + \delta + \mathcal{O}(f(N)), m + \alpha + \mathcal{O}(f(N))).$$

Применяя лемму 6, заключаем, что существует *стандартная полугроздь* U_j (с теми же параметрами) такая, что

$$|A \cap U_j| \geq 2^{m-\delta+\mathcal{O}(f(N))},$$

и номер j не превосходит $2^{\alpha+\delta+\mathcal{O}(f(N))}$. Таким образом, колмогоровская сложность j не превосходит $\alpha + \delta + \mathcal{O}(f(N))$.

Далее, для слов x_i $(i = 1, 2)$ выполнены два свойства:

- для каждой пары $\bar{v} \in A \cap U_j$ выполнено $K(x_i|\bar{v}) \leq K(x_i|y, w)$ (по определению $A = ST(y, w|\bar{x})$);
- для каждой пары $\bar{v} \in A \cap U_j$ выполнено неравенство $K(\bar{v}|j) \leq \log |U_j| + \mathcal{O}(\log N) \leq m$ (зная номер j, можно алгоритмически перечислять список элементов полугрозди U_j).

Это означает, что x_i принадлежит множеству

$$X(i) = \{\hat{x} \quad | \quad \text{существует не менее } 2^{m-\delta+\mathcal{O}(f(N))} \text{ слов } \bar{v} \text{ таких, что } K(\hat{x}|\bar{v}) \leq K(x_i|y, w) \leq K(x_i) - \alpha + f(N) \text{ и } K(\bar{v}|j) \leq m\}.$$

Чтобы перечислять элементы $X(i)$, нужно знать число j и дополнительные $\mathcal{O}(\log N)$ битов информации (чтобы задать параметры полугрозди). Далее, можно оценить сверху размер $X(i)$. Действительно, для фиксированного j имеется не более 2^{m+1} различных наборов $\bar{v}$, для которых $K(\bar{v}|j) \leq m$; для каждого $\bar{v}$ существует не более $2^{K(x_i)-\alpha+f(N)}$ различных $\hat{x}$, для которых $K(\hat{x}|\bar{v}) \leq K(x_i) - \alpha + f(N)$. Поскольку для каждого $\hat{x} \in X(i)$ должно существовать *не менее* $2^{m-\delta+\mathcal{O}(f(N))}$ разных $\bar{v}$, мы получаем

$$\log |X(i)| \leq \log \frac{2^m \cdot 2^{K(x_i)-\alpha+f(N)}}{2^{m-\delta+O(f(N))}} \leq K(x_i) - \alpha + \delta + \mathcal{O}(f(N)).$$

Следовательно, $K(x_i|j) \leq K(x_i) - \alpha + \delta + \mathcal{O}(f(N))$ (другими словами, взаимная информация между j и x_i не меньше $\alpha - \delta - \mathcal{O}(f(N))$). Из симметрии взаимной информации получаем

$$K(j|x_i) = K(x_i|j) + K(j) - K(x_i) + \mathcal{O}(\log N) \leq 2\delta + \mathcal{O}(f(N)).$$

Положим $z = j$. Поскольку $K(z) \geq I(z : x_i) \geq \alpha - \delta - \mathcal{O}(f(N))$, для определённой выше функции $g(n)$ мы получаем $K(z) \geq \alpha - g(N)$ и $K(z|x_i) \leq g(N)$, что и требовалось доказать.

Случай 2^0: Для некоторых пар $\langle y', w' \rangle \in A$ и для большинства $\langle y'', w'' \rangle \in A$ выполнено неравенство

$$I(y'w' : y''w'') < \alpha - \delta.$$

Это означает, что

$$K(y'y''w'w'') \geq 2m + \alpha + \delta - \mathcal{O}(\log N) \tag{4}$$

Поскольку это неравенство выполнено для большинства пар $\langle y'', w''\rangle \in A$, мы можем выбрать среди них них одну, для которой $\langle y', w'\rangle$ и $\langle y'', w''\rangle$ независимы относительно $\bar{x}$. В частности, слова y' и y'' также независимы относительно $\bar{x}$ (т.е. $I(y' : y''|\bar{x}) = \mathcal{O}(\log N)$). Далее, для всех $\bar{x}, y', y''$ выполнено неравенство

$$I(y'y'' : \bar{x}) \le I(y' : \bar{x}) + I(y'' : \bar{x}) + I(y' : y''|\bar{x}) + \mathcal{O}(\log N)$$

(по существу это сумма двух элементарных соотношений:

$$K(y'y'') \le K(y') + K(y'') + \mathcal{O}(\log N),$$
$$K(y'|x) + K(y''|x) = K(y'y''|x) + I(y' : y''|x) + \mathcal{O}(\log N),$$

которые немедленно вытекают из теоремы Колмогорова–Левина о сложности пары [1]). Для данных слов величины $I(y' : \bar{x})$ и $I(y'' : \bar{x})$ ограничены сверху $f(N)$ ($\bar{x}$ и y независимы), и $I(y' : y''|\bar{x}) = \mathcal{O}(\log N) \ll f(N)$. Таким образом, получаем

$$I(y'y'' : \bar{x}) \le 3f(N) \tag{5}$$

Также оценим сверху (очень грубо) сложность пары $\langle y', y''\rangle$: $K(y'y'') \le 2K(y) + 3f(N) \le 3N$.

Из (4) и (5) получаем для $y^1 = \langle y', y''\rangle$ и $w^1 = \langle w', w''\rangle$ следующее неравенство:

$$K(w^1|y^1) \ge \alpha + \delta - 3f(N) - \mathcal{O}(\log N) \ge \alpha + \delta/2.$$

Следовательно, мы построили слово y^1 такое, что $I(y^1 : \bar{x}) \le 3f(N)$ и

$$\exists w^1 :\ K(w^1|y^1) \ge \alpha + \delta/2,\ K(w^1|x_i, y^1) \le 3f(N)\ (i = 1, 2).$$

Подведём итог. Мы получили вместо исходной пары $\langle y, w\rangle$ новую пару слов $\langle y^1, w^1\rangle$. По построению, слово y^1 независимо с $\bar{x}$ (хотя точность 'независимости' стала втрое хуже: $I(y^1 : \bar{x}) \le 3f(N)$). Относительно оракула y^1 слово w^1 оказывается просто относительно x_i ('простота' означает, что соответствующая относительная колмогоровская сложность также не превосходит $3f(N)$). Сложность w^1 относительно y^1 не может быть меньше $\alpha+\delta/2$. Таким образом, $\alpha+\delta/2$ битов общей информации могут быть выделены из слов x_1, x_2 при уровне точности $3f(N)$, если y^1 дано в качестве оракула. Отметим, что сложности слов w^1, y^1 заведомо не превосходят $3N$.

Далее мы итерируем приведённое выше рассуждение. Мы повторяем ту же процедуру с парой w^1, y^1. Обозначим $\alpha^1 = \alpha+\delta/2$, $m_1 = K(y^1)$ и $f_1(N) = 3f(N)$. Рассмотрим строгую типизацию пары $\langle y^1, w^1\rangle$ относительно $\bar{x}$:

$$A^1 = ST(y^1, w^1|\bar{x}).$$

Как и раньше, возможны два случая.

Случай 1^1: для каждой пары $\langle y', w'\rangle \in A^1$ для большинства $\langle y'', w''\rangle \in A^1$

$$I(y'w' : y''w'') \ge \alpha_1 - \delta.$$

В этом случае A^1 является полугроздью с параметрами

$$(m_1 - \mathcal{O}(f_1(N)), m_1 + \delta + \mathcal{O}(f_1(N)), m_1 + \alpha_1 + \mathcal{O}(f_1(N))).$$

По лемме 6 существует номер j такой, что для $i = 1, 2$

$$K(j|x_i) \le 2\delta + \mathcal{O}(f_1(N)), \quad I(j : x_i) \ge \alpha_1 - \delta + \mathcal{O}(f_1(N)).$$

Так же, как и в Случае 1^0, положим $z := j$, и доказательство закончено.

Случай 2^1: предположим, что для каждой пары $\langle y', w' \rangle \in A^1$ и для большинства $\langle y'', w' \rangle \in A^1$ выполнено $I(y'w' : y''w'') < \alpha_1 - \delta$. Тогда существует такая пара $\langle y^2, w^2 \rangle$, что

1. $K(y^2) = m_2 < 3m_1$,
2. $I(y^2 : \bar{x}) \le f_2(N) := 3f_1(N)$,
3. $K(w^2|y^2, x_i) \le f_2(N)$,
4. $K(w^2|y^2) = \alpha_2 \ge \alpha_1 + \delta/2$.

Повторяя данную конструкцию снова и снова, на каждом шаге s мы будем получать слова w^s, y^s, для которых

1. $K(y^s) = m_s = 3m_{s-1}$,
2. $I(y^s : \bar{x}) \le f_s(N) := 3f_{s-1}(N) = 3^s f(N)$,
3. $K(w^s|y^s, x_i) \le f_s(N)$,
4. $K(w^s|y^s) = \alpha_s > \alpha_{s-1} + \delta/2 = \alpha + s\delta/2$.

Мы итерируем конструкцию для случаев $2^1, 2^2, 2^3 \ldots, 2^j, \ldots$, пока на некотором шаге s_{max} не возникнет *случай* $1^{j_{max}}$.

Описанная итерация не может длиться слишком долго. Действительно, через $s = D\sqrt{\log \frac{N}{f(N)}}$ шагов (для достаточно большой константы D) мы получаем противоречие с неравенством

$$K(w^s|y^s) \le K(w^s|x_1, y^s) + K(w^s|x_2, y^s) + I(x_1 : x_2|y^s) + \mathcal{O}(\log N)$$

(нетрудно проверить, что это неравенство выполнено для любых слов; см., например, доказательство неравенства (6) в [8]): значение в левой части данного неравенства не меньше $DN/2$, а значение в правой не больше

$$2f_s(N) + I(x_1 : x_2|y_s) + \mathcal{O}(\log N) = \mathcal{O}(N).$$

Замечание. Во всех приведённых рассуждениях мы игнорировали аддитивные члены порядка $\mathcal{O}(\log K(y^s, w^s))$. Мы имели на это право, поскольку $\log K(y^s, w^s) \ll f(N)$. Эта оценка выполнена, т.к. $K(y^s), K(w^s) < N^2$ для $s \ll \log N$.

Таким образом, после некоторого числа повторений *Случая* 2^s, на шаге $s_{max} < D\sqrt{\log \frac{N}{f(N)}}$ мы переходим к *случаю* $1^{s_{max}}$. Это значит, что мы получаем такое слово z, что

$$K(z) \ge \alpha + s_{max}\delta/2 - \mathcal{O}(f_{s_{max}}(N)) > \alpha - g(N)$$

и

$$K(z|x_i) \le 2\delta + f_{s_{max}} < 2\delta + 3^{D\sqrt{\log \frac{N}{f(N)}}} f(N) < g(N) \ (i = 1, 2).$$

Другими словами, не менее α битов общей информации нам удаётся выделить из слов x_i для уровня точности $g(N)$. ▲

7 Заключение

Полученные нами результаты не дают полного ответа на поставленные вопросы. Остаётся недоказанной наша основная гипотеза 3 для нестохастических кортежей. Остаётся открытым даже довольно специальный случай этого вопроса — гипотеза 2 (формулировка доказанной нами теоремы 2 выглядит несколько искусственно). Наконец, было бы интересно получить подтверждения нашей основной гипотезы 1 для свойств более общего вида, требующих в формулировке нескольких перемен кванторов. Мы предполагаем, что доказательство этих результатов требует развития новой комбинаторной техники.

Список литературы

[1] Звонкин, А.К., Левин, Л.А.: Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов. УМН, 25:6(156) (1970), 85–127.

[2] Gács, P., Körner, J.: Common Information Is far Less Than Mutual Information. Problems of Control and Information Theory, 2, 49-62 (1973).

[3] Ahlswede, R., Körner, J.: On Common Information and Related Characteristics of Correlated Information Sources. Presented at the 7th Prague Conf. on Inf. Th., Stat. Dec. Fct's and Rand. Proc. (1974)

[4] Li, M., Vitányi, P.: An introduction to Kolmogorov complexity and its applications. 2nd ed., Springer-Verlag, New York (1997).

[5] Zhang, Z., Yeung, R.W.: On Characterization of Entropy Functions via Information Inequalities. IEEE Trans. on Information Theory, 44 1440-1452 (1998).

[6] Muchnik, An.A: On Common Information. Theoretical Computer Science, 207, 319–328 (1998).

[7] Ромащенко А.: Пары слов с нематериализуемой взаимной информацией. Пробл. передачи информ., 36:1 (2000), 3–20.

[8] Hammer, D., Romashchenko, A., Shen, A., Vereshchagin, N.: Inequalities for Shannon Entropy and Kolmogorov Complexity. Journal of Computer and System Sciences. 60, 442–464 (2000).

[9] Makarychev, K., Makarychev, Yu., Romashchenko, A., Vereshchagin, N., A New Class of non Shannon Type Inequalities for Entropies. Communications in Information and Systems. 2:2, 147-166 (2002).

[10] Romashchenko, A. Shen, A. Vereshchagin, N.: Combinatorial Interpretation of Kolmogorov Complexity. Theoretical Computer Science. 271, 111–123 (2002).

[11] Chernov, A., Muchnik, An.A., Shen, A., Romashchenko, A., Vereshchagin, N.K.: Upper Semi-Lattice of Binary Strings with the Relation "x Is Simple Conditional to y" Theoretical Computer Science. 271, 69–95 (2002).

[12] Romashchenko, A.: Extracting the Mutual Information for a Triple of Binary Strings. Proc. 18th Annual IEEE Conference on Computational Complexity (2003).

[13] Шень, А.: Понятие (α, β)-стохастичностипо Колмогорову и его свойства. ДАН СССР, 1983. 271:6, 1337–1349.

[14] Uspensky, V.A., Shen, A.: Relations Between Varieties of Kolmogorov Complexities. Mathematical Systems Theory 29(3): 271-292 (1996).

8 Приложение

В этом разделе для полноты изложения мы приводим доказательства технических лемм, использовавшихся в основном тексте.

Доказательство леммы 1. Прежде всего, для любого $\bar{x}' \in T(\bar{x}|\bar{y})$

$$K(\bar{x}'|\bar{y}) \le K(\bar{x}|\bar{y}).$$

Следовательно, число таких $\bar{x}'$ не превосходит $2^{K(\bar{x}|\bar{y})+1}$. Далее, оценим размер $T(\bar{x}|\bar{y})$ снизу. Заметим, что зная $\bar{y}$ и все числа из профиля $\vec{K}'(\bar{x}|\bar{y})$, можно перечислять список всех элементов множества $T(\bar{x}|\bar{y})$ (разумеется, не имея очень большой дополнительной информации, нельзя определить, когда данный процесс перечисления закончится). Таким образом, чтобы получить $\bar{x}$ из $\bar{y}$, достаточно знать все компоненты расширенного сложностного профиля $\vec{K}(\bar{x}|\bar{y})$, а также номер кортежа $\bar{x}$ в указанном списке (в порядке перечисления). Следовательно,

$$K(\bar{x}|\bar{y}) \le \log |T(\bar{x}, \bar{y})| + \mathcal{O}(\log N),$$

что и даёт требуемую оценку на размер $T(\bar{x}|\bar{y})$. ▲

Доказательство леммы 2. Согласно лемме 1

$$|T(\bar{x}|\bar{y})| \ge 2^{K(\bar{x}|\bar{y}) - C \log N}$$

для некоторой константы C. Следовательно, не меньше половины $\bar{x}' \in T(\bar{x}|\bar{y})$ имеют сложность относительно $\bar{y}$ большую или равную

$$K(\bar{x}|\bar{y}) - C \log N - 1.$$

Именно эти $\bar{x}' \in T(\bar{x}|\bar{y})$ мы и включим в $ST(\bar{x}|\bar{y})$.

Поскольку каждый элемент $\bar{x}' \in ST(\bar{x}|\bar{y})$ также принадлежит и $T(\bar{x}|\bar{y})$, мы имеем

$$\vec{K}'(\bar{x}', \bar{y}) \le \vec{K}'(\bar{x}, \bar{y}).$$

Остаётся доказать, что с точностью до $\mathcal{O}(\log N)$ выполнено обратное неравенство. Иначе говоря, для любых $V_1, V_2 \subset \{1, \ldots, n\}$, $W_1, W_2 \subset \{1, \ldots, m\}$ мы должны показать, что

$$K(\bar{x}'_{V_1}, \bar{y}_{W_1} | \bar{x}'_{V_2}, \bar{y}_{W_2}) \ge K(\bar{x}_{V_1}, \bar{y}_{W_1} | \bar{x}_{V_2}, \bar{y}_{W_2}) - \mathcal{O}(\log N). \tag{6}$$

Для этого заметим, что

$$K(\bar{x}', \bar{y}) = K(\bar{x}'_{V_2}, \bar{y}_{W_2}) + K(\bar{x}'_{V_1}, y_{W_1} | \bar{x}'_{V_2}, y_{W_2}) + K(\bar{x}', \bar{y}' | \bar{x}'_{V_1 \cup V_2}, \bar{y}_{W_1 \cup W_2}) + \mathcal{O}(\log N).$$

Правая часть этого равенства не превосходит

$$K(\bar{x}_{V_2}, \bar{y}_{W_2}) + K(\bar{x}'_{V_1}, y_{W_1}|\bar{x}'_{V_2}, y_{W_2}) + K(\bar{x}, \bar{y}|\bar{x}_{V_1\cup V_2}, \bar{y}_{W_1\cup W_2}) + \mathcal{O}(\log N),$$

поскольку $\bar{x}' \in T(\bar{x}|\bar{y})$. Далее, учитывая равенство

$$\begin{aligned} K(\bar{x}', \bar{y}) + \mathcal{O}(\log N) &= K(\bar{x}, \bar{y}) = \\ &= K(\bar{x}_{V_2}, \bar{y}_{W_2}) + K(\bar{x}_{V_1}, y_{W_1}|\bar{x}_{V_2}, y_{W_2}) + \\ &\quad + K(\bar{x}, \bar{y}'|\bar{x}_{V_1\cup V_2}, \bar{y}_{W_1\cup W_2}) + \mathcal{O}(\log N), \end{aligned}$$

получаем (6). ▲

Доказательство леммы 3. Доказательство аналогично доказательству предыдущей леммы. Сразу приступим к доказательству более общего утверждения пункта (2). Для любых U_1, U_2, U_3, U_4 имеем

$$K(\bar{x}'_{U_1}, \bar{y}_{U_3}|\bar{x}'_{U_2}, \bar{y}_{U_4}, z) \le K(\bar{x}'_{U_1}, \bar{y}_{U_3}|\bar{x}'_{U_2}, \bar{y}_{U_4}) + \mathcal{O}(1) \le K(\bar{x}_{U_1}, \bar{y}_{U_3}|\bar{x}_{U_2}, \bar{y}_{U_4}) + \delta_1.$$

Если теперь предположить, что для некоторых V_1, V_2, W_1, W_2

$$K(\bar{x}'_{V_1}, \bar{y}_{W_1}|\bar{x}'_{V_2}, \bar{y}_{W_2}, z) < K(\bar{x}_{V_1}\bar{y}_{W_1}|\bar{x}_{V_2}, \bar{y}_{W_2}) - \delta - D\log N,$$

то (аналогично рассуждению в доказательстве леммы 2) получаем

$$K(\bar{x}', \bar{y}|z) < K(\bar{x}, \bar{y}) - \delta + 2\delta_1 - D\log N + \mathcal{O}(\log N).$$

Положив $\delta = 2\delta_1 + \delta_2$, получаем противоречие с условием леммы при достаточно большой константе D. ▲

Доказательство леммы 6: Зафиксируем алгоритм, которые получает на вход числа α, β, γ и перечисляет список *всех* (α, β, γ)-полугроздей. Будем называть этот алгоритм *стандартным перечислителем*. Хотя число полугроздей (для любой тройки параметров) конечно, стандартный перечислитель никогда не останавливается, поскольку у нас нет эффективного средства решить, найдены ли к данному моменту все существующие полугрозди с заданными параметрами. Мы можем лишь гарантировать, что каждая полугроздь будет рано или поздно обнаружена и включена в список.

Далее мы опишем другой перечислитель, который выбирает некоторую подпоследовательность из списка всех полугроздей, выдаваемого стандартным перечислителем. Для этого мы запускаем *стандартный перечислитель*, последовательно рассматриваем порождаемые им полугрозди, и *выбираем* некоторые из них следующим образом. Предположим, что полугрозди $U_0, \ldots, U_s$ уже *выбраны*; пусть теперь стандартный перечислитель находит очередную полугроздь V. Обозначим $\epsilon = 2(\beta - \alpha + 2)$. Если $|V \cap U_i| < 2^{\beta-\epsilon}$ для всех $i = 0, \ldots, s$, то мы *выбираем* данную полугроздь и полагаем $U_{s+1} = V$. В противном случае мы пропускаем V и ожидаем появления следующей полугрозди в стандартном перечислении.

Пусть $U_0, \ldots, U_q$ есть список всех полугроздей, которые выбираются описанной процедурой (при заданных значениях параметров α, β, γ) в порядке их обнаружения стандартным перечислителем. Из конструкции очевидно, что для каждой полугрозди V выполнено либо $V = U_i$, либо хотя бы $|V \cap U_i| \ge 2^{\beta-\epsilon}$ (для некоторого $i \le q$). Также из конструкции следует, что $|U_i \cap U_j| < 2^{\beta-\varepsilon}$ для любых двух *выбранных* полугроздей U_i, U_j. Остаётся доказать, что q не слишком велико.

Достаточно показать, что каждая точка x принадлежит менее чем $2^{\beta-\alpha+2}$ выбранным полугроздям. Действительно, существует не более чем 2^γ слов x таких, что $K(x) < \gamma$. Если каждое слово x принадлежит не более, чем $2^{\beta-\alpha+2}$ выбранным полугроздям, и каждая полугроздь U_i содержит не больше 2^α слов, то общее число выбранных полугроздей не может быть больше

$$\frac{2^\gamma \cdot 2^{\beta-\alpha+2}}{2^\alpha} = 2^{\beta+\gamma-2\alpha+2}$$

Итак, остаётся оценить сверху число выбранных полугроздей, содержащих одну фиксированную точку x.

Предположим, что найдётся $N = 2^{\beta-\alpha+2}$ различных выбранных полугроздей U_i, содержащих некотое слово фиксированное x. Обозначим

$$U_i' = U_i \cap \{y \mid K(y|x) < \beta\}$$

для всех этих полугроздей U_i. Из определения полугрозди следует, что каждое из U_i' содержит не менее $2^{\alpha-1}$ элементов. С одной стороны, мы имеем

$$\left|\bigcup U_i'\right| \leq |\{y \mid K(y|x) < \beta\}| < 2^\beta$$

С другой стороны,

$$\left|\bigcup U_i'\right| \geq \sum_i |U_i'| - \sum_{i<j} \left|U_i' \cap U_j'\right|$$

Из $|U_i'| \geq 2^{\alpha-1}$, и $\left|U_i' \cap U_j'\right| \leq |U_i \cap U_j| \leq 2^{\beta-\varepsilon}$, следует

$$\left|\bigcup U_i'\right| \geq N \cdot 2^{\alpha-1} - N^2 \cdot 2^{\beta-\epsilon} = 2^\beta,$$

и мы получаем противоречие. ▲