МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.ЛОМОНОСОВА
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи
УДК 519.722+510.5

Ромащенко
Андрей Евгеньевич

НЕРАВЕНСТВА ДЛЯ КОЛМОГОРОВСКОЙ СЛОЖНОСТИ И ОБЩАЯ ИНФОРМАЦИЯ

Специальность 01.01.06 – математическая логика, алгебра и теория чисел

ДИССЕРТАЦИЯ
на соискание ученой степени кандидата физико-математических наук

Научный руководитель:
проф., д.ф.-м.н. Н. К. Верещагин

МОСКВА — 2000

Оглавление

Введение

Актуальность темы. Одна из первых работ А. Н. Колмогорова по теории алгоритмической сложности называлась «Три подхода к определению понятия "количество информации"». Двумя из трех рассмотренных подходов были энтропия Шеннона и алгоритмическая (или колмогоровская) сложность. В этой, а также в нескольких последующих работах ([3, 4]) Колмогоров указывал на связь данных понятий и параллелизм их свойств. Один из простейших примеров параллелизма свойств шенноновской энтропии и колмогоровской сложности – симметричность шенноновской взаимной информации и симметричность взаимной информации по Колмогорову. Другим, нетривиальным примером является аналогичность свойств двух родственных понятий – введенных П. Гачем и Я. Кёрнером общей информации пары случайных величин и общей информации пары слов [9]. Гач и Кёрнер исследовали свойства общей информации пар случайных величин и пар слов методами теории вероятностей. В ряде других работ [5, 13, 15] свойства общей информации пар слов изучались алгоритмическими методами.

Многие естественные свойства колмогоровской сложности и шенноновской энтропии формулируются с помощью линейных неравенств. Ряд нетривиальных неравенств для колмогоровской сложности, а также их приложения изучались в [10, 11]. В [16] рассматривалась связь между выразимыми с помощью линейных неравенств свойствами колмогоровской сложности и шенноновской энтропии.

Таким образом, к различным вопросам теории информации разработаны вероятностный и алгоритмический подходы. Многие параллельные понятия из классической и алгоритмической теории информации имеют аналогичные свойства. Изучение данного параллелизма является важной задачей, находящейся на границе двух дисциплин – теории вероятностей и математической логики.

Целью данной работы является изучение классов линейных нера-

венств, выполняющихся для колмогоровских сложностей произвольных наборов слов и для шенноновских энтропий произвольных случайных величин, а также усиление результатов Гача и Кёрнера [9] и Ан. А. Мучника [5, 13], касающихся алгоритмического варианта понятия общей информации.

Методы исследования. В работе применяются методы теории алгоритмов и теории вероятностей.

Научная новизна. Все основные результаты работы являются новыми и состоят в следующем:

1) Доказано совпадение классов линейных неравенств, выполняющихся для колмогоровской сложности слов и шенноновской энтропии дискретных случайных величин.

2) Показано, что неравенство Инглетона не выполняется для колмогоровской сложности и шенноновской энтропии.

3) Доказано, что определенные в [15] отношения условной простоты на последовательностях слов задают частичные порядки, являющиеся верхними полурешетками, но не являющиеся нижними полурешетками.

4) Построены два новых класса примеров пар слов, имеющих большую взаимную информацию и нулевую общую информацию. Получен положительный ответ на вопрос Ан. А. Мучника [13] о возможности дополнить любое слово до пары с большой взаимной и нулевой общей информацией.

Приложения. Работа носит теоретический характер. Полученные результаты относятся к теории колмогоровской сложности и могут применяться в классической и алгоритмической теории информации.

Апробация работы. Результаты диссертации докладывались на Научно-исследовательском семинаре по математической логике в МГУ (руководители академик РАН проф. С. И. Адян и проф. В. А. Успенский), а также на Колмогоровском семинаре кафедры математической логики и теории алгоритмов мех.-мат. факультета МГУ (руководители проф. Н. К. Верещагин, д. ф.-м. н. А. Л. Семенов и к. ф.-м. н. А. Х. Шень).

Публикации. Основные результаты диссертации опубликованы в работах [15, 16, 17, 18].

Для полноты изложения в диссертации приводятся теоремы 2 и 3, не принадлежащие автору. Данные результаты доказаны Н. К. Верещагиным, А. Х. Шенем и Д. Хаммером (и опубликованы в совместной работе [16]). Исследуемая в диссертации формализация отношения условной простоты была предложена А. Х. Шенем.

Структура работы. Работа состоит из введения, 4 глав и списка литературы, содержащего 18 наименований. Общий объем диссертации – 88 страниц.

Опишем кратко содержание работы. Во введении дан краткий обзор работ, связанных с темой диссертации, и сформулированы основные результаты диссертации.

Колмогоровская, или алгоритмическая, сложность слов была введена в 1965 году в работе А. Н. Колмогорова [2]. Пусть x и y – двоичные слова (конечные последовательности нулей и единиц). Неформально колмогоровскую сложность y относительно x можно определить как длину самой короткой программы, которая получает на вход слово x и выдает слово y. Договоримся называть способом программирования любую частичную вычислимую функцию двух аргументов φ (аргументы и значения φ – двоичные слова). Значение $\varphi(p,x)$ будем интерпретировать как результат работы программы p на входе x. Если $\varphi(p,x)$ не определено, то будем говорить, что при способе программирования φ программа p на входе x не выдает никакого результата. Теперь мы можем дать определение колмогоровской сложности слова y относительно слова x при способе программирования φ:

$$K_\varphi(y|x) = \begin{cases} \min\limits_{\varphi(p,x)=y} |p| & , \quad \text{если существуют такие } p, \text{ что } \varphi(p,x)=y \\ \infty & , \quad \text{если нет такого } p, \text{ что } \varphi(p,x)=y \end{cases}$$

Разумеется, определяемая таким образом сложность зависит от способа программирования. Важным открытием Колмогорова было существование *оптимального* способа программирования. Сформулируем данное утверждение точно. Будем говорить, что способ программирования ψ не хуже, чем способ программирования φ, если

$$\exists C\ \forall x,y\ K_\psi(y|x) \le K_\varphi(y|x) + C.$$

Согласно [2, Основная Теорема] существует оптимальный способ программирования, т. е. способ программирования φ_0, который не хуже любого другого.

Оптимальный способ программирования не единственен. Если φ_0 и ψ_0 – два оптимальных способа программирования, то они отличаются лишь на ограниченную величину:

$$\exists C\ \forall x, y |K_{\psi_0}(y|x) - K_{\varphi_0}(y|x)| \leq C,$$

т. е. асимптотически колмогоровские сложности относительно разных оптимальный способов программирования ведут себя одинаково. Некоторые оптимальные способы программирования могут иметь те или иные особые свойства. Однако в данной работе мы не будем интересоваться различиями между оптимальными способами программирования. Поэтому мы можем зафиксировать любой из них (назовем его φ_0). В дальнейшем будем рассматривать только колмогоровскую сложность относительно выбранного оптимального способа программирования: $K_{\varphi_0}(y|x)$. При этом мы будем опускать индекс φ_0, т. е. будем пользоваться обозначением

$$K(y|x) := K_{\varphi_0}(y|x).$$

Величину $K(y|x)$ будем называть *колмогоровской сложностью слова y относительно слова x*. Заметим, что для любых x, y условная сложность $K(y|x)$ конечна.

Колмогоровской сложностью слова x назовем сложность x относительно пустого слова. Колмогоровскую сложность x будем обозначать $K(x)$:

$$K(x) := K(x|\epsilon).$$

Кроме колмогоровской сложности слов будем рассматривать колмогоровскую сложность кортежей слов. Для этого мы должны выбрать некоторую вычислимую нумерацию всех кортежей слов, т. е. вычислимую биекцию между множеством всех кортежей двоичных слов и множеством двоичных слов. В выборе такой нумерации, так же как и в выборе оптимального способа программирования, имеется значительный произвол. Зафиксировав некоторую нумерацию кортежей, назовем колмогоровской сложностью кортежа слов $\langle y_1, \ldots, y_n \rangle$ относительно кортежа слов $\langle x_1, \ldots, x_m \rangle$ колмогоровскую сложность *номера* первого кортежа относительно *номера* второго кортежа. Данную сложность мы будем обозначать $K(y_1, \ldots, y_n | x_1, \ldots, x_m)$. Ясно, что смена нумерации кортежей изменит колмогоровские сложности кортежей лишь на ограниченную величину.

Заметим, что для любого конечного алфавита $\mathcal{A}$ можно определить колмогоровскую сложность слов в данном алфавите (а также колмогоровскую сложность кортежей слов в данном алфавите), выбрав произвольную нумерацию букв алфавита $\mathcal{A}$. Аналогично, можно говорить о колмогоровской сложности произвольных конструктивных объектов: конечных графов, элементов конечных полей, подпространств в конечномерном линейном пространстве над конечным полем и т. д., если выбрать вычислимую нумерацию данных объектов. При этом изменение нумерации приведет к изменению колмогоровских сложностей рассматриваемых объектов лишь на ограниченную величину. Говоря о колмогоровской сложности тех или иных конструктивных объектов, мы всегда будем подразумевать, что для них зафиксирована некоторая вычислимая нумерация.

В данной работе рассматриваются соотношения колмогоровских сложностей кортежей и составляющих их слов. Простейшие соотношения такого рода изучались в самых первых работах по колмогоровской сложности. Так, в [1] показано, что сложность пары слов выражается через сложность первого слова в паре и условную сложность второго:

$$K(x,y) = K(x) + K(y|x) + \mathcal{O}(\log(|x| + |y|)). \tag{0.1}$$

Важно отметить, что логарифмическое слагаемое в равенстве (0.1) нельзя заменить на член меньшего порядка[1].

Взаимная информация слов x и y определяется как разность между колмогоровской сложностью слова y и условной сложностью слова y относительно слова x:

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

Таким образом, взаимная информация x и y показывает, насколько знание слова x упрощает задачу нахождения y. Из (0.1) и определения взаимной информации следует равенство

$$I(x:y) = K(x) + K(y) - K(x,y) + \mathcal{O}(\log(|x| + |y|)). \tag{0.2}$$

Из (0.2) видно, что взаимная информация пары слов симметрична (с точностью до логарифического слагаемого), т. е.

$$|I(x:y) - I(y:x)| = \mathcal{O}(\log(|x| + |y|)). \tag{0.3}$$

[1]Большинство двоичных слов длины n имеют колмогоровскую сложность, близкую к n. Поэтому поправка $\mathcal{O}(\log(|x| + |y|))$ в (0.1) невелика.

Как показано в [1], в равенствах (0.2) и (0.3), как и в (0.1), нельзя избавиться от логарифмического слагаемого. Также лишь с точностью до $\mathcal{O}(\log(|x|+|y|))$ выполнено неравенство, выражающее свойство субаддитивности

$$K(x,y) \leq K(x) + K(y) + \mathcal{O}(\log(|x|+|y|)). \tag{0.4}$$

Отметим, что свойство монотонности

$$K(x) \leq K(x,y) + \mathcal{O}(1) \tag{0.5}$$

выполнено с точностью аддитивной константы.

Таким образом, многие простейшие и наиболее естественные неравенства для колмогоровских сложностей выполнены только с точностью до аддитивного логарифмического члена. В данной работе мы всегда будем рассматривать соотношения между колмогоровскими сложностями слов с точностью до логарифмического слагаемого.

Отметим, что кроме определенной выше колмогоровской сложности (называемой также *простой колмогоровской сложностью*), используются родственные ей понятия *префиксной*, *монотонной* сложности, сложности *разрешения* и *априорной энтропии*. Эти варианты сложностей имеют различные специальные свойства [14]. Однако для каждого слова x все перечисленные выше виды сложностей имеют значения, отличающиеся лишь на $\mathcal{O}(\log|x|)$. Поэтому, поскольку все дальнейшие утверждения мы будем формулировать «с точностью до логарифма», результаты данной работе могут быть перенесены на все варианты колмогоровской сложности.

Одна из первых работ Колмогорова по алгоритмической сложности называлась «Три подхода к определению понятия “количество информации”». Двумя из трех рассмотренных в статье подходов были энтропия Шеннона и алгоритмическая сложность. Колмогоров показал, что алгоритмический подход является уточнением вероятностного, и обратил внимание на параллелизм свойств энтропий случайных величин и алгоритмической сложности слов. Простейшие примеры данного параллелизма – соотношения (0.1), (0.2), (0.3) для колмогоровской сложности и аналогичные им соотношения

$$H(\alpha,\beta) = H(\alpha) + H(\alpha|\beta) \tag{0.6}$$

$$\mathcal{I}(\alpha:\beta) = H(\alpha) + H(\beta) - H(\alpha,\beta) \tag{0.7}$$

$$\mathcal{I}(\alpha:\beta) = \mathcal{I}(\beta:\alpha) \tag{0.8}$$

для шенноновской энтропии. Для шенноновских энтропий выполнены также свойства монотонности

$$H(\alpha) \leq H(\alpha, \beta) \tag{0.9}$$

и субаддитивности

$$H(\alpha, \beta) \leq H(\alpha) + H(\beta). \tag{0.10}$$

Некоторые более сложные неравенства для колмогоровской сложности, а также их применения рассматривались в [10]. Отметим, что для всех неравенств для колмогоровской сложности, рассмотренных в [10], нетрудно доказать шенноновские аналоги (аналогичные неравенства, выполненные для шенноновской энтропии).

В данной работе мы докажем, что все свойства, выразимые с помощью линейных равенств и неравенств, одинаковы для колмогоровской сложности и шенноновской энтропии, т. е. классы линейных неравенств, выполненных для колмогоровских сложностей и для шенноновских энтропий, совпадают. Также мы покажем, что существует линейное неравенство (неравенство Инглетона), выполненное для размерностей линейных подпространств, но не выполненное для шенноновской энтропии и колмогоровской сложности.

Еще одним примером параллелизма между классической и алгоритмической теорией информации является понятие общей информации. В [9] Гач и Кёрнер определили *общую информацию* пары случайных величин, а также дали в терминах колмогоровской сложности неформальное определение аналогичного понятия для пары слов. Интуитивно величина общей информации двух объектов (двух случайных величин или двух слов) есть размер их «общей части», т. е. максимум информации, которую можно выделить одновременно из обоих объектов (см. также [6], где понятие общей информации пары случайных величин обсуждается с точки зрения теории кодирования). Легко показать, что как в шенноновском, так и в колмогоровском случае величина общей информации не превосходит взаимной информации объектов. Основным результатом работы [9] было доказательство того, что общая информация случайных величин может быть намного меньше взаимной информации. При этом Гач и Кёрнер указали эффективный способ вычисления общей информации пары случайных величин по их совместному распределению. Также в [9] было построено семейство примеров пар слов, у которых

общая информация намного меньше их взаимной информации. Доказательство Гача и Кёрнера использовало методы теории вероятностей и было довольно сложным.

Ан. А. Мучник [5, 13] предложил более простое и использующее только алгоритмические методы доказательство существования пар слов, у которых общая информация много меньше взаимной. В [13] был поставлен вопрос: для всякого ли слова x существует такое y, что взаимная информация x и y имеет заданную величину, а общая информация нулевая? В данной работе мы получим положительный ответ на вопрос Мучника.

Вернемся к вопросу о строгом определении понятия общей информации пары слов. Неформально мы говорим, что x и y имеют u битов общей информации, если есть слово z сложности u, которое просто относительно x и относительно y. При попытке дать точное определение возникают технические трудности. Дело в том, что для определения общей информации требуется формализовать интуитивное отношение «слово z просто относительно слова x». Однако кажется невозможным отделить слова «сложные» относительно x от слов «простых» относительно x. Чтобы преодолеть это затруднение, в [15] было предложено перейти от рассмотрения индивидуальных слов к бесконечным последовательностям слов. На множестве последовательностей слов можно определить отношение «условной простоты». А именно, рассмотрим последовательности слов $x_1, x_2, \ldots$ и $y_1, y_2, \ldots$, предполагая $|x_n| = \mathcal{O}(n)$ и $|y_n| = \mathcal{O}(n)$. Будем говорить, что последовательность $y_1, y_2, \ldots$ проста относительно последовательности $x_1, x_2, \ldots$, если скорость роста условной сложности $K(y_n|x_n)$ асимптотически мала по сравнению с n. Возможны различные варианты уточнения понятия асимптотической малости. Наиболее естественным кажется говорить, что $K(y_n|x_n) = \mathcal{O}(\log n)$. Но возможны и другие варианты формализации отношения условной простоты: для любой функция $f(n)$ такой, что

$$f(n) = \Omega(\log n),\ \ f(n) = o(n) \text{ при } n \to \infty,$$

можно определить простоту $\{y_n\}$ относительно $\{x_n\}$ условием

$$K(y_n|x_n) = \mathcal{O}(f(n)).$$

Таким образом, мы определили семейство отношений частичного предпорядка на пространстве последовательностей слов. После факториза-

ции по соответствующим отношениям эквивалентности получаем семейство частичных порядков. В данной работе мы покажем, что все указанные частично упорядоченные множества являются верхними полурешетками, но не являются нижними полурешетками.

Определяемое семейство полурешеток можно рассматривать как финитный аналог тьюринговых степеней неразрешимости, и изучение их алгебраических свойств кажется интересной самостоятельной задачей. Некоторые свойства данных полурешеток рассматривались в [15]. Но для нас эти полурешетки важны прежде всего как инструмент для формализации понятия общей информации. Зафиксировав один из указанных вариантов определения условной простоты, можно определить общую информацию для пары *последовательностей* слов как их точную нижнюю грань относительно введенного отношения предпорядка. Мы докажем существование пар последовательностей с большой взаимной информацией и нулевой точной нижней гранью, что в терминологии Гача и Кёрнера соответствует существованию пар слов с большой взаимной информацией и нулевой общей информацией. Мы рассмотрим два семейства примеров пар, имеющих большую взаимную и нулевую общую информацию. Первое из рассматриваемых нами семейств примеров было предложено в [9]. Мы приведем для него новое, более простое доказательство того, что величина общей информации много меньше, чем взаимная информация; при этом мы получим более сильную оценку на величину общей информации. Второе семейство рассматриваемых нами примеров является обобщением конструкции из работы [13]. Оба рассматриваемых класса примеров позволят нам дать положительный ответ на вопрос Ан. А. Мучника.

Далее мы кратко опишем содержание глав 1-4.

В **главе 1** мы дадим несколько определений и докажем основные технические утверждения, необходимые для получения результатов, связывающих колмогоровскую сложность и энтропию Шеннона.

Пусть случайная величина φ принимает значения $a_1, \ldots, a_m$ с вероятностями $p_1, \ldots, p_m$. Рассмотрим n копий случайных величин φ. Точнее, пусть $\varphi_1, \ldots, \varphi_n$ независимы и каждая из них распределена как φ. Будем интересоваться «типичными» значениями n-ки случайных величин $\langle \varphi_1, \ldots, \varphi_n \rangle$. Значениями данного кортежа случайных величин являются слова длины n в алфавите $a_1 \ldots, a_m$. По закону больших чисел с близкой к единице вероятностью доли букв a_i в этом слове будут близки к веро-

ятностям p_i. Заметим, что согласно закону больших чисел среди значений φ_j доля букв a_i с большой вероятностью будет равно $p_i \cdot n + \mathcal{O}(\sqrt{n})$, т. е. уклонение частот от вероятностей обратно пропорционально корню из n. Для нас будут интересны слова, в которых частоты букв уклоняются от вероятностей меньше, чем можно ожидать в соответствии с законом больших чисел. Будем говорить, что слово длины n *типично* относительно распределения φ, если оно содержит $p_i \cdot n + \mathcal{O}(1)$ букв a_i $(i = 1, \ldots, m)$. Чтобы придать точный смысл малому члену $\mathcal{O}(1)$, мы должны определять типичность не для индивидуальных слов, а для бесконечных последовательностей.

Аналогичное определение имеет смысл также и для многомерных случайных величин φ. Дадим формальное определение сразу для многомерного случая.

Определение. *Пусть случайные величины $\varphi^1, \varphi^2, \ldots, \varphi^k$ принимают значения в конечных алфавитах $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ и имеют совместное распределение*

$$P(a^1, a^2, \ldots, a^k) = \mathrm{Prob}[\varphi^1 = a^1, \varphi^2 = a^2, \ldots, \varphi^k = a^k].$$

Назовем кортеж $\langle \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k \rangle$, состоящий из k бесконечных последовательностей слов, P-типичным, если каждое слово x_n^i имеет длину n, и для любого набора значений $\langle a^1, a^2, \ldots, a^k \rangle$ число таких позиций i, что в каждом слове x_n^j на i-м месте стоит буква a^j, равно $nP(a^1, a^2, \ldots, a^k) + \mathcal{O}(1)$.

Важным свойством P-типичных последовательностей является связь колмогоровской сложности слов таких последовательности и шенноновской энтропии распределения P. Обратимся к одномерному случаю: пусть P – распределение случайной величины φ, а $\mathbf{x} = x_1, x_2, \ldots$ – типичная относительно распределения P последовательность слов. Тогда скорость роста колмогоровской сложности x_n ограничена энтропией φ:

$$K(x_n) \leq nH(\varphi) + \mathcal{O}(\log n).$$

Заметим, что если бы мы считали типичными слова, у которых частоты букв уклоняются от вероятностей на величину $\mathcal{O}(1/\sqrt{n})$, то имела бы место более грубая оценка $K(x_n) \leq nH(\varphi) + \mathcal{O}(\sqrt{n})$.

Указанная связь между шенноновской энтропией и колмогоровской сложностью слов с соответствующими частотами букв была отмечена в статье Колмогорова [4].

Нам потребуется очевидное обобщение данного утверждения на многомерный случай.

Утверждение 1. *Если случайные величины $\varphi^1, \varphi^2, \ldots, \varphi^k$ имеют некоторое совместное распределение P, то для любого P-типичного кортежа последовательностей $\langle \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k \rangle$, любого непустого набора индексов $V \subseteq \{1, 2, \ldots, k\}$ и любого (быть может пустого) набора индексов $W \subseteq \{1, 2, \ldots, k\}$ выполнено неравенство*

$$K(x_n^V | x_n^W) \leq nH(\varphi^V | \varphi^W) + \mathcal{O}(\log n) \tag{0.11}$$

(В случае, когда набор индексов W пуст, условная колмогоровская сложность и условная шенноновская энтропия в равенстве (0.11) *обращаются в безусловные.)*

Таким образом, мы определили типичность кортежа последовательностей слов относительно совместно распределенных случайных величин. При этом скорости роста колмогоровских сложностей слов данной последовательности не превосходят соответствующих шенноновских энтропий случайных величин. Наш следующий шаг – определение *случайных* кортежей относительно данного распределения. Для случайного (относительно данного распределения вероятностей) кортежа последовательностей слов скорости роста колмогоровских сложностей *равны* соответствующим шенноновским энтропиям.

Определение. *Будем называть P-типичный кортеж последовательностей слов $\langle \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k \rangle$ P-случайным, если для любого непустого набора индексов $V \subseteq \{1, 2 \ldots, k\}$ и любого (быть может пустого) набора индексов $W \subseteq \{1, 2 \ldots, k\}$ выполнено равенство*

$$K(x_n^V | x_n^W) = nH(\varphi^V | \varphi^W) + \mathcal{O}(\log n). \tag{0.12}$$

Для одномерного случая существование P-случайных кортежей было показано в [4]. Доказательство в общем случай является простым обобщением и использует те же комбинаторные рассуждения.

Кроме факта существования P-случайных кортежей для любого распределения P нам потребуется следующее более общее утверждение: пусть P' является проекцией r-мерного распределения P на первые s координат, и задан P'-случайный кортеж $\mathbf{x}^1, \ldots, \mathbf{x}^s$; тогда данный кортеж можно расширить до P-случайного кортежа $\mathbf{x}^1, \ldots, \mathbf{x}^s, \mathbf{x}^{s+1}, \ldots, \mathbf{x}^r$.

Утверждение 2. *Пусть P – совместное распределение r случайных величин, P' – проекция распределения P на первые s координат ($s < r$), и кортеж $\langle \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^s \rangle$ является P'-случайным. Тогда существуют такие последовательности $\mathbf{x}^{s+1}, \mathbf{x}^{s+2}, \ldots, \mathbf{x}^r$, что кортеж $\langle \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^r \rangle$ является P-случайным.*

Отметим, что существование P-случайных кортежей является следствием возможности расширять P'-случайные кортежи до P-случайных. Действительно, достаточно взять $s = 0$; тогда P'-типичный кортеж является пустым, и любой P-случайный кортеж расширяет его.

Таким образом, в главе 1 мы даем определение P-типичных и P-случайных кортежей последовательностей слов и доказываем их основные свойства. Особенностью наших определений является то, что мы требуем большей близости частот к вероятностям, чем может обеспечить закон больших чисел. Это позволяет нам получить более точные оценки для колмогоровских сложностей слов типичных и случайных последовательностей. Результаты главы 1 мы будем использовать в главе 2 и в главе 4.

В **главе 2** мы рассмотрим линейные неравенства, выполненные для колмогоровских сложностей слов. Мы докажем, что класс неравенств, выполненных для колмогоровских сложностей слов, совпадает с классом неравенств, выполненных для шенноновской энтропии. Кроме того, мы докажем, что неравенство Инглетона не выполнено для колмогоровской сложности и энтропии Шеннона.

Пусть даны слова $x_1, x_2, \ldots, x_n$. Тогда мы можем рассмотреть колмогоровские сложности данных слов ($K(x_1)$, $K(x_2)$, ...), колмогоровские сложности пар, составленных из данных слов ($K(x_1, x_2)$, $K(x_1, x_3)$, ...), колмогоровские сложности троек слов и т. д. Всего из n слов можно выбрать $(2^n - 1)$ подмножеств (не считая пустого). Колмогоровские сложности кортежей, отличающихся только порядком членов, равны с точностью до $\mathcal{O}(1)$ (константа определяется длиной кортежа, но не зависит от длины слов, входящих в кортеж). Таким образом, для n-ки слов можно ограничиться рассмотрением колмогоровских сложностей $(2^n - 1)$ кортежей, составленных из данных слов. Будем интересоваться классом линейных неравенств, выполненных для данных сложностей независимо от выбора слов $x_1, \ldots, x_n$.

Запишем общий вид линейного неравенства для колмогоровских

сложностей троек слов.

$$\lambda_1 K(x_1) + \lambda_2 K(x_2) + \lambda_3 K(x_3) + \lambda_{12} K(x_1, x_2) + \lambda_{23} K(x_2, x_3) + \\ + \lambda_{13} K(x_1, x_3) + \lambda_{123} K(x_1, x_2, x_3) \geq \mathcal{O}(\log(|x_1| + |x_2| + |x_n|))$$

Более формально, мы будем говорить, что неравенство с коэффициентами $\lambda_1, \lambda_2, \lambda_3, \lambda_{12}, \lambda_{23}, \lambda_{13}, \lambda_{123}$ выполнено для колмогоровских сложностей троек слов, если существует такая константа C, что для любых слов x_1, x_2, x_3

$$\lambda_1 K(x_1) + \lambda_2 K(x_2) + \lambda_3 K(x_3) \lambda_{12} K(x_1, x_2) + \lambda_{23} K(x_2, x_3) + \\ + \lambda_{13} K(x_1, x_3) + \lambda_{123} K(x_1 x_2 x_3) \geq -C \log(|x_1| + |x_2| + |x_n|) - C.$$

Аналогично можно записать общий вид неравенства и для произвольной n-ки слов. Неравенство для колмогоровских сложностей n-ки слов будет содержать $(2^n - 1)$ слагаемых в левой части и логарифмический член в правой.

В правой части всех рассматриваемых неравенств для колмогоровских сложностей будет стоять логарифм суммы длин слов, входящих в неравенства. Кажется естественным рассматривать неравенства для колмогоровской сложности именно «с точностью до логарифма», поскольку такие важные соотношения как (0.1), (0.4) выполняются с точностью до аддитивного логарифмического члена.

Мы не ограничиваем общности, не рассматривая неравенств, включающих условные сложности, поскольку ввиду (0.1) условные колмогоровские сложности можно заменить на линейную комбинацию безусловных сложностей.

Простейшими примерами линейных неравенств для колмогоровских сложностей являются (0.5) и (0.4). Рассмотрим пример неравенства для тройки слов. Прежде всего запишем релятивизованный аналог неравенства (0.4).

$$K(y, z|x) \leq K(y|x) + K(z|x) + \mathcal{O}(\log(|x| + |y| + |z|))$$

Прибавляя к обеим частям неравенства $2K(x)$ и учитывая (0.1), получим

$$K(x) + K(x, y, z) \leq K(x, y) + K(x, z) + \mathcal{O}(\log(|x| + |y| + |z|)). \quad (0.13)$$

Очевидно, неравенство (0.13) останется верным, если в нем слова x, y, z заменить на произвольные кортежи слов X, Y, Z. Неравенства, которые можно получить из (0.13) подстановкой некоторых кортежей вместо слов x, y и z, будем называть *базисными* неравенствами.

Рассмотрим также класс линейных неравенств, выполненных для шенноновской энтропии. Будем рассматривать случайные величины с конечной областью значений. Неравенство для n-ки случайных величин $\varphi_1 \ldots, \varphi_n$ может содержать энтропии каждой из данных случайных величин $H(\varphi_1)$,..., $H(\varphi_n)$, энтропии пар $H(\varphi_1, \varphi_2)$,... и т. д. Общий вид неравенства для тройки случайных величин можно записать как

$$\lambda_1 H(\varphi_1) + \lambda_2 H(\varphi_2) + \lambda_3 H(\varphi_3) + \lambda_{12} H(\varphi_1, \varphi_2) + \\ + \lambda_{23} H(\varphi_2, \varphi_3) + \lambda_{13} H(\varphi_1, \varphi_3) + \lambda_{123} H(\varphi_1, \varphi_2, \varphi_3) \geq 0$$

Аналогично записывается общий вид неравенства для произвольной n-ки случайных величин. Мы интересуемся классом неравенств указанного вида, выполненных для произвольной n-ки совместно распределенных случайных величин.

Простейшие примеры неравенств для энтропии Шеннона – неравенства (0.9) и (0.10). Нетрудно также проверить, что верен и аналог неравенства (0.13)

$$H(\varphi_1) + H(\varphi_1, \varphi_2, \varphi_3) \leq H(\varphi_1, \varphi_2) + H(\varphi_1, \varphi_3). \qquad (0.14)$$

Так же как и в колмогоровском случае, можно определить *базисные* неравенства как результат подстановки произвольных кортежей случайных величин вместо φ_1, φ_2 и φ_3 в неравенстве (0.14).

Наконец, также можно рассмотреть класс линейных неравенств, выполняющихся для размерностей любых линейных подпространств в конечномерных пространствах над $\mathbb{R}$ и над конечными полями.

Таким образом, для каждого n мы рассматриваем четыре класса линейных неравенств:

- неравенства, выполненные для колмогоровских сложностей n-ок слов,
- неравенства, выполненные для шенноновских энтропий n-ок случайных величин,
- положительные линейные комбинации базисных неравенств,
- неравенства, выполненные для размерностей n-ок подпространств в конечномерных линейных пространствах.

Верещагин, Хаммер и Шень показали в [16] (см. также [11]), что для $n = 3$ все четыре класса неравенств совпадают. Кроме того, очевидно, что

для любого n третий класс содержится в первом и во втором. В данной работе мы покажем, что первые два класса неравенств совпадают при любом n, а четвертый класс строго больше первых двух при $n \geq 4$.

Сформулируем полученные результаты более точно.

- Классы неравенств, выполненных для колмогоровской сложности и для шенноновской энтропии, совпадают для всех n. Заметим, что данный результат соответствует замечанию А. Н. Колмогорова о параллелизме свойств алгоритмической и шенноновской энтропий.

- Для $n = 4$ существуют неравенства, выполненные для размерностей подпространств, но неверные для колмогоровской сложности и шенноновской энтропии. Таким образом, классы неравенств для шенноновской энтропии и колмогоровской сложности являются собственным подмножеством класса неравенств, выполненных для размерностей подпространств.

Обсудим последнее утверждение более подробно. Чтобы показать, что класс неравенств, выполненных для размерностей подпространств, строго больше класса неравенств, выполненных для колмогоровской сложности и шенноновской энтропии, мы укажем неравенство, верное для размерностей любой четверки подпространств, но нарушающееся для некоторых четверок случайных величин.

Примером такого неравенства является неравенство Инглетона

$$\dim(a) + \dim(b) + \dim(c \oplus d) + \dim(a \oplus b \oplus c) + \dim(a \oplus b \oplus d) \leq \\ \dim(a \oplus b) + \dim(a \oplus c) + \dim(a \oplus d) + \dim(b \oplus c) + \dim(b \oplus d), \quad (0.15)$$

которое выполнено для любых подпространств a, b, c, d [12]. Мы предъявим распределение вероятностей четверки случайных величин, для энтропий которых аналог неравенства Инглетона

$$H(\alpha) + H(\beta) + H(\gamma, \delta) + H(\alpha, \beta, \gamma) + H(\alpha, \beta, \delta) \leq \\ H(\alpha, \beta) + H(\alpha, \gamma) + H(\alpha, \delta) + H(\beta, \gamma) + H(\beta, \delta) \quad (0.16)$$

не выполнен.

Замечание. Неравенство Инглетона довольно громоздко. Используя обычные сокращения, (0.16) можно переписать в более кратком виде

$$\mathcal{I}(\alpha : \beta) \leq \mathcal{I}(\alpha : \beta | \gamma) + \mathcal{I}(\alpha : \beta | \delta) + \mathcal{I}(\gamma : \delta). \quad (0.17)$$

Опишем кратко план главы 2. Прежде всего мы определим формально классы неравенств, выполненных для колмогоровских сложностей n-ок слов, для шенноновских энтропий n-ок случайных величин и для размерностей n-ок линейных подпространств. Затем мы докажем теорему о совпадении первых двух классов.

Теорема 1. *Всякое линейное неравенство, выполненное для колмогоровской сложности, выполняется также и для шенноновской энтропии, и наоборот, всякое неравенство, выполненное для шенноновской энтропии, выполняется для колмогоровской сложности.*

Далее мы приведем две теоремы Верещагина, Хаммера и Шеня из [16]. Первая теорема – о соотношении классов неравенств для шенноновских энтропий и для размерностей подпространств.

Теорема 2. [16] *Всякое линейное неравенство, выполненное для шенноновской энтропии, выполняется также и для размерностей подпространств.*

Вторая теорема (мы приведем ее без доказательства) – о совпадении четырех классов неравенств для $n = 3$.

Теорема 3. [16] *Классы неравенств, выполненных для энтропий троек случайных величин, для колмогоровских сложностей троек слов и для размерностей троек подпространств в конечномерных линейных пространствах над конечными полями и над $\mathbb{R}$ совпадают с классом неотрицательных линейных комбинаций неравенств вида (0.14).*

Наконец, мы покажем, что неравенство Инглетона не выполнено для шенноновской энтропии и колмогоровской сложности. Для этого достаточно доказать следующее утверждение.

Утверждение 3. *Существует четверка случайных величин, для которых неравенство (0.17) не выполнено.*

Таким образом, для четырех рассмотренных классов линейных неравенств остается неизвестным только соотношение между классом линейных комбинаций базисных неравенств и классом неравенств, выполненных для колмогоровской сложности и шенноновской энтропии. Приведем гипотезу, сформулированную в [16].

Гипотеза. *Всякое линейное неравенство, выполненное для колмогоровских сложностей произвольной n-ки слов, является неотрицательной линейной комбинацией базисных неравенств.*

В **главе 3** мы рассмотрим предложенное А. Х. Шенем семейство отношений частичного предпорядка на последовательностях слов (см. [15]). Эти отношения являются различными вариантами формализации интуитивного отношения «слово y просто относительно слова x». Мы покажем, что каждое из рассматриваемых отношений задает структуру, являющуюся верхней полурешеткой, но не являющуюся нижней полурешеткой. Полученные верхние полурешетки могут представлять интерес как финитный аналог тьюринговых степеней неразрешимости. Кроме того, они позволяют формализовать понятие общей информации слов и представляют собой удобный инструмент для изучения свойств колмогоровской сложности «с точностью до логарифмического члена» (или с более грубой погрешностью).

Пусть α и β — бесконечные двоичные последовательности. Говорят, что последовательность β *сводится по Тьюрингу* к последовательности α, если существует такая многоленточная машина Тьюринга, которая печатает на выходной ленте последовательность β, если на входной ленте записана последовательность α. Отношение сводимости по Тьюрингу (обозначаемое $\leq_T$) рефлексивно и транзитивно. Таким образом, на множестве бесконечных двоичных последовательностей определяется предпорядок. Классы эквивалентности $((\mu \sim \nu) \leftrightarrow (\mu \leq_T \nu) \wedge (\nu \leq_T \mu))$ называют тьюринговыми степенями. Они образуют верхнюю полурешетку. Эта полурешетка подробна изучалась в теории рекурсии (см., например, [7]).

Определим финитный аналог тьюринговых степеней. Заменим бесконечные последовательности α и β двоичными словами x и y. Разумеется, для любых x и y найдется машина Тьюринга M, которая печатает слово y, получив на вход слово x. Чтобы получить нетривиальное определение, необходимо ввести ограничения на машину M. Потребуем, чтобы M была простой (точнее, чтобы ее программа была короткой по сравнению со словами x и y). Другими словами, потребуем, чтобы условная колмогоровская сложность $K(y|x)$ была мала. Неформально говоря, мы хотим ввести на двоичных словах отношение, имеющее смысл «слово y просто относительно слова x». Но что значит, что величина колмогоровской сложности мала? По-видимому, невозможно определить отно-

шение условной простоты для индивидуальных слов: нельзя провести точную границу между словами «простыми относительно x» и словами «сложными относительно x». Будем интересоваться асимптотическими свойствами и вместо индивидуальных двоичных слов рассмотрим бесконечные последовательности слов. Ограничимся последовательностями слов, длина которых растет не более чем линейно.

Определение. *Обозначим R класс всех последовательностей двоичных слов $\mathbf{x} = \{x_n\}$, для которых существует такая константа c, что $|x_n| \leq cn$.*

Далее мы будем предполагать, что все рассматриваемые последовательности принадлежат R.

Теперь нужно определить «простоту» последовательности $\mathbf{y} = \{y_n\}$ относительно последовательности $\mathbf{x} = \{x_n\}$. Может показаться разумным говорить, что $\mathbf{y}$ просто относительно $\mathbf{x}$, если для некоторой константы c и для всех n выполняется $K(y_n|x_n) \leq c$. Однако такой подход не очень интересен. Во-первых, хотелось бы считать, что если каждое y_n есть подслово соответствующего x_n, то последовательность $\mathbf{y}$ проста относительно $\mathbf{x}$. Но для того, чтобы выделить подслово, нужно указать его начало и конец. А для этого нужна дополнительная информация, логарифмически зависящая от n. Во-вторых, такие естественные свойства колмогоровской сложности, как симметричность взаимной информации (0.3) или субаддитивность (0.4) выполняются только точностью до аддитивного члена $\mathcal{O}(\log(|x|+|y|))$. Так что более естественно ограничивать условную сложность не константой, а логарифмически растущей величиной: $K(y_n|x_n) = \mathcal{O}(\log n)$. Данное отношение предпорядка было определено в [15].

Можно также рассматривать отношение «условной простоты», соответствующее более слабому ограничению на условную сложность. Например, при изучении связей между колмогоровской сложностью и энтропией Шеннона может быть полезно заменить логарифм на величину $\mathcal{O}(\sqrt{n})$. Мы рассмотрим самую общую ситуацию. Пусть дана некоторая функция $f(n) : \mathbb{N} \to \mathbb{N}$, которая растет медленнее линейной, но не медленнее логарифма:

$$f(n) = o(n) \text{ и } f(n) = \Omega(\log n) \text{ при } n \to \infty. \tag{0.18}$$

Определение. *Пусть функция $f : \mathbb{N} \to \mathbb{N}$ удовлетворяет условиям (0.18). Тогда будем говорить, что последовательность $\mathbf{y} = \{y_n\}$*

f-проста относительно последовательности $\mathbf{x} = \{x_n\}$ (*обозначаем* $\mathbf{y} \leq_f \mathbf{x}$), *если* $K(y_n|x_n) = \mathcal{O}(f(n))$.

Очевидно, отношение $\leq_f$ транзитивно и симметрично. Мы получаем семейство множеств с отношениями частичного предпорядка $\langle R, \leq_f \rangle$ для всевозможных функций f.

Определение. *Назовем* $\mathcal{S}_f$ *частично упорядоченное множество, являющееся факторизацией* $\langle R, \leq_f \rangle$ *по отношению эквивалентности* $(\mathbf{x} \sim \mathbf{y}) \leftrightarrow ((\mathbf{x} \leq_f \mathbf{y}) \wedge (\mathbf{y} \leq_f \mathbf{x}))$.

Замечание. Наиболее важен случай $f(n) = \log n$. Для краткости мы будем использовать обозначения $\mathcal{S}$ и $\langle R, \leq \rangle$ вместо $\mathcal{S}_{\log n}$ и $\langle R, \leq_{\log n} \rangle$ соответственно. Мы также будем говорить «последовательность $\mathbf{y} = \{x_n\}$ проста относительно последовательности $\mathbf{x} = \{x_n\}$» вместо «последовательность $\mathbf{y} = \{y_n\}$ $(\log n)$-проста относительно последовательности $\mathbf{x} = \{x_n\}$».

Отметим, что в S_f имеется наименьший элемент. Это класс эквивалентности, содержащий последовательность пустых слов $\Lambda = \epsilon, \epsilon, \ldots$

Мы докажем, что для любой функции f, удовлетворяющей (0.18), S_f является верхней полурешеткой, но не является решеткой. Доказательство того, что любые два элемента в S_f имеют точную верхнюю грань (т. е. S_f является верхней полурешеткой), тривиально. Тем не менее, сформулируем данный факт в виде утверждения.

Утверждение 4. *Частично упорядоченное множество* S_f *является верхней полурешеткой.*

Основным результатом данной главы является доказательство того, что некоторые пары элементов из S_f не имеют точной нижней грани. Мы рассмотрим две конструкции, позволяющие получать пары последовательностей, не имеющие точной нижней грани. Первая конструкция будет использована в доказательстве теоремы 4. Она позволит строить пары последовательностей, не имеющий точной нижней грани в S_f, при условии $f(n) = o(\sqrt{n})$.

Теорема 4. *Если* $f(n) = o(\sqrt{n})$ *и* $f(n) = \Omega(\log n)$, *то в* S_f *существуют пары элементов, не имеющие точной нижней грани.*

Вторая конструкция универсальна, т. е. она позволяет получать последовательности, не имеющие точной нижней грани, для любой S_f. Мы воспользуемся ей в доказательстве теоремы 5.

Теорема 5. *Для любой функции $f : \mathbb{N} \to \mathbb{N}$, удовлетворяющей (0.18), в S_f существует пара элементов, не имеющих точной нижней грани.*

Формально утверждение теоремы 4 является следствием теоремы 5. Но мы приводим отдельное доказательство первой, менее общей теоремы, поскольку используемая в ее доказательстве конструкция может представлять самостоятельный интерес. Кроме того, предварительное рассмотрение первой, более простой конструкции позволит сделать более ясным доказательство общей теоремы 5.

Таким образом, мы покажем, что для любого $f(n)$, удовлетворяющего (0.18), S_f является верхней полурешеткой, но не является нижней полурешеткой. В следующей главе мы будем использовать только полурешетку S, поскольку наиболее естественным кажется изучение свойств колмогоровской сложности «с точностью до логарифмического слагаемого». Заметим, однако, что все дальнейшие рассуждения могут быть перенесены на случай произвольной S_f почти дословно.

В дальнейшем мы определим общую информацию последовательностей $\mathbf{x}$, $\mathbf{y}$ из R как точную нижнюю грань для классов эквивалентности $\mathbf{x}$ и $\mathbf{y}$ в S. Как показывают теоремы 4 и 5, пара последовательностей может не иметь точной нижней грани. В главе 4 мы подробно рассмотрим другую ситуацию: будем интересоваться парами последовательностей $\mathbf{x}$, $\mathbf{y}$, которые имеют точную нижнюю грань Λ (нулевую нижнюю грань). При этом взаимная информация $I(x_n : y_n)$ может быть очень велика.

Прежде чем сформулировать основные результаты **главы 4**, приведем мотивировку введения понятия общей информации. Сначала рассмотрим простейший пример. Пусть слово x является конкатенацией слов p и r, а слово y – конкатенацией слов q и r.

$$x = pr, \quad y = qr$$

Предположим, что все три слова p, q и r выбраны случайными и независимыми, и их длины равны n. Тогда сложности слов x и y равны $2n$, а их взаимная информация равна n. Это соответствует интуиции: у слов x и y есть общая часть r, и взаимная информация равна сложности этой общей части. Слово r «материализует» взаимную информацию x и y.

Неформально можно определить *общую информацию* слов x и y как максимальную сложность такого z, что z имеет малую сложность относительно x и относительно y. В рассмотренном выше примере общая информация была равна сложности r и совпадала с величиной взаимной информации.

Нетрудно показать, что общая информация пары не может быть больше взаимной информации. Действительно, для любых слов x, y, z выполнено неравенство

$$K(z) \leq K(z|x) + K(z|y) + I(x:y) + \mathcal{O}(\log n)$$

(мы докажем данное неравенство в лемме 12). Следовательно, если условные сложности z относительно x и относительно y малы, то сложность z не может быть намного больше $I(x:y)$. Таким образом, общая информация пары с точностью до малого слагаемого не превосходит ее взаимной информации.

В [9] был рассмотрен вопрос: может ли общая информация двух слов быть много меньше взаимной? Гач и Кёрнер дали положительный ответ на данный вопрос. Однако чтобы сформулировать точное утверждение, необходимо уточнить понятие общей информации. Для этого прежде всего нужно пояснить, что значит, что некоторое слово z легко получить по слову x и по слову y (т. е. условные колмогоровские сложности $K(z|x)$ и $K(z|y)$ малы). Мы перейдем от индивидуальных слов к бесконечным последовательностям. Сформулируем результат Гача и Кёрнера в терминах последовательностей.

Теорема 6. [9] *Существуют такие последовательности слов x_n, y_n, что*

$$K(x_n) = n + o(n), \quad K(y_n) = n + o(n) \quad I(x_n : y_n) = an + o(n),$$

(a – положительная константа), и для любой последовательности слов z_n, удовлетворяющей условию

$$K(z_n|x_n) = o(n), \quad K(z_n|y_n) = o(n),$$

выполнено $K(z_n) = o(n)$.

Таким образом, существуют последовательности слов, у которых нельзя материализовать взаимную информацию. Более того, теорема 6 утверждает, что существуют такие x_n, y_n, у которых нельзя материализовать

даже часть взаимной информации x_n и y_n. Точнее, величина материализуемой взаимной информации бесконечно мала по сравнению с n.

В [9] описывается некоторый класс примеров пар $\langle x_n, y_n\rangle$, обладающих сформулированным выше свойством. При этом конструкция позволяет строить такие последовательности x_n, y_n для любых значений параметра a $(0 < a < 1)$, т. е. можно указать такие x_n и y_n, взаимная информация которых очень велика (a близко к единице), но даже ее малая часть не может быть материализована.

В работе [9] не проводилась точная оценка остаточных членов. Но, анализируя доказательство, можно проверить, что теорема 6 останется верной, если в формулировке заменить члены $o(n)$ на $\mathcal{O}(\sqrt{n})$ (или на $\mathcal{O}(f(n))$, где $f(n)$ – любая функция, растущая быстрее $\sqrt{n}$, но медленнее n: $f(n) = o(n)$, $f(n) = \Omega(\sqrt{n})$). Однако, как мы видели выше, в утверждениях о колмогоровской сложности естественно формулировать равенства с точностью до логарифмического члена. Поэтому кажется интересным рассмотреть усиление теоремы 6, а именно, доказать ее, заменив в формулировке члены $o(n)$ на $\mathcal{O}(\log n)$. Более формально, естественным усилением теоремы Гача и Кёрнера является

Теорема 7. [5, 13] *Для любой функции $f(n)$ такой, что $f(n) = o(n)$ и $f(n) = \Omega(\log n)$, существуют такие последовательности слов x_n, y_n, что*

$$K(x_n) = n + \mathcal{O}(f(n)), \quad K(y_n) = n + \mathcal{O}(f(n)), \quad I(x_n : y_n) = an + \mathcal{O}(f(n))$$

(a – некоторая положительная константа), и для любой последовательности слов z_n, которая f-проста относительно $\mathbf{x}$ *и* $\mathbf{y}$*, выполнено $K(z_n) = \mathcal{O}(f(n))$.*

В работах [5, 13] было получено доказательство теоремы 7 для произвольного значения параметра a из интервала $(0, 1)$. В [15, 16] рассматривались другие доказательства данной теоремы для некоторых специальных значений a.

Таким образом, для любого $a < 1$ можно найти такие слова x_n и y_n, что их сложности примерно равны n, взаимная информация примерно равна an, и их взаимную информацию нельзя материализовать. Ан. А. Мучником был поставлен вопрос: при каких значениях параметра a для каждого x_n сложности n можно подобрать y_n сложности n такое, что взаимная информация $I(x_n : y_n)$ примерно равна an, но ее нельзя

материализовать? Более точно, для каких значений параметра a имеет место следующее усиление теоремы 7.

Теорема 8. *Пусть $f(n)$ – такая функция, что $f(n) = o(n)$ и $f(n) = \Omega(\log n)$, и x_n – такая последовательность, что $K(x_n) = n + \mathcal{O}(f(n))$. Тогда существует последовательность y_n такая, что*

$$K(y_n) = n + \mathcal{O}(f(n)), \quad I(x_n : y_n) = an + \mathcal{O}(f(n)),$$

и для любой последовательности слов z_n, которая f-проста относительно **x** *и* **y**, *выполнено $K(z_n) = \mathcal{O}(f(n))$.*

Для $a = 1/2$ данная теорема была доказана в [13].

В формулировках теорем 7 и 8 мы избегали использования понятия «общая информация». Вместо этого мы говорили о «нематериализуемости взаимной информации». Рассмотрим одну из возможных формализаций понятия общей информации и переформулируем рассмотренные теоремы в новых терминах. Для этого мы воспользуемся полурешеткой S, введенной главе 3.

Определение. *Назовем общей информацией последовательностей* **x** *и* **y** *их точную нижнюю грань в полурешетке S.*

(Аналогичные определения можно дать, заменив функцию $\log n$ на произвольную $f(n)$, удовлетворяющую (0.18), а полурешетку S на S_f.) В новых терминах теорема 4 (для случая $f(n) = \log n$) утверждает, что не всякая пара последовательностей имеет общую информацию. Далее, будем говорить, что **x** и **y** имеют *нулевую общую информацию*, если их точной нижней гранью в S является Λ (наименьший элемент в S). Таким образом, теорему 7 для случая $f(n) = \log n$ можно переформулировать следующим образом: существуют последовательности **x** и **y** такие, что $I(x_n : y_n) = an + \mathcal{O}(\log n)$ $(a > 0)$, но общая информация **x** и **y** нулевая. Соответственно, теорема 8 утверждает, что для любой последовательности **x** такой, что $K(x_n) = n + \mathcal{O}(\log n)$, и любого $a \in (0, 1)$ найдется такая последовательность **y**, что

$$K(y_n) = n + \mathcal{O}(\log n), \; I(x_n : y_n) = an + \mathcal{O}(\log n),$$

а общая информация **x** и **y** нулевая.

В данной работе мы рассмотрим два метода, позволяющие доказывать теорему 8 для сколь угодно близких к единице значений параметра

a. В разделе 4.1 мы рассмотрим первый из них. Он основан на свойствах P-случайных пар для двумерных распределений P особого вида. Данное семейство пар является частным случаем класса примеров, рассматривавшихся в [9]. Мы дадим для данных пар новое, более простое доказательство нематериализуемости взаимной информации. Новый метод позволяет улучшить оценку на величину выделяемой информации (и, тем самым, получить доказательства теоремы 7 и теоремы 8).

Кратко опишем данную конструкцию. Будем брать в качестве **x**, **y** пару, случайную относительно двумерного распределения P специального вида. Мы покажем, что для любого $a \in (0,1)$ можно предъявить такое двумерное распределение P, что всякая P-случайная пара удовлетворяет требованию теоремы 7 при данном значении параметра a. Поскольку для любого распределения P существуют P-случайные кортежи, мы получаем доказательство теоремы 7.

С помощью данной конструкции и утверждения 2 из главы 1 мы докажем также и теорему 8 (для любых значений параметра $a \in (0,1)$).

В разделе 4.2 мы рассматриваем второй метод доказательства теорем 7 и 8. Он основан на алгебраической конструкции, обобщающей метод, использованный в [13]. В качестве x_n и y_n мы будем брать пары случайных ортогональных k-мерных подпространств в m-мерном пространстве над конечным полем F_n (в [13] рассматривалась данная конструкция для $k = 1$, $m = 2$). Мы докажем, что при $k < m/2$ данные пары удовлетворяют требованию теоремы 7. Выбирая значения параметров k и m, можно сделать значение параметра a сколь угодно близким к единице. С помощью данной конструкции мы получим также еще одно доказательство теоремы 8.

Замечание. Конструкции из разделов 4.1 и 4.2 могут давать пары $\langle x_n, y_n \rangle$, имеющие одинаковые значения колмогоровских сложностей $K(x_n)$, $K(y_n)$ и $K(x_n, y_n)$. И те, и другие пары имеют нулевую общую информацию. Однако можно предположить, что эти пары отличаются другими, более тонкими свойствами, связанными с выделяемостью взаимной информации. Возможно, данные пары имеют разные сложностные профили (понятие сложностного профиля пары определяется в [15]). В [15] доказываются некоторые оценки для сложностного профиля пары ортогональных пространств. К сожалению, нетривиальные оценки для сложностных профилей P-случайных пар неизвестны.

Подведем итог. В главе 4 мы формулируем в терминах последователь-

ностей слов теорему Гача и Кёрнера о существовании пар слов с невыделяемой взаимной информацией. Мы приводим две новые конструкции, позволяющие доказывать данную теорему с логарифмической оценкой на величину общей информацией. При этом мы получаем положительный ответ на вопрос Ан. А. Мучника.

Автор пользуется случаем, чтобы выразить свою благодарность проф. Н. К. Верещагину за постановку задач, научное руководство и внимание к работе, а также А. Х. Шеню и Ан. А. Мучнику за полезные обсуждения.

Используемые обозначения

- Будем обозначать слова (конечные последовательности букв заданного конечного алфавита) строчными латинскими буквами x, y, z,... Длину слова x (количество букв в x) будем обозначать $|x|$. Пустое слово будем обозначать ε.
- $\mathbf{x} = \{x_n\}$, $\mathbf{y} = \{y_n\}$, $\mathbf{z} = \{z_n\}$, ... – бесконечные последовательности слов.
- $\langle x_1, x_2, \ldots, x_n \rangle$ – кортеж двоичных слов; считаем фиксированной некоторую вычислимую нумерацию всех конечных кортежей слов.
- α, β, γ, ... – случайные величины. Все случайные величины, рассматриваемые в работе, имеют конечную область значений.
- $K(x)$ – колмогоровская сложность слова x.
- $K(x_1, x_2, \ldots, x_n)$ – колмогоровская сложность кортежа слов

$$\langle x_1, x_2, \ldots, x_n \rangle.$$

- $K(x|y)$ – условная колмогоровская сложность слова x относительно слова y.
- $K(x_1, x_2, \ldots, x_n | y_1, y_2, \ldots, y_m)$ – колмогоровская сложность кортежа $\langle x_1, x_2, \ldots, x_n \rangle$ относительно кортежа $\langle y_1, y_2, \ldots, y_m \rangle$.
- $I(x : y) := K(y) - K(y|x)$ – взаимная информация слов x и y.
- $I(x : y|z) := K(y|z) - K(y|x, z)$ – взаимная информация слов x и y относительно z; аналогично обозначается взаимная информация кортежей слов (относительно кортежей слов).

- $H(\alpha)$ – энтропия Шеннона случайной величины α; если случайная величина α принимает значения $a_1,\ldots,a_n$, то

$$H(\alpha) := -\sum_{i=1}^{n} \mathrm{Prob}[\alpha = a_i] \cdot \log\left(\mathrm{Prob}[\alpha = a_i]\right).$$

  Если случайные величины α, β имеют некоторое совместное распределение, причем α принимает значения $a_1, \ldots, a_n$, а β принимает значения $b_1, \ldots, b_m$, то условная энтропия β относительно α определяется как

$$H(\beta|\alpha) := -\sum_{i,j} \mathrm{Prob}[\alpha = a_i, \beta = b_j] \cdot \log\left(\mathrm{Prob}[\beta = b_j | \alpha = a_i]\right).$$

- $\mathcal{I}(\alpha : \beta) := H(\beta) - H(\beta|\alpha)$ – взаимная информация случайных величин α и β;

- Пусть $\langle x^1, x^2 \ldots, x^k \rangle$ – кортеж слов, $V = \{i_1, \ldots, i_r\} \subseteq \{1, \ldots, k\}$ – набор индексов; обозначим через x^V кортеж $\langle x^{i_1}, x^{i_2}, \ldots, x^{i_r} \rangle$. Аналогичное обозначение будем использовать для кортежей случайных величин и кортежей последовательностей слов.

- Если a и b линейные подпространства в линейном пространстве L, будем обозначать $(a \oplus b)$ сумму (линейную оболочку) данных подпространств.

- Пусть $\{a_i\}$ – семейство подпространств в линейном пространстве L. Обозначим $\bigoplus\limits_i a_i$ сумму (линейную оболочку) всех подпространств данного семейства.

- Для функций $f : \mathbb{N} \to \mathbb{N}$, $g : \mathbb{N} \to \mathbb{N}$, будем использовать стандартные асимптотические обозначения:

$$\begin{aligned}
f(n) = \mathcal{O}(g(n)) &\rightleftharpoons \exists C\, \exists N\, \forall n \geq N\ |f(n)| \leq C|g(n)|, \\
f(n) = \Omega(g(n)) &\rightleftharpoons \exists c > 0\, \exists N\, \forall n \geq N\ |f(n)| \geq c|g(n)|, \\
f(n) = \Theta(g(n)) &\rightleftharpoons \exists c, C > 0\, \exists N\, \forall n \geq N\ c|g(x)| \leq |f(n)| \leq C|g(n)|, \\
f(n) = o(g(n)) &\rightleftharpoons \forall c\, \exists N\, \forall n \geq N\ |f(n)| \leq c|g(n)|
\end{aligned}$$

- Целую часть числа r снизу и сверху будем обозначать $\lfloor r \rfloor$ и $\lceil r \rceil$ соответственно.

- Все логарифмы в статье берутся по основанию 2.

Глава 1.

P-типичность и P-случайность

В данной главе мы дадим основные определения и докажем несколько технических лемм. Прежде всего мы докажем лемму 1. Утверждение этой леммы в той или иной форме встречается практически во всех работах по колмогоровской сложности. На этом простом комбинаторном факте в конечном итоге основано любое применение колмогоровской сложности. Лемма утверждает, что если множество слов A имеет простое описание, то все слова из данного множества имеют сложность не намного больше $\log|A|$; кроме того, в A найдется хотя бы одно слово со сложностью не менее $\log|A|$. Чтобы сформулировать утверждение строго, необходимо уточнить, что значит, что множество A *имеет простое описание*. Нам будет удобно предполагать, что задано целое семейство множеств A_n, и для получения списка элементов n-ого множества достаточно иметь логарифмическую по n информацию.

Лемма 1. *Пусть задано семейство конечных множеств A_n, $n = 1, 2, \ldots$, элементами которых являются слова в некотором алфавите. Пусть при этом для каждого n список слов, входящих в A_n, может быть порожден алгоритмом размера $\mathcal{O}(\log n)$ (иначе говоря, колмогоровская сложность кортежа, составленного из всех слов множества A_n в лексикографическом порядке, равна $\mathcal{O}(\log n)$). Тогда сложность любого слова из A_n не превосходит $\log|A_n| + \mathcal{O}(\log n)$; кроме того, для каждого n найдется слово $x_n \in A_n$ сложности не меньше $\log|A_n|$.*

Доказательство. Сначала докажем первое утверждение леммы. Пусть слово x_n принадлежит A_n. Чтобы получить x_n можно действовать следующим образом: сначала найдем список всех элементов множества A_n, а затем выберем из этого списка нужное слово. По условию леммы сложность порождения списка всех элементов A_n равна $\mathcal{O}(\log n)$. Далее

остается указать номер слова x_n в списке элементов A_n. Следовательно, $K(x_n) \leq \log |A_n| + \mathcal{O}(\log n)$.

Докажем второе утверждение леммы. Нужно показать, что во множестве A_n найдется элемент сложности не менее $\log |A_n|$. Заметим, что число слов, имеющих сложность меньше некоторого числа C, не больше, чем число программ длины менее C. А число программ длины менее C складывается из числа программ длины 0, 1, ..., $(C-1)$ и равно

$$1 + 2 + 2^2 + \ldots + 2^{C-1} = 2^C - 1.$$

Таким образом, число слов сложности меньше $\log |A_n|$ не превосходит $|A_n| - 1$. Следовательно, хотя бы одно слово из A_n имеет сложность не меньше $\log |A_n|$. □

Рассмотрим несколько примеров применения леммы 1.

Пример 1. Пусть множество A_n состоит из всех двоичных слов длины n. Применяя лемму 1, заключаем, что каждое слово длины n имеет сложность не более $n + \mathcal{O}(\log n)$, и для каждого n найдется слово длины n со сложностью не менее n.

Отметим, что в данном примере первое утверждение можно усилить, заменив логарифмическое слагаемое на константу, т. е. можно доказать, что сложность любого двоичного слова длины n не превосходит $n + \mathcal{O}(1)$.

Пример 2. Пусть $p_1, p_2, \ldots, p_k$ – положительные числа, и их сумма равна 1. Пусть также C – некоторая положительная константа. Определим семейство множеств A_n. Каждое A_n будет содержать слова в алфавите из k букв. Будем обозначать через $m_1(x), \ldots, m_k(x)$ число вхождений в слово x соответственно 1-ой,..., k-ой букв алфавита. Множество A_n будет содержать те и только те слова x длины n, для которых

$$|m_i(x) - np_i| \leq C, i = 1, 2 \ldots, k. \tag{1.1}$$

То есть A_n состоит из слов, в которых доли букв близки $p_1, \ldots, p_k$.

Для каждого n имеется конечное число наборов чисел m_1,...,m_k, дающих в сумме n и удовлетворяющих (1.1). Заметим, что число таких наборов не превосходит некоторой константы, определяемой числами C и k (но не зависящей от n). Зафиксируем один из таких наборов $\langle m_1, \ldots, m_k \rangle$. Количество слов длины n, содержащих m_i вхождений i-ой буквы алфавита (для $i = 1, \ldots, k$), равно

$$S = \frac{n!}{m_1! m_2! \ldots m_k!} \tag{1.2}$$

Используя формулу Стирлинга, нетрудно проверить, что

$$\log S = -n \sum \frac{m_i}{n} \log \frac{m_i}{n} + \mathcal{O}(\log n).$$

(Константа перед логарифмическим членом зависит от C и k.) Учитывая (1.1), получаем

$$\log S = -(\Sigma p_i \log p_i)n + \mathcal{O}(\log n). \tag{1.3}$$

Поскольку для каждого n имеется $\mathcal{O}(1)$ наборов чисел $\langle m_1, \ldots, m_k \rangle$, удовлетворяющих (1.1), логарифм числа элементов в A_n может отличаться от (1.3) лишь на аддитивную константу. Таким образом,

$$\log |A_n| = -(\Sigma p_i \log p_i)n + \mathcal{O}(\log n).$$

Теперь, пользуясь леммой 1, мы можем сделать два вывода:

а) сложность любого слова из A_n не превосходит $-(\sum p_i \log p_i)n + \mathcal{O}(\log n)$.

б) найдется слово $x_n \in A_n$, для которого $K(x_n) = -(\Sigma p_i \log p_i)n + \mathcal{O}(\log n)$.

Отметим, что выражение $-(\Sigma p_i \log p_i)$ есть шенноновская энтропия случайной величины, принимающей k значений с вероятностями $p_1, \ldots, p_k$.

Определение 1. *Пусть случайные величины $\varphi^1, \varphi^2, \ldots, \varphi^k$ принимают значения в конечных алфавитах $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ и имеют совместное распределение*

$$P(a^1, a^2, \ldots, a^k) = \mathrm{Prob}[\varphi^1 = a^1, \varphi^2 = a^2, \ldots, \varphi^k = a^k].$$

Назовем кортеж $\langle \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k \rangle$, состоящий из k бесконечных последовательностей слов, P-типичным, если каждое слово x_n^i имеет длину n, и для любого набора значений $\langle a^1, a^2, \ldots, a^k \rangle$ число таких позиций i, что в каждом слове x_n^j на i-м месте стоит буква a^j, равно $nP(a^1, a^2, \ldots, a^k) + \mathcal{O}(1)$.

Пример 3. Рассмотрим совместное распределение пары случайных величин $\langle \varphi, \psi \rangle$ со следующими свойствами: обе величины φ, ψ принимают значения 0 и 1 с вероятностью $1/2$ (т. е. являются двоичными равномерно распределенными); при этом φ и ψ принимают разные значения с

вероятностью α (и, соответственно, совпадают с вероятностью $(1-\alpha)$). Таким образом,

$$\begin{aligned}\mathrm{Prob}[\varphi=0,\psi=0]=\mathrm{Prob}[\varphi=1,\psi=1] &= \tfrac{1-\alpha}{2},\\ \mathrm{Prob}[\varphi=1,\psi=0]=\mathrm{Prob}[\varphi=0,\psi=1] &= \tfrac{\alpha}{2}.\end{aligned}$$

Пусть пара последовательностей $\langle \mathbf{x}, \mathbf{y}\rangle$ P-типична относительно указанного распределения. Тогда слова x_n, y_n имеют длину n, причем каждое из них содержит по $n/2+\mathcal{O}(1)$ нулей и единиц. Кроме того, x_n и y_n отличаются друг от друга в $\alpha n+\mathcal{O}(1)$ битов.

Отметим, что если пара $\langle \mathbf{x}, \mathbf{y}\rangle$ типична относительно совместного распределения пары случайных величин $\langle\varphi,\psi\rangle$, то последовательность $\mathbf{x}$ типична относительно распределения случайной величины φ, а последовательность $\mathbf{y}$ типична относительно распределения ψ.

Утверждение 1. *Если случайные величины $\varphi^1,\varphi^2,\ldots,\varphi^k$ имеют некоторое совместное распределение P, то для любых P-типичных последовательностей $\mathbf{x}^1,\mathbf{x}^2,\ldots,\mathbf{x}^k$, любого непустого набора индексов $V\subseteq\{1,2,\ldots,k\}$ и любого (быть может пустого) набора индексов $W\subseteq\{1,2,\ldots,k\}$ выполнено неравенство*

$$K(x_n^V|x_n^W)\le nH(\varphi^V|\varphi^W)+\mathcal{O}(\log n) \tag{1.4}$$

(В случае, когда набор индексов W пуст, условная колмогоровская сложность и условная шенноновская энтропия в равенстве (1.4) обращаются в безусловные.)

Доказательство. Сначала рассмотрим случай, когда набор индексов W пуст. Требуется доказать, что для любого V выполнено равенство $K(x_n^V)\le nH(\varphi^V)+\mathcal{O}(\log n)$. Если V содержит ровно один индекс, то мы можем сослаться на вывод (а) из примера 2.

Пусть теперь V содержит $l>1$ индексов. Мы закодируем кортежи x_n^V последовательностью слов y_n и сведем l-мерный случай к уже рассмотренному одномерному. Для этого нам потребуется ввести несколько новых обозначений.

Обозначим l-мерную случайную величину φ^V через α. Пусть конечное множество $\mathcal{A}$ – область значений α (значениями α являются кортежи длины l, задающие набор значений φ^V, т. е. $\mathcal{A}$ есть декартово произведение областей значений функций из кортежа φ^V). Разумеется, $H(\alpha)=H(\varphi^V)$.

Далее, определим последовательность слов $\mathbf{y}$. Слова y_n будут состоять из n букв алфавита $\mathcal{A}$. При этом слова y_n будут кодировать кортежи слов x_n^V в следующем смысле: буква на i-ом месте в слове y_n будет соответствовать набору букв, стоящих на i-ых местах в словах кортежа x_n^V. Очевидно, $K(y_n) = K(x_n^V) + \mathcal{O}(1)$.

Поскольку кортеж $\mathbf{x}^V$ типичен относительно распределения φ^V, построенная последовательность $\mathbf{y}$ типична относительно распределения α. К последовательности $\mathbf{y}$ и случайной величине α применим вывод (a) из примера 2. Таким образом, $K(y_n) \leq nH(\alpha) + \mathcal{O}(\log n)$. А значит,

$$K(x_n^V) \leq nH(\varphi^V) + \mathcal{O}(\log n).$$

Перейдем к рассмотрению основного случая. Пусть набор индексов W непуст. Пусть $V' = V \setminus W$. Очевидно, $H(\varphi^{V'}|\varphi^W) = H(\varphi^V|\varphi^W)$ и $K(x_n^{V'}|x_n^W) = K(x_n^V|x_n^W) + \mathcal{O}(1)$. Следовательно, не ограничивая общности, можно считать, что наборы индексов V и W не пересекаются.

Рассмотрим случайные величины $\alpha = \varphi^V$ и $\beta = \varphi^W$. Они принимают значения в некоторых конечных алфавитах $\mathcal{A}$ и $\mathcal{B}$ (множество $\mathcal{A}$ является декартовым произведением множеств значений случайных величин φ^V, а $\mathcal{B}$, соответственно, декартовым произведением множеств значений величин φ^W). Будем считать, что $\mathcal{A} = \{a_1, a_2, \ldots, a_s\}$ и $\mathcal{B} = \{b_1, b_2, \ldots, b_r\}$.

Определим новые последовательности слов $\mathbf{v}$ и $\mathbf{w}$. Слова v_n будут состоять из букв алфавита $\mathcal{A}$, а слова w_n – из букв алфавита $\mathcal{B}$. Слово v_n будет иметь длину n, и каждая его буква определяется буквами, стоящими в соответствующей позиции в словах кортежа x_n^V. Аналогично, w_n также будет иметь длину n, и каждая буква в слове w_n определяется буквами, стоящими в соответствующей позиции в словах кортежа x_n^W. Таким образом, v_n и w_n кодируют кортежи x_n^V и x_n^W. Очевидно, что $K(v_n|w_n) = K(x_n^V|x_n^W) + \mathcal{O}(1)$. При этом пара $\langle \mathbf{v}, \mathbf{w} \rangle$ является типичной относительно распределения $\langle \alpha, \beta \rangle$.

Осталось доказать, что если пара последовательностей $\langle \mathbf{v}, \mathbf{w} \rangle$ типична относительно распределения пары случайных величин $\langle \alpha, \beta \rangle$, то

$$K(v_n|w_n) \leq nH(\alpha|\beta) + \mathcal{O}(\log n).$$

Зафиксируем некоторый номер n. Обозначим m_{ij} число таких позиций, в которых в слове v_n стоит буква a_i, а в слове w_n – буква b_j. Поскольку пара $\langle \mathbf{v}, \mathbf{w} \rangle$ типична относительно распределения $\langle \alpha, \beta \rangle$,

$$m_{ij} = \mathrm{Prob}[\alpha = a_i, \beta = b_j] \cdot n + \mathcal{O}(1). \tag{1.5}$$

Далее, последовательность $\mathbf{w}$ типична относительно распределения β. Это значит, что для каждого $j = 1, \ldots, r$ слово w_n содержит

$$m_j = \mathrm{Prob}[\beta = b_j] \cdot n + \mathcal{O}(1) \tag{1.6}$$

букв b_j. Заметим, что

$$m_j = \sum_i m_{ij}.$$

Предположим, что слово w_n известно и требуется найти v_n. Сложность нахождения всех чисел m_{ij} есть $\mathcal{O}(\log n)$ (константа перед логарифмом зависит от размеров алфавитов $\mathcal{A}$ и $\mathcal{B}$, но не зависит от n). Пусть величины m_{ij} уже найдены. Тогда для получения слова v_n достаточно для каждого $j = 1, 2 \ldots, r$ указать разбиение тех m_j позиций, в которых в слове w_n стоит буква b_j, на s классов (содержащих $m_{1j}, \ldots, m_{sj}$ позиций), в которых в слове v_n стоят буквы $a_1, \ldots, a_s$ соответственно.

Для каждого j имеется

$$\frac{m_j!}{m_{1j}! m_{2j}! \ldots m_{sj}!}$$

способов разбить m_j позиций на непересекающиеся подмножества размера $m_{1j}, \ldots, m_{sj}$. Следовательно,

$$K(v_n|w_n) \le \sum_j \log \left(\frac{m_j!}{m_{1j}! m_{2j}! \ldots m_{sj}!} \right) + \mathcal{O}(\log n). \tag{1.7}$$

Оценивая факториалы с помощью формулы Стирлинга, нетрудно преобразовать правую часть (1.7) в выражение

$$-n \sum_{i,j} \frac{m_{ij}}{n} \log \frac{m_{ij}}{m_j} + \mathcal{O}(\log n).$$

Из (1.5) и (1.6) следует, что $\frac{m_{ij}}{n} = \mathrm{Prob}[\alpha = a_i, \beta = b_j] + \mathcal{O}(1/n)$ и $\frac{m_{ij}}{m_j} = \mathrm{Prob}[\alpha = a_i | \beta = b_j] + \mathcal{O}(1/n)$. Значит правая часть неравенства (1.7) равна

$$-n \sum_{ij} \mathrm{Prob}[\alpha = a_i, \beta = b_j] \log \mathrm{Prob}[\alpha = a_i | \beta = b_j] + \mathcal{O}(\log n) =$$
$$= n \cdot H(\alpha|\beta) + \mathcal{O}(\log n).$$

□

Определение 2. *Будем называть P-типичный кортеж последовательностей слов $\langle \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k \rangle$ P-случайным, если для любого непустого набора индексов $V \subseteq \{1, 2 \ldots, k\}$ и любого (быть может пустого) набора индексов $W \subseteq \{1, 2 \ldots, k\}$ выполнено равенство*

$$K(x_n^V | x_n^W) = nH(\varphi^V | \varphi^W) + \mathcal{O}(\log n). \tag{1.8}$$

(В случае, когда набор индексов W пуст, условная колмогоровская сложность и условная шенноновская энтропия в равенстве (1.8) обращаются в безусловные.)

Ниже мы докажем, что для любого распределения P существует P-случайный кортеж.

Пример 4. Пусть α – случайная величина, равная нулю с вероятностью 1, а слова x_n состоят из n нулей. Тогда $H(\alpha) = 0$ и $K(x_n) = \mathcal{O}(\log n)$. При этом последовательность $\mathbf{x}$ случайна относительно распределения величины α.

Пример 5. Рассмотрим следующее распределение P случайной величины α:

$$\mathrm{Prob}[\alpha = 0] = 1/2, \ \mathrm{Prob}[\alpha = 1] = 1/2.$$

Нетрудно проверить, что $H(\alpha) = 1$.

Опишем соответствующие данному распределению P-случайные последовательности $\mathbf{x}$. Прежде всего они должны быть P-типичными, т. е. каждое слово x_n должно должно иметь длину n и состоять из $n/2 + \mathcal{O}(1)$ нулей и $n/2 + \mathcal{O}(1)$ единиц.

Далее, P-случайность последовательности значит, что

$$K(x_n) = nH(\alpha) + \mathcal{O}(\log n) = n + \mathcal{O}(\log n).$$

Таким образом, P-случайными будут последовательности слов, содержащих примерно равное число нулей и единиц и имеющих сложность близкую к максимальной (среди всех двоичных слов данной длины).

Замечание 1. У P-случайных последовательностей слов не только условные и безусловные колмогоровские сложности, но и величины взаимной информации пропорциональны соответствующим шенноновским величинам. Пусть, например, пара последовательностей $\langle \mathbf{x}, \mathbf{y} \rangle$ случайна относительно распределения P пары случайных величин $\langle \varphi, \psi \rangle$. Согласно определению P-случайности величины $K(x_n)$, $K(y_n)$ и $K(x_n, y_n)$

равны (с точностью до логарифмического слагаемого) $nH(\varphi)$, $nH(\psi)$ и $nH(\varphi,\psi)$ соответственно. Тогда аналогичное равенство выполняется и для взаимной информации. В самом деле, сравнивая равенства

$$\mathcal{I}(\varphi:\psi) = H(\varphi) + H(\psi) - H(\varphi,\psi)$$

и

$$I(x_n:y_n) = K(x_n) + K(y_n) - K(x_n,y_n) + \mathcal{O}(\log n)$$

получаем

$$I(x_n:y_n) = n\mathcal{I}(\varphi:\psi) + \mathcal{O}(\log n).$$

Лемма 2. *Пусть кортеж последовательностей* $\langle \mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_k}\rangle$ *типичен относительно совместного распределения* P *случайных величин* $\langle\varphi_1,\varphi_2,\ldots,\varphi_k\rangle$, *и* $K(x_n,\ldots,x_k) = nH(\varphi_1,\varphi_2,\ldots,\varphi_k) + \mathcal{O}(\log n)$. *Тогда данный кортеж последовательностей* P*-случаен.*

Доказательство. Пусть V – произвольный набор индексов. Сначала рассмотрим случай $W = \emptyset$ и докажем, что $K(x_n^V) = nH(\varphi^V)$. Согласно лемме 1 выполнено неравенство $K(x_n^V) \le nH(\varphi^V) + \mathcal{O}(\log n)$. Докажем обратное неравенство.

Обозначим $U = \{1,2,\ldots,k\}$ набор всех индексов. Пусть $V' = U \setminus V$. Из леммы 1 следует неравенство

$$K(x_n^{V'}|x_n^V) \le nH(\varphi^{V'}|\varphi^V) + \mathcal{O}(\log n).$$

Складывая его с равенствами

$$\begin{aligned} &K(x_n^U) = K(x_n^V) + K(x_n^{V'}|x_n^V) + \mathcal{O}(\log n),\\ &nH(\varphi^U) = K(x_n^U) + \mathcal{O}(\log n),\\ &nH(\varphi^{V'}|\varphi^V) + nH(\varphi^V) = nH(\varphi^U), \end{aligned}$$

получаем $K(x_n^V) \ge nH(\varphi^V) + \mathcal{O}(\log n)$. Таким образом, для любого набора индексов V выполнено равенство

$$K(x_n^V) = nH(\varphi^V) + \mathcal{O}(\log n).$$

Теперь пусть W – произвольный набор индексов; обозначим $\tilde{U} = V \cup W$. Вычитая из равенства

$$K(x_n^{\tilde{U}}) = nH(\varphi^{\tilde{U}}) + \mathcal{O}(\log n)$$

равенство

$$K(x_n^W) = nH(\varphi^W) + \mathcal{O}(\log n),$$

получаем

$$K(x_n^V|x_n^W) = nH(\varphi^V|\varphi^W) + \mathcal{O}(\log n).$$

□

В дальнейшем нам потребуется конструкция, позволяющая расширять P-случайные кортежи. А именно, пусть s, r – неотрицательные целые числа, и $s < r$. Будем считать, что случайные величины $\varphi^1, \varphi^2, \ldots, \varphi^r$ имеют совместное распределение P. Обозначим через P' проекцию распределения P на первые s координат, т. е. совместное распределение величин $\varphi^1, \varphi^2, \ldots, \varphi^s$. Тогда любой P'-случайный кортеж можно дополнить до P-случайного. Более точно, выполнена следующая лемма.

Утверждение 2. *Пусть P – совместное распределение r случайных величин, P' – проекция распределения P на первые s координат, и кортеж $\langle \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^s \rangle$ является P'-случайным. Тогда существуют такие последовательности $\mathbf{x}^{s+1}, \mathbf{x}^{s+2}, \ldots, \mathbf{x}^r$, что кортеж $\langle \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^r \rangle$ является P-случайным.*

Доказательство. Как и в доказательстве леммы 1, мы будем заменять многомерные распределения одномерными и кодировать кортежи слов одним словом. А именно, мы будем рассматривать случайные величины $\alpha = \langle \varphi^1 \ldots, \varphi^s \rangle$ и $\beta = \langle \varphi^{s+1}, \ldots, \varphi^r \rangle$, а кортежи $\langle x_n^1 \ldots, x_n^s \rangle$ и $\langle x_n^{s+1}, \ldots, x_n^r \rangle$ заменим на пару слов v_n, w_n. Обозначим множества, в которых α и β принимают значения, через $\mathcal{A} = \{a_1, a_2, \ldots, a_k\}$ и $\mathcal{B} = \{b_1, b_2, \ldots, b_l\}$ соответственно ($\mathcal{A}$ есть декартово произведение областей значений случайных величин $\varphi^1, \ldots, \varphi^s$, и $\mathcal{B}$ есть декартово произведение областей значений $\varphi^{s+1}, \ldots, \varphi^r$). Слова v_n будут состоять из n букв алфавита $\mathcal{A}$, слова w_n – из n букв алфавита $\mathcal{B}$.

Преобразуем заданные в условии леммы кортежи $\langle x_n^1 \ldots, x_n^s \rangle$ в слова v_n. Буква слова v_n в позиции i будет соответствовать набору букв, стоящих в i-ых позициях в словах кортежа $\langle x_n^1, \ldots, x_n^s \rangle$ $(i = 1, \ldots, n)$. Поскольку исходный кортеж P'-случаен, построенная последовательность $\mathbf{v}$ случайна относительно распределения α.

Следует отдельно рассмотреть случай, когда $s = 0$. Тогда будем считать, что $\mathcal{A} = \{0\}$ $(k = 1)$, случайная величина α равна нулю с вероятностью 1, а в качестве v_n мы будем брать слова, состоящие из n нулей (см. пример 4).

Мы докажем, что для каждой случайной относительно α последовательности $\mathbf{v}$ найдется такая последовательность $\mathbf{w}$ (в алфавите $\mathcal{B}$), что пара $\langle \mathbf{v}, \mathbf{w} \rangle$ является случайной относительно совместного распределения $\langle \alpha, \beta \rangle$.

По найденной последовательности $\mathbf{w}$ можно будет построить кортеж $\langle \mathbf{x}^{s+1}, \ldots, \mathbf{x}^r \rangle$: буквы в слове w_n кодируют набор букв, стоящих в соответствующих позициях в словах кортежа $\langle x_n^{s+1}, \ldots, x_n^r \rangle$. И если пара $\langle \mathbf{v}, \mathbf{w} \rangle$ случайна относительно распределения $\langle \alpha, \beta \rangle$, то соответствующий ей кортеж $\langle \mathbf{x}^1, \ldots, \mathbf{x}^r \rangle$ случаен относительно распределения $\langle \varphi^1, \ldots, \varphi^r \rangle$.

Итак, остается доказать существование требуемой последовательности $\mathbf{w}$. Рассмотрим совместное распределение величин α и β:

$$\mathrm{Prob}[\alpha = a_i, \beta = b_j] = p_{ij},$$

где $i = 1, \ldots, k, j = 1, \ldots, l$. Проекция этого распределения на первую координату дает распределение α:

$$\mathrm{Prob}[\alpha = a_i] = \sum_j p_{ij}.$$

Пусть в слове v_n буквы a_i встречаются m_i раз ($i = 1, \ldots, k$). Поскольку последовательность $\mathbf{v}$ является случайной относительно α,

$$m_i = n \sum_j p_{ij} + \mathcal{O}(1).$$

Следовательно, можно представить числа m_i в виде сумм $m_i = \sum_j m_{ij}$ так, что

$$m_{ij} = np_{ij} + \mathcal{O}(1).$$

Выбор набора чисел m_{ij} неоднозначен; зафиксируем для каждого n один из таких наборов.

Будем называть двоичное слово $\hat{w}_n$ длины n допустимым, если для всех i, j найдется ровно m_{ij} позиций, в которых в слове v_n стоит буква a_i, а в слове $\hat{w}_n$ стоит буква b_j. Если $\hat{w}_n$ допустимо, то частоты появления букв в этом слове соответствуют распределению вероятностей β, и пара $\langle \{v_n\}, \{\hat{w}_n\} \rangle$ является типичной относительно $\langle \alpha, \beta \rangle$. Остается выбрать для каждого n такое допустимое $\hat{w}_n$, чтобы полученная пара была не только типичной, но и случайной относительно $\langle \alpha, \beta \rangle$. Покажем, что для этого достаточно для каждого n брать допустимое слово, имеющее самую большую сложность относительно v_n.

Для фиксированного v_n количество соответствующих ему допустимых слов $\hat{w}_n$ равно

$$\prod_i \frac{m_i!}{m_{i1}!m_{i2}!\ldots m_{il}!}$$

Если в качестве w_n взять допустимое слово, имеющее максимальную сложность относительно v_n, то

$$K(w_n|v_n) = -\sum_i \log\left(\frac{m_i!}{m_{i1}!m_{i2}!\ldots m_{il}!}\right) + \mathcal{O}(\log n)$$

Как и в доказательстве утверждения 1, правая часть полученного равенства есть $nH(\beta|\alpha) + \mathcal{O}(\log n)$.

Далее, из случайности $\mathbf{v}$ относительно α следует $K(v_n) = nH(\alpha) + \mathcal{O}(\log n)$. Учитывая равенство

$$K(v_n, w_n) = K(v_n) + K(w_n|v_n) + \mathcal{O}(\log n),$$

мы получаем

$$K(v_n, w_n) = nH(\alpha) + nH(\beta|\alpha) + \mathcal{O}(\log n) = nH(\alpha, \beta) + \mathcal{O}(\log n). \quad (1.9)$$

Согласно лемме 2 из равенства (1.9) вытекает P-случайность построенной пары $\langle \mathbf{v}, \mathbf{w}\rangle$ относительно распределения $\langle \alpha, \beta\rangle$. Следовательно, соответствующий полученной последовательности $\mathbf{w}$ кортеж $\langle \mathbf{x}^{s+1}, \ldots, \mathbf{x}^r\rangle$ дополняет $\mathbf{x}^1, \ldots, \mathbf{x}^s$ до P-случайного кортежа. □

Следствие 1. *Для каждого распределения P существует P-типичный кортеж.*

Доказательство. Применим утверждение 2 при $s = 0$.

Глава 2.

Неравенства для колмогоровской сложности и шенноновской энтропии

В этой главе мы рассмотрим классы линейных неравенств, выполняющихся для колмогоровских сложностей n-ок слов, шенноновских энтропий n-ок случайных величин и размерностей n-ок линейных подпространств. Мы докажем, что классы неравенств для колмогоровской сложности и шенноновской энтропии совпадают. В [11, 16] было показано, что всякое линейное неравенство, выполненное для шенноновской энтропии, верно также и для линейных пространств, причем для $n = 3$ все три класса неравенств (для колмогоровских сложностей, шенноновских энтропий и размерностей подпространств) совпадают. Мы покажем, что для $n \geq 4$ класс линейных неравенств для размерностей подпространств строго больше, чем совпадающие классы неравенств для колмогоровской сложности и шенноновской энтропии. Точнее, аналоги неравенства Инглетона, верного для размерностей подпространств, не выполняются для энтропии Шеннона и для колмогоровской сложности.

Будем рассматривать линейные неравенства для шенноновской энтропии и колмогоровской сложности. Сначала определим класс неравенств, выполненных для энтропии Шеннона. Пусть имеются n совместно распределенных случайных величин $\varphi_1 \ldots, \varphi_n$. Тогда мы можем рассмотреть энтропии каждой из этих величин, а также энтропии пар случайных величин, энтропии троек, и т. д. То есть для каждого непустого набора индексов $W \subseteq \{1, 2, \ldots, n\}$ можно рассмотреть энтропию кортежа φ^W. Всего имеется $(2^n - 1)$ кортеж случайных величин, для каждого из которых определена энтропия. Будем интересоваться, какие линейные неравенства выполняются для данного набора энтропий для любых

$\varphi_1, \ldots, \varphi_n$. Чтобы задать линейное неравенство для энтропий n случайных величин необходимо указать $(2^n - 1)$ коэффициентов. Например, для $n = 3$ линейное неравенство для энтропий имеет вид

$$\lambda_1 H(\varphi_1) + \lambda_2 H(\varphi_2) + \lambda_3 H(\varphi_3) + \lambda_{1,2} H(\varphi_1, \varphi_2) + \\ + \lambda_{1,3} H(\varphi_1, \varphi_3) + \lambda_{2,3} H(\varphi_2, \varphi_3) + \lambda_{1,2,3} H(\varphi_1, \varphi_2, \varphi_3) \geq 0,$$

где $\lambda_1, \lambda_2, \lambda_3, \lambda_{1,2}, \lambda_{1,3}, \lambda_{2,3}, \lambda_{1,2,3}$ – произвольные вещественные коэффициенты.

Для произвольного n общий вид линейного неравенства для энтропий n случайных величин можно записать как

$$\sum_W \lambda_W H(\varphi^W) \geq 0, \tag{2.1}$$

где сумма берется по всем непустым $W \subseteq \{1, 2, \ldots, n\}$.

Определение 3. *Обозначим $\Lambda_S(n)$ множество всех наборов из $(2^n - 1)$ коэффициентов $\{\lambda_W\}$, для которых неравенство (2.1) верно для любых случайных величин $\varphi_1, \ldots, \varphi_n$.*

Будем говорить, что $\Lambda_S(n)$ – это класс линейных неравенств, выполненных для шенноновских энтропий n-ок случайных величин.

Дадим аналогичное определение для колмогоровских сложностей. Для любых n слов $x^1, x^2, \ldots, x^n$ можно рассмотреть колмогоровские сложности каждого из этих слов, колмогоровские сложности пар слов, троек слов и т. д., т. е. для любого непустого набора индексов $W \subseteq \{1, 2, \ldots, n\}$ определена колмогоровская сложность соответствующего кортежа $K(x^W)$. Если два кортежа составлены из одних и тех же слов и отличаются только их порядком, то колмогоровские сложности этих кортежей могут отличаться лишь на ограниченную величину (зависящую от длин кортежей, но не от сложности входящих в них слов). Например,

$$K(x, y) = K(y, x) + \mathcal{O}(1).$$

Поэтому для n-ки слов $x^1, \ldots, x^n$ мы ограничимся рассмотрением колмогоровских сложностей $(2^n - 1)$ различных кортежей x^W, соответствующих разным наборам индексов W.

Так же как и для энтропий Шеннона, мы хотим определить класс неравенств, выполненных для колмогоровских сложностей n-ок слов. Имеется существенное отличие неравенств для колмогоровской сложности

от неравенств для шенноновской энтропии. В колмогоровском случае мы будем рассматривать неравенства только с точностью до логарифмического слагаемого (логарифм берется от суммы длин всех слов, включенных в рассматриваемое неравенство).

Определение 4. *Обозначим $\Lambda_K(n)$ множество всех наборов из $(2^n - 1)$ коэффициентов $\{\lambda_W\}$, для которых существует такая константа $C > 0$, что неравенство (2.2) верно для любых n слов $x_1, \ldots, x_n$ (сумма берется по всем непустым $W \subseteq \{1, 2, \ldots, n\}$).*

$$\sum_W \lambda_W K(x^W) \geq -C \log(|x_1| + |x_2| + \ldots + |x_n|) - C. \tag{2.2}$$

Множество $\Lambda_K(n)$ мы будем называть классом линейных неравенств, выполненных для колмогоровских сложностей n-ок слов.

Теорема 1. *Для любого n классы $\Lambda_S(n)$ и $\Lambda_K(n)$ совпадают, т. е. всякое линейное неравенство, выполненное для шенноновской энтропии, выполняется также и для колмогоровской сложности, и наоборот.*

Доказательство. $[\Lambda_K(n) \subseteq \Lambda_S(n)]$ Пусть для некоторого набора $\{\lambda_W\}$ существует такая константа C, что неравенство (2.2) верно для любой n-ки слов. Докажем, что для любой n-ки случайных величин выполнено неравенство (2.1).

Зафиксируем произвольную n-ку случайных величин $\varphi_1, \ldots, \varphi_n$. Обозначим совместное распределение данных случайных величин через P. Согласно следствию 1 существует P-случайный кортеж последовательностей слов $\langle \mathbf{x^1}, \ldots, \mathbf{x^n} \rangle$, где каждая последовательность $\mathbf{x}^i$ состоит из слов $x_1^i, x_2^i, \ldots$, и длина слова x_m^i равна m. Согласно определению P-случайности для любого непустого $W \subseteq \{1, \ldots, n\}$ выполнено равенство

$$K(x_m^W) = mH(\varphi^W) + \mathcal{O}(\log m). \tag{2.3}$$

По предположению для колмогоровских сложностей слов $x_m^1, x_m^2, \ldots, x_m^n$ выполнено неравенство

$$\sum_W \lambda_W K(x_m^W) \geq -C \log m - C.$$

Согласно (2.3) в данном неравенстве можно заменить колмогоровские сложности на соответствующие шенноновские энтропии:

$$\sum_W \lambda_W \left(mH(\varphi^W) + \mathcal{O}(\log m) \right) \geq -C \log m - C. \tag{2.4}$$

Чтобы получить (2.1), остается разделить обе части неравенства (2.4) на m и устремить m к бесконечности.

$[\Lambda_S(n) \subseteq \Lambda_K(n)]$ Теперь докажем обратное включение. Предположим, что для некоторого набора коэффициентов λ_W неравенство (2.1) выполнено для любой n-ки случайных величин. Докажем, что аналогичное неравенство верно и для колмогоровских сложностей.

Пусть $x = \langle x^1, x^2, \ldots, x^n \rangle$ – произвольная n-ка слов. Обозначим m сумму длин данных слов: $m = |x_1| + \ldots + |x_n|$. Докажем, что для колмогоровских сложностей слов $x_1, x_2, \ldots, x_n$ выполняется неравенство (2.2). Для этого мы определим некоторое n-мерное распределение $\langle \varphi_1, \varphi_2, \ldots, \varphi_n \rangle$ и воспользуемся неравенством (2.1) для полученных случайных величин.

Рассмотрим множество A, которое будет состоять из всех n-ок $y = \langle y_1, y_2, \ldots, y_n \rangle$ таких, что

$$K(y^V | y^W) \leq K(x^V | x^W)$$

для любого непустого $V \subseteq \{1, 2, \ldots, n\}$ и любого (быть может, пустого) $W \subseteq \{1, 2 \ldots, n\}$. Заметим, что множество A заведомо непусто, поскольку n-ка x в нем содержится. Покажем, что число элементов в A равно $2^{K(x)+\mathcal{O}(\log m)}$. Действительно, согласно определению A все элементы $y \in A$ имеют сложность не более $K(x)$. Но число кортежей, имеющих сложность не больше $K(x)$, не превосходит $2^{K(x)+1}$, т. е. $\log |A| \leq K(x) + 1$. Теперь докажем, что с точностью до $\mathcal{O}(\log m)$ выполнено обратное неравенство.

Лемма 3. $\log |A| \geq K(x) - \mathcal{O}(\log m)$.

Доказательство леммы. Для того, чтобы построить алгоритм, перечисляющий множество A, достаточно знать все величины сложностей $K(x^V | x^W)$. Чтобы указать все эти числа, требуется $\mathcal{O}(\log m)$ битов. Если данный перечисляющий алгоритм построен, то для нахождения кортежа x нужны дополнительные $\log |A|$ битов, указывающие порядковый номер x в перечислении списка элементов A. Следовательно, $K(x) \leq \log |A| + \mathcal{O}(\log m)$, и лемма доказана. □

Рассмотрим случайную величину $\varphi = \langle \varphi_1, \ldots, \varphi_n \rangle$, равномерно распределенную на множестве A (значениями случайной величины φ являются n-ки слов, и $\varphi_1, \ldots, \varphi_n$ – компоненты этих n-ок). Покажем, что для любого набора индексов W шенноновская энтропия случайной величины φ^W мало отличаются от колмогоровской сложности $K(x^W)$.

Лемма 4. *Для любого непустого* $W \subseteq \{1, \ldots, n\}$ $H(\varphi^W) = K(x^W) + \mathcal{O}(\log m)$.

Доказательство леммы. Если случайная величина φ^W принимает с положительной вероятностью значение y^W, то по построению множества A выполнено неравенство $K(y^W) \leq K(x^W)$. Число кортежей y^W, сложность которых не больше $K(x^W)$, не превосходит $2^{K(x^W)+1}$. Таким образом, случайная величина φ^W принимает не более $2^{K(x^W)+1}$ значений. Шенноновская энтропия случайной величины не превосходит логарифма числа ее значений. Следовательно, $H(\varphi^W) \leq K(x^W) + 1$.

Теперь докажем обратное неравенство. Обозначим

$$\neg W = \{1, 2, \ldots, n\} \setminus W.$$

Пусть $W = \{i_1, i_2 \ldots, i_k\}$, и $z = \langle z^{i_1}, \ldots, z^{i_k} \rangle$ – некоторое значение случайной величины φ^W. Оценим сверху вероятность $\mathrm{Prob}[\varphi^W = z]$. Согласно определению φ

$$\mathrm{Prob}[\varphi^W = z] = \frac{|\{y \in A : y^W = z\}|}{|A|}.$$

По лемме 3 имеем $|A| \geq 2^{K(x)} - \mathcal{O}(\log m)$. Оценим $|\{y \in A : y^W = z\}|$.

Для всех $y \in A$ должно выполняться неравенство $K(y^{\neg W}|y^W) \leq K(x^{\neg W}|x^W)$. Поэтому если кортеж y принадлежит A и $y^W = z$, то

$$K(y^{\neg W}|y^W) = K(y^{\neg W}|z) \leq K(x^{\neg W}|x^W).$$

Число кортежей y, сложность которых относительно z не превосходит $K(x^{\neg W}|x^W)$, не больше $2^{K(x^{\neg W}|x^W)+1}$. Таким образом,

$$|\{y \in A : y^W = z\}| \leq 2^{K(x^{\neg W}|x^W)+1}.$$

Получаем

$$\mathrm{Prob}[\varphi^W = z^W] \leq \frac{2^{K(x^{\neg W}|x^W)+1}}{2^{K(x)+\mathcal{O}(\log m)}}.$$

Следовательно,

$$\begin{aligned}
&\log\left(\mathrm{Prob}[\varphi^W = z]\right) \leq K(x^{\neg W}|x^W) - K(x) + \mathcal{O}(\log m) = \\
&= K(x^{\neg W}|x^W) - (K(x^W) + K(x^{\neg W}|x^W)) + \mathcal{O}(\log m) = \\
&= -K(x^W) + \mathcal{O}(\log m).
\end{aligned}$$

Теперь мы можем получит нижнюю оценку для $H(\varphi^W)$.

$$H(\varphi^W) = -\sum_z \mathrm{Prob}[\varphi^W = z] \cdot \log\left(\mathrm{Prob}[\varphi^W = z]\right) \geq$$
$$\geq \sum_z \mathrm{Prob}[\varphi^W = z](K(x^W) + \mathcal{O}(\log m)) = K(x^W) + \mathcal{O}(\log m).$$

С другой стороны, как мы отмечали выше,

$$H(\varphi^W) \leq K(x^W) + 1.$$

Таким образом, $H(\varphi^W) = K(x^W) + \mathcal{O}(\log m)$. □

Для доказательства теоремы остается записать для случайной величины φ неравенство (2.1) и заменить в нем все величины $H(\varphi^W)$ на $K(x^W) + \mathcal{O}(\log m)$. Мы получим неравенство

$$\sum \lambda_W K(x^W) + \mathcal{O}(\log m) \geq 0,$$

и теорема доказана. □

Далее рассмотрим неравенства для размерностей линейных подпространств. Пусть задано конечномерное линейное пространство L над некоторым конечным полем или над $\mathbb{R}$ и линейный подпространства $a_1, a_2, \ldots, a_n$ в L. Будем интересоваться линейными неравенствами, связывающими размерности данных подпространств, размерности сумм пар подпространств, сумм троек подпространств и т. д.

Определение 5. *Обозначим $\Lambda_L(n)$ множество всех таких наборов из $(2^n - 1)$ коэффициентов $\{\lambda_W\}$, что для любого конечномерного линейного пространства L над произвольным конечным полем или над $\mathbb{R}$ и для любых линейных подпространств $a_1, \ldots, a_n$ в L выполнено неравенство*

$$\sum_W \lambda_W \dim\left(\bigoplus_W x^W\right) \geq 0, \tag{2.5}$$

(сумма берется по всем непустым $W \subseteq \{1, \ldots, n\}$).

Множество $\Lambda_L(n)$ мы будем называть классом линейных неравенств, выполненных для размерностей n-ок линейных подпространств.

Согласно следующей теореме (Верещагин, Хаммер, Шень [16]) всякое неравенство для шенноновской энтропии верно и для размерностей подпространств.

Теорема 2. *Для любого n $\Lambda_S(n) \subseteq \Lambda_L(n)$, т. е. всякое линейное неравенство, выполненное для шенноновской энтропии, выполняется также и для размерностей подпространств.*

Доказательство. Пусть для некоторого набора коэффициентов неравенство (2.1) верно для любой n-ки случайных величин. Покажем, что неравенство (2.5) с теми же коэффициентами верно для любой n-ки линейных подпространств из конечномерного линейного пространства над конечным полем или над $\mathbb{R}$.

Сначала рассмотрим линейные пространства над конечными полями. Пусть $a_1, \ldots, a_n$ – подпространства в конечномерном пространстве L над некоторым конечным полем F. Для доказательства теоремы достаточно построить такие случайные величины $\varphi_1, \ldots, \varphi_n$, что энтропии $H(\varphi_i)$ пропорциональны соответствующим размерностям $\dim(a_i)$, энтропии пар $H(\varphi_i, \varphi_j)$ пропорциональны размерностям соответствующих сумм $\dim(a_i \bigoplus a_j)$, и т. д. В самом деле, если мы построим такие случайные величины, то для доказательства неравенства (2.5) достаточно будет подставить в (2.1) вместо энтропий $H(\varphi^W)$ пропорциональные им величины $\dim(\bigoplus\limits_{i \in W} a_i)$.

Рассмотрим равномерное распределение на множестве всех линейных функционалов $\alpha : L \to F$. Пусть a – некоторое линейное подпространство в L. Тогда ограничение α на подпространство a будет равномерно распределенной случайной величиной, принимающей $|F|^{\dim(a)}$ значений (т. к. существует $|F|^{\dim(a)}$ линейных отображений из a в F). Таким образом,

$$H(\alpha|_a) = \dim(a) \cdot \log |F|.$$

Если a и b – два линейных подпространства в L, то пара случайных величин $\langle \alpha|_a, \alpha|_b \rangle$ эквивалентна случайной величине $\alpha|_{a \oplus b}$. Следовательно, энтропия пары $\langle \alpha|_a, \alpha|_b \rangle$ равна $H(\alpha|_{a \oplus b})$. То же верно и для троек подпространств, четверок и т. д.

Определим φ_i как ограничение случайного линейного оператора α на подпространство a_i. Тогда φ^W эквивалентна ограничению α на $\bigoplus\limits_{i \in W} a_i$. Следовательно,

$$H(\varphi^W) = \dim\left(\bigoplus_{i \in W} a_i\right) \cdot \log F,$$

и утверждение теоремы доказано.

Пусть теперь L – n-мерное линейное пространство над $\mathbb{R}$. Можно считать, что на L введена структура евклидова пространства. Рассмотрим случайную величину α, равномерно распределенную на единичном шаре B в L ($B \subset L$ – это множество точек L, находящихся на расстоянии не

более 1 от нуля).

Для любого подпространства a в L можно рассмотреть случайную величину α_a, являющуюся ортогональной проекцией α на a. Данная случайная величина непрерывна, и мы должны перейти к ее дискретизации. Зафиксируем некоторое $\varepsilon > 0$ и покроем пересечение шара B с подпространством a равными непересекающимися кубиками размерности $\dim(a)$ и размера $\varepsilon \times \varepsilon \times \ldots \times \varepsilon$. Обозначим $\alpha_{a,\varepsilon}$ случайную величину, значением которой будет кубик, содержащий точку α_a. Покажем, что

$$H(\alpha_{a,\varepsilon}) = \log(1/\varepsilon) \cdot \dim(a) + \mathcal{O}(1)$$

при $\varepsilon \to 0$.

Очевидно, число кубиков $k_{a,\varepsilon}$, покрывающих пересечение B с подпространством a, есть $\mathcal{O}((1/\varepsilon)^{\dim(a)})$. Следовательно,

$$H(\alpha_{a,\varepsilon}) \leq \log(1/\varepsilon) \cdot \dim(a) + \mathcal{O}(1).$$

Докажем обратное неравенство. Пусть q – кубик из рассмотренного покрытия пересечения единичного шара с подпространством a. Оценим вероятность того, что значение α_a лежит в q.

Пусть x_0 – произвольная точка из пересечения B и подпространства a. Рассмотрим сечение шара B, перпендикулярное подпространству a и проходящее через точку x_0. Такое сечение будет шаром размерности $(\dim(L) - \dim(a))$ с радиусом не более 1.

Теперь рассмотрим множество всех точек из B, которые при ортогональном проектировании на подпространство a отображаются в куб q. Объем этого множества не превосходит произведения $(\dim(L) - \dim(a))$-мерного объема $(\dim(L) - \dim(a))$-мерного шара радиуса 1 и $\dim(a)$-мерного объема кубика q. Таким образом, объем полученного множества не превосходит $\mathcal{O}(\varepsilon^{\dim(a)})$.

Следовательно, вероятность того, что α_a лежит в кубике q не превосходит $\mathcal{O}(\varepsilon^{\dim(a)})$. Теперь мы можем оценить снизу энтропию случайной величины $\alpha_{a,\varepsilon}$.

$$H(\alpha_{a,\varepsilon}) = -\sum_q \mathrm{Prob}[\alpha_a \in q] \log \mathrm{Prob}[\alpha_a \in q] \geq$$
$$= \sum_q \mathrm{Prob}[\alpha_a \in q] \cdot (\log(1/\varepsilon) \cdot \dim(a) + \mathcal{O}(1)) = \log(1/\varepsilon) \cdot \dim(a) + \mathcal{O}(1).$$

Далее заметим, что как и в рассмотренном выше случае пространств над конечным полем, проекция $\alpha_{a_i \bigoplus a_j}$ эквивалентна паре случайных величин $\langle \alpha_{a_i}, \alpha_{a_j} \rangle$. Для ε-дискретизаций это уже не будет верно, т. к. случайная величина $\alpha_{a_i \bigoplus a_j, \varepsilon}$ и пара $\langle \alpha_{a_i,\varepsilon}, \alpha_{a_j,\varepsilon} \rangle$ не определяют друг друга

однозначно. Однако для любого фиксированного значения $\alpha_{a_i \bigoplus a_j,\varepsilon}$ имеется лишь конечное число значений $\langle \alpha_{a_i,\varepsilon}, \alpha_{a_j,\varepsilon} \rangle$ и наоборот. Следовательно, условные энтропии $H(\langle \alpha_{a_i,\varepsilon}, \alpha_{a_j,\varepsilon} \rangle | \alpha_{a_i \bigoplus a_j,\varepsilon})$ и $H(\alpha_{a_i \bigoplus a_j,\varepsilon} | \langle \alpha_{a_i,\varepsilon}, \alpha_{a_j,\varepsilon} \rangle)$ не превосходят $\mathcal{O}(1)$. Поэтому

$$H(\alpha_{a_i,\varepsilon}, \alpha_{a_j,\varepsilon}) = H(\alpha_{a_i \bigoplus a_j,\varepsilon}) + \mathcal{O}(1).$$

Аналогичные рассуждения можно провести для троек подпространств, четверок, и т. д.

Запишем теперь неравенство (2.1) для случайных величин $\varphi_i = \alpha_{a_i,\varepsilon}$. Как мы доказали, в данном неравенстве можно заменить энтропии $H(\varphi^W)$ на величины

$$\log(1/\varepsilon) \cdot \dim \left(\bigoplus_{i \in W} a_i \right) + \mathcal{O}(1).$$

Остается устремить ε к нулю. $\square$

Согласно теореме 1 классы неравенств, выполненных для колмогоровских сложностей и для шенноновских энтропий, совпадают. Для каждого n все известные неравенства из класса $\Lambda_S(n) = \Lambda_K(n)$ являются следствиями (т. е. линейными комбинациями) неравенств вида

$$H(\varphi^{U \cup W}) + H(\varphi^{V \cup W}) \geq H(\varphi^{U \cup V \cup W}) + H(\varphi^W) \tag{2.6}$$

($U, V, W \subseteq \{1, 2, \ldots, n\}$ – произвольные множества индексов) в шенноновском случае и

$$K(x^{U \cup W}) + K(x^{V \cup W}) \geq K(x^{U \cup V \cup W}) + K(x^W) - \mathcal{O}(\log m) \tag{2.7}$$

в колмогоровском случае (здесь $U, V, W \subseteq \{1, 2, \ldots, n\}$ – также произвольные множества индексов, а m – сумма длин всех слов, входящих в неравенство). Используя обычные сокращения, неравенства (2.6) и (2.7) можно переписать в виде

$$\mathcal{I}(\varphi^U : \varphi^V | \varphi^W) \geq 0$$

и

$$I(x^U : x^V | x^W) + \mathcal{O}(\log m) \geq 0$$

соответственно. Такие неравенства будем называть *базисными*.

Определение 6. *Обозначим $\Lambda_B(n)$ множество всех таких наборов из (2^n-1) коэффициентов $\{\lambda_W\}$, что неравенство*

$$\sum_W \lambda_W H(\varphi^W) \geq 0, \tag{2.8}$$

является линейной комбинацией неравенств вида (2.6).

Проверим, что всякое базисное неравенство выполняется для шенноновских энтропий и колмогоровских сложностей, т. е. $\Lambda_B(n)$ содержится в $\Lambda_S(n) = \Lambda_K(n)$.

Лемма 5. *Для любых наборов индексов U, V, W и для любых случайных величин $\varphi^1, \ldots, \varphi^n$ выполняется неравенство (2.6).*

Доказательство. Пусть α, β, γ – произвольные случайные величины. Рассмотрим релятивизованный вариант неравенства (0.10)

$$H(\alpha|\gamma) + H(\beta|\gamma) \geq H(\alpha, \beta|\gamma).$$

Прибавляя к обеим частям неравенства $2H(\gamma)$, получим

$$H(\alpha, \gamma) + H(\beta, \gamma) \geq H(\alpha, \beta, \gamma) + H(\delta).$$

Остается подставить вместо α, β и γ кортежи случайных величин φ^U, φ^V, φ^W. □

Используя теорему 1 и теорему 2, мы получаем следствие леммы 5.

Следствие 2. *Неравенство (2.7) выполнено для любых слов $x^1, \ldots, x^n$ и для любых $U, V, W \subseteq \{1, 2, \ldots, n\}$, а неравенство*

$$\dim(P \oplus R) + \dim(Q \oplus R) \geq \dim(P \oplus Q \oplus R) + \dim(R)$$

выполнено для любых кортежей линейных подпространств P, Q, R из конечномерного линейного пространства над произвольным конечным полем или над $\mathbb{R}$. (Здесь $\dim(P \oplus R)$ – размерность суммы подпространств из кортежей P и R, $\dim(Q \oplus R)$ – размерность суммы подпространств из кортежей Q и R и т. д.)

Заметим, что базисное неравенство для колмогоровских сложностей, а также указанное неравенство для размерностей подпространств легко доказать непосредственно.

Пример 6. Покажем, как из базисных неравенств (2.6) можно получить неравенства монотонности (0.9) и субаддитивности (0.10).
Обозначим $\langle \varphi^1, \varphi^2 \rangle = \langle \alpha, \beta \rangle$. Рассмотрим неравенство (2.6) для данной пары случайных величин для разных наборов индексов U, V, W.

1) Пусть $U = V = \{2\}$, а $W = \{1\}$. Тогда неравенство (2.6) принимает вид

$$H(\alpha, \beta) + H(\alpha, \beta) \geq H(\alpha, \beta) + H(\alpha),$$

и мы получаем неравенство (0.9).
1) Пусть $U = \{1\}$, $V = \{2\}$, а набор индексов W пуст. Тогда неравенство (2.6) можно переписать как

$$H(\alpha) + H(\beta) \geq H(\alpha, \beta),$$

и мы получаем неравенство (0.10).

Приведем без доказательства теорему Верещагина, Хаммера и Шеня из работы [16].

Теорема 3. *Для $n = 3$ совпадают 4 класса неравенств:*

$$\Lambda_B(n) = \Lambda_S(n) = \Lambda_K(n) = \Lambda_L(n).$$

Таким образом, для $n = 3$ классы неравенств, выполненных для шенноновской энтропии, для колмогоровской сложности и для размерностей линейных подпространств совпадают. Однако для $n = 4$ ситуация становится более сложной. А именно, имеются неравенства, выполненные для размерностей линейных подпространств, но не для шенноновской энтропии и колмогоровской сложности. Рассмотрим пример такого неравенства.

Инглетон [12] показал, что для любой четверки a, b, c, d подпространств произвольного конечномерного линейного пространства выполнено неравенство

$$\begin{aligned}\dim(a) + \dim(b) + \dim(c \oplus d) + \dim(a \oplus b \oplus c) + \dim(a \oplus b \oplus d) \leq \\ \dim(a \oplus b) + \dim(a \oplus c) + \dim(a \oplus d) + \dim(b \oplus c) + \dim(b \oplus d). \quad (2.9)\end{aligned}$$

Аналогичное неравенство для шенноновской энтропии имеет вид

$$\begin{aligned}H(\alpha) + H(\beta) + H(\gamma, \delta) + H(\alpha, \beta, \gamma) + H(\alpha, \beta, \delta) \leq \\ H(\alpha, \beta) + H(\alpha, \gamma) + H(\alpha, \delta) + H(\beta, \gamma) + H(\beta, \delta). \quad (2.10)\end{aligned}$$

Используя обычные сокращения, неравенство (2.10) можно переписать как

$$\mathcal{I}(\alpha:\beta) \le \mathcal{I}(\alpha:\beta|\gamma) + \mathcal{I}(\alpha:\beta|\delta) + \mathcal{I}(\gamma:\delta). \qquad (2.11)$$

Покажем, что неравенство Инглетона не выполняется для шенноновской энтропии.

Утверждение 3. *Существует четверка случайных величин, для которых неравенство (2.11) не выполнено.*

Доказательство. Достаточно указать четверку случайных величин $\alpha, \beta, \gamma, \delta$ таких, что

$$I(\alpha:\beta) > 0,\ I(\alpha:\beta|\gamma) = 0,\ I(\alpha:\beta|\delta) = 0,\ I(\gamma:\delta) = 0,$$

т. е. случайные величины γ и δ должны быть независимыми, а случайные величины α и β – зависимыми, но независимыми при любом фиксированном значении γ и при любом фиксированном значении δ. Предъявим пример такого совместного распределения четверки случайных величин. Каждая из величин α, β, γ, δ будет принимать значения 0 и 1. При этом γ и δ будут независимы и равномерно распределены. Условные распределения вероятностей α и β будут такими, как на рис. 2.1. Нужно проверить,

$\gamma=0$ и $\delta=0$		
α \ β	0	1
0	0	0
1	0	1

$\gamma=1$ и $\delta=0$		
α \ β	0	1
0	1/8	3/8
1	3/8	1/8

$\gamma=0$ и $\delta=1$		
α \ β	0	1
0	1/8	3/8
1	3/8	1/8

$\gamma=1$ и $\delta=1$		
α \ β	0	1
0	1	0
1	0	0

Рис. 2.1. Условные распределения вероятностей $\langle\alpha, \beta\rangle$ при известных значениях γ и δ.

что α и β будут независимы относительно γ и относительно δ. Доказательства независимости α и β при каждом из условий $\gamma = 0$, $\gamma = 1$, $\delta = 0$, $\delta = 1$ аналогичны. Рассмотрим подробно только случай $\gamma = 0$. Легко видеть, что совместное распределение α и β при условии $\gamma = 0$ будет таким, как на рис. 2.2. Матрица распределения вероятностей вы-

$\alpha \diagdown \beta$	0	1
0	1/16	3/16
1	3/16	9/16

Рис. 2.2. Распределение вероятностей $\langle \alpha, \beta \rangle$ при условии $\gamma = 0$.

рождена. Следовательно, α и β независимы при условии $\gamma = 0$.

Осталось показать, что случайные величины α и β зависимы. Легко проверить, что они имеют совместное распределение, указанное на рис. 2.3. Данная таблица распределения невырождена, а значит α и β

$\alpha \diagdown \beta$	0	1
0	5/16	3/16
1	3/16	5/16

Рис. 2.3. Распределение вероятностей $\langle \alpha, \beta \rangle$.

зависимы и $\mathcal{I}(\alpha : \beta) > 0$. $\square$

Глава 3.

Полурешетки с отношением условной простоты

В этой главе мы определим на пространстве последовательностей слов семейство отношений частичного предпорядка $\leq_f$. Данные отношения формализуют интуитивное отношение «слово y просто относительно слова x». Факторизуя пространство последовательностей по отношению эквивалентности

$$\mathbf{x} \sim_f \mathbf{y} \leftrightarrow \mathbf{x} \leq_f \mathbf{y} \wedge \mathbf{y} \leq_f \mathbf{x},$$

мы получим частично упорядоченные множества S_f. Далее мы докажем, что каждое S_f является верхней полурешеткой, но не является нижней полурешеткой. Полурешетки S_f можно рассматривать как финитный аналог степеней неразрешимости Тьюринга. В дальнейшим мы воспользуемся данными решетками как инструментом для изучения понятия общей информации.

Определение 7. *Обозначим R класс всех последовательностей двоичных слов $\mathbf{x} = \{x_n\}$, для которых существует такая константа c, что $|x_n| \leq cn$.*

Далее в этой главе мы будем предполагать, что все рассматриваемые последовательности принадлежат R.

Определим «простоту» последовательности $\mathbf{y} = \{y_n\}$ относительно последовательности $\mathbf{x} = \{x_n\}$. Наиболее естественным кажется считать, что $\mathbf{y}$ проста относительно $\mathbf{x}$, если условная сложность $K(y_n|x_n)$ ограничена логарифмически растущей величиной:

$$K(x_n|y_n) = \mathcal{O}(\log n).$$

Можно также рассматривать отношение «условной простоты», соответствующее более слабому ограничению на условную сложность. Пусть дана некоторая функция $f(n) : \mathbb{N} \to \mathbb{N}$, которая растет медленнее линейной, но не медленнее логарифма:

$$f(n) = o(n) \text{ и } f(n) = \Omega(\log n) \text{ при } n \to \infty. \tag{3.1}$$

Определение 8. *Пусть функция $f : \mathbb{N} \to \mathbb{N}$ удовлетворяет условиям (3.1). Тогда будем говорить, что последовательность $\mathbf{y} = \{y_n\}$ f-проста относительно последовательности $\mathbf{x} = \{x_n\}$ (обозначаем $\mathbf{y} \leq_f \mathbf{x}$), если $K(y_n|x_n) = \mathcal{O}(f(n))$.*

Мы получаем семейство множеств с отношениями частичного предпорядка $\langle R, \leq_f \rangle$ для всевозможных функций f.

Определение 9. *Назовем $\mathcal{S}_f$ частично упорядоченное множество, являющееся факторизацией $\langle R, \leq_f \rangle$ по отношению эквивалентности $(\mathbf{x} \sim \mathbf{y}) \leftrightarrow ((\mathbf{x} \leq_f \mathbf{y}) \wedge (\mathbf{y} \leq_f \mathbf{x}))$.*

Замечание 2. Наиболее важен случай $f(n) = \log n$. Мы будем использовать обозначения $\mathcal{S}$ и $\langle R, \leq \rangle$ вместо $\mathcal{S}_{\log n}$ и $\langle R, \leq_{\log n} \rangle$ соответственно. Мы также будем говорить «последовательность $\mathbf{y} = \{y_n\}$ проста относительно последовательности $\mathbf{x} = \{x_n\}$» вместо «последовательность $\mathbf{y} = \{y_n\}$ log-проста относительно последовательности $\mathbf{x} = \{x_n\}$».

Рассмотрим несколько простых примеров эквивалентных в S_f последовательностей.

Пример 7. 1) Рассмотрим произвольную последовательность $\mathbf{x}$ из R. Зная слово x_n и его сложность $K(x_n)$, можно перечислять множество программ, имеющих длину $K(x_n)$ и печатающих x_n. Эти программы будут оптимальными для слова x_n. (Заведомо существует хотя бы одна оптимальная программа. В общем случае таких программ может быть несколько.) Назовем p_n первую программу, получаемую в данном перечислении. Образуем из данных программ последовательность $\mathbf{p}$. Очевидно, $\mathbf{x} \sim_f \mathbf{p}$ для любой функции $f(n)$, удовлетворяющей (3.1).

2) Пусть $f(n)$ удовлетворяет (3.1). Рассмотрим произвольные последовательности $\mathbf{x}$ и $\mathbf{y}$ такие, что для любого n одно из слов x_n, y_n является началом другого, причем длины x_n и y_n отличаются на величину $\mathcal{O}(f(n))$. Иначе говоря, слово x_n получается из y_n приписыванием или отбрасыванием $\mathcal{O}(f(n))$ битов справа. Тогда $\mathbf{x} \sim_f \mathbf{y}$.

3) Пусть $f(n)$ удовлетворяет (3.1) и дана такая последовательность $\mathbf{x}$, что $K(x_n) = n + \mathcal{O}(f(n))$. Пусть также задано семейство конструктивных объектов A_n такое, что

$$|A_n| = 2^{n+\mathcal{O}(f(n))},$$

и для каждого n список элементов A_n можно породить программой длины $\mathcal{O}(f(n))$.

Тогда существует такая последовательность $\mathbf{x}'$, что

$$\forall n\ x'_n \in A_n \wedge K(x'_n) = \log |A_n| + \mathcal{O}(f(n)),$$

и $\mathbf{x}' \sim_f \mathbf{x}$.

Действительно, как и в рассмотренном выше примере 1), от последовательности $\mathbf{x}$ можно перейти к эквивалентной ей последовательности кратчайших программ $\mathbf{p}$. Каждая программа p_n имеет длину $K(x_n)$, и $K(p_n) = K(x_n) + \mathcal{O}(1)$. Далее, отрезая от слова p_n справа $\mathcal{O}(f(n))$ битов, можно получить слово x'_n, длина которого не превосходит $\log |A_n|$. Сложность полученного слова x'_n будет не более чем на $\mathcal{O}(f(n))$ отличаться от его длины. Данное x'_n можно рассматривать как номер элемента из множества A_n. Очевидно, полученная последовательность $\mathbf{x}'$ эквивалентна $\mathbf{x}$ в $\langle R, \leq_f \rangle$.

Докажем, что для любой функции f, удовлетворяющей (3.1), S_f является верхней полурешеткой, но не является решеткой. Тот факт, что любые два элемента в S_f имеют точную верхнюю грань (т. е. S_f является верхней полурешеткой), тривиален (Утверждение 4). Основным результатом данной главы является доказательство того, что некоторые пары элементов из S_f не имеют точной нижней грани. Мы рассмотрим две конструкции, позволяющие получать пары последовательностей, не имеющие точной нижней грани. Первая конструкция будет использована в доказательстве теоремы 4. Она позволит строить пары последовательностей, не имеющий точной нижней грани в S_f, при условии $f(n) = o(\sqrt{n})$. Вторая конструкция универсальна, т. е. она позволяет получать последовательности, не имеющие точной нижней грани, для любой S_f. Мы воспользуемся ей в доказательстве теоремы 5. Утверждение теоремы 4 является следствием теоремы 5. Однако мы приводим отдельное доказательство первой, менее общей теоремы, поскольку используемая в ее доказательстве конструкция может представлять самостоятельный интерес. Кроме того, предварительное рассмотрение

первой, более простой конструкции позволит сделать более ясным доказательство общей теоремы 5.

Утверждение 4. *Частично упорядоченное множество S_f является верхней полурешеткой.*

Доказательство. Достаточно доказать, что для любых последовательностей $\mathbf{x}, \mathbf{y} \in R$ существует последовательность $\mathbf{z}$, являющаяся точной верхней гранью для $\mathbf{x}$ и $\mathbf{y}$ в $\langle R, \leq_f \rangle$. Согласно определению $\mathbf{z}$ будет точной верхней гранью для $\mathbf{x}$ и $\mathbf{y}$, если

- $\mathbf{z}$ является верхней гранью для $\mathbf{x}$ и $\mathbf{y}$, т. е. $\mathbf{x} \leq_f \mathbf{z}$ и $\mathbf{y} \leq_f \mathbf{z}$, и
- $\mathbf{z} \leq_f \mathbf{u}$ для любой другой верхней грани $\mathbf{u}$ последовательностей $\mathbf{x}$ и $\mathbf{y}$.

Если $\mathbf{x} = x_1, x_2, \ldots$ и $\mathbf{y} = y_1, y_2, \ldots$, то в качестве $\mathbf{z}$ можно взять последовательность пар $z_n = \langle x_n, y_n \rangle$. Легко видеть, что последовательность $\mathbf{z} = z_1, z_2, \ldots$ будет точной верхней гранью для $\mathbf{x}$ и $\mathbf{y}$. □

Теорема 4. *Если $f(n) = o(\sqrt{n})$ и $f(n) = \Omega(\log n)$, то в S_f существуют пары элементов, не имеющие точной нижней грани.*

Доказательство. Для доказательства теоремы достаточно предъявить пару последовательностей $\mathbf{x}, \mathbf{y} \in R$, для которых нет точной нижней грани в $\langle R, \leq_f \rangle$. Чтобы определить слова x_n и y_n, мы построим множество A_n, состоящее из пар слов определенного вида. Затем возьмем в качестве $\langle x_n, y_n \rangle$ пару из данного множества с максимальной колмогоровской сложностью.

Зафиксируем n и опишем множество A_n. Введем обозначение: $m = \lfloor \sqrt{n} \rfloor$. Пусть нам даны $2m$ двоичных слов длины m

$$a_1^0, a_2^0, \ldots, a_m^0, a_1^1, a_2^1, \ldots, a_m^1, \;\; a_j^i \in \{0,1\}^m.$$

и последовательность из m нулей и единиц

$$\epsilon = \epsilon_1 \epsilon_2 \ldots \epsilon_m, \;\; \epsilon_i \in \{0,1\}.$$

Будем называть слова a_j^i *блоками*. Составим из всех $2m$ блоков слово $\bar{x}$

$$\bar{x} = \langle a_1^0, \ldots, a_m^0, a_1^1, \ldots, a_m^1 \rangle$$

а из блоков, соответствующих битам слова ϵ, слово $\bar{y}$

$$\bar{y} = \langle a_1^{\epsilon_1}, a_2^{\epsilon_2}, \ldots, a_m^{\epsilon_m} \rangle$$

Отметим, что $K(\bar{x}) \leq 2m^2 + \mathcal{O}(\log n)$ и $K(\bar{y}) \leq m^2 + \mathcal{O}(\log n)$.

Пусть A_n – множество всех пар $\langle \bar{x}, \bar{y} \rangle$, которые можно получить указанным способом (т. е. все пары слов, которые дает описанная конструкция при некотором выборе блоков a_j^i и слова ϵ). Подсчитаем количество элементов в A_n.

Для каждого $i = 1, \ldots, m$ имеется по 2^{2m} вариантов для выбора пары блоков a_i^0 и a_i^1. При этом в 2^m случаях блоки оказываются одинаковыми, и в $(2^{2m} - 2^m)$ случаях – различными. Если $a_i^0 = a_i^1$, то i-ый блок слова $\bar{y}$ определен однозначно. Если же $a_i^0 \neq a_i^1$, то для выбора i-ого блока слова $\bar{y}$ имеется два варианта; какой именно блок будет стоять на i-ом месте в $\bar{y}$, зависит от значения ϵ_i. Таким образом, для каждого i есть $(2(2^{2m} - 2^m) + 2^m)$ способов выбрать a_i^0, a_i^1 и i-ый блок слова $\bar{y}$. Следовательно, A_n состоит из

$$(2 \cdot (2^{2m} - 2^m) + 2^m)^m$$

пар.

Если пара $\langle x_n, y_n \rangle$ является элементом A_n, имеющим максимальную сложность, то согласно лемме 1

$$\begin{aligned} K(x_n, y_n) &= \log[(2 \cdot (2^{2m} - 2^m) + 2^m)^m] + \mathcal{O}(\log n) = \\ &= m(2m+1) + \mathcal{O}(\log n) = 2m^2 + m + \mathcal{O}(\log n). \end{aligned} \tag{3.2}$$

Таким образом, пара последовательностей $\mathbf{x}$, $\mathbf{y}$ определена. Докажем, что она не имеет точной нижней грани. Предположим противное: пусть последовательность $\mathbf{z}$ является точной нижней гранью $\mathbf{x}$ и $\mathbf{y}$. Чтобы получить противоречие, нам потребуется рассмотреть некоторый класс нижних граней $\mathbf{x}$ и $\mathbf{y}$.

Для каждого n слово x_n состоит из $2\lfloor\sqrt{n}\rfloor$ блоков длины $\lfloor\sqrt{n}\rfloor$. Чтобы различать блоки, соответствующие разным n, будем называть блоки, составляющие слово x_n, *n-блоками*. При этом будем называть n-блоки, входящие в слово y_n, *выделенными*, а блоки, входящие в x_n, но не входящие в y_n, – *невыделенными*. Отметим, что все блоки, входящие в слово x_n, различны (иначе сложность пары $\langle x_n, y_n \rangle$ была бы слишком мала).

Очевидно, любой выделенный n-блок имеет логарифмическую сложность относительно x_n и y_n. Рассмотрим последовательность $\mathbf{u}$, в кото-

рой каждый член u_n есть некоторый выделенный n-блок. Тогда последовательность $\mathbf{u}$ будет нижней гранью $\mathbf{x}$ и $\mathbf{y}$, а значит, будет f-проста относительно $\mathbf{z}$.

Таким образом, все выделенные n-блоки имеют сложность $Cf(n)$ относительно z_n. Вообще говоря, множители C перед $f(n)$ могут быть разными для разных выделенных блоков. Докажем, что среди этих констант можно выбрать наибольшую.

Лемма 6. *Существует такая константа C, что для любого выделенного n-блока b*

$$K(b|z_n)) \leq C \cdot f(n).$$

Доказательство леммы. Для каждого n выберем среди всех выделенных n-блоков тот, который имеет максимальную сложность относительно z_n. Обозначим его $b(n)$. Последовательность $\mathbf{u} = b(1), b(2), \ldots$ является нижней гранью $\mathbf{x}$ и $\mathbf{y}$. А следовательно, $\mathbf{u}$ проста относительно $\mathbf{z}$. Это значит, что для некоторой константы C

$$\forall n \; K(b(n)|z_n) \leq C \cdot f(n).$$

Поскольку $b(n)$ имеет максимальную сложность относительно z_n среди всех выделенных n-блоков, данная константа C годится и для всех остальных выделенных блоков, т. е. для любого выделенного n-блока b выполнено неравенство $K(b|z_n) \leq C \cdot f(n)$. □

Лемма 7. *Существует такая константа C, что для всех n и для любого невыделенного n-блока b*

$$K(b|y_n)) \geq \sqrt{n} - C \cdot f(n).$$

Доказательство леммы. Для каждого n зафиксируем некоторый невыделенный n-блок $b(n)$. Заметим, что слово x_n можно получить из y_n в 3 этапа:

- находим слово $\epsilon(n)$, соответствующее паре $\langle x_n, y_n \rangle$;
- находим блок $b(n)$;
- находим $(\lfloor \sqrt{n} \rfloor - 1)$ оставшихся невыделенных n-блоков.

Сложность слова $\epsilon(n)$, а также сложность любого невыделенного блока не превосходят $\lfloor\sqrt{n}\rfloor$. Следовательно,

$$K(x_n|y_n) \le \sqrt{n} + K(b(n)|y_n) + \lfloor\sqrt{n}\rfloor(\lfloor\sqrt{n}\rfloor - 1) + \mathcal{O}(\log n) = \\ = \lfloor\sqrt{n}\rfloor^2 + K(b(n)|y_n) + \mathcal{O}(f(n)).$$

Поскольку $K(y_n) \le \lfloor\sqrt{n}\rfloor^2 + \mathcal{O}(f(n))$, имеем неравенство

$$K(x_n, y_n) = K(x_n|y_n) + K(y_n) + \mathcal{O}(\log n) \le 2\lfloor\sqrt{n}\rfloor^2 + K(b(n)|y_n) + \mathcal{O}(f(n))$$

Используя условие (3.2), получаем $K(b(n)|y_n) \ge \sqrt{n} - \mathcal{O}(f(n))$. □

Лемма 8. *Существует такая константа C, что для любого невыделенного n-блока b*

$$K(b(n)|z_n)) \ge \sqrt{n} - C \cdot f(n).$$

Доказательство леммы. По предположению $K(z_n|y_n) = \mathcal{O}(f(n))$. Следовательно, для любого невыделенного блока b

$$K(b|y_n) \le K(b|z_n) + K(z_n|y_n) + \mathcal{O}(\log n) = K(b|z_n) + \mathcal{O}(f(n)).$$

Применяя лемму 7, получаем требуемое неравенство. □

Теперь мы можем закончить доказательство теоремы. Согласно лемме 6 для достаточно больших n все выделенные n-блоки имеют сложность намного меньше $\sqrt{n}/2$ относительно z_n. В то же время, согласно лемме 8 для достаточно больших n все невыделенные n-блоки имеют сложность много больше $\sqrt{n}/2$ относительно z_n. Следовательно, применяя к z_n все программы длины не более $\sqrt{n}/2$, мы рано или поздно получим все выделенные n-блоки и не получим ни одного невыделенного.

Данное наблюдение дает нам способ построения слова y_n по x_n со сложностью $\mathcal{O}(f(n))$. В самом деле, из x_n со сложностью $\mathcal{O}(f(n))$ можно получить z_n. Далее применяем к z_n все программы длины не более $\sqrt{n}/2$, пока не выясним, какие из блоков, образующих x_n, являются выделенными. После этого составляем из выделенных блоков y_n. Таким образом,

$$K(x_n, y_n) = K(x_n) + K(y_n|x_n) + \mathcal{O}(\log n) = \\ = K(x_n) + \mathcal{O}(f(n)) \le 2(\lfloor n\rfloor)^2 + \mathcal{O}(f(n)),$$

что противоречит (3.2). □

Теорема 5. *Для любой функции $f : \mathbb{N} \to \mathbb{N}$, удовлетворяющей (3.1), в S_f существует пара элементов, не имеющих точной нижней грани.*

Доказательство. Достаточно построить последовательности слов $\mathbf{x}, \mathbf{y} \in R$, которые не имеют точной нижней грани в $\langle R, \leq_f \rangle$. Выберем функцию $g(n)$, которая растет медленнее $n/f(n)$, но быстрее $\log(n/f(n))$:

$$g(n) = o(n/f(n)),\ \log(n/f(n)) = o(g(n)) \text{ при } n \to \infty. \tag{3.3}$$

Пусть k_n – ближайшее к $n/g(n)$ число, делящееся нацело на $f(n)$. Отметим, что k_n растет быстрее, чем $f(n)$.

Зафиксируем n и построим множество пар слов A_n. Затем в качества $\langle x_n, y_n \rangle$ возьмем пару из A_n, имеющую максимальную сложность. Для простоты обозначений будем записывать слова в алфавите $\mathcal{A}_n$, состоящем из $2^{g(n)}$ букв. (Для перехода к двоичному алфавиту достаточно закодировать каждую букву из $\mathcal{A}_n$ набором из $g(n)$ нулей и единиц.)

Опишем пары $\langle \bar{x}, \bar{y} \rangle$, которые мы будем включать в A_n. Слово $\bar{x}$ будет состоять из $2k_n$ букв алфавита $\mathcal{A}_n$. Будем представлять себе $\bar{x}$ состоящим из двух половин, каждая по k_n букв:

$$x_n = \langle a^{01}a^{02}\dots a^{0k_n}, a^{11}a^{12}\dots a^{1k_n} \rangle,\ a^{ij} \in \mathcal{A}_n.$$

Слово $\bar{y}$ будет состоять из k_n букв, причем каждая i-ая его буква должна совпадать либо с i-ой буквой из левой половины $\bar{x}$, либо с i-ой буквой из правой половины $\bar{x}$:

$$y_n = \langle b^1 b^2 \dots b^{k_n} \rangle,\ \text{где } b^i = a^{0i} \text{ или } b^i = a^{1i},\ i = 1, 2, \dots, k_n.$$

Все пары $\langle \bar{x}, \bar{y} \rangle$ указанного вида и образуют A_n. Заметим, что для любой пары $\langle \bar{x}, \bar{y} \rangle \in A_n$

$$K(\bar{x}) \leq 2k_n g(n) + \mathcal{O}(\log n) \text{ и } K(\bar{y}) \leq k_n g(n) + \mathcal{O}(\log n).$$

Теперь подсчитаем число элементов в множестве A_n. Прежде всего найдем для каждого $i = 1, \dots, k_n$ число способов согласованно выбрать буквы a^{i0} и a^{i1} для слова $\bar{x}$ и букву b^i для слова $\bar{y}$. Во-первых, буквы a^{i0} и a^{i1} можно выбрать несовпадающим ($|\mathcal{A}_n|(|\mathcal{A}|_n - 1)$ вариантов). При этом для выбора буквы b^i имеется два варианта: она должна совпадать либо с a^{i0}, либо с a^{i1}. Во-вторых, буквы a^{i0} и a^{i1} можно выбрать одинаковыми ($|\mathcal{A}|$ вариантов). Тогда для выбора буквы b^i имеется только один способ. Таким образом, для выбора всех букв в словах $\bar{x}$, $\bar{y}$ есть $(2|\mathcal{A}_n|(|\mathcal{A}_n| - 1) + |\mathcal{A}_n|)^{k_n}$ вариантов. Используя (3.3), нетрудно проверить, что

$$(2|\mathcal{A}_n|(|\mathcal{A}_n| - 1) + |\mathcal{A}_n|)^{k_n} = 2^{2k_n g(n) + k_n + \mathcal{O}(f(n))}.$$

Мы выбираем из A_n пару $\langle x_n, y_n \rangle$, имеющую максимальную возможную сложность. Следовательно,

$$K(x_n, y_n) = 2k_n g(n) + k_n + \mathcal{O}(f(n)). \tag{3.4}$$

Докажем, что у построенных последовательностей **x** и **y** нет точной нижней грани.

Как и в доказательстве теоремы 4, нам потребуется рассмотреть нижние грани **x** и **y** специального вида. Для этого разрежем слово y_n на $m_n = k_n/f(n)$ подслов длины $f(n)$. Назовем эти подслова $t_n(1), \ldots, t_n(m_n)$.

Определение 10. *Будем называть (n, j)-блоками слова b в алфавите $\mathcal{A}_n$, имеющие длину $f(n)$ и удовлетворяющие следующему условию: для каждого i $(1 \le i \le f(n))$ i-ая буква слова b совпадает либо с $((j-1)f(n)+i)$-ой буквой из левой половины слова x_n, либо с $((j-1)f(n)+i)$-ой буквой из правой половины слова x_n.*

В частности, $t_n(j)$ является (n, j)-блоком. Если мы знаем слово x_n, но не знаем y_n, то мы не можем сказать, из каких именно блоков $t_n(j)$ состоит слово y_n. Однако для любого j мы можем указать список всех (n, j)-блоков (в этом списке обязательно содержится и $t_n(j)$). Заметим, что для любых n, j имеется не более $2^{f(n)}$ различных (n, j)-блоков.

Каждый блок $t_n(j)$ легко получить как по x_n, так и по y_n. Точнее, $K(t_n(j)|x_n) = \mathcal{O}(f(n))$, $K(t_n(j)|y_n) = \mathcal{O}(\log n)$. Пусть **u** такая последовательность, что каждое слово u_n есть некоторый блок $t_n(j_n)$. Тогда $\mathbf{u} \le_f \mathbf{x}$ и $\mathbf{u} \le_f \mathbf{y}$.

Предположим теперь, что у последовательностей **x** и **y** есть точная нижняя грань **z**. Тогда $\mathbf{u} \le_f \mathbf{z}$. Значит существует такая константа C_u, что $K(t_n(j_n)|z_n) \le C_u f(n)$. Константа C_u зависит от выбора последовательности **u**, но мы покажем, что среди всех таких констант существует наибольшая. Доказательство будет аналогично доказательству леммы 6.

Лемма 9. *Существует такая константа C, что для любого блока $t_n(j)$ из слова y_n выполнено неравенство $K(t_n(j)|z_n) \le C \cdot f(n)$.*

Доказательство леммы. Для каждого n выберем в слове y_n блок $t_n(j_n)$, сложность которого относительно z_n максимальна. Последовательность выбранных блоков **u** является нижней гранью **x** и **y**, а потому f-проста относительно **z**. Следовательно, для некоторой константы

C выполнено неравенство $K(u_n|z_n) \le Cf(n)$. В силу выбора **u** данная константа C является универсальной, т. е. для любого номера n и для любого блока $t_n(j)$ из y_n $K(t_n(j)|z_n) \le Cf(n)$. □

Таким образом, применяя к слову z_n программы длины не более $C \cdot f(n)$, можно получить все блоки $t_n(j)$ из слова y_n.

Далее мы докажем две леммы, аналогичные леммам 7 и 8 из доказательства теоремы 4.

Лемма 10. *Существует такая константа C, что для любого (n,j)-блока $\tilde{t}_n(j)$, который отличается от $t_n(j)$ не менее, чем в $f(n)/10$ буквах, выполнено неравенство*

$$K(\tilde{t}_n(j)|y_n)) \ge f(n)g(n)/10 - C \cdot f(n).$$

Доказательство леммы. Пусть $\tilde{t}_n(j)$ – некоторый (n,j)-блок, отличающийся от $t_n(j)$ не менее чем в $f(n)/10$ буквах. Тогда слово x_n можно получить из y_n в 4 этапа:

- установим, из какой половины слова x_n взята каждая буква слова y_n (для этого требуется k_n битов); после выполнения этого этапа будут известны k_n букв в слове x_n;
- построим блок $\tilde{t}_n(j)$;
- укажем $f(n)/10$ букв в слове $\tilde{t}_n(j)$, которые отличаются от соответствующих букв в блоке $t_n(j)$; при этом мы получим еще $f(n)/10$ букв слова x_n;
- получим оставшиеся $(k_n - f(n)/10)$ букв слова x_n (для этого требуется не более $(k_n - f(n)/10)g(n)$ битов).

Таким образом,

$$K(x_n|y_n) \le k_n + K(\tilde{t}_n(j)|y_n) + (k_n - f(n)/10)g(n) + \mathcal{O}(f(n)).$$

Поскольку $K(y_n) \le k_n g(n) + \mathcal{O}(\log n)$, получаем

$$K(x_n, y_n) = K(x_n|y_n) + K(y_n) + \mathcal{O}(\log n) \le$$
$$2k_n g(n) - f(n)g(n)/10 + k_n + K(b(n)|y_n) + \mathcal{O}(f(n)).$$

Учитывая (3.4), получаем утверждение леммы. □

Лемма 11. *Существует такая константа C, что если (n,j)-блок $\tilde{t}_n(j)$ отличается от $t_n(j)$ не менее, чем в $f(n)/10$ буквах, то*

$$K(\tilde{t}_n(j)|z_n)) \geq f(n)g(n)/10 - C \cdot f(n).$$

Доказательство леммы. Поскольку последовательность **z** f-проста относительно **y**,

$$K(\tilde{t}_n(j)|y_n) \leq K(\tilde{t}_n(j)|z_n) + K(z_n|y_n) + \mathcal{O}(\log n) = K(\tilde{t}_n(j)|z_n) + \mathcal{O}(f(n)).$$

Остается применить лемму 10. □

Согласно лемме 9 для достаточно больших n все блоки $t_n(j)$ имеют сложность намного меньше $f(n)g(n)/20$ относительно z_n. В то же время, согласно лемме 11 для достаточно больших n все (n,j)-блоки $\tilde{t}_n(j)$, отличающиеся от соответствующего $t_n(j)$ не менее чем в $f(n)/10$ буквах, имеют сложность много больше $f(n)g(n)/20$ относительно z_n. Следовательно, если применять к z_n программы длины не более $f(n)g(n)/20$, то можно получить все блоки $t_n(j)$, но нельзя получить ни одного (n,j)-блока, отличающиеся от соответствующего $t_n(j)$ более чем в $f(n)/10$ буквах. Пусть $l_n(j)$ – число всех (n,j)-блоков, имеющих сложность не более $f(n)g(n)/20$ относительно z_n. Очевидно,

$$l_n(j) \leq C^0_{f(n)} + C^1_{f(n)} + C^2_{f(n)} + \ldots + C^{f(n)/10}_{f(n)}.$$

Оценим полученную сумму сверху. Рассмотрим случайную величину ξ, равную сумме $s = f(n)$ независимых одинаково распределенных случайных величин, каждая из которых принимает значения 0 и 1 с вероятностью $1/2$. Тогда

$$\mathrm{Prob}[\xi \leq s/10] = (C^0_s + C^1_s + \ldots + C^{s/10}_s)/2^s.$$

При этом математическое ожидание ξ равно $s/2$. Воспользуемся оценкой Бернштейна для уклонения ξ от среднего значения (см. [8], неравенство (43)):

$$\mathrm{Prob}\left[\left|\frac{\xi}{s} - \frac{1}{2}\right| \geq \frac{4}{10}\right] \leq 2e^{-(\frac{4}{10})^2 s}.$$

Огрубляя оценку, получаем

$$C^0_s + C^1_s + \ldots + C^{s/10}_s = 2^s \cdot \mathrm{Prob}\left[\frac{\xi}{s} \leq \frac{1}{2} - \frac{4}{10}\right] \leq 2^{\frac{9}{10}s}.$$

Следовательно, $l_n(j) \leq 2^{9/10 f(n)}$.

Теперь оценим сложность слова y_n относительно x_n. Прежде всего из слова x_n со сложностью $\mathcal{O}(f(n))$ можно получить z_n. Далее, будем применять к слову z_n все программы длины не более $f(n)g(n)/20$ и выписывать все получаемые в результате (n,j)-блоки. Рано или поздно все блоки $t_n(j)$, входящие в y_n, будут получены. Для того, чтобы выбрать из списка всех получаемых (n,j)-блоков соответствующие слову y_n блоки $t_n(j)$, потребуется дополнительная информация, не превосходящая

$$m_n \cdot [\max_j \log l_n(j)] \leq (k_n/f(n)) \cdot 9/10 f(n) = 9/10 k_n.$$

Зная x_n и все блоки $t_n(j)$, мы можем восстановить y_n. Таким образом, получаем $K(y_n|x_n) \leq 9/10k_n + \mathcal{O}(f(n))$. Поскольку $K(x_n) \leq 2g(n)k_n + \mathcal{O}(f(n))$, имеем

$$K(x_n, y_n) = K(x_n) + K(y_n|x_n) + \mathcal{O}(\log n) \leq 2g(n)k_n + 9/10k_n + \mathcal{O}(f(n)),$$

что противоречит (3.4). Таким образом, мы получили противоречие с предположением о существовании точной нижней грани у последовательностей $\mathbf{x}$ и $\mathbf{y}$. $\square$

Глава 4.

Пары с нематериализуемой взаимной информацией

В [9] рассматривался вопрос: всегда ли для двух слов можно найти третье, которое материализовало бы их взаимную информацию? Слово z, материализующее взаимную информацию x и y, должно легко вычисляться по каждому из них. Таким образом, мы интересуемся, всегда ли есть такое слово, которое имеет малую сложность относительно каждого из двух данных, и сложность которого равна их взаимной информации.

Гач и Кёрнер дали в [9] отрицательный ответ на данный вопрос. Однако чтобы сформулировать точное утверждение, необходимо пояснить, что значит, что некоторое слово z легко получить по слову x и по слову y. Перейдем от индивидуальных слов к бесконечным последовательностям слов и будем рассматривать асимптотические свойства их сложностей. В таких терминах можно сформулировать результат Гача и Кёрнера.

Теорема 6. [9] *Существуют такие последовательности слов* **x**, **y**, *что*

$$K(x_n) = n + o(n), \quad K(y_n) = n + o(n) \quad I(x_n : y_n) = an + o(n),$$

(a – положительная константа), и для любой последовательности слов **z**, *удовлетворяющей условию*

$$K(z_n|x_n) = o(n), \quad K(z_n|y_n) = o(n),$$

выполнено $K(z_n) = o(n)$.

Иначе говоря, существуют x_n и y_n такие, что если сложность z_n мала относительно x_n и относительно y_n, то и сложность самого z_n мала. При

этом взаимная информация слов x_n и y_n растет линейно по n. Таким образом, существуют последовательности слов, у которых нельзя материализовать взаимную информацию. Более того, теорема 6 утверждает, что нельзя материализовать даже часть взаимной информации x_n и y_n. Точнее, величина материализуемой взаимной информации бесконечно мала по сравнению с n.

В работе [9] указывается семейство примеров пар $\langle x_n, y_n \rangle$, удовлетворяющие требованию теоремы 6. При этом конструкция позволяет строить такие последовательности x_n, y_n для любых значений параметра a $(0 < a < 1)$. Т. е. можно указать такие x_n и y_n, взаимная информация которых очень велика (a близко к единице), но даже ее малая часть не может быть материализована.

В работе Гача и Кёрнер не оцениваются остаточные члены. Но анализируя доказательство в [9], можно проверить, что теорема 6 останется верной, если в формулировке заменить члены $o(n)$ на $\mathcal{O}(\sqrt{n})$ (или на $\mathcal{O}(f(n))$, где $f(n)$ – любая функция, растущая быстрее $\sqrt{n}$, но медленнее линейной: $f(n) = o(n)$, $f(n) = \Omega(\sqrt{n})$). Однако в утверждениях о колмогоровской сложности естественно формулировать равенства с точностью до логарифмического члена. Поэтому кажется интересным рассмотреть усиление теоремы 6, а именно, доказать ее, заменив в формулировке члены $o(n)$ на $\mathcal{O}(\log n)$. Таким образом, естественным усилением теоремы Гача и Кёрнера является следующая теорема.

Теорема 7. [5, 13] *Для любой функции $f(n)$ такой, что $f(n) = o(n)$ и $f(n) = \Omega(\log n)$, существуют такие последовательности слов* $\mathbf{x}$, $\mathbf{y}$, *что*

$$K(x_n) = n + \mathcal{O}(f(n)), \quad K(y_n) = n + \mathcal{O}(f(n)), \quad I(x_n : y_n) = an + \mathcal{O}(f(n))$$

(a – некоторая положительная константа), и для любой последовательности слов $\mathbf{z}$, *которая f-проста относительно* $\mathbf{x}$ *и относительно* $\mathbf{y}$, *выполнено $K(z_n) = \mathcal{O}(f(n))$.*

В работах [5, 13] было получено доказательство теоремы 7 для произвольного значения параметра a $(0 < a < 1)$. Другие доказательства теоремы 7 для некоторых специальных значений a приводились также в [15, 16].

Таким образом, для любого положительного $a < 1$ можно найти такие x_n и y_n, что сложности данных слов примерно равны n, взаимная инфор-

мация примерно равна an, и их взаимную информацию нельзя материализовать. Ан. А. Мучником был поставлен вопрос: при каких значениях параметра a для каждого x_n сложности n можно подобрать y_n сложности n такое, что взаимная информация $I(x_n : y_n)$ примерно равна an, но ее нельзя материализовать? Более точно, для каких значений параметра a имеет место следующее усиление теоремы 7.

Теорема 8. *Пусть $f(n)$ – такая функция, что $f(n) = o(n)$ и $f(n) = \Omega(\log n)$, и* $\mathbf{x}$ *– такая последовательность, что $K(x_n) = n + \mathcal{O}(f(n))$. Тогда существует последовательность* $\mathbf{y}$ *такая, что*

$$K(y_n) = n + \mathcal{O}(f(n)), \quad I(x_n : y_n) = an + \mathcal{O}(f(n)),$$

и для любой последовательности слов $\mathbf{z}$*, которая f-проста относительно* $\mathbf{x}$ *и относительно* $\mathbf{y}$*, выполнено $K(z_n) = \mathcal{O}(f(n))$.*

Для $a = 1/2$ данная теорема была доказана в [13].

В данной работе рассматриваются два рассуждения, позволяющие доказывать теорему 8 для сколь угодно близких к единице значений параметра a. В первом из них используются пары слов, случайные относительно распределения особого вида. Это семейство пар является частным случаем класса примеров, предложенных в [9]. Новый метод доказательства позволяет улучшить оценку на величину материализуемой общей информации (и, тем самым, дать доказать теоремы 7 и 8). Второе рассуждение основано на новой алгебраической конструкции, обобщающей метод, использованный в [13].

Замечание 3. Для простоты будем доказывать теорему 8 только для $f(n) = \log n$. На случай произвольной $f(n)$ все рассуждения переносятся дословно.

4.1. Стохастические пары

Рассмотрим совместное распределение пары случайных величин $\langle \varphi, \psi \rangle$ со следующими свойствами: обе величины φ, ψ принимают значения 0 и 1 с вероятностью $1/2$ (т. е. являются двоичными равномерно распределенными); при этом φ и ψ принимают разные значения с вероятностью α (и, соответственно, совпадают с вероятностью $(1 - \alpha)$). Таким образом,

$$\begin{aligned} \mathrm{Prob}[\varphi = 0, \psi = 0] = \mathrm{Prob}[\varphi = 1, \psi = 1] &= \tfrac{1-\alpha}{2}, \\ \mathrm{Prob}[\varphi = 1, \psi = 0] = \mathrm{Prob}[\varphi = 0, \psi = 1] &= \tfrac{\alpha}{2}. \end{aligned}$$

Данное распределение задается таблицей на рис. 4.1.

$\varphi \backslash \psi$	0	1
0	$\frac{1-\alpha}{2}$	$\frac{\alpha}{2}$
1	$\frac{\alpha}{2}$	$\frac{1-\alpha}{2}$

Рис. 4.1. Распределение P пары случайных величин φ, ψ

Определение 11. *Назовем последовательности слов* **x**, **y** α*-парой, если они являются* P*-случайной парой относительно распределения* P*, заданного на рис.4.1.*

Напомним, что для любого распределения P существуют P-случайные кортежи.

Пусть случайные величины φ и ψ имеют совместное распределение P, указанное на 4.1, а последовательности **x** и **y** образуют α-пару. Тогда x_n, y_n являются случайными словами длины n, и отличаются друг от друга в $\alpha n + \mathcal{O}(1)$ битов.

Далее, нетрудно вычислить шенноновские энтропии случайных величин φ и ψ, а также энтропию пары $\langle \varphi, \psi \rangle$

$$H(\varphi) = H(\psi) = 1,$$
$$H(\varphi, \psi) = 1 - (\alpha \log \alpha + (1-\alpha)\log(1-\alpha)).$$

Получаем взаимную информацию данных случайных величин

$$\mathcal{I}(\varphi : \psi) = 1 + (\alpha \log \alpha + (1-\alpha)\log(1-\alpha)).$$

Обозначим величину взаимной информации

$$c(\alpha) = 1 + (\alpha \log \alpha + (1-\alpha)\log(1-\alpha)). \tag{4.1}$$

Очевидно, $c(\alpha) > 0$ при $\alpha \neq 1/2$.

Если **x** и **y** образуют α-пару, то

$$\begin{aligned} K(x_n) &= n + \mathcal{O}(\log n), \\ K(y_n) &= n + \mathcal{O}(\log n), \\ I(x_n, y_n) &= c(\alpha)n + \mathcal{O}(\log n). \end{aligned}$$

Таким образом, при $\alpha \neq 1/2$ взаимная информация x_n и y_n растет линейно по n.

Следует выделить случай $\alpha = 1/2$. Поскольку $c(1/2) = 0$, взаимная информация слов x_n и y_n равна $\mathcal{O}(\log n)$. Это соответствует интуиции: если два слова выбраны случайно и независимо, то они отличаются примерно в половине битов. Но согласно определению $1/2$-пары слова x_n и y_n как раз и должны быть парой случайных слов, отличающихся примерно в половине битов.

Именно α-пары оказываются примерами слов, удовлетворяющих теореме 7. Мы будем доказывать, что для любого $\alpha \in (0,1)$ взаимная информация случайных α-пар не материализуется. Для этого нам понадобятся следующие технические леммы.

Лемма 12. *Пусть последовательность* $\mathbf{z}$ *проста относительно* $\mathbf{x}$ *и относительно* $\mathbf{y}$, *т. е.* $K(z_n|x_n) = \mathcal{O}(\log n)$ *и* $K(z_n|y_n) = \mathcal{O}(\log n)$. *Тогда* $K(z_n) \le I(x_n : y_n) + \mathcal{O}(\log n)$.

Доказательство. Для любых x_n, y_n, z_n имеют место соотношения

$$\begin{aligned}
&K(x_n, z_n) = K(x_n) + K(z_n|x_n) + \mathcal{O}(\log n),\\
&K(x_n) - K(x_n|y_n) = I(x_n : y_n) + \mathcal{O}(\log n),\\
&K(x_n|y_n) \le K(x_n|z_n) + K(z_n|y_n) + \mathcal{O}(\log n),\\
&K(x_n|z_n) + K(z_n) = K(x_n, z_n) + \mathcal{O}(\log n).
\end{aligned}$$

Складывая их, получаем

$$K(z_n) \le K(z_n|x_n) + K(z_n|y_n) + I(x_n : y_n) + \mathcal{O}(\log n). \qquad (4.2)$$

Учитывая $K(z_n|x_n) = \mathcal{O}(\log n)$ и $K(z_n|y_n) = \mathcal{O}(\log n)$, получаем $K(z_n) \le I(x_n : y_n) + \mathcal{O}(\log n)$. $\square$

Лемма 13. *Пусть даны четыре последовательности* $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x}'$, $\mathbf{y}'$ *такие, что* x_n *и* y_n *независимы относительно* x'_n *и относительно* y'_n:

$$I(x_n : y_n|x'_n) = \mathcal{O}(\log n), \quad I(x_n : y_n|y'_n) = \mathcal{O}(\log n).$$

Тогда всякая последовательность $\mathbf{z}$, *простая относительно* $\mathbf{x}$ *и* $\mathbf{y}$ $(K(z_n|x_n) = \mathcal{O}(\log n)$, $K(z_n|y_n) = \mathcal{O}(\log n))$, *проста также и относительно* $\mathbf{x}'$ *и* $\mathbf{y}'$ $(K(z_n|x'_n) = \mathcal{O}(\log n)$, $K(z_n|y'_n) = \mathcal{O}(\log n))$.

Доказательство. Пусть $\mathbf{z}$ проста относительно $\mathbf{x}$ и $\mathbf{y}$. Покажем, что $\mathbf{z}$ проста также относительно $\mathbf{x}'$ и $\mathbf{y}'$. Рассмотрим релятивизованный вариант неравенства (4.2)

$$K(z_n|x'_n) \le K(z_n|x_n, x'_n) + K(z_n|y_n, x'_n) + I(x_n : y_n|x'_n) + \mathcal{O}(\log n).$$

Ослабляя его, получаем

$$K(z_n|x'_n) \le K(z_n|x_n) + K(z_n|y_n) + I(x_n : y_n|x'_n) + \mathcal{O}(\log n).$$

Поскольку $\mathbf{z}$ проста относительно $\mathbf{x}$ и $\mathbf{y}$, и последовательности $\mathbf{x}$ и $\mathbf{y}$ независимы относительно $\mathbf{x}'$, получаем $K(z_n|x'_n) = \mathcal{O}(\log n)$. Аналогично $K(z_n|y'_n) = \mathcal{O}(\log n)$. $\square$

Следующая лемма позволит сводить задачу о нематериализуемости взаимной информации α-пары к задаче о нематериализуемости взаимной информации некоторой β-пары, причем $\beta > \alpha$.

Лемма 14. *Пусть $\alpha < 1/2$, и последовательности $\mathbf{x}$, $\mathbf{y}$ образуют α-пару. Тогда для любого β, удовлетворяющего неравенству*

$$\alpha < \beta \le \min\{1 - \sqrt{1-2\alpha}, 1/2\},$$

существует такая β-пара $\mathbf{x}'$, $\mathbf{y}'$, что любая последовательность $\mathbf{z}$, простая относительно $\mathbf{x}$ и $\mathbf{y}$, является также простой относительно $\mathbf{x}'$ и $\mathbf{y}'$.

Доказательство. Построим последовательности $\mathbf{x}'$, $\mathbf{y}'$, которые являются случайной β-парой и

$$I(x_n : y_n|x'_n) = \mathcal{O}(\log n), \quad I(x_n : y_n|y'_n) = \mathcal{O}(\log n).$$

Тогда согласно лемме 13 всякая последовательность, простая относительно $\mathbf{x}$ и $\mathbf{y}$, будет также простой и относительно $\mathbf{x}'$ и $\mathbf{y}'$. Последовательности $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x}'$, $\mathbf{y}'$ будут образовывать четверку случайную относительно некоторого распределения P специального вида.

Рассмотрим четверку двоичных случайных величин φ_1, φ_2, φ_3 и φ_4, имеющих следующее совместное распределение P. Во-первых, потребуем, чтобы совместные распределения пар случайных величин $\langle\varphi_1, \varphi_2\rangle$ и $\langle\varphi_3, \varphi_4\rangle$ были такими, как показано на рис. 4.2.

Распределение φ_1 и φ_2		
$\varphi_1 \backslash \varphi_2$	0	1
0	$\frac{1-\alpha}{2}$	$\frac{\alpha}{2}$
1	$\frac{\alpha}{2}$	$\frac{1-\alpha}{2}$

Распределение φ_3 и φ_4		
$\varphi_3 \backslash \varphi_4$	0	1
0	$\frac{1-\beta}{2}$	$\frac{\beta}{2}$
1	$\frac{\beta}{2}$	$\frac{1-\beta}{2}$

Рис. 4.2. Проекции распределения P

$\varphi_3=0,\varphi_4=0$		
φ_1 \ φ_2	0	1
0	$\frac{1-\beta-t}{1-\beta}$	0
1	0	$\frac{t}{1-\beta}$

$\varphi_3=0,\varphi_4=1$		
φ_1 \ φ_2	0	1
0	$\frac{\beta-\alpha}{2\beta}$	$\frac{\alpha}{2\beta}$
1	$\frac{\alpha}{2\beta}$	$\frac{\beta-\alpha}{2\beta}$

$\varphi_3=1,\varphi_4=0$		
φ_1 \ φ_2	0	1
0	$\frac{\beta-\alpha}{2\beta}$	$\frac{\alpha}{2\beta}$
1	$\frac{\alpha}{2\beta}$	$\frac{\beta-\alpha}{2\beta}$

$\varphi_3=1,\varphi_4=1$		
φ_1 \ φ_2	0	1
0	$\frac{t}{1-\beta}$	0
1	0	$\frac{1-\beta-t}{1-\beta}$

Рис. 4.3. Распределения φ_1, φ_2 при фиксированных значениях φ_3, φ_4

Во-вторых, условные распределения вероятностей пары $\langle \varphi_1, \varphi_2 \rangle$ при известных значениях φ_3 и φ_4 должны быть такими, как на рис. 4.3.

В качестве значения параметра t возьмем $\frac{1-\beta-\sqrt{1-2\alpha}}{2}$. Выражение для величины t имеет смысл (подкоренное выражение неотрицательно), т. к. по условию $\alpha \leq 1/2$.

Данное определение корректно, т. е. рис. 4.2 и 4.3 действительно задают совместное распределение четверки случайных величин, если все числа, стоящие в таблицах, неотрицательны. Это условие выполнено, если $t \geq 0$ и $1-\beta-t \geq 0$. Легко проверить, что оба неравенства выполнены для выбранного значения параметра. В самом деле, первое неравенство вытекает из ограничения $\beta \leq 1 - \sqrt{1-2\alpha}$ в условии леммы, а второе неравенство выполнено для любых α и β из интервала $(0, \frac{1}{2})$.

Докажем, что при выбранном значении t случайные величины φ_1 и φ_2 независимы относительно φ_3 и φ_4, а значит

$$\mathcal{I}(\varphi_1 : \varphi_2 | \varphi_3) = 0, \quad \mathcal{I}(\varphi_1 : \varphi_2 | \varphi_4) = 0.$$

Доказательства независимости φ_1 и φ_2 при условиях $\varphi_3 = 0$, $\varphi_4 = 0$, $\varphi_3 = 1$, и $\varphi_4 = 1$ совершенно аналогичны, и мы рассмотрим только случай $\varphi_3 = 0$.

Итак, докажем, что при указанном выборе параметра t случайные величины φ_1 и φ_2 независимы при условии $\varphi_3 = 0$. Легко видеть, что совместное распределение φ_1 и φ_2 при условии $\varphi_3 = 0$ будет таким, как на рис. 4.4. Независимость пары двоичных случайных величин означает, что задающая распределение матрица из четырех чисел имеет ранг один. Остается найти такое значение t, при котором определитель матрицы, заданной на рис. 4.4, равен нулю. Мы получаем квадратное урав-

$\varphi_1 \backslash \varphi_2$	0	1
0	$1-\beta-t+\frac{\beta-\alpha}{2}$	$\frac{\alpha}{2}$
1	$\frac{\alpha}{2}$	$t+\frac{\beta-\alpha}{2}$

Рис. 4.4. Распределение φ_1 и φ_2 при условии $\varphi_3 = 0$

нение

$$(1-\beta-t+\frac{\beta-\alpha}{2})(t+\frac{\beta-\alpha}{2})-\frac{\alpha^2}{4}=0,$$

одним из корней которого и будет число $\frac{1-\beta-\sqrt{1-2\alpha}}{2}$.

Мы доказали, что случайные величины φ_1 и φ_2 независимы относительно φ_3 и φ_4. Следовательно, если последовательности слов $\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}'$ образуют P-случайную четверку, то

$$I(x_n : y_n | x'_n) = \mathcal{O}(\log n), \quad I(x_n : y_n | y'_n) = \mathcal{O}(\log n).$$

Кроме того, $\mathbf{x}$ и $\mathbf{y}$ образуют α-пару, а $\mathbf{x}'$ и $\mathbf{y}'$ – β-пару.

Для доказательства леммы остается заметить, что если дана произвольная α-пара $\langle \mathbf{x}, \mathbf{y} \rangle$, то согласно утверждению 2 ее можно расширить до P-случайной четверки $\langle \mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}' \rangle$. □

Следствие 3. *Пусть* $\frac{3}{8} < \alpha < \frac{1}{2}$, *и* $\mathbf{x}$, $\mathbf{y}$ – α*-пара. Тогда всякая последовательность* $\mathbf{z}$, *простая относительно* $\mathbf{x}$ *и* $\mathbf{y}$, *имеет логарифмическую сложность* $(K(z_n) = \mathcal{O}(\log n))$.

Доказательство. Пусть $\mathbf{z}$ – последовательность, простая относительно $\mathbf{x}$ и $\mathbf{y}$. Нетрудно проверить, что $\frac{1}{2} < 1 - \sqrt{1-2\alpha}$, т. е. для α, заданного в условии леммы, и $\beta = 1/2$ выполнено условие леммы 14. Следовательно, $\mathbf{z}$ также является простой относительно некоторых $\mathbf{x}'$ и $\mathbf{y}'$, образующих случайную $1/2-$пару. Но взаимная информация x'_n и y'_n не превосходит $\mathcal{O}(\log n)$. Из леммы 12 получаем $K(z_n) = \mathcal{O}(\log n)$. □

Пусть теперь α – произвольное число из интервала $(0, 1)$. Для того чтобы доказать, что для всякой α-пары взаимная информация не материализуется, достаточно несколько раз повторить прием из доказательства следствия 3. Проведем это рассуждение формально.

Утверждение 5. *Пусть* $0 < \alpha < 1$, *а* $\mathbf{x}$ *и* $\mathbf{y}$ – α*-пара. Тогда для всякой последовательности* $\mathbf{z}$, *простой относительно* $\mathbf{x}$ *и относительно* $\mathbf{y}$,

$$K(z_n) = \mathcal{O}(\log n).$$

Доказательство. Пусть последовательность $\mathbf{z}$ проста относительно $\mathbf{x}$ и относительно $\mathbf{y}$. Сначала отметим, что если $\mathbf{x}$ и $\mathbf{y}$ являются α-парой, то, заменив все биты слов последовательности $\mathbf{y}$ на противоположные, мы получим $(\frac{1}{2}-\alpha)-$пару с такими же свойствами материализуемости взаимной информацией. Поэтому достаточно рассмотреть $\alpha \le 1/2$.

Случай $\alpha = 1/2$ тривиален. Слова случайной $1/2-$пары независимы, т. е. $I(x_n : y_n) = \mathcal{O}(\log n)$, а по лемме 12 сложность слова z_n не больше, $I(x_n : y_n) + \mathcal{O}(\log n)$. Таким образом, остается рассмотреть $\alpha < 1/2$.

Пусть теперь $0 < \alpha < 1/2$. Выберем параметр α_1:

$$\alpha^1 = \min\{1 - \sqrt{1-2\alpha}, 1/2\}.$$

Согласно лемме 13 существует такая α^1-пара $\langle \mathbf{x}^1, \mathbf{y}^1 \rangle$, что всякая последовательность $\mathbf{z}$, простая относительно $\mathbf{x}$ и $\mathbf{y}$, является также простой и относительно $\mathbf{x}^1$, $\mathbf{y}^1$. Если $\alpha^1 = 1/2$, то $K(z_n)$ не превосходит $I(x_n^1 : y_n^1) + \mathcal{O}(\log n)$, и все доказано. В противном случае снова применим лемму 13, согласно которой существует такая α^2-пара $\langle \mathbf{x}^2, \mathbf{y}^2 \rangle$,

$$\alpha^2 = \min\{1 - \sqrt{1-2\alpha^1}, 1/2\},$$

что всякая $\mathbf{z}$, простая относительно $\mathbf{x}^1$ и $\mathbf{y}^1$, является простой и относительно $\mathbf{x}^2$, $\mathbf{y}^2$. Повторяя применение леммы 13, получим последовательность пар

$$\langle \mathbf{x}^1, \mathbf{y}^1 \rangle, \langle \mathbf{x}^2, \mathbf{y}^2 \rangle, \ldots, \langle \mathbf{x}^n, \mathbf{y}^n \rangle, \ldots,$$

где для каждого n последовательности $\mathbf{x}^n$ и $\mathbf{y}^n$ образуют α^n-пару,

$$\alpha^{n+1} = \min\{1 - \sqrt{1-2\alpha^n}, 1/2\}, \ n = 1, 2, \ldots \tag{4.3}$$

При этом всякая простая относительно $\mathbf{x}$ и $\mathbf{y}$ последовательность $\mathbf{z}$ является также простой относительно каждой из последовательностей $\mathbf{x}^n$ и $\mathbf{y}^n$. Остается доказать, что на некотором шаге будет получена $1/2$-пара:

$$\exists N \ \alpha^N = 1/2.$$

Предположим противное. Тогда $\{\alpha^n\}$ является бесконечной строго возрастающей последовательностью, все члены которой меньше $1/2$. Значит последовательность имеет некоторый предел α^∞. Подставляя α^∞ в рекуррентное соотношение (4.3), получаем

$$\alpha^\infty = 1 - \sqrt{1-2\alpha^\infty},$$

откуда $\alpha^\infty = 0$. Но это противоречит возрастанию α_n. $\square$

Утверждение 5 позволяет доказывать теорему 7 для любого значения параметра a. Действительно, согласно (4.1) для любого a из интервала $(0,1)$ существует такое α, что $c(\alpha) = a$; тогда для α-пары $\mathbf{x}$, $\mathbf{y}$

$$K(x_n) = n + \mathcal{O}(\log n),\ K(y_n) = n + \mathcal{O}(\log n),\ I(x_n : y_n) = a \cdot n + \mathcal{O}(\log n).$$

Но согласно утверждению 5 для всякой α-пары общая информация не материализуется.

Теперь легко доказать теорему 8. Пусть дана последовательность $\mathbf{x}$ такая, что $K(x_n) = n + \mathcal{O}(\log n)$. Без ограничения общности можно считать, что слова x_n имеют длину n и содержат по $n/2 + \mathcal{O}(1)$ нулей и единиц (см. пример 7). Согласно утверждению 2 для любого $\alpha \in (0,1)$ можно подобрать такую последовательность $\mathbf{y}$, что $\mathbf{x}$ и $\mathbf{y}$ образуют α-пару. Остается применить утверждение 5.

4.2. Пары ортогональных подпространств

В этом разделе мы рассмотрим вторую конструкцию, позволяющую получать последовательности слов $\mathbf{x}$, $\mathbf{y}$ с нематериализуемой взаимной информацией. Зафиксируем два параметра – натуральные числа m и k такие, что $2k < m$. Для каждого $n \in \mathbb{N}$ выберем некоторое конечное поле F_n (поле F_n будет содержать $2^{\Theta(n)}$ элементов). Обозначим через V_n линейное m-мерное пространство над F_n. Считаем, что в V_n выбран некоторый базис. Будем говорить, что векторы v и w из V_n *ортогональны*, если в фиксированной системе координат

$$v = (v^1, v^2, \ldots, v^m),\ w = (w^1, w^2, \ldots, w^m)$$

и

$$v^1 w^1 + v^2 w^2 + \ldots + v^m w^m = 0.$$

Далее, будем называть линейные подпространства $A, B \subseteq V_n$ *ортогональными*, если любой вектор из A ортогонален любому вектору из B.

В качестве x_n и y_n будем брать пары ортогональных k-мерных подпространств из V_n. Если P_n – число всех пар ортогональных k-мерных подпространств в V_n, то, очевидно, $K(x_n, y_n) \le \log P_n + \mathcal{O}(\log n)$. Будем интересоваться случайными парами $\langle x_n, y_n \rangle$, т. е. такими парами, что $K(x_n, y_n) = \log P_n + \mathcal{O}(\log n)$. Отметим, что в случае $k = 1$, $m = 3$ мы получаем конструкцию из [13].

Параметрами данной конструкции являются числа m и k, а также размеры полей F_n. Далее мы выберем такие значения параметров, что сложности x_n и y_n будут близки к n; величина $I(x_n : y_n)$ будет зависеть от отношения k и m. Наиболее интересен случай, когда k выбирается близким к $m/2$, т. к. при этом взаимная информация x_n и y_n оказывается близкой к n.

При фиксированных значениях параметров всякое слово x_n сложности n можно рассматривать как код случайного k-мерного подпространства из V_n. Для любого k-мерного подпространства x_n найдется ортогональное ему k-мерное пространство y_n такое, что условная сложность $K(y_n|x_n)$ имеет максимальную возможную величину (точнее, $K(y_n|x_n)$ есть логарифм числа k-мерных подпространств в V_n, ортогональных подпространству x_n). Имеет место однородность: все k-мерные подпространства ортогональны одинаковому числу k-мерных подпространств. Поэтому полученная пара $\langle x_n, y_n\rangle$ будет случайной, т. е. будет иметь сложность $\log P_n + \mathcal{O}(\log n)$. Чтобы получить доказательство теоремы 8, остается показать, что для случайных пар ортогональных подпространств выполняется теорема 7 (нельзя материализовать взаимную информацию). Доказательство этого факта (утверждение 6) и является основным результатом данного раздела.

Доказательство основано на следующем свойстве ортогональных подпространств. Рассмотрим граф G_n, вершинами которого являются все k-мерные подпространства из V_n. Ребрами в графе будут соединены ортогональные подпространства. Зафиксируем некоторую вершину графа v_0. Рассмотрим случайное блуждание на графе, начинающееся в v_0. Пусть v_i – вершина графа, полученная после i шагов случайного блуждания. Для каждого i случайная величина v_i распределена на множестве вершин G_n (т. е. на множестве k-мерных подпространств V_n). Мы покажем, что для некоторого s распределение v_s близко к равномерному. При этом s зависит от m и k, но не от n.

Итак, пусть для каждого n слова x_n и y_n кодируют случайную пару ортогональных k-мерных подпространств из V_n. Перейдем к формальному доказательству.

Лемма 15. *Коллмогоровские сложности и взаимная информация* **x**, **y** *растут линейно по* n*:*

$$K(x_n) = (mk - k^2)|F_n| + \mathcal{O}(\log n), \tag{4.4}$$

$$K(y_n) = (mk - k^2)|F_n| + \mathcal{O}(\log n), \tag{4.5}$$
$$I(x_n : y_n) = k^2|F_n| + \mathcal{O}(\log n). \tag{4.6}$$

Доказательство. Пусть W – линейное пространство над F_n. Найдем число последовательностей $e_1, e_2, \ldots, e_k$, состоящих из k линейно независимых векторов пространства W.

Обозначим $s = \dim(W)$ и $N = |F_n|$. В качестве e_1 можно взять любой ненулевой вектор W. Таким образом, для выбора первого вектора в последовательности имеется $N^s - 1$ вариант. Пусть вектор e_1 уже выбран. Тогда для выбора второго вектора последовательности имеется $(N^s - N)$ вариантов, поскольку e_2 не должен линейно зависеть от e_1. Далее, если выбраны первые i векторов последовательности, то в качестве вектора e_{i+1} может быть взят любой вектор, не лежащий в линейной оболочке $e_1, \ldots, e_i$. Т. е. для выбора e_{i+1} имеется $N^s - N^i$ вариантов. Следовательно, в пространстве W имеется

$$(N^s - 1)(N^s - N)\ldots(N^s - N^{k-1}) = N^{ks}(1 + \mathcal{O}(1/N))$$

последовательностей из k линейно независимых векторов. Подставляя вместо s число m, находим количество последовательностей из k линейно независимых векторов во всем пространстве V_n. Далее, подставляя вместо s число k, находим количество последовательностей из k линейно независимых векторов в каждом k-мерном подпространстве V_n. Отношение этих величин

$$Q_n = N^{mk-k^2}(1 + \mathcal{O}(1/N))$$

дает количество k-мерных подпространств в V_n. Заметим, что

$$K(x_n) \leq \log Q_n + \mathcal{O}(\log n), \quad K(y_n) \leq \log Q_n + \mathcal{O}(\log n). \tag{4.7}$$

Далее мы покажем, что для случайной пары ортогональных пространств данные неравенства обращаются в равенства.

Если подпространство x_n уже задано, то y_n лежит в подпространстве векторов V_n, ортогональных x_n. Размерность этого подпространства равна $(m - k)$. Но во всяком $(m - k)$-мерном пространстве имеется

$$T_n = N^{(m-k)k-k^2}(1 + \mathcal{O}(1/N))$$

k-мерных подпространств. Следовательно, для любой пары ортогональных подпространств $\langle x_n, y_n \rangle$

$$K(x_n|y_n) \leq \log T_n + \mathcal{O}(\log n), \; K(x_n|y_n) \leq \log T_n + \mathcal{O}(\log n). \tag{4.8}$$

В пространстве V_n имеется $(Q_n \cdot T_n)$ пар ортогональных k-мерных пространств. Поскольку пара $\langle x_n, y_n \rangle$ выбирается случайной,

$$K(x_n, y_n) = \log Q_n + \log T_n + \mathcal{O}(\log n).$$

Далее, поскольку сложность пары $K(x_n, y_n)$ с точностью до логарифмического слагаемого равна $K(x_n) + K(y_n|x_n)$ и $K(y_n) + K(x_n|y_n)$, неравенства (4.7) и (4.8) для случайной пары $\langle x_n, y_n \rangle$ обращаются в равенство.

Следовательно,

$$I(x_n : y_n) = K(x_n) + K(y_n) - K(x_n, y_n) + \mathcal{O}(\log n) = \log(\frac{Q_n}{T_n}) + \mathcal{O}(\log n).$$

Непосредственные вычисления логарифмов Q_n и T_n доказывают требуемое утверждение. □

Будем считать m, k и F_n ($|F_n| = 2^{\Theta(n)}$) выбранными такими, что

$$K(x_n) = n + \mathcal{O}(\log n), \tag{4.9}$$

$$K(y_n) = n + \mathcal{O}(\log n), \tag{4.10}$$

$$I(x_n : y_n) = an + \mathcal{O}(\log n), \tag{4.11}$$

где a – некоторая положительная константа. Таким образом, взаимная информация x_n и y_n растет линейно по n. Нетрудно заметить, что когда отношение k/m стремится к $1/2$, соответствующее значение a стремится к единице. Следовательно, выбирая значения параметров m и k, можно сделать величину a сколь угодно близкой к единице.

Покажем, что у построенных последовательностей x_n, y_n нельзя материализовать взаимную информацию, т. е. для них выполняется утверждение теоремы 7.

Утверждение 6. *Для построенных последовательностей* **x**, **y** *и для всякой последовательности* **z**, *которая проста относительно* **x** *и относительно* **y**, *выполнено равенство* $K(z_n) = \mathcal{O}(\log n)$.

Доказательство. Пусть z_n – последовательность слов простых относительно x_n и y_n, т. е. $K(z_n|x_n) = \mathcal{O}(\log n)$ и $K(z_n|y_n) = \mathcal{O}(\log n)$. Докажем, что $K(z_n) = \mathcal{O}(\log n)$. Зафиксируем натуральное n. Далее для простоты обозначений будем опускать нижний индекс n всюду, где это не приведет к путанице.

Поскольку $\mathbf{z}$ просто относительно $\mathbf{x}$ и $\mathbf{y}$,

$$\begin{aligned} K(x|z) = K(x,z) - K(z) + \mathcal{O}(\log n) &= \\ = K(x) + K(z|x) - K(z) + \mathcal{O}(\log n) &= \\ = K(x) - K(z) + \mathcal{O}(\log n). & \end{aligned} \tag{4.12}$$

Аналогичные вычисления можно провести для y. Положим

$$D = \max\{K(x|z), K(y|z)\}.$$

Отметим, что $|K(x|z) - K(y|z)| = \mathcal{O}(\log n)$. В новых обозначениях

$$K(x|z) \leq D,\ K(y|z) \leq D.$$

Далее мы докажем, что существует достаточно много слов, сложность которых относительно z не превосходит D (а значит, число D достаточно велико). Более точно, покажем, что $D = K(x) - \mathcal{O}(\log n)$. Это будет значить, что условная сложность $K(x|z)$ очень мало отличается от безусловной сложности $K(x)$. Далее, используя (4.12), мы получим логарифмическую оценку на сложность z.

В доказательстве будут рассматриваться цепочки подпространств – конечные последовательности

$$x^0 - y^1 - x^1 - y^2 - \ldots - y^r - x^r, \tag{4.13}$$

где x^i, y^i – k-мерные подпространства V_n, причем любые два соседних в цепочке подпространства ортогональны. Подпространство x_0 будем называть левым концом, а подпространство x^r – правым концом цепочки. Число r назовем длиной цепочки (r не зависит от n). Будем интересоваться только такими цепочками, в которых $x^0 = x$. Такая цепочка является траекторией случайного блуждания на графе G_n. Отметим, что число шагов блуждания четно, нечетные шаги обозначены x^i $(i = 0, 1, \ldots, r)$, а четные шаги, соответственно, y^i $(i = 1, \ldots, r)$.

Случайное блуждание на графе соответствует равномерному распределению на множестве цепочек с фиксированным левым концом. Равномерное распределение на цепочках индуцирует некоторое распределение на множестве их правых концов. Мы подберем такое значение параметра r, что получаемое распределение на множестве правых концов цепочек окажется близким к равномерному. В то же время, мы покажем, что с достаточно большой вероятностью правый конец случайно выбранной

цепочки имеет сложность не более D относительно z. Используя это, мы докажем, что число правых концов цепочек, имеющих сложность не более D относительно z, велико.

Прежде всего покажем, что для полиномиальной доли всех цепочек вида (4.13) сложность правого конца x^r мала относительно z.

Лемма 16. *Пусть X^r – множество всех цепочек вида* (4.13)*. Тогда число таких цепочек, для правых концов которых выполнено неравенство*

$$K(x^r|z) \le D,$$

не меньше, чем $\frac{|X^r|}{\mathrm{poly}(n)}$ (где $\mathrm{poly}(n)$ – некоторый многочлен).

Доказательство. Докажем более сильное утверждение. А именно, покажем, что не менее чем полиномиальную долю составляют цепочки подпространств, для которых выполняются следующие два условия:

а) *все* элементы цепочки x^i, y^i имеют сложность не более D относительно z;

б) каждая пара соседних (в цепочке) подпространств $\langle y^j, x^j\rangle$ или $\langle x^j, y^{j+1}\rangle$ случайна, т. е. имеет сложность не меньше $\log P_n - \mathcal{O}(\log n)$ (здесь и далее мы пользуемся обозначениями из доказательства утверждения 15).

Отметим, что пара $\langle y^j, x^j\rangle$ имеет сложность близкую к P_n тогда и только тогда, когда сложность $K(x^j)$ близка к $\log Q_n$, а условная сложность $K(y^j|x^j)$ близка к $\log T_n$. Точнее, случайность пар $\langle y^j, x^j\rangle$ эквивалентна тому, что для некоторой константы C выполнены неравенства

$$\begin{aligned} K(x^j) &\ge \log Q_n - C\log n, \\ K(y^j|x^j) &\ge \log T_n - C\log n. \end{aligned}$$

Доказательство леммы проведем по индукции по длине цепочки. Пусть имеется не менее $\frac{|X^i|}{n^c}$ цепочек длины i, удовлетворяющих условию леммы. Выберем любую из них и рассмотрим все возможные ее продолжения $\ldots - y^{i+1} - x^{i+1}$. При этом подпространство y^{i+1} должно быть ортогонально x^i, а x^{i+1} – ортогонально y^{i+1}. Всего имеется T_n^2 таких продолжений. Достаточно доказать, что среди этих продолжений по крайней мере полиномиальная доля удовлетворяет условиям (а) и (б).

По предположению $K(x^i|z)$ и $K(y^i|z)$ не превосходят D, и пара $\langle x^i, y^i\rangle$ случайна. (При $i = 0$ будем считать $y^0 = y$.) Учитывая определение D и

соотношение (4.12), получаем

$$K(z|x^i) = K(x^i, z) - K(x^i) + \mathcal{O}(\log n) =$$
$$= K(x^i|z) + K(z) - K(x^i) + \mathcal{O}(\log n) \leq$$
$$\leq D + K(z) - K(x) + \mathcal{O}(\log n) = \mathcal{O}(\log n).$$

Рассмотрим множество L всех k-мерных подпространств $\hat{y}$, которые ортогональны x^i и $K(\hat{y}|z) \leq D$. Заметим, что зная слово x^i можно с логарифмической сложностью получить z, а затем запустить процесс перечисления множества L. Подпространство y^i лежит в L. Поэтому для нахождения y^i достаточно иметь программу, перечисляющую L, и знать номер y^i в этом перечислении. Таким образом,

$$K(y^i|x^i) \leq \log |L| + \mathcal{O}(\log n).$$

По предположению индукции пара $\langle x^i, y^i \rangle$ случайна, и сложность y^i относительно x^i не меньше $\log T_n - C \log n$. Следовательно, $|L| \geq T_n/\mathrm{poly}(n)$.

Теперь выберем некоторую константу $C' > C$ и отбросим те подпространства из L, сложность которых относительно x^i меньше $T_n - C' \log n$. Более точно, пусть $L' \subset L$ состоит из всех таких подпространств $\hat{y}$, что

$$K(\hat{y}|x^i) \geq T_n - C' \log n.$$

Если константа C' достаточно велика, то $|L'| \geq T(n)/\mathrm{poly}(n)$. Любое подпространство из L' можно взять в качестве y^{i+1}. Действительно, для любого $\hat{y} \in L'$ пара $\langle x^i, \hat{y} \rangle$ случайна, и $K(\hat{y}|z) \leq D$.

Аналогично можно доказать, что если $y^{i+1} \in L'$ выбрано, то имеется не менее $T_n/\mathrm{poly}(n)$ подпространств $\hat{x}$, каждое из которых можно взять в качестве x^{i+1}.

Таким образом, среди T_n^2 продолжений выбранной цепочки длины i имеется не менее $\frac{T_n^2}{\mathrm{poly}(n)}$, удовлетворяющих условиям (а) и (б). (Отметим, что степень получаемого многочлена $\mathrm{poly}(n)$ зависит от r.) $\square$

Определим последовательность чисел l_i следующим рекуррентным соотношением:

$$l_0 = k, \tag{4.14}$$
$$l_{i+1} = \max\{l_i + 2k - m; 0\}. \tag{4.15}$$

Заметим, что начиная с некоторого номера все числа l_i равны нулю.

Назовем цепочку подпространств *правильной*, если для $i = 1, 2, \ldots, r$

$$\dim(x^0 \cap x^i) = l_i \tag{4.16}$$

Далее покажем, что случайно выбранная цепочка с экспоненциально близкой к единице вероятностью является правильной. Более точно, имеет место следующая лемма.

Лемма 17. *Среди цепочек подпространств вида* (4.13) *доля правильных не меньше, чем* $1 - 2^{-cn}$, *для некоторого* $c > 0$.

Доказательство леммы. Прежде всего докажем две простые комбинаторные сублеммы.

Сублемма 1. *Пусть дана система из* q_1 *линейно независимых однородных линейных уравнений с* s *переменными над полем* F_n. *Выберем случайно (относительно равномерного распределения) еще* q_2 *уравнений с теми же переменными. Тогда с вероятностью не меньше* $1 - 2^{-cn}$ *(для некоторой константы* $c > 0$*) ранг расширенной системы из* $q_1 + q_2$ *уравнений будет равен* $\min\{s; q_1 + q_2\}$.

Доказательство сублеммы. Обозначим $q = q_1 + q_2$. Пусть $q \leq s$. Покажем, что с экспоненциально близкой к единице вероятностью все q уравнений расширенной системы линейно независимы. Действительно, если уравнения оказались линейно зависимы, то хотя бы одно из q_2 случайно выбранных уравнений является линейной комбинацией остальных $(q-1)$ уравнений. Из $(q-1)$ уравнения можно составить не более $|F_n|^{q-1}$ линейных комбинаций. А для выбора случайного уравнения имеется $|F_n|^s$ способов. Таким образом, для каждого i вероятность того, что i-е уравнение системы есть линейная комбинация остальных, не превосходит $|F_n|^{(q-1)-s}$. Остается просуммировать данные вероятности по всем номерам i добавленных уравнений (от 1 до q_2). Поскольку $q_1 \leq q \leq s$, доля линейно зависимых систем уравнений не превосходит $q|F_n|^{-1}$. Но $|F_n| = 2^{\Theta(n)}$, и утверждение доказано.

Если $q > s$, то с вероятностью экспоненциально близкой к единице первые s уравнений расширенной системы линейно независимы, и ранг системы равен s. □

Замечание 4. Отметим очевидное следствие доказанной сублеммы. Пусть дана система из q_1 линейных уравнений с s переменными над

конечным полем F. Добавим к данной системе еще q_2 случайно выбранных линейных уравнений. Предположим, о некоторых совокупностях уравнений расширенной системы известно, что они оказались линейно независимыми. (Например, известно, что 1-ое, 2-ое и (q_1+1)-ое уравнение, а также 3-е, 4-ое и (q_1+2)-ое уравнение расширенной системы линейно независимы.) Очевидно, что при данном условии вероятность события "ранг всей расширенной системы равен $\min\{s; q\}$" тем более экспоненциально близка к единице.

Сублемма 2. а) *Пусть W – линейное пространство над конечным полем F, a – некоторое s_1-мерное подпространство в W. Выберем случайно (относительно равномерного распределения) s_2-мерное подпространство b в W. Тогда вероятность события $\dim(a \cap b) = r$ зависит только от величин $r, s_1, s_2, \dim(W), |F|$ (но не зависит от выбора s_1-мерного подпространства a).*

б) *Пусть W – линейное пространство над конечным полем F, a и b – s-мерные линейные подпространства W. Выберем случайно s-мерное подпространство a_1 из ортогонального дополнения a и s-мерное подпространство b_1 из ортогонального дополнения b. Тогда для любого l вероятности событий $\dim(a_1 \cap b) = l$ и $\dim(a \cap b_1) = l$ равны.*

Доказательство сублеммы. а) Пусть a' – произвольное линейное подпространство W размерности s_1. Покажем, что вероятности событий $\dim(a \cap b) = r$ и $\dim(a' \cap b) = r$ равны.

Очевидно, существует автоморфизм φ пространства W такой, что $\varphi a = a'$. Тогда для любого линейного подпространства b из пространства W имеем $\dim(a \cap b) = \dim(a' \cap \varphi b)$. Следовательно,

$$\mathrm{Prob}_b[\dim(a \cap b) = l] = \mathrm{Prob}_b[\dim(a' \cap \varphi b) = l] = \mathrm{Prob}_b[\dim(a' \cap b) = l].$$

б) Заметим, что $\dim(a^\perp \cap b) = \dim(a \cap b^\perp)$. В самом деле,

$$\dim(W) - \dim(a \cap b^\perp) = \dim((a \cap b^\perp)^\perp) = \dim(a^\perp \oplus b) =$$
$$\dim(a^\perp) + \dim(b) - \dim(a^\perp \cap b) = \dim(W) - \dim(a^\perp \cap b).$$

Остается применить первое утверждение сублеммы к пространству $a^\perp$, в котором лежат $b \cap a^\perp$ и случайно выбранное a_1, и к пространству $b^\perp$, в котором лежат $a \cap b^\perp$ и случайно выбранное b_1. □

Будем доказывать лемму индукцией по длине цепочки. База индукции очевидна. Для выполнения шага индукции достаточно показать, что

если цепочка

$$x^0 - y^1 - x^1 - y^2 - \ldots - y^i - x^i$$

правильна, то все ее продолжения

$$x^0 - y^1 - x^1 - y^2 - \ldots - y^i - x^i - y^{i+1} - x^{i+1}$$

кроме экспоненциально малой доли также правильны. Рассмотрим заключительный фрагмент данной цепочки

$$x^i - y^{i+1} - x^{i+1}.$$

Для выбора y^{i+1} и x^{i+1} имеется T_n^2 различных возможностей (в обозначениях из доказательства леммы 15). Подсчитаем, с какой вероятностью $\dim(x^0 \cap x^{i+1}) = l_{i+1}$.

Предположим, что y^{i+1} уже выбрано. Далее можно рассматривать только пару подпространств x^0 и y^{i+1}: требуется узнать, какова вероятность того, что случайно выбранное третье подпространство (мы назовем его x^{i+1}) имеет l_{i+1}-мерное пересечение с x^0 при условии, что y^{i+1} и x^{i+1} оказались ортогональными. Согласно сублемме 2(б) последняя задача эквивалентна другой: какова вероятность того, что случайно выбранное подпространство y^0 будет иметь l_{i+1}-мерное пересечение с y^{i+1} при условии, что x^0 и y^0 ортогональны.

Таким образом, для того чтобы вычислить вероятность, с которой выполняется равенство $\dim(x^0 \cap x^{i+1}) = l_{i+1}$, можно решить другую задачу: пусть y^0 – случайно выбранное k-мерное подпространство в V_n, ортогональное x^0, а y^{i+1} – случайно выбранное k-мерное подпространство в V_n, ортогональное x^i; какова вероятность того, что $\dim(y^0 \cap y^{i+1}) = l_{i+1}$?

Подпространства y^0 и y^{i+1} задаются системами из $m-k$ линейных уравнений. Без ограничения общности можно считать, что первые k уравнений в первой системе соответствуют ортогональности y^0 пространству x^0, а первые k уравнений во второй системе – ортогональности y^{i+1} пространству x^i. Можно также считать, что первые l_i уравнений в обеих системах одинаковы и соответствуют требованию ортогональности обоих подпространств y^0 и y^{i+1} подпространству $x^0 \cap x^i$.

Объединяя системы уравнений, задающие y^0 и y^{i+1}, мы получаем систему из $2(m-k) - l_i$ уравнений. Согласно сублемме 1 с вероятностью экспоненциально близкой к единице данная система уравнений имеет ранг $\min\{m, 2(m-k) - l_i\}$, а значит выполнено равенство

$$\dim(y^0 \cap y^{i+1}) = \max\{0; 2k - (m - l_i)\} = l_{i+1}.$$

Следовательно, с такой же вероятностью и $\dim(x^0 \cap x^{i+1}) = l_{i+1}$. Таким образом, с экспоненциально близкой к единице вероятностью для цепочки выполняется условие (4.16). □

Пусть A^r – множество всех k-мерных подпространств в V_n, пересечение которых с x имеет размерность l_r. Согласно лемме 17 для большинства цепочек вида (4.13) x^r принадлежит A^r. Теперь докажем, что имеет место однородность: все подпространства из A^r являются правыми концами одинакового числа правильных цепочек длины r (по-прежнему рассматриваем только такие цепочки, левый конец которых совпадает с x).

Лемма 18. *Если $v, w \in A^r$, то число правильных цепочек, правый конец которых совпадает с v, равно числу правильных цепочек, правый конец которых совпадает с w.*

Доказательство. Доказательство проведем индукцией по r. Нужно вычислить для каждого подпространства $x^r \in A^r$ количество таких цепочек

$$x^{r-1} - y^r - x^r,$$

что $x^{r-1} \in A^{r-1}$. Точнее, нужно показать, что количество таких цепочек не зависит от выбора $x^r \in A^r$, а зависит только от номера r. Зафиксируем подпространство $x^r \in A^r$ и будем выбирать случайное ортогональное ему y^r. Далее выберем случайное x^{r-1}, ортогональное y^r. Нас интересует вероятность события $x^{r-1} \in A^{r-1}$ (т. е. $\dim(x^0 \cap x^{r-1}) = l_{r-1}$).

Пусть подпространство y^r уже выбрано. Требуется определить, с какой вероятностью случайно выбранное подпространство x^{r-1} имеет l_{r-1}-мерное пересечение с x^0 при условии, что x^{r-1} и y^r ортогональны. Согласно сублемме 2(б) данная вероятность равна вероятности того, что случайно выбранное подпространство y^0 имеет l_{r-1}-мерное пересечение с y^r при условии, что y^0 и x^0 ортогональны.

Таким образом, мы можем перейти к решению новой задачи: найти вероятность того, что пара случайно выбранных k-мерных подпространств y^0, y^r, первое из которых ортогонально x^0, а второе ортогонально x^r, имеют l_{r-1}-мерное пересечение.

Подпространства y^0 и y^r задаются системами $n-k$ линейно независимых уравнений. Будем считать, что первые k уравнений первой системы соответствуют ортогональности y^0 подпространству x^0, а первые k уравнений второй системы – ортогональности y^r подпространству x^r.

Кроме того, можно считать, что первые l_r уравнений обеих систем совпадают (и соответствуют ортогональности обоих подпространств y^0, y^r подпространству $x^0 \cap x^r$). Объединим две данные системы уравнений. Пространство решений новой системы (из $2(n-k)-l_r$ уравнений) есть пересечение y^0 и y^r. Нас интересует вероятность того, что его размерность равна l_{r-1}. Очевидно, эта вероятность определяется значениями n, k и r, но не зависит от выбора конкретного x^r. Нам не требуется значение данной вероятности. Важно лишь, что оно однозначно определяется номером r (при данном n) и не зависит от выбора $x^r \in A^r$. $\square$

Закончим доказательство утверждения 6. Рассмотрим множество X^r цепочек подпространств (4.13) длины r. Согласно лемме 17 все такие цепочки за исключением экспоненциально малой доли являются правильными, а значит, их правые концы лежат в A^r. Далее, по лемме 16 по крайней мере $\frac{|X^r|}{\mathrm{poly}(n)}$ цепочек имеют правые концы, простые относительно z (т. е. их сложность относительно z не превосходит D). Значит не менее чем $\frac{|X^r|}{\mathrm{poly}(n)}$ цепочек одновременно являются правильными и имеют правый конец сложности не более D относительно z. Но в силу однородности (лемма 18) все подпространства из A^r являются правыми концами одинакового числа правильных цепочек. Следовательно, не менее чем $\frac{|A^r|}{\mathrm{poly}(n)}$ элементов множества A^r имеют сложность не больше D относительно z.

Выберем минимальное r, для которого $l_r = 0$. Тогда A^r состоит из всех k-мерных подпространств, имеющих нулевое пересечение с x. В этом случае все k-мерные подпространства из V_n кроме экспоненциально малой доли лежат в A^r. Действительно, подпространство x в V_n задается системой $(m-k)$ линейных уравнений. Выберем случайно еще $(m-k)$ линейно независимых уравнений (задающих некоторое k-мерное подпространство в V_n). Объединим две данные системы уравнений. Новая система содержит $2(m-k)$ уравнений. Поскольку $2k < m$, с экспоненциально близкой к 1 вероятностью ранг данной системы равен m (сублемма 1), и она не имеет нетривиальных решений. Это значит, что случайно выбранное подпространство в V_n с экспоненциально близкой к 1 вероятностью имеет нулевое пересечение с x.

Итак, A^r содержит все k-мерные подпространства из V_n кроме экспоненциально малой доли. Мы знаем, что не менее чем полиномиальная доля всех подпространств из A^r имеет сложность не более D_n относительно z. Следовательно, и среди всех k-мерных подпространств V_n по

крайней мере полиномиальная доля имеет сложность не более D относительно z. Таким образом, если Q_n – число всех k-мерных подпространств в V_n, то

$$2^D \geq \frac{Q_n}{\mathrm{poly}(n)}.$$

Напомним, что $K(x) = \log Q_n + \mathcal{O}(\log n)$. Далее, $D = K(x|z) + \mathcal{O}(\log n) = K(x) - K(z) + \mathcal{O}(\log n)$. Следовательно,

$$K(x) - K(z) \geq K(x) - \mathcal{O}(\log n).$$

Таким образом, $K(z) = \mathcal{O}(\log n)$. $\square$

Литература

[1] *Звонкин А.К., Левин Л.А.* Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов. // УМН. 1970. Т. 25, №6, С. 85-127.

[2] *Колмогоров А.Н.* Три подхода к определению понятия «количество информации». // Проблемы передачи информации, 1965. Т. 1, №1, С. 3-11.

[3] *Колмогоров А.Н.* К логическим основам теории информации и теории вероятностей. // Проблемы передачи информации. 1969. Т. 5, №3, С. 3-7.

[4] *Колмогоров А.Н.* Комбинаторные основания теории инфориации и исчисления вероятностей. // УМН. 1983. Т. 38, №4, С. 27-36.

[5] *Мучник Ан.А.* О выделении общей информации двух слов. // Тез. докл. Первого Всемирного Конгресса Общ-ва матем. статист. и теории вероятностей им. Бернулли. М.: Наука. 1986. С. 453.

[6] *Чисар И., Кёрнер Я.* Теория информации. Теоремы кодирования для дискретных систем без памяти. // М.: Мир. 1985.

[7] *Шёнфилд Дж.* Степени неразрешимости. // М.: Наука. 1977.

[8] *Ширяев А.Н.* Вероятность. // М.: Наука. 1989.

[9] *Gács P., Körner J.* Common Information is Far Less Then Mutual Information. // Probl. of Control and Inform. Theory. 1973. V. 2, №2, P. 149-162.

[10] *Hammer D., Shen A.* A Strange Application of Kolmogorov Complexity. // Theory of Computing Systems. 1998. V. 31, P. 1-4.

[11] *Hammer D.* Complexity inequalities. Wissenschaft & Technik Verlag, Berlin, ISBN 3–89685–479–8, 1998. 143 pp.

[12] *Ingleton A.W.* Representation of matroids. In: Welsh D.J.A., editor. *Combinatorial Mathematics and its applications.* Academic Press (London), 1971. P. 149-167.

[13] *Muchnik An.A.* On Common Information. // Theoretical Computer Science. 1998. V. 207, P. 319-328

[14] *Uspensky V.A., Shen A.* Relations between varieties of Kolmogorov complexities. // Mathematical system theory. 1996. V. 29, №3, P. 271-292.

[15] *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. // Preprint DIMACS TR 97-74, Rutgers University, 1997

[16] *Hammer D., Romashchenko A., Shen A., Vereshchagin N.* Inequalities for Shannon Entropy and Kolmogorov Complexity. // Journal of Computer and System Sciences. 2000. V. 60, P. 442-464.

[17] *Ромащенко А.Е.* Пары слов с нематериализуемой общей информацией. Проблемы передачи информации. 2000. Т. 36, Вып. 1, С. 3-20.

[18] *Ромащенко А.Е.* Полурешетка последовательности слов с отношением условной простоты. // Вестник Московского Университета, 2000. Принято в печать.