ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том 56 **2020** **Вып. 1**

УДК 621.391.1 : 519.72

© 2020 г. **Р. Арагона[1], Ф. Марци, Ф. Миньози, М. Специалетти**

ЭНТРОПИЯ И СЖАТИЕ: ПРОСТОЕ ДОКАЗАТЕЛЬСТВО НЕРАВЕНСТВА ХИНЧИНА – ОРНШТЕЙНА – ШИЛДСА

Статья посвящена фольклорному утверждению "энтропия является нижней гранью возможного сжатия данных". Точнее, используя энтропийную теорему, получено простое доказательство поточечного неравенства, впервые сформулированного Орнштейном и Шилдсом, которое является почти наверное версией неравенства в среднем, впервые доказанного Хинчиным в 1953 году. Далее дается элементарное доказательство первоначального неравенства Хинчина, которое можно использовать в качестве упражнения для студентов, изучающих теорию информации. В заключение приведены исторические и технические замечания об этом неравенстве.

Ключевые слова: эргодические источники, энтропия, сжатие без потери данных, дешифруемое кодирование, теорема Шеннона – Макмиллана.

DOI: 10.31857/S0555292320010027

Посвящается памяти профессора Альдо де Лука

§ 1. Введение и обозначения

Эта статья посвящена фольклорному утверждению "энтропия является нижней гранью возможного сжатия данных". В то время как почти каждый специалист в области теории информации знаком с этим утверждением, и его история, и наиболее общая математическая формулировка, а именно неравенство Орнштейна и Шилдса из работы [1] 1990 года, намного менее известны. Главная цель статьи – дать простое доказательство неравенства Орнштейна – Шилдса. Это простое доказательство, как и другие известные доказательства этого результата, использует известную теорему Шеннона – Макмиллана – Бреймана. Название этой теоремы относится к трем разным видам сходимости в *энтропийной теореме*. Так как мы хотим сразу перейти к делу, в этом параграфе мы просто введем необходимые обозначения и сформулируем основные результаты, которые будут использоваться в дальнейшем, а затем посвятим целый раздел историческому обзору этого неравенства, который мы считаем одной из наиболее важных частей этой статьи. Сразу хотим отметить, что Шилдс в [2, § I.1] дает два доказательства неравенства Орнштейна – Шилдса: одно из них предполагает, как и мы, выполнение энтропийной теоремы; второе доказательство довольно длинное, но не использует никаких известных глубоких результатов. Шилдс в своей книге также показал, что энтропийная теорема может быть легко выведена из неравенства Орнштейна – Шилдса 1990 года и обратного утверждения о существовании универсальных кодов, достигающих энтропийной границы.

Все обозначения, которые не определены явно в данной статье, можно найти в [2].

[1] Член Национальной группы по алгебраическим, геометрическим структурам и их приложениям (INdAM-GNSAGA) Национального института высшей математики им. Франческо Севери, Италия.

Напомним определение *типичного множества* и, так как мы планируем использовать ее в дальнейшем, теорему 3 Шеннона из работы [3], доказанную Шенноном для независимых одинаково распределенных (н.о.р.) источников, в том виде, в котором она представлена в [4]. Макмиллан в [5] называет ее свойством асимптотической равнораспределенности (САР). Для любого множества A через $|A|$ обозначается мощность A, а $\mathcal{X}^n$ – множество всех последовательностей длины n с элементами из $\mathcal{X}$.

Теперь дадим определение типичных множеств и приведем формулировку теоремы 3 Шеннона для случая н.о.р. источников.

Определение 1. Множество $A_\varepsilon^{(n)}$ последовательностей $(x_1, x_2, \dots, x_n) \in \mathcal{X}^n$ называется *типичным множеством* относительно распределения $p(x)$, если выполняется неравенство

$$2^{-n(H(X)+\varepsilon)} \leqslant p(x_1, x_2, \dots, x_n) \leqslant 2^{-n(H(X)-\varepsilon)}.$$

Теорема 1 [3, теорема 3]. 1) *Для* $(x_1, \dots, x_n) \in A_\varepsilon^{(n)}$ *выполняется неравенство*

$$H(X) - \varepsilon \leqslant -\frac{1}{n}\log_2 p(x_1, \dots, x_n) \leqslant H(X) + \varepsilon;$$

2) $\mathbf{P}\{A_\varepsilon^{(n)}\} > 1 - \varepsilon$ *для достаточно больших* n;

3) $|A_\varepsilon^{(n)}| \leqslant 2^{n(H(X)+\varepsilon)}$ *для достаточно больших* n;

4) $|A_\varepsilon^{(n)}| \geqslant (1-\varepsilon)2^{n(H(X)-\varepsilon)}$ *для достаточно больших* n.

Стоит отметить, что пункты 1) и 2) представляют собой утверждение теоремы 3 из [3], точная формулировка которой будет приведена в историческом обзоре в § 3; пункты 3) и 4) являются непосредственными следствиями пунктов 1) и 2).

Теорема Шеннона доказана в [3] для сходимости по вероятности: в самой статье только для н.о.р. источников, а в Приложении 3 – для эргодических и стационарных марковских источников. Впоследствии Макмиллан [5] сформулировал и доказал *энтропийную теорему* для стационарных(не обязательно эргодических) процессов для сходимости в среднем, т.е. в L_1. Наконец, Брейман [6, 7] получил такой же результат для стационарных и эргодических процессов и конечных алфавитов в случае сходимости почти наверное. Позднее этот результат был обобщен Чангом [8, 9] на случай счетного алфавита. Как упоминалось выше, эти версии энтропийных теорем для разных типов сходимости называются теоремой Шеннона – Макмиллана – Бреймана, или энтропийной теоремой, как в [2].

Далее сформулирована версия энтропийной теоремы для сходимости почти наверное, или поточечная версия, в том виде, в каком она была представлена в [2, теорема I.7.1].

Теорема 2. *Для любого стационарного эргодического источника и для любого* $\varepsilon > 0$ *выполнено* $(x_1, \dots, x_n) \in A_\varepsilon^{(n)}$ *асимптотически почти наверное.*

Очевидно, что из теоремы 2 следует теорема 1.

Для полноты изложения мы приводим следующее определение из [2] и лемму Бореля – Кантелли(см. [2, лемма I.1.14]).

Определение 2. Свойство P называется *измеримым*, если множество всех $\boldsymbol{x}$, для которых истинно $P(\boldsymbol{x})$, является измеримым множеством. Для последовательности $\{P_n\}$ измеримых свойств $P_n(\boldsymbol{x})$ выполняется *асимптотически почти наверное*, если для почти всех $\boldsymbol{x}$ существует число $N = N(\boldsymbol{x})$, такое что $P_n(\boldsymbol{x})$ истинно при $n \geqslant N$.

В теореме 2 в качестве P выступает свойство вектора $\boldsymbol{x} = (x_1, \dots, x_n)$ принадлежать множеству $A_\varepsilon^{(n)}$.

Лемма (Бореля – Кантелли). *Если последовательность* $\{C_n\}$ *измеримых множеств в вероятностном пространстве* (X, Σ, μ) *такова, что* $\sum \mu(C_n) < \infty$, *то для почти всех* x *найдется* $N = N(x)$, *такое что* $x \notin C_n$ *для* $n \geqslant N$.

Теперь дадим главное определение этой статьи.

Для начала напомним, что для конечного множества символов (или слов) χ через χ^* обозначается множество всевозможных последовательностей символов (или слов) из χ произвольной длины, т.е. $\chi^* = \bigcup\limits_{n=0}^{+\infty} \chi^n$.

Определение 3.

1) Двоичным *дешифруемым кодированием* (*faithful-code sequence* или *one-to-one sequence*) называется произвольная функция γ из χ^* в $\{0,1\}^*$, такая что для любого натурального n ее ограничение на χ^n инъективно.
2) Двоичным *префиксным кодированием* называется однозначно дешифруемое кодирование γ, чье ограничение на χ^n является префиксным кодом для любого натурального n.
3) Двоичным *кодированием* называется произвольная инъективная функция γ из χ^* в $\{0,1\}^*$.

Отметим, что термин *faithful code sequence* из первого пункта определения 3 используется в [1] и в [2], а термин *one-to-one code sequence* используется в [10] и во многих других работах (см. [11, 12], а также цитируемые там и цитирующие их работы). Этот же объект также классически называется *non-singular code sequence* (см., например, [4] и снова [10]).

Для простоты изложения мы ограничиваемся рассмотрением двоичных кодов. Все представленные здесь обозначения и результаты могут быть обобщены для общего случая конечных алфавитов по аналогии с работой Хинчина [13].

По определению класс дешифруемых кодирований строго включает класс префиксных кодирований и класс кодирований, поэтому любой результат, выполняющийся для класса дешифруемых кодирований также выполнен и для двух других классов, и является, с точки зрения логики, более сильным результатом. Так или иначе, добавляя некие дополнительные рассуждения и доказательства, иногда возможно получить из слабого результата более сильный. Именно это и было проделано с исторической точки зрения, и об этом рассказано в п. 3.2 настоящей статьи.

Как мог заметить читатель, для нас кодирование – это просто инъективная функция, что обеспечивает нам возможность однозначного декодирования. Если длина кодируемого сообщения известна декодеру, то дешифруемое кодирование тоже гарантируют однозначное декодирование. Отметим также, что кодирование не обязано обеспечивать сжатие в обычном смысле этого слова, так как, исходя из определения 3, длина текста может и увеличиться.

Статья организована следующим образом. В § 2 мы приводим новое доказательство неравенства Орнштейна – Шилдса, использующее теорему Шеннона – Макмиллана – Бреймана, и которое, по нашему мнению, проще любого другого, представленного в известной нам литературе. Хотя неравенство Хинчина, впервые сформулированное в [13], и следует из неравенства Орнштейна – Шилдса, в п. 2.1 мы, отталкиваясь от теоремы 3 Шеннона [3], даем новое элементарное доказательство этого неравенства без использования утверждений из теории меры, таких как лемма Бореля – Кантелли, и которое может быть использовано в качестве упражнения для студентов, изучающих теорию информации.

Параграф 3 полностью посвящен исторической справке и некоторым техническим наблюдениям, касающимся этого неравенства. Приведя небольшой обзор представленных в литературе доказательств неравенства Орнштейна – Шилдса, мы объясняем в п. 3.3, почему считаем наше доказательство более простым, чем другие.

Параграф 5 посвящен некоторым воспоминаниям третьего автора о профессоре Альдо де Лука.

§ 2. Простые доказательства

Сформулированная далее теорема доказана Орнштейном и Шилдсом в [1], но здесь мы дадим новое, более простое доказательство. Оно работает для любого стационарного и эргодического источника энтропии H.

Во избежание путаницы в обозначениях мы используем символ $\|\cdot\|$ в качестве обозначения длины последовательности символов (или слов) вместо обычно используемого символа $|\cdot|$, который здесь обозначает мощность множества.

Теорема 3. *Для любого дешифруемого кодирования γ почти наверное выполняется*

$$\liminf_{n\in\mathbb{N}} \frac{\|\gamma((x_1,\cdots,x_n))\|}{n} \geqslant H.$$

Доказательство. Для произвольного $\varepsilon > 0$ и произвольного натурального n определим множества

$$C_\varepsilon^{(n)} = \big\{(x_1,\dots,x_n) \in A_\varepsilon^{(n)} :\ \|\gamma(x_1,\dots,x_n)\| \leqslant \log_2(|A_\varepsilon^{(n)}|) - 3\varepsilon n - 1\big\}.$$

Так как γ – дешифруемое кодирование, то $|C_\varepsilon^{(n)}| = |\gamma C_\varepsilon^{(n)}|$. Множество $\gamma C_\varepsilon^{(n)}$ является подмножеством $\{0,1\}^*$, поэтому оно содержит менее 2^{t+1} строк длины не более t. Следовательно, $|C_\varepsilon^{(n)}| \leqslant |A_\varepsilon^{(n)}|2^{-3\varepsilon n}$.

Так как $C_\varepsilon^{(n)} \subseteq A_\varepsilon^{(n)}$, то вероятность каждого элемента $C_\varepsilon^{(n)}$ ограничена неравенством из определения 1. Используя этот факт, а также оценку мощности множества $A_\varepsilon^{(n)}$, данную в п. 3) теоремы 1, мы получаем, что $\mathbf{P}(C_\varepsilon^{(n)}) \leqslant 2^{-\varepsilon n}$ для достаточно больших n. Для произвольного фиксированного $\varepsilon > 0$ мы применяем лемму Бореля – Кантелли к последовательности $C_\varepsilon^{(n)}$ и, используя теорему 2, получаем, что асимптотически почти наверное $(x_1,\dots,x_n)$ принадлежит $A_\varepsilon^{(n)}$ и не принадлежит $C_\varepsilon^{(n)}$, т.е. $\|\gamma(x_1,\dots,x_n)\| \geqslant \log_2(|A_\varepsilon^{(n)}|) - 3\varepsilon n$.

Применяя оценку мощности $A_\varepsilon^{(n)}$ из п. 4) теоремы 1, получаем, что для любого ε почти наверное выполняется

$$\liminf_{n\in\mathbb{N}} \frac{\|\gamma(x_1,\dots,x_n)\|}{n} \geqslant \liminf_{n\in\mathbb{N}} \frac{\log_2(|A_\varepsilon^{(n)}|) - 3\varepsilon n}{n} \quad \geqslant H(X) - 4\varepsilon,$$

где $\dfrac{\log_2(1-\varepsilon)}{n}$ исчезает в пределе из-за того, что ε фиксировано, а $\log_2(1-\varepsilon)$ – константа. Так как это неравенство выполнено для любого ε, в том числе и для последовательности $\varepsilon_m = \dfrac{1}{m}$, $m = 1,\dots,+\infty$, то завершение доказательства является простым упражнением по анализу.

2.1. Неравенства в среднем. Здесь мы даем элементарное доказательство результата Хинчина о том, что энтропия является нижней оценкой для усредненного коэффициента сжатия, не используя лемму Бореля – Кантелли (лемма в § 1). Мы получаем его непосредственно из теоремы Шеннона (теорема 1 в § 1), которую можно найти в любой книге по теории информации (см., например, [4, 14–16]).

Очевидно, что любой результат о среднем следует из аналогичного поточечного результата, и в частности, результат этого пункта следует из теоремы 3. Как бы то ни было, мы решили оставить оба доказательства, так как доказательство результата в среднем использует только элементарные математические понятия и может

быть использовано как упражнение для изучающих теорию информации студентов по аналогии с результатом Шеннона [3, теорема 4] (см. [4, глава 3, упражнение 11]). Из результата о среднем вытекают в качестве следствий классические результаты теории информации, такие как, например, факт о том, что энтропия является нижней оценкой для средней длины однозначно декодируемых блочных кодов или нижней оценкой для коэффициента сжатия арифметического кодирования.

Отметим, что мы используем ту же идею, что и в доказательстве теоремы 3, и первые шесть строк доказательств полностью совпадают.

Теорема 4. *Для любых н.о.р. источников энтропии H и любого дешифруемого кодирования γ*

$$\liminf_{n\in\mathbb{N}} \frac{1}{n} \sum_{\boldsymbol{x}\in\mathcal{X}^n} \|\gamma(\boldsymbol{x})\| \cdot p(\boldsymbol{x}) \geqslant H.$$

Доказательство. Начало рассуждения повторяет доказательство теоремы 3 вплоть до утверждения, что $\mathbf{P}(C_\varepsilon^{(n)}) \leqslant 2^{-\varepsilon n} \leqslant \varepsilon$ при достаточно большом n.

Для любого n

$$\frac{1}{n} \sum_{\boldsymbol{x}\in\mathcal{X}^n} \|\gamma(\boldsymbol{x})\| \cdot p(\boldsymbol{x}) \geqslant \frac{1}{n} \sum_{\boldsymbol{x}\in A_\varepsilon^{(n)}\setminus C_\varepsilon^{(n)}} \|\gamma(\boldsymbol{x})\| \cdot p(\boldsymbol{x}) \geqslant$$

$$\geqslant \frac{\log_2(|A_\varepsilon^{(n)}|) - 3\varepsilon n}{n} \sum_{\boldsymbol{x}\in A_\varepsilon^{(n)}\setminus C_\varepsilon^{(n)}} p(\boldsymbol{x}) =$$

$$= \left[\frac{\log_2(|A_\varepsilon^{(n)}|)}{n} - 3\varepsilon\right] \left[\mathbf{P}(A_\varepsilon^{(n)}) - \mathbf{P}(C_\varepsilon^{(n)})\right].$$

Используя п. 4) теоремы 1, получаем $\dfrac{\log_2(|A_\varepsilon^{(n)}|)}{n} \geqslant H - 2\varepsilon$ для достаточно больших n. Из п. 2) теоремы 1 и полученной ранее оценки для $\mathbf{P}(C_\varepsilon^{(n)})$ для достаточно больших n и произвольного $\varepsilon < \dfrac{1}{2}$ получаем

$$\liminf_{n\in\mathbb{N}} \frac{1}{n} \sum_{\boldsymbol{x}\in\mathcal{X}^n} \|\gamma(\boldsymbol{x})\| \cdot p(\boldsymbol{x}) \geqslant (H - 5\varepsilon)(1 - 2\varepsilon).$$

Так как неравенство выполняется для всех $\varepsilon < \dfrac{1}{2}$, то завершение доказательства является простым упражнением по математическому анализу.

Из предыдущего доказательства мы можем получить, применяя теорему 1, желаемую нижнюю оценку для среднего коэффициента сжатия. Стоит подчеркнуть, что предыдущее доказательство может быть использовано для доказательства такой же нижней оценки в среднем также для стационарных и эргодических источников по аналогии с первоначальным доказательством Хинчина, так как теорема 1 может быть обобщена для таких типов источников. Отметим, наконец, что изначально результат Хинчина был получен для стационарных марковских цепей.

§ 3. Исторические и технические замечания

3.1. Исторический обзор. Теорема 3 была впервые доказана в [1] в 1990 году Орнштейном и Шилдсом. Точнее, она была частью их теоремы 1 в обратимом случае, доказанной в [1, § 2]. Другое простое доказательство дано в книге Шилдса [2] 1996 года вместе со вторым доказательством, похожим на первоначальное доказательство из [1, § 2] и не использующим энтропийную теорему. Неравенство сформулировано

в теореме II.1.2 в [2], и два доказательства приведены, соответственно, в пп. II.1.b и II.1.c. Третье доказательство дано в [12] Контояннисом и Верду в п. i) их теоремы 12. Упомянутые три доказательства теоремы 3 исчерпывают, насколько нам известно, все представленные в литературе доказательства.

Как отмечали Контояннис и Верду в [12] перед теоремой 12, аналогичный более слабый результат для префиксных кодирований вместо дешифруемых был доказан в [11, 17, 18]. Точнее, в диссертации Бэррона [17] доказана лемма, элементарным следствием которой является аналог теоремы 3 для префиксного кодирования. Контояннис использует ее в своей работе [11] 1997 года и утверждает: "Это неопубликованный результат, который появлялся в [17], а также в [19] в более общей форме", а затем дает доказательство леммы в Приложении. На самом деле, видимо, несколько раньше лемма Бэррона была сформулирована как неопубликованный результат его диссертации и доказана в [2]. Стоит отметить, что диссертации Бэррона [17] и Алго [19] были защищены в одном и том же году, обе в Стэнфордском университете под руководством профессора Томаса М. Ковера.

Отметим далее, что более слабый аналогичный результат для кодирований был сформулирован и доказан впервые Хинчиным в [13] в 1953 году (см. также [20]). Этот результат более слабый не только потому, что он доказан лишь для кодирований, но и потому, что он сформулирован "в среднем", а не поточечно. Заметим, что ни одна из процитированных ранее работ не цитирует, в свою очередь, Хинчина. В том числе и по этой причине мы считаем этот исторический раздел одним из наиболее важных в данной статье.

Стоит подчеркнуть, что по аналогии с теоремой Шеннона – Макмиллана – Бреймана мы решили называть теорему 3 "неравенством Хинчина – Орнштейна – Шилдса", несмотря на то, что нужно отдать должное также и Имре Чисару, о чем будет рассказано далее в п. 3.2.

Результат Хинчина восходит к золотому веку начала теории информации. Действительно, Хинчин связывает его с одной из первых теорем основополагающей работы Шеннона [3]. Следующая теорема является первоначальной формулировкой теоремы Шеннона – Макмиллана – Бреймана, в то время как Хинчин вывел свое неравенство из следующей версии Шеннона [3, теорема 4]. Обе теоремы часто присутствуют в учебниках, таких как [4, глава 3].

Теорема 5 [3, теорема 3]. *Для любых $\varepsilon > 0$ и $\delta > 0$ существует N_0, такое что последовательности произвольной длины $N \geqslant N_0$ попадают в одну из двух категорий*:

1. *Множество последовательностей, чья вероятность меньше ε*;
2. *Все остальные, для вероятностей которых выполняется неравенство*

$$\left| \frac{\log p^{-1}}{N} - H \right| < \delta.$$

Отметим, что при $\varepsilon = \delta$ из этой теоремы получается теорема 1.

Теорема 5 доказана в [3] для сходимости по вероятности: в основной части статьи только для н.о.р. источников, а в Приложении 3 – для эргодических и стационарных марковских источников. Позднее Макмиллан [5], используя теорию меры и технику мартингалов, вывел для стационарных (не обязательно эргодических) процессов для сходимости в L_1 простое следствие энтропийной эргодической теоремы, связывающее размеры кодовых книг с соответствующими энтропиями. Наконец, Брейман [6, 7] доказал такой же результат для стационарных и эргодических процессов с конечными алфавитами в случае сходимости почти наверное, также использовав теорию мартингалов. Результат Бреймана был обобщен Чангом на случай счетных алфавитов в [8, 9].

Как уже упоминалось ранее, эти три варианта энтропийной эргодической теоремы для разных типов сходимости называются теоремой Шеннона – Макмиллана – Бреймана.

Существует обширная литература, посвященная разработке более простых доказательств энтропийных эргодических теорем и их обобщений. Обычно эти доказательства не являются короткими, но зачастую используют достаточно элементарную математику. В качестве примера простого доказательства обобщения теоремы Шеннона – Макмиллана упомянем результат Киффера из [21], где приведено короткое доказательство наиболее общего из известных на тот момент не связанных с мартингалами результатов, обобщающее доказательство, представленное в [22, теорема 3.5.3]. Эти результаты успешно решают задачу для случая сходимости в L_1.

Однако задача описать историю теоремы Шеннона – Макмиллана – Бреймана, ее обобщений и следствий находится далеко за пределами данного параграфа; заинтересованного читателя мы отсылаем к работам [23–25] и цитирующим их работам.

В общем, также далеко за пределами настоящей статьи находится задача дать простое доказательство энтропийной теоремы, даже несмотря на то, что, как отмечал Шилдс в [2] и упоминали мы в § 1, неравенство Орнштейна – Шилдса тесно связано с энтропийной теоремой. Мы хотим дать новое простое доказательство неравенства Орнштейна – Шилдса, которое, по нашему мнению, проще любого другого существующего доказательства, использующего энтропийную теорему, что обсуждается далее в п. 3.3.

Возвращаясь к четвертой теореме Шеннона, расположим все последовательности произвольной длины n в порядке убывания вероятности. Для любого q, $0 < q < 1$, определим $n(q)$ как число последовательностей, которое нужно взять из этого упорядоченного множества, начиная с наиболее вероятной, чтобы набрать суммарную вероятность q для взятых последовательностей. В [3, теорема 4] утверждается, что $\lim\limits_{n\to\infty} \dfrac{\log_2 n(q)}{n} = H$.

Шеннон после формулировки теоремы 4 без каких-либо доказательств заявляет: “Мы можем интерпретировать $\log_2(n(q))$ как количество битов, необходимых для указания последовательности в случае, когда мы рассматриваем только наиболее вероятные последовательности с суммарной вероятностью q. Тогда $\dfrac{\log_2 n(q)}{n} = H$ будет количеством битов на символ, необходимым для указания последовательности. Теорема утверждает, что для больших n эта величина не зависит от q и равна H”.

Данная интерпретация Шеннона является правильной, и на основе этого ключевого заявления было доказано несколько формальных утверждений, рассматривающих функции, кодирующие блоки фиксированной длины n в блоки фиксированной длины k (см., например, [15, 26]).

Хинчин, следуя интерпретации Шеннона как исследовательскому направлению и используя теорему 4 из [3], вместо этого дает в [13] формальное определение *среднего сжатия* последовательностей фиксированной длины, а также *коэффициента сжатия* как верхнего предела среднего сжатия. Затем он доказывает в [13, теорема 4], насколько нам известно, впервые, что энтропия ограничивает снизу коэффициент сжатия любой инъективной функции. Доказательство Хинчина работает также и в случае, если в определении среднего сжатия использовать нижний предел, а не верхний, как делаем мы в теореме 4. Логически, использование нижнего предела дает более сильный результат.

В другом контексте, а именно в случае лингвистических источников, это неравенство Хинчина было доказано в [27].

3.2. От слабого неравенства к сильному. В § 2 работы Орнштейна и Шилдса 1990 года [1], в которой впервые была сформулирована и доказана теорема 3, описана тех-

ника преобразования дешифруемого кодирования в префиксное кодирование путем добавления "маленького" дополнительного заголовка. Описание этой техники занимает существенную часть их § 2, а также она описана в книге Шилдса 1996 года [2] в п. I.7.d; "маленький" дополнительный заголовок имеет размер $O(\log(\|\gamma(w)\|))$ для любого $w \in \chi^*$ и "не влияет на асимптотическое поведение", как говорится в книге Шилдса.

Мы не можем в данном коротком пункте детально описать эту технику, которая в основном состоит в добавлении к $\gamma(w)$ дельта-кода Элайеса числа $\|w\|$ [28]. Мы лишь отметим здесь, что в [1, § 2] этот метод приписывается Имре Чисару, и более того, в благодарностях авторы пишут: "Мы хотим отдельно поблагодарить Имре Чисара, который исправил несколько наших ошибок и сделал много предложений по улучшению наших рассуждений".

Более простое из двух доказательств, представленных в книге Шилдса 1996 года [2], в точности состоит в соединении техники Чисара и результата Бэррона 1985 года [17]. О нем пойдет речь в п. 3.3.

3.3. Сходства и различия доказательств. Что такое "простое" доказательство? Можем ли мы сказать, что простое и изящное доказательство, данное в конце математической книги и использующее все предыдущие результаты этой книги, действительно "простое"?

Мы считаем, что правильный ответ на второй вопрос – "нет", а ответ на первый вопрос нам неизвестен. Возможно, хорошим является вариант, предлагаемый принципом (или бритвой) Оккама, обсуждавшийся также в диссертации Бэррона [17], и позволяющий решить, какое из доказательств "проще", чем другое. В этом пункте мы анализируем три доказательства, которые были известны до нашего, и сравниваем их всех.

Как уже упоминалось выше, в [2, пп. II.1.b, II.1.c] приведены два доказательства теоремы 3. Одно из них короткое и элегантное, другое, более длинное, по словам Шилдса "было разработано в [1] и не использует теорему Шеннона – Макмиллана – Бреймана [...]".

Сначала обсудим второе, более длинное доказательство. Оно занимает более четырех страниц в книге Шилдса и использует некоторые предыдущие обозначения и результаты, но не использует никаких глубоких результатов. Даже несмотря на то, что оно не использует теорему Шеннона – Макмиллана – Бреймана, мы не можем сказать, что оно проще нашего доказательства. Мы считаем, что нельзя утверждать, что одно из этих доказательств проще другого. Мы подчеркиваем, что красота этого длинного доказательства заключается среди прочего в том, что в книге Шилдса сразу после него приведено короткое доказательство, в котором теорема Шеннона – Макмиллана – Бреймана выводится из теоремы 3 и обратного неравенства, сформулированного в [2, теорема II.1.1]. Это рассуждение демонстрирует общность и силу теоремы 3.

Теперь проанализируем три оставшихся доказательства: первое изящное доказательство, написанное в книге Шилдса 1996 года, доказательство Контоянниса и Верду 2014 года и наше доказательство теоремы 3. Все три используют лемму Бореля – Кантелли (лемма в § 1) и теорему Шеннона – Макмиллана – Бреймана.

По поводу первого изящного доказательства из [2] Шилдс пишет, что теорема 3 "следует из энтропийной теоремы вместе с удивительно простой нижней оценкой длины слов префиксного кода". Это доказательство действительно состоит из нескольких компонентов:

1) Факт, что для любого n существует основанное на технике Имре Чисара (см. п. 3.2) преобразование дешифруемого кода в префиксный код, которое "не влияет на асимптотическое поведение";

2) Применение леммы Бэррона, доказанной в [17] (см. также [2, лемма II.1.3]), которая, в свою очередь, использует

2a) неравенство Крафта для префиксных кодов [29];

2b) лемму Бореля – Кантелли (лемма в § 1);

3) Теорема Шеннона – Макмиллана – Бреймана (теорема 2).

Мы хотим подчеркнуть, что наше простое доказательство теоремы 3 использует из этих пунктов только 2b) и 3), так же как и доказательство Контоянниса и Верду, и в целом является более коротким, даже если не включать длину доказательства неравенства Крафта. Поэтому мы считаем свое доказательство более простым по сравнению с первым доказательством из книги Шилдса. Последний аргумент в пользу нашей точки зрения приведен в конце данного пункта.

Рассмотрим теперь доказательство Контоянниса и Верду части i) теоремы 12 из [12], которая в точности совпадает с нашей теоремой 3. Их доказательство использует теорему 11 из [12], в которой, в свою очередь, применяется теорема 5 из [12], являющаяся "естественным аналогом соответствующей обратной теоремы, доказанным для префиксного кодирования в [17]" Бэрроном по словам авторов. Их изящное доказательство теоремы 5 из [12] использует метод двойного подсчета для обобщения леммы Бэррона, избегая применения неравенства Крафта. Получающееся в итоге доказательство п. i) теоремы 12 из [12], даже если учитывать все эти применяемые теоремы и их доказательства, оказывается немного длиннее элегантного доказательства из книги Шилдса [2], но, по крайней мере, не использует неравенство Крафта.

В качестве последнего аргумента мы отметим, что оба разобранных доказательства [2, п. II.1.b] и [12, теорема 12], в отличие от нашего, в первой своей части показывают, что сумма по n вероятностей множеств последовательностей длины n, имеющих "малую" длину после сжатия, сходится, и это позволяет применить лемму Бореля – Кантелли. Далее эти доказательства используют теорему Шеннона – Макмиллана – Бреймана. В некотором смысле эта процедура аналогична тому, что делает Хинчин в среднем случае, когда применяет [3, теорема 4].

Наше доказательство вместо этого в своей первой части показывает, что сумма по n вероятностей множеств последовательностей длины n, которые 1) имеют малую длину после сжатия, и 2) являются типичными, сходится, и этот факт позволяет нам использовать оценки на вероятность каждого элемента типичного множества, и за счет этого упростить доказательство.

Несомненно, мы доказали более слабый результат более простым способом, но этот слабый результат все же позволяет нам получить желаемое поточечное неравенство. Мы полагаем, что интерпретация Шеннона, приведенная в конце п. 3.1, могла направлять два других доказательства по пути к более сильному результату; по этому же пути следовал и Хинчин в доказательстве для среднего случая.

§ 4. Заключение

В этой статье мы получили из энтропийной теоремы для н.о.р. источников в случае сходимости по вероятности, т.е. из теоремы 3 Шеннона в [3], а также из более общей энтропийной теоремы для стационарных и эргодических процессов в случае сходимости почти наверное, также называемой теоремой Макмиллана – Бреймана [5–7], элементарное доказательство неравенства Хинчина и простое доказательство неравенства Орнштейна – Шилдса соответственно. В частности, мы используем более глубокий классический результат для доказательства второго неравенства, которое является версией для сходимости почти наверное первого неравенства, выполненного для сходимости в среднем.

§ 5. Памяти профессора Альдо де Лука

Третий автор вспоминает рассказы и объяснения, данные ему профессором Альдо де Лука около 30 лет назад во время прогулки по бульвару Сен-Мишель в Париже. Альдо очень ценил усилия Хинчина по формализации, и даже использовал в своей исследовательской работе [30] термин “standard sequences” из английской версии работы Хинчина, а не более распространенный термин “typical sequences”.

Голос Альдо звучал очаровывающе, как голос отца, читающего красивую сказку своим сыновьям, или историю о храбром рыцаре, сражающемся за честь и математическую точность.

Авторы благодарят профессора Д. Перрена за указание на работу [2] во время конференции, проходившей в Риме 11–12 июля 2019 года, и посвященной памяти профессора Альдо де Лука, на которой была представлена предварительная версия данной статьи. Авторы также благодарят рецензентов за их замечания.

СПИСОК ЛИТЕРАТУРЫ

1. *Ornstein D., Shields P.C.* Universal Almost Sure Data Compression // Ann. Probab. 1990. V. 18. № 2. P. 441–452.
2. *Shields P.C.* The Ergodic Theory of Discrete Sample Paths. Providence, R.I.: Amer. Math. Soc., 1996.
3. *Shannon C.E.* A Mathematical Theory of Communication // Bell Syst. Tech. J. 1948. V. 27. № 3. P. 379–423.
4. *Cover T.M., Thomas J.A.* Elements of Information Theory. Hoboken, NJ: Wiley, 2006.
5. *McMillan B.* The Basic Theorems of Information Theory // Ann. Math. Statist. 1953. V. 24. № 2. P. 196–219.
6. *Breiman L.* The Individual Ergodic Theorem of Information Theory // Ann. Math. Statist. 1957. V. 28. № 3. P. 809–811.
7. *Breiman L.* Correction Notes: Correction to “The Individual Ergodic Theorem of Information Theory” // Ann. Math. Statist. 1960. V. 31. № 3. P. 809–810.
8. *Chung K.L.* A Note on the Ergodic Theorem of Information Theory // Ann. Math. Statist. 1961. V. 2. № 2. P. 612–614.
9. *Chung K.L.* The Ergodic Theorem of Information Theory // Recent Developments in Information and Decision Processes (Proc. 3rd Sympos. on Information and Decision Processes. Purdue Univ., Lafayette, IN, USA. April 12–13, 1961). New York: Macmillan, 1962. P. 141–148.
10. *Blundo C., De Prisco R.* New Bounds on the Expected Length of One-to-One Codes // IEEE Trans. Inform. Theory 1996. V. 42. № 1. P. 246–250.
11. *Kontoyiannis I.* Second-Order Noiseless Source Coding Theorems // IEEE Trans. Inform. Theory. 1997. V. 43. № 4. P. 1339–1341.
12. *Kontoyiannis I., Verdú S.* Optimal Lossless Data Compression: Non-asymptotics and Asymptotics // IEEE Trans. Inform. Theory. 2014. V. 60. № 2. P. 777–795.
13. *Хинчин А.Я.* Понятие энтропии в теории вероятностей // УМН. 1953. Т. 8. № 3 (55). С. 3–20.
14. *Yeung R.W.* Information Theory and Network Coding. Boston: Springer, 2008.
15. *Csiszár I., Körner J.* Information Theory: Coding Theorems for Discrete Memoryless Systems. Cambridge, UK: Cambridge Univ. Press, 2011.
16. *MacKay D.J.C.* Information Theory, Inference and Learning Algorithms. New York: Cambridge Univ. Press, 2003.
17. *Barron A.R.* Logically Smooth Density Estimation: PhD Thesis. Stanford Univ., CA, USA, 1985.
18. *Kieffer J.C.* Sample Converses in Source Coding Theory // IEEE Trans. Inform. Theory 1991. V. 37. № 2. P. 263–268.

19. *Algoet P.H.* Log-Optimum Investment: PhD Thesis. Stanford Univ., CA, USA, 1985.
20. *Khinchin A.I.* Mathematical Foundations of Information Theory. New York: Dover Publ., 1957.
21. *Kieffer J.C.* A Simple Proof of the Moy–Perez Generalization of the Shannon–McMillan Theorem // Pacific J. Math. 1974. V. 51. № 1. P. 203–206.
22. *Gallager R.G.* Information Theory and Reliable Communication. New York: John Wiley & Sons, 1968.
23. *Algoet P.H., Cover T.M.* A Sandwich Proof of the Shannon–McMillan–Breiman Theorem // Ann. Probab. 1988. V. 16. № 2. P. 899–909.
24. *Barron A.R.* The Strong Ergodic Theorem for Densities: Generalized Shannon–McMillan–Breiman Theorem // Ann. Probab. 1985. V. 13. № 4. P. 1292–1303.
25. *Bjelaković I., Krüger T., Siegmund-Schultze R., Szkoła A.* The Shannon–McMillan Theorem for Ergodic Quantum Lattice Systems // Invent. Math. 2004. V. 155. № 1. P. 203–222.
26. *Longo G., Sgarro A.* The Source Coding Theorem Revisited: A Combinatorial Approach // IEEE Trans. Inform. Theory. 1979. V. 25. № 5. P. 544–548.
27. *Hansel G., Perrin D., Simon I.* Compression and Entropy // Proc. 9th Annual Sympos. on Theoretical Aspects of Computer Science (STACS'92). Cachan, France. Feb. 13–15, 1992. Lect. Notes Comp. Sci. V. 577. Berlin: Springer, 1992. P. 515–528.
28. *Elias P.* Universal Codeword Sets and Representations of the Integers // IEEE Trans. Inform. Theory. 1975. V. 21. № 2. P. 194–203.
29. *Kraft L.G.* A Device for Quantizing, Grouping, and Coding Amplitude-Modulated Pulses: Ph.D. Thesis. MIT, Cambridge, USA, 1949.
30. *de Luca A.* On the Entropy of a Formal Language // Automata Theory and Formal Languages (Proc. 2nd GI Conf. Kaiserslautern, Germany. May 20–23, 1975). Lect. Notes Comp. Sci. V. 33. Berlin: Springer, 1975. P. 103–109.

Арагона Риккардо
Марци Франческа
Отделение инженерных и информационных наук и математики,
Университет Л'Аквила, Италия
riccardo.aragona@univaq.it
francesca.marzi@univaq.it
Миньози Филиппо
Отделение инженерных и информационных наук и математики,
Университет Л'Аквила, Италия
Институт высокопроизводительных вычислений и сетей,
Национальный исследовательский совет, Палермо, Италия
filippo.mignosi@univaq.it
Специалетти Маттео
Неаполитанский университет им. Фридриха II, Неаполь, Италия
matteo.spezialetti@guest.univaq.it

Поступила в редакцию
12.12.2019
После доработки
08.01.2020
Принята к публикации
15.01.2020