

ФИВТ МФТИ. Краткий конспект лекций курса
Введение в теорию информации.
(А.Е. Ромашенко, весна 2013).

1 Лекция 1, 19 февраля.

1.1 Определение информации по Хартли

Определим *количество информации* в конечном множестве $A \subset \{0, 1\}^*$ как $\mathcal{X}(A) = \log |A|$. Для многомерных множеств определяем количество информации в каждой его проекции. Например, если $A \subset \mathbb{N} \times \mathbb{N}$, то

$$\mathcal{X}(A) = \log |A|, \quad \mathcal{X}_1(A) = \log |\pi_1 A|, \quad \mathcal{X}_2(A) = \log |\pi_2 A|$$

(здесь $\pi_i A$ обозначает проекцию множества A на i -ую координату). Заметим, что $\mathcal{X}(A) \leq \mathcal{X}_1(A) + \mathcal{X}_2(A)$, причём равенство достигается, только если A есть в точности прямое произведение его проекций на первую и вторую координаты (т.е., значения проекций A на первую и вторую координату в естественном смысле независимы).

Определим количество информации во второе компоненте (проекции) A при известной первой компоненте как логарифм максимального количества значений второй координаты элементов из A , соответствующих некоторому фиксированному значению первой координаты. Для 2-мерных A это значит, что мы рассматриваем всевозможные вертикальные сечения, и берём логарифм от самого большого из них:

$$\mathcal{X}_{2|1}(A) = \max_{a \in \pi_1(A)} |\{b : (a, b) \in A\}|$$

Аналогично можно определить “количество информации” и для большего количества координат. Нетрудно заметить, что

$$\mathcal{X}(A) \leq \mathcal{X}_1(A) + \mathcal{X}_{1|2}(A)$$

(по существу, это утверждение о том, что размер множества не больше, чем произведение размера проекции на первую координату и максимального размера вертикального сечения).

Домашнее задание 1.1. Докажите, что для любого 3-мерного множества A выполняется неравенство $2\mathcal{X}(A) \leq \mathcal{X}_{12}(A) + \mathcal{X}_{23}(A) + \mathcal{X}_{13}(A)$.

1.2 Информация по Хартли в детских задачах.

Пример 1. Сколько нужно задать вопросов, подразумевающих ответ *да* или *нет*, чтобы отгадать задуманное число из интервала $1..100$? Тот же вопрос для числа из интервала $1..N$ для произвольного N . Можно ли не задавать вопросы один за другим (адаптивная стратегия), а прислать список сразу со всеми вопросами (неадаптивная стратегия)?

Те же вопросы для отгадывания задуманной пары различных чисел из интервала $1..N$.

В решении этой Задачи 1 доказательство нижней оценки на минимальное число вопросов можно изложить тремя альтернативными способами: (а) подсчет числа листьев в дереве–стратегии фиксированной высоты, (б) рассуждение со “злонамеренным противником” (adversary argument), и (в) “информационная” оценка на языке информации Хартли.

Домашнее задание 1.2. *Сколько нужно задать вопросов, чтобы отгадать целое число из интервала $1..100$, если отвечающему разрешается солгать в ответ на один из вопросов?*

Домашнее задание 1.3. *Потребуем оплачивать получаемые ответы: за каждый полученный ответ “да” нужно заплатить рубль, а за каждый полученный ответ “нет” нужно заплатить два рубля. Сколько денег нужно иметь, чтобы отгадать целое число из интервала $1..100$?*

Пример 2. Имеется 25 монет одинаковых на вид. Одна из монет фальшивая. Все настоящие монеты имеют одинаковый вес, фальшивая легче. Есть чашечные весы без гирь. Сколько нужно произвести взвешиваний, чтобы найти фальшивую монету? Тот же вопрос для N монет (среди которых есть одна фальшивая).

Пример 3. Имеется 12 монет, одна из них фальшивая. Все настоящие монеты имеют одинаковый вес, а фальшивая легче или тяжелее. Есть чашечные весы без гирь, с помощью которых можно сравнить веса двух любых групп монет. Сколько нужно произвести взвешиваний, чтобы найти фальшивую монету и определить, легче она или тяжелее?

Домашнее задание 1.4. *Решите аналогичную задачу для 13 монет.*

Пример 4. Имеется 15 монет, одна из них фальшивая. Все настоящие монеты имеют одинаковый вес, а фальшивая легче или тяжелее. Есть чашечные весы без гирь. Сколько нужно произвести взвешиваний, чтобы найти фальшивую монету (не требуется определять, легче она или тяжелее)?

Домашнее задание 1.5. *Решите аналогичную задачу для 14 монет.*

Домашнее задание 1.6. *Имеется N камней разного веса и чашечные весы, которые позволяют сравнить веса любых двух шаров. (а) Сколько нужно взвешиваний, чтобы найти самый тяжелый камень?*

(б) Сколько нужно взвешиваний, чтобы найти самый тяжелый и второй по весу камни?

(в) Сколько нужно взвешиваний, чтобы упорядочить N камни по весу? Найдите точный ответ на этот вопрос для $N = 2, 3, 4, 5$. Оцените асимптотику роста числа взвешиваний с ростом числа камней: докажете, что для произвольного N требуется $\Theta(N \log N)$ взвешиваний.

2 Лекция 2, 26 февраля

2.1 Энтропия Шеннона: определение и простейшие свойства энтропии дискретной случайной величины

Определение. Шенноновской энтропией случайной функции α , которая принимает k значений с вероятностями p_1, \dots, p_k ($\sum_{i=1}^k p_i = 1$), называется число $H(\alpha) = -\sum p_i \log p_i$.

Мы будем применять это определение в том числе и для распределений, в которых некоторые p_i равны нулю. Как обычно, мы по непрерывности доопределяем функцию $x \log x$ в нуле и полагаем $0 \cdot \log \frac{1}{0} = 0$.

Простейшие свойства энтропии Шеннона:

- $H(\alpha) \geq 0$, причём равенство достигается, если и только если величина α является вырожденной (вероятность p_i одного из значений равна единице, а вероятности всех остальных значений равны нулю).
- для случайной величины с k значениями $H(\alpha) \leq \log k$, причём равенство достигается, если и только если распределение равномерное, т.е., $p_1 = \dots = p_k = \frac{1}{k}$ (доказательство: воспользуемся вогнутостью логарифма и неравенством Йенсена).

Для пары совместно определённых случайных величин α, β мы имеем энтропии $H(\alpha), H(\beta), H(\alpha, \beta)$.

- $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$, причём равенство достигается тогда и только тогда, когда α и β независимы (в обычном смысле теории вероятностей).

Также для пары совместно определённых случайных величин α, β при каждом фиксированном значении b величины β мы имеем некоторое условное распределение вероятностей на значениях α . Обозначим энтропию этого условного распределения $H(\alpha | \beta = b)$. Условная (или *относительная*) энтропия Шеннона величины α относительно величины β определяется как усреднение энтропий $H(\alpha | \beta = b)$ по всем значениям β :

$$H(\alpha | \beta) = \sum_b H(\alpha | \beta = b) \cdot \text{Prob}[\beta = b]$$

Простейшие свойства относительной энтропии:

- $H(\alpha, \beta) = H(\alpha | \beta) + H(\beta)$
- $H(\alpha | \beta) \geq 0$, причём равенство достигается тогда и только тогда, когда α есть детерминированная функция β (т.е., по значению β с вероятностью один можно однозначно восстановить значение α).

- $H(\alpha | \beta) \leq H(\alpha)$, причём равенство достигается только при независимости α и β

Определим *информацию в α о величине β* как разницу между простой и относительной энтропиями:

$$I(\alpha : \beta) = H(\beta) - H(\beta | \alpha)$$

Основные свойства взаимной информации:

- $I(\alpha : \beta) = I(\beta : \alpha) = H(\alpha) + H(\beta) - H(\alpha, \beta)$,
- $I(\alpha : \beta) \leq H(\alpha)$, $I(\alpha : \beta) \leq H(\beta)$,
- $I(\alpha : \beta) \geq 0$, равенство достигается, если и только если величины α и β независимы,
- $I(\alpha : \beta) = H(\alpha)$, если и только если α есть детерминированная функция β ,
- $I(\alpha : \alpha) = H(\alpha)$.

Аналогично определим *относительную взаимную информацию* (взаимную информацию между α и β при известном значении γ). Мы рассмотрим три варианта этого определения.

Первый вариант определения:

$$I(\alpha : \beta | \gamma) := \sum_c I(\alpha : \beta | \gamma = c) \cdot \text{Prob}[\gamma = c]$$

(здесь $I(\alpha : \beta | \gamma = c)$ обозначает взаимную информацию между α и β в условном распределении информации на парах (α, β) при зафиксированном значении $\gamma = c$).

Второй вариант определения:

$$I(\alpha : \beta | \gamma) := H(\beta | \gamma) - H(\beta | \alpha, \gamma).$$

Третий вариант определения:

$$I(\alpha : \beta | \gamma) := H(\alpha | \gamma) + H(\beta | \gamma) - H(\alpha, \beta | \gamma).$$

Домашнее задание 2.1. Докажите, что три определения относительной взаимной информации эквивалентны.

Домашнее задание 2.2. Докажите, что

- $H(\alpha, \beta, \gamma) + H(\gamma) \leq H(\alpha, \gamma) + H(\beta, \gamma)$,
- $2H(\alpha, \beta, \gamma) \leq H(\alpha, \beta) + H(\alpha, \gamma) + H(\beta, \gamma)$.

Домашнее задание 2.3. Докажите, что

- если α, β, γ образуют цепь Маркова (относительные распределения вероятностей γ при условии $\alpha = a$ и $\beta = b$ такое же, как и при условии $\beta = b$), то $I(\alpha : \gamma) \leq I(\alpha : \beta)$ и $I(\alpha : \gamma) \leq I(\beta : \gamma)$;
- если четвёрка случайных величин $\alpha, \beta, \gamma, \delta$ образуют цепь Маркова, то $I(\alpha : \delta) \leq I(\beta : \gamma)$.

2.2 Использование энтропии в “детских” задачах

Пример 1: Энтропийная эвристика (жадный “энтропийный” алгоритм) в задаче о поиске одной фальшивой монеты и её относительного веса из 12 (см. Пример 3 из Лекции 1).

Пример 2: Энтропийное доказательство нижней оценки для числа взвешиваний при поиске одной фальшивой монеты из 14 без обязательно определения относительного веса (см. домашнее задание 5 из Лекции 1).

Пример 3: Энтропийная эвристика (жадный “энтропийный” алгоритм) в задаче о сортировке 5 камней по весу (см. домашнее задание 6 из Лекции 1).

3 Лекция 3, 5 марта: энтропии пар и троек случайных функций

3.1 Относительная взаимная информация: определения и простейшие свойства

Напомним, что для распределение тройки случайных величин α, β, γ мы тремя эквивалентными способами определили относительную взаимную информацию $I(\alpha : \beta | \gamma)$. Из этих определений немедленно вытекают следующие свойства:

- $I(\alpha : \beta | \gamma) \geq 0$,
- $I(\alpha : \beta | \gamma) = I(\beta : \alpha | \gamma)$,
- $I(\alpha : \beta | \gamma) = H(\alpha, \gamma) + H(\beta, \gamma) - H(\alpha, \beta, \gamma) - H(\gamma)$.

Замечание. Существуют такие распределения вероятностей (α, β, γ) , что $I(\alpha : \beta) = 0$, но $I(\alpha : \beta | \gamma) > 0$. И наоборот, для некоторых распределений $I(\alpha : \beta | \gamma) = 0$, но $I(\alpha : \beta) > 0$. (Приведите примеры таких распределений!)

3.2 Энтропийные профили распределений

Далее мы рассмотрим вопрос: какими могут быть *энтропийные профили* — значения энтропий набора совместно распределённых случайных величин. Начнем с самого простого вопроса: какой может быть энтропия одной случайной величины?

Утверждение 1. Для любого вещественного $h \geq 0$ найдётся случайная величина α такая, что $H(\alpha) = h$.

Для совместно распределённой пары (α_1, α_2) у нас имеется уже довольно много информационных величин: энтропии $H(\alpha_1), H(\alpha_2), H(\alpha_1, \alpha_2)$, относительные энтропии $H(\alpha_1 | \alpha_2)$ и $H(\alpha_2 | \alpha_1)$, а также взаимная информация $I(\alpha_1 : \alpha_2)$. Однако не все эти величины независимы. Например, зная

значения $H(\alpha_1), H(\alpha_2), H(\alpha_1, \alpha_2)$, можно вычислить значения обеих относительных энтропий и взаимной информации. И наоборот, зная $H(\alpha_1 | \alpha_2), H(\alpha_2 | \alpha_1)$ и $I(\alpha_1 : \alpha_2)$, мы можем восстановить $H(\alpha_1), H(\alpha_2), H(\alpha_1, \alpha_2)$.

Все энтропии принимают неотрицательные значения. Кроме того, мы знаем, что энтропия пары $H(\alpha_1, \alpha_2)$ не меньше каждой из энтропий $H(\alpha_1), H(\alpha_2)$ и не больше суммы этих двух энтропий. Следующее утверждение показывает, что никаких других ограничений на значения энтропий пары случайных величин нет.

Утверждение 2. (а) Если h_1, h_2, h_{12} удовлетворяют неравенствам

$$\begin{aligned} 0 &\leq h_1, h_2, \\ h_1 &\leq h_{12}, \\ h_2 &\leq h_{12}, \\ h_{12} &\leq h_1 + h_2, \end{aligned}$$

то найдутся случайные величины α_1, α_2 такие, что

$$H(\alpha_1) = h_1, \quad H(\alpha_2) = h_2, \quad H(\alpha_1, \alpha_2) = h_{12}.$$

(б) Для любых трёх неотрицательных вещественных чисел $h_{1|2}, h_{2|1}, h_{1:2}$ найдутся случайные величины α_1, α_2 такие, что

$$H(\alpha_1 | \alpha_2) = h_{1|2}, \quad H(\alpha_2 | \alpha_1) = h_{2|1}, \quad I(\alpha_1 : \alpha_2) = h_{1:2}.$$

Замечание 1: Пункты (а) и (б) этого утверждения эквивалентны. По существу, это одно и то же утверждение, записанное в двух разных системах координат.

Замечание 2: Данное утверждение говорит, что множество троек чисел, представляющих энтропии всевозможных распределений (α_1, α_2) , является замкнутым и выпуклым конусом в \mathbb{R}^3 . Этот конус ограничивается тремя плоскостями (является пересечением трёх полупространств в \mathbb{R}^3). Пункты (а) и (б) утверждения описывают этот конус в двух разных системах координат. Во второй системе координат (пункт (б)) данный конус выглядит совсем просто — это первый координатный октант, т.е., множество точек, все три координаты которых неотрицательны.

Для тройки совместно распределённых случайных величин $\alpha_1, \alpha_2, \alpha_3$ все энтропии можно задать набором из $2^3 - 1 = 7$ параметров. В самом деле, чтобы однозначно определить все энтропии (условные и безусловные) и взаимные информации (условные и безусловные), включающие $\alpha_1, \alpha_2, \alpha_3$, достаточно указать значения

$$H(\alpha_1), H(\alpha_2), H(\alpha_3), H(\alpha_1, \alpha_2), H(\alpha_1, \alpha_3), H(\alpha_2, \alpha_3), H(\alpha_1, \alpha_2, \alpha_3).$$

Иногда удобнее пользоваться другой система координат и описывать энтропии тройки $(\alpha_1, \alpha_2, \alpha_3)$ другим набором из 7 параметров:

$$\begin{aligned} &H(\alpha_1 | \alpha_2, \alpha_3), H(\alpha_2 | \alpha_1, \alpha_3), H(\alpha_3 | \alpha_1, \alpha_2), \\ &I(\alpha_1 : \alpha_2 | \alpha_3), I(\alpha_1 : \alpha_3 | \alpha_2), I(\alpha_2 : \alpha_3 | \alpha_1), \\ &I(\alpha_1 : \alpha_2 : \alpha_3), \end{aligned}$$

где взаимная информация тройки $I(\alpha_1 : \alpha_2 : \alpha_3)$ определяется следующим образом:

$$I(\alpha_1 : \alpha_2 : \alpha_3) = I(\alpha_1 : \alpha_2) - I(\alpha_1 : \alpha_2 | \alpha_3).$$

Заметим, что величина $I(\alpha_1 : \alpha_2 : \alpha_3)$ симметрична по своим трём аргументам. Взаимная информация тройки не имеет наглядного смысла. Ниже мы увидим, что для некоторых распределений вероятностей величина $I(\alpha_1 : \alpha_2 : \alpha_3)$ отрицательна.

Для энтропий троек случайных величин выполняются следующие базисные ограничения:

$$\begin{aligned} H(\alpha_i | \alpha_j, \alpha_k) &\geq 0, \\ I(\alpha_i : \alpha_j | \alpha_k) &\geq 0, \\ I(\alpha_i : \alpha_j) &\geq 0 \end{aligned}$$

для всех наборов попарно различных i, j, k . Отметим, что последнее неравенство можно переписать как

$$I(\alpha_1 : \alpha_2 : \alpha_3) + I(\alpha_i : \alpha_j | \alpha_k) \geq 0.$$

Эти 9 базисных неравенств задают выпуклый замкнутый конус в 7-мерном вещественном пространстве. Чтобы набор из семи чисел представлял энтропии какой-то тройки случайных величин, необходимо, чтобы данная точка лежала в описанном конусе. Однако, как мы увидим ниже, данное условие не является достаточным.

Домашнее задание 3.1 (необязательное). *Докажите, что всякое линейное неравенство, выполненное для энтропий всех троек случайных величин, есть комбинация девяти указанных выше базисных неравенств с неотрицательными коэффициентами.*

Эквивалентная формулировка: если набор из семи чисел удовлетворяет указанным базисным неравенствам, то для сколь угодно малого $\varepsilon > 0$ можно найти распределение вероятностей $(\alpha_1, \alpha_2, \alpha_3)$, соответствующие энтропии которого ε -близки к заданным семи числам. (В отличие от случая двумерных распределений, множество всех реализуемых наборов чисел не совпадает в точности с указанным замкнутым и выпуклым конусом, а лишь плотно в нём.)

Следующие две задачи составляют наименее техническую (и наиболее содержательную) часть решения необязательного Домашнего задания 3.1.

Домашнее задание 3.2. Пусть векторы \vec{v} и \vec{w} из \mathbb{R}^7 представляют наборы энтропий каких-то распределений (α, β, γ) и $(\alpha', \beta', \gamma')$. Докажите, что существует распределение $(\alpha'', \beta'', \gamma'')$, энтропии которого представлены координатами вектора $\vec{v} + \vec{w}$.

Другими словами, множество 7-мерных векторов, представляющих наборы энтропий троек всевозможных случайных величин, замкнуто относительно покомпонентного сложения.

Домашнее задание 3.3. Пусть вектор \vec{v} из \mathbb{R}^7 представляет набор энтропий какого-то распределения (α, β, γ) . Докажите, что для любого $\lambda > 0$ и любого $\varepsilon > 0$ найдется распределение $(\alpha', \beta', \gamma')$, энтропии которого ε -близки к соответствующим координатами вектора $\lambda \cdot \vec{v}$.

Другими словами, множество 7-мерных векторов, представляющих наборы энтропий троек всевозможных случайных величин, “почти” замкнуто относительно умножения на константу.

Домашнее задание 3.4. Постройте пример распределения (α, β, γ) , для которого

$$\begin{aligned} H(\alpha | \beta, \gamma) &= H(\beta | \alpha, \gamma) = H(\gamma | \alpha, \beta) = 0, \\ I(\alpha : \beta | \gamma) &= I(\alpha : \gamma | \beta) = I(\beta : \gamma | \alpha) = 1, \\ I(\alpha : \beta : \gamma) &= -1. \end{aligned}$$

Домашнее задание 3.5. Докажите, что для любого распределения тройки (α, β, γ) выполнено неравенство

$$H(\gamma) \leq H(\gamma | \alpha) + H(\gamma | \beta) + I(\alpha : \beta).$$

Замечание: Это неравенство показывает, что если γ является детерминированной функцией α и детерминированной функцией β , то энтропия γ не превосходит $I(\alpha : \beta)$.

4 Лекция 4, 12 марта

4.1 Информационные неравенства и энтропии для распределений троек случайных величин

Информационные диаграммы для троек случайных величин. Доказательство информационных неравенств с помощью диаграмм. Перевод доказательств с языка диаграмм на язык формальных неравенств.

Домашнее задание 4.1. Если $H(\alpha_1) = H(\alpha_2) = H(\alpha_3) = h$, $H(\alpha_1, \alpha_2) = H(\alpha_2, \alpha_3) = H(\alpha_1, \alpha_3) = 2h$, $H(\alpha_1, \alpha_2, \alpha_3) = 2h$, то величина h есть двоичный логарифм некоторого целого числа N .

Следующая классическая теорема Шеннона показывает, насколько длинным должен быть секретный ключ, чтобы схема шифрования была абсолютно надежна.

Теорема 1 (Шеннона о надежном шифровании с секретным ключом). Опишем схему шифрования с секретным ключом. Отправитель хочет переслать Получателю сообщение M (выбранное по некоторому заранее известному распределению вероятностей). Отправитель и Получатель заранее договорились о значении секретного ключа K . Отправитель производит «шифрование» с использованием секретного ключа и вычисляет значение $C = C(M, K)$. Затем он пересылает по открытому каналу связи

полученное зашифрованное сообщение C . Получатель производит расшифровку и восстанавливает по зашифрованному сообщению C и значению ключа K исходное сообщение M . При этом требуется, чтобы передаваемое по открытому каналу зашифрованное сообщение C не имело никакой взаимной информации с сообщением M (свойство секретности). Для любой схемы такого вида энтропия секретного ключа K не может быть меньше энтропии передаваемого сообщения M .

Условие теоремы можно переписать в виде следующих информационных неравенств:

$$\begin{cases} H(C|M, K) = 0, \\ H(M|C, K) = 0, \\ I(C : M) = 0. \end{cases}$$

Из этих трех условий и базисных неравенств для энтропии нетрудно вывести $H(K) \geq H(M)$.

Замечание 1: Утверждение теоремы остается верным и без первого условия $H(C|M, K) = 0$ (достаточно потребовать $H(M|C, K) = 0$ и $I(C : M) = 0$). Содержательно это означает, что размер (энтропия) ключа должна быть большей не только для детерминированных схем шифрования, но и для схем шифрования с вероятностной процедурой кодирования $(M, K) \mapsto C$.

Замечание 2: В известной схеме Вернама (Gilbert Vernam) шифрование состоит в побитовом XOR n -битового сообщения M с n -битовым же ключом K , т.е., $C_i = M_i \oplus K_i$, $i = 1, \dots, n$. (Декодирование выполняется аналогично: $M_i = C_i \oplus K_i$, $i = 1, \dots, n$.) Если исходное сообщение M равномерно распределено на всех n -битовых строках и биты ключа K также выбираются случайно и равномерно (независимо от M), то $H(M) = H(K) = n$. В данном случае энтропии сообщения и ключа совпадают.

4.2 Схемы разделения секрета.

Определение совершенной схемы разделения секрета.

Утверждение 3. *Предположим, что не один участник схемы разделения секрета не может восстановить секрет в одиночку. Тогда любую совершенную схему разделения секрета для такой схемы можно модифицировать так, чтобы распределение на множестве секретов стало равномерным, а энтропии долей секрета каждого участника схемы не изменились.*

Участник номер i схемы разделения секрета называется существенным, если некоторый набор участников $\{j_1, \dots, j_r\}$ не имеет никакой информации о секрете, но участники $\{i, j_1, \dots, j_r\}$ вместе могут однозначно восстановить секрет (добавление i -го участника превращает некоторую группу из неавторизованной в авторизованную).

Домашнее задание 4.2. *Докажите, что для каждого существенного участника совершенной схемы разделения секрета выполнено неравенство $H(S_i) \geq H(S_0)$ (энтропия доли такого участника не может быть меньше, чем энтропия самого секрета).*

Определение идеальной схемы разделения секрета. Идеальное совершенное разделение секрета для пороговой структуры доступа (полиномиальная схема Шамира).

Пример структуры доступа, не имеющей совершенной идеальной схемы разделения секрета: структура из 4 участников, минимальными авторизованными группами в которой являются пары $\{\{1, 2\}, \{2, 3\}, \{3, 4\}\}$.

Домашнее задание 4.3. Докажите, что для описанной выше структуры доступа существует схема разделения секрета $(S_0, S_1, S_2, S_3, S_4)$, в которой

$$\max_{i>0} \frac{H(S_i)}{H(S_0)} = 3/2.$$

Домашнее задание 4.4. Рассмотрим структуру доступа для 4 участников, минимальными авторизованными группами в которой являются пары $\{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{2, 4\}\}$. Докажите, что для данной схемы не существует совершенной идеальной схемы разделения секрета.

4.3 Оптимальное кодирование дискретного распределения.

Определение 1. Набор слов $c_1, \dots, c_n \in \{0, 1\}^*$ называется однозначно декодируемым кодом, если никакое слово $x \in \{0, 1\}^*$ нельзя двумя разными способами представить в виде конкатенации слов c_i .

Стандартным примером однозначно декодируемого кода является класс (бес)префиксных кодов.

Теорема 2 (неравенство Крафта). Если слова $c_1, \dots, c_n \in \{0, 1\}^*$ образуют однозначно декодируемый код, то $\sum 2^{-|c_i|} \leq 1$.

Теорема 3. Если для набора чисел $l_1, \dots, l_n \in \mathbb{N}$ выполнено неравенство $\sum 2^{-l_i} \leq 1$, то существует префиксный код $c_1, \dots, c_n \in \{0, 1\}^*$ с такими длинами кодовых слов (т.е. $|c_i| = l_i$ для $i = 1, \dots, n$).

Из этих двух теорем следует, что всякий однозначно декодируемый код можно переделать в код префиксный, не меняя длин кодовых слов.

Теорема 4. (а) Для любого распределения вероятностей p_1, \dots, p_n и любого однозначно декодируемого кода c_1, \dots, c_n

$$\sum p_i |c_i| \geq \sum p_i \log \frac{1}{p_i}.$$

(б) Для любого распределения вероятностей p_1, \dots, p_n найдётся такой префиксный код c_1, \dots, c_n , что

$$\sum p_i |c_i| < \sum p_i \log \frac{1}{p_i} + 1$$

Доказательство: пользуемся вогнутостью логарифма и неравенством Крафта.

4.4 Конструкции кодов для сжатия информации.

Обсуждение классических конструкций кодов: код Шеннона–Фано, код Хаффмана (и доказательство его оптимальности), арифметическое кодирование (с оценкой средней длины кодового слова).

Домашнее задание 4.5. Приведите пример распределения вероятностей, для которого код Шеннона–Фано не является оптимальным.

5 Лекция 5, 19 марта: блочное кодирование для канала без шума.

Теорема 5 (Шеннона о блочном кодировании источника). Пусть случайная величина α распределена на конечном множестве $\{a_1, \dots, a_k\}$. Рассмотрим последовательность независимых одинаково распределенных копий этой случайной величины и кодирование блоков из n таких случайных величин.

(1) Для всякого $L > h$ существуют функции кодирования и декодирования

$$C_n : A^n \rightarrow \{0, 1\}^{\lfloor L \cdot n \rfloor}$$

и

$$D_n : \{0, 1\}^{\lfloor L \cdot n \rfloor} \rightarrow A^n$$

такие, что вероятность ошибки

$$\varepsilon_n = \text{Prob}_{a_{i_1} \dots a_{i_n}} [D_n(C_n(a_{i_1} \dots a_{i_n})) \neq (a_{i_1} \dots a_{i_n})]$$

стремится к нулю при $n \rightarrow \infty$ (буквы a_{i_s} для каждой позиции $s = 1 \dots n$ выбираются по распределению α , независимо для всех i).

(2) [слабое обращение] Для всякого $L < h$ и для любой последовательности функции кодирования и декодирования

$$C_n : A^n \rightarrow \{0, 1\}^{\lceil L \cdot n \rceil}$$

и

$$D_n : \{0, 1\}^{\lceil L \cdot n \rceil} \rightarrow A^n$$

вероятность ошибки

$$\varepsilon_n = \text{Prob}_{a_{i_1} \dots a_{i_n}} [D_n(C_n(a_{i_1} \dots a_{i_n})) \neq (a_{i_1} \dots a_{i_n})]$$

(буквы a_{i_s} для каждой позиции $s = 1 \dots n$ выбираются независимо друг от друга по распределению α) не стремится к нулю при $n \rightarrow \infty$.

(2') [сильное обращение] Для всякого $L < h$ и для любой последовательности функции кодирования и декодирования

$$C_n : A^n \rightarrow \{0, 1\}^{\lceil L \cdot n \rceil}$$

и

$$D_n : \{0, 1\}^{[L \cdot n]} \rightarrow A^n$$

вероятность ошибки

$$\varepsilon_n = \text{Prob}_{a_{i_1} \dots a_{i_n}} [D_n(C_n(a_{i_1} \dots a_{i_n})) \neq (a_{i_1} \dots a_{i_n})]$$

(буквы a_{i_s} для каждой позиции $s = 1 \dots n$ выбираются независимо друг от друга по распределению α) стремится к единице при $n \rightarrow \infty$.

Домашнее задание 5.1. Если $H(\alpha_1) = H(\alpha_2) = H(\alpha_3) = h$, $H(\alpha_1, \alpha_2) = H(\alpha_2, \alpha_3) = H(\alpha_1, \alpha_3) = 2h$, $H(\alpha_1, \alpha_2, \alpha_3) = 2h$, то величина h есть двоичный логарифм некоторого целого числа N .

Домашнее задание 5.2. Докажите, что для каждого фиксированного алфавита $A = \{a_1, \dots, a_k\}$ существует последовательность блочных кодов (C_n, D_n)

$$C_n : A^n \rightarrow \{0, 1\}^* \text{ и } D_n : \{0, 1\}^* \rightarrow A^n$$

таким, что для любого распределения α на алфавите A

$$\lim_{n \rightarrow \infty} \text{Prob}_{a_{i_1} \dots a_{i_n}} [D_n(C_n(a_{i_1} \dots a_{i_n})) \neq (a_{i_1} \dots a_{i_n})] = 0$$

(вероятность ошибки стремится к нулю) и

$$\lim_{n \rightarrow \infty} \frac{|E(C_n(a_{i_1} \dots a_{i_n}))|}{n} = H(\alpha)$$

(средняя длина кодового слова составляет $H(\alpha) + o(1)$ битов на один символ), где буквы a_{i_s} для каждой позиции $s = 1 \dots n$ выбираются независимо друг от друга по распределению α .

Замечание: Верно и немного более сильное утверждение: множество всех кодовых слов (образ функции C_n) можно сделать префиксным, а вероятность ошибки декодирования можно сделать равной нулю.

6 Лекция 6, 26 марта: простейшие конструкции комбинаторной теории кодирования

Комбинаторная модель канала с шумом. Определение кода длины n над алфавитом Σ , исправляющего e ошибок. Расстояние кода d (минимальное расстояние между двумя несовпадающими кодовыми словами); связь кодового расстояния и числа исправляемых ошибок.

Утверждение 4 (граница Хэмминга (volume bound)). Для любого кода с параметрами (n, k, d) выполнено неравенство

$$2^n \geq 2^k \cdot (C_n^0 + \dots + C_n^e),$$

где $e = \lfloor \frac{d-1}{2} \rfloor$.

Утверждение 5 (граница Гилберта (Gilbert)). . Если

$$2^n \geq 2^k \cdot (C_n^0 + \dots + C_n^{d-1}),$$

то существует код с параметрами (n, k, d) .

Асимптотическое поведение границы Хэмминга и границы Гилберта для кодов с расстоянием $d = \alpha n$ при $n \rightarrow \infty$. Определение линейных кодов над алфавитом $\Sigma = \mathbb{F}_q$ (алфавит Σ состоит из элементов конечного поля размера q).

Утверждение 6. Расстояние линейного кода C равно

$$\min_{0 \neq x \in C} \text{dist}(x, 0),$$

что совпадает с

$$\min_{0 \neq x \in C} \omega(x),$$

где $\omega(x)$ (вес слова x) есть число ненулевых элементов в x .

Порождающая и проверочная матрицы линейного кода. Замечание: в линейном коде с расстоянием d любые $d - 1$ столбцов проверочной матрицы линейно независимы.

Утверждение 7 (Граница Варшамова–Гилберта). Если

$$2^n > 2^{k-1} (C_n^0 + \dots + C_n^{d-1}),$$

то существует линейный код над полем из двух элементов с параметрами $[n, k, \geq d]$.

Домашнее задание 6.1. Сформулируйте и докажите теорему Варшамова–Гилберта для произвольного размера поля q .

Построение совершенного кода Хэмминга для параметров

$$[n = 2^s - 1, k = 2^s - s - 1, d = 3]_2$$

для произвольных s . Алгоритмы кодирования и декодирования для кода Хэмминга. Замечание: шары радиуса 1, описанные вокруг кодовых слов кода Хэмминга, покрывают весь булев куб (без пересечений)

Домашнее задание 6.2. Опишите оптимальные стратегии игры угадай число с одним неверным ответом (первый игрок задумывает целое число от 1 до n ; второй игрок задаёт вопросы, предполагающие ответы да или нет; второй игрок отвечая на эти вопросы может солгать не более одного раза) для $n = 500, 1000, 200$.

Домашнее задание 6.3. Пусть некоторый код длины n имеет расстояние $d > n/2$. Докажите, что число кодовых слов не превосходит $n + 1$.

Указание: Представьте все кодовые слова в виде векторов в \mathbb{R}^n , каждая координата которых равна $+1$ или -1 ; заметьте, что угол между любыми двумя из этих векторов оказывается тупым.

Домашнее задание 6.4. Пусть некоторый код длины n имеет расстояние $d > 0.9n$. Сколько кодовых слов может быть в таком коде?

7 Лекция 7, 2 апреля.

Утверждение 8. В любом двоичном коде с длиной кодовых слов n и кодовым расстоянием $d \leq \alpha n$ для $\alpha > 1/2$ число кодовых слов не превосходит $1 + \frac{1}{2\alpha-1}$.

Определение кода Адамара, его кодовое расстояние.

Домашнее задание 7.1. Докажите, что в двоичном коде с длиной кодовых слов n и кодовым расстоянием $d \geq n/2$ число кодовых слов не превосходит $2n$. Замечание: Код Адамара достигает этой границы: он содержит 2^{k+1} кодовых слов длины $n = 2^k$.

Домашнее задание 7.2. Постройте порождающую и проверочную матрицы кода Адамара.

Домашнее задание 7.3. Опишите оптимальную стратегию игры угадай число от 1 до 150 с одним неверным ответом (как обычно, первый игрок задумывает целое число от 1 до 150; второй игрок задаёт вопросы, предполагающие ответы да или нет; второй игрок отвечая на эти вопросы может солгать не более одного раза). Замечание: стандартная конструкция, использующая линейный код, и нижняя оценка Хэмминга дают в этой задаче для $n = 150$ два несовпадающих числа.

Код Рида–Соломона. Оценка Синглтона. Полиномиальный алгоритм декодирования кода Рида–Соломона. Каскадные коды (конкатенация кодов). Оценка для параметров каскадного кода. Наивный алгоритм декодирования каскадного кода, исправляющий $\approx d/4$ ошибок.

Домашнее задание 7.4. Нарисуйте график границы Синглтона для двухбуквенного алфавита. Сравните её поведение при $n \rightarrow \infty$ с границей Хэмминга и границей Варшавова–Гилберта.

Домашнее задание 7.5. Рассмотрим следующий линейный код с параметрами $(n, r, q = 2^t)$ (n, r и t — натуральные числа). В качестве кодовых слов мы возьмем графики (таблицы всех значений) многочленов степени $\leq r$ от t переменных над полем \mathbb{F}_q с $q = 2^t$ элементами. Для найдите длину кодового слова, число кодовых слов и расстояние этого кода.

Замечание 1: Такой код можно рассматривать для произвольного размера поля q , а не только для полей характеристики 2.

Замечание 2: Код Адамара и код Рида–Соломона являются частными случаями этого кода.

8 Лекция 8, 9 апреля.

Вероятностная модель канала с шумом. Дискретный канал без памяти (канал задаётся входным алфавитом A , выходным алфавитом B и набором

условных вероятностей p_{ij} для $i = 1, \dots, |A|$ и $j = 1, \dots, |B|$). Формальное определение пропускной способности дискретного канала без памяти.

Примеры: (1) двоичный симметричный канал, в котором пересылаемый бит меняется на противоположный с вероятностью ε ; (2) двоичный несимметричный канал, в котором пересылаемый ноль всегда передается без ошибок, а пересылаемая единица с вероятностью ε превращается в ноль; (3) канал с алфавитом $\{0, 1\}$ на входе и $\{0, 1, *\}$ на выходе: ноль и единица с вероятностью $(1 - \varepsilon)$ пересылаются без ошибок, и с вероятностью ε превращаются в $*$.

Вычисление пропускной способности для двоичного симметричного канала.

Домашнее задание 8.1. Вычислите пропускную способность двух других каналов из приведенного примера.

Теорема 6 (Шеннона о кодировании для дискретного канала с шумом). Пусть пропускная способность канала без памяти (с входным алфавитом A и выходным алфавитом B) равна R .

(а) Для всякого $L < R$ существуют функции кодирования и декодирования

$$C_k : \{0, 1\}^k \rightarrow A^{\lfloor k/L \rfloor}$$

и

$$D_k : B^{\lfloor k/L \rfloor} \rightarrow \{0, 1\}^k$$

такие, что вероятность ошибки σ_k при кодировании блоков из k битов

$$\{0, 1\}^k \xrightarrow{\text{кодирование } C_k} A^{\lfloor k/L \rfloor} \xrightarrow{\text{искажение в канале}} B^{\lfloor k/L \rfloor} \xrightarrow{\text{декодирование } D_k} \{0, 1\}^k$$

стремится к нулю при $k \rightarrow \infty$.

(б) Если $L > R$, то для любых функции кодирования и декодирования

$$C_k : \{0, 1\}^k \rightarrow A^{\lfloor k/L \rfloor}$$

и

$$D_k : B^{\lfloor k/L \rfloor} \rightarrow \{0, 1\}^k$$

вероятность ошибки σ_k не стремится к нулю при $k \rightarrow \infty$.

(б') При тех же предположениях, что и в (б), вероятность ошибки σ_k стремится к единице.

Доказательство пункта (а) теоремы для двоичного симметричного канала.

9 Лекция 9, 16 апреля.

9.1 Вероятностная модель канала с шумом (окончание)

Доказательство обращения и сильного обращения теоремы Шеннона о кодировании для дискретного канала с шумом (сильное обращение для двучленного симметричного канала).

9.2 Колмогоровская сложность

Определение 2. Пусть $U : \{0, 1\}^* \rightarrow \{0, 1\}^*$ есть (частичная) вычислимая функция. Определим $K_U(x) = \min\{|p| : U(p) = x\}$. (Минимум пустого множества считается равным бесконечности.)

Теорема 7. Существует такой способ описания (частичная вычислимая функция) U , что для любой другой V и для всех слов x $K_U(x) \leq K_V(x) + O(1)$

Фиксируем какой-либо оптимальный способ описания U и обозначаем соответствующую сложность $K(x)$. Называем эту величину *колмогоровской сложностью* слова x .

Простейшие свойства колмогоровской сложности:

- $K(x) \leq |x| + O(1)$;
- $K(xx) \leq |x| + O(1)$;
- $K(f(x)) \leq K(x) + O(1)$ для всякой вычислимой f ;
- для всякого n существует слово длины n и сложности не менее n ;
- существует такая константа C , что для всякого n не менее, чем для 99% слов длины n выполнено $n - c \leq K(x) \leq n + c$.

Теорема 8. Не существует алгоритма, который по заданному n находит бы слово с колмогоровской сложностью не менее n .

Следствие: Колмогоровская сложность $K(x)$ не является вычислимой функцией.

Следствие: Оптимальный способ описания не может быть всюду определенной функцией.

10 Лекция 10, 23 апреля.

Утверждение 9. (а) $K(x, y) \leq K(x) + K(y) + O(\log(|x| + |y|))$. (б) Для любого $C > 0$ найдутся такие слова x и y , что $K(x, y) > K(x) + K(y) + C$.

Определение относительной колмогоровской сложности.

Простейшие свойства колмогоровской сложности:

- $K(x|y) \leq |x| + O(1)$;
- $K(x|\Lambda) = K(x) + O(1)$;
- $K(x|y) \leq K(x) + O(1)$ для всякой вычислимой f ;
- для всякого n существует слово длины n и сложности не менее n ;
- существует такая константа C , что для всякого n не менее, чем для 99% слов x длины n выполнено $n - c \leq K(x|y) \leq n + c$.

Теорема 9 (Колмогорова–Левина).

$$K(x, y) = K(x) + K(y|x) + O(\log K(x, y))$$

Замечание: логарифмический член в теореме Колмогорова–Левина устранить нельзя.

Домашнее задание 10.1. Докажите, что существует такое число $C > 0$, что

$$K(x, y) \leq |x| + |y| + \log(|x| + |y|) + C$$

для всех x, y .

Домашнее задание 10.2. Пусть слово x длины n состоит из pn единиц и $(1-p)n$ нулей. Тогда

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

Домашнее задание 10.3. Докажите, что (а) $K(x, y, z) + K(z) \leq K(x, z) + K(y, z) + O(\log K(x, y, z))$,

$$(б) K(x, y, z) + K(z) \leq K(x, z) + K(y, z) + O(\log K(x, y, z)),$$

$$(в) K(z) \leq K(z|x) + K(z|y) + I(x : y).$$

Домашнее задание 10.4. $K(x, K(x)) = K(x) + O(1)$.

Домашнее задание 10.5. Пусть U оптимальный способ описания (из определения колмогоровской сложности), и p кратчайшее в смысле U описание для некоторого слова x . Докажите, что $K(p) = K(x) + O(1)$.

11 Лекция 11, 30 апреля: применение колмогоровской сложности

Теорема 10. Для того чтобы распознать языка палиндромов, одноленточной машине Тьюринга (с одной головкой) требуется время $\Omega(n^2)$.

Домашнее задание 11.1. Докажите, что одноленточная машина Тьюринга, удваивающее входное слово (машина начинает работу на ленте, на которой записано слово x , и заканчивает работу, оставив на ней xx), работает $\Omega(n^2)$ шагов на словах длины n .

Теорема 11. Для любого k существует язык L_k , распознаваемый конечным автоматом с k читающими головками, но не распознаваемый автоматом с меньшим числом головок (все головки движется вдоль входного слова слева направо).

Утверждение 10. Для всякой бесконечной двоичной последовательности $\omega_0\omega_1\omega_2\dots$ найдутся сколь угодно большие номера n такие, что

$$K(\omega_0\dots\omega_{n-1}) \leq n - \log n + O(1)$$

Определение префиксной колмогоровской сложности $KP(x)$. Простейшие свойства префиксной сложности.

Утверждение 11.

$$\sum_{x \in \{0,1\}^*} 2^{-KP(x)} \leq 1.$$

Домашнее задание 11.2. Докажите, что для любого числа C существует такое слово x , что

$$KP(x) > |x| + \log |x| + C.$$

Домашнее задание 11.3. Докажите, что

$$KP(x, y) \leq KP(x) + KP(y) + O(1).$$

Случайность по Мартин-Лёфу: Бесконечная двоичная последовательность $\omega_0\omega_1\omega_2\dots$ называется случайной, если существует такая константа C , что для всех n

$$KP(\omega_0\dots\omega_{n-1}) \geq n - C.$$

Утверждение 12. Почти все (по равномерной бернуллиевской мере) бесконечные двоичные последовательности $\omega_0\dots\omega_n\dots$ случайны по Мартин-Лёфу.

Домашнее задание 11.4. (а) Докажите, что случайная по Мартин-Лёфу последовательность не может быть вычислимой.

(б) Докажите, что последовательность вида $\omega_0\omega_0\omega_1\omega_1\dots\omega_n\omega_n\dots$ не может быть случайной по Мартин-Лёфу.

Теорема 12 (Закон больших чисел в форме Харди-Литлвуда). Для почти всех двоичных последовательностей $\omega_0\omega_1\omega_2\dots$

$$[\text{доля единиц среди первых } n \text{ битов последовательности}] = \frac{1}{2} + O\left(\sqrt{\frac{\ln n}{n}}\right).$$

12 Лекция 12, 7 мая

12.1 Коммуникационная сложность

Определение детерминированного коммуникационного протокола для двух участников. Определение детерминированной коммуникационной сложности $CC(f)$ функции $f : A \times B \rightarrow C$ для конечных множеств A, B, C . Доказательство оценок $CC(f) \leq \lceil \log |A| \rceil + \lceil \log |B| \rceil$ и $CC(f) \leq \lceil \log |A| \rceil + \lceil \log |C| \rceil$.

Теорема 13. *Детерминированная коммуникационная сложность предиката равенства $EQ_n : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$, определяемого как*

$$EQ_n(a, b) = \begin{cases} 1, & \text{если } a \neq b, \\ 0, & \text{иначе,} \end{cases}$$

равна $n + 1$.

Схема доказательства теоремы: Верхняя оценка очевидна. Для получения нижней оценки замечаем, что каждому листу в коммуникационном протоколе соответствует «комбинаторный прямоугольник».

Вероятностные коммуникационные протоколы с общими и отдельными источниками случайных битов Алисы и Боба, определение двух видов вероятностной коммуникационной сложности $RCC_\varepsilon(f)$ (для модели с отдельными случайными битами Алисы и Боба) и $RCC_\varepsilon^{\text{pub}}(f)$ (для модели с общими случайными битами для Алисы и Боба). Доказательство оценок $RCC_\varepsilon(EQ_n) = O(\log \frac{n}{\varepsilon})$ и $RCC_\varepsilon^{\text{pub}}(EQ_n) = O(\log \frac{1}{\varepsilon})$.

Теорема 14. *Для любого $\varepsilon < 1/2$ и для любого предиката $f(x, y)$*

$$RCC_\varepsilon(f) = \Omega(\log CC(f)).$$

Следствие: $RCC_\varepsilon^{\text{pub}}(EQ_n) = \Theta(\log n)$ для каждого $\varepsilon < 1/2$.

Домашнее задание 12.1. *Докажите, что $GT_n = n + 1$, где предикат $EQ_n : \{1, \dots, 2^n\}^n \times \{1, \dots, 2^n\}^n \rightarrow \{0, 1\}$ определяется как*

$$GT_n(a, b) = \begin{cases} 1, & \text{если } a > b, \\ 0, & \text{иначе.} \end{cases}$$

Домашнее задание 12.2. *Докажите, что $CC(\text{DISJ}_n) = n + 1$ для предиката $\text{DISJ}_n : 2^{\{1, \dots, n\}} \times 2^{\{1, \dots, n\}} \rightarrow \{0, 1\}$ (предикат дизъюнктивности пары множеств $A, B \subset \{1, \dots, n\}$), определяемого как*

$$\text{DISJ}_n(A, B) = \begin{cases} 1, & \text{если } A \cap B = \emptyset, \\ 0, & \text{иначе.} \end{cases}$$

Домашнее задание 12.3. *Рассмотрим функцию*

$$\text{MAX}_n : \{1, \dots, n\} \times \{1, \dots, n\} \rightarrow \{1, \dots, n\}$$

(максимум из двух целых чисел из интервала от 1 до 2^n). Докажите, что

- (a) $CC(\text{MAX}_n) \leq 2n$,
- (б) $CC(\text{MAX}_n) \leq \frac{3}{2}n + O(1)$,
- (в) $CC(\text{MAX}_n) \leq \frac{4}{3}n + O(1)$,
- (г) $CC(\text{MAX}_n) \leq n + O(\sqrt{n})$.

Домашнее задание 12.4. Докажите, что для предиката GT_n из домашнего задания 12.1 и любого $\varepsilon > 0$

- (a) $\text{RCC}_\varepsilon^{\text{pub}}(\text{GT}_n) = O(\log^2 n)$,
- (б) $\text{RCC}_\varepsilon^{\text{pub}}(\text{GT}_n) = O((\log n) \cdot (\log \log n))$,
- (в)* $\text{RCC}_\varepsilon^{\text{pub}}(\text{GT}_n) = O(\log n)$.

13 Лекция 13, 14 мая.

13.1 Коммуникационная сложность (продолжение)

Теорема 15. Для любых $\varepsilon > 0$, $\delta > 0$ и любой функции $f : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}$

$$\text{RCC}_{\varepsilon+\delta}(f) \leq \text{RCC}_\varepsilon^{\text{pub}}(f) + O(\log n + \log(1/\delta)).$$

13.2 Принцип кратчайшего описания.

Принцип кратчайшего описания (MDL, minimum description length principle) в машинном обучении и анализе данных. Примеры использования метода MDL.

Список литературы

- [1] Т.М. Cover, J.A. Thomas. Elements of information Theory, 2006.
- [2] R.W. Yeung. A First Course in Information Theory, 2002.
- [3] Н.К. Верещагин, Е.В. Шепин. Информация, кодирование, предсказание, 2012.
- [4] И. Чисар, Я. Кернер. Теория информации, 1985.
- [5] Г.А. Кабатянский. Математика разделения секрета, *Математическое просвещение*, т. 2, No 3, стр. 115–126, 1998.
- [6] M. Li, and P. Vitanyi. An Introduction to Kolmogorov Complexity and Its Applications, 2008.
- [7] В. А. Успенский, Н. К. Верещагин, А. Шень. Введение в колмогоровскую сложность, 2012.

- [8] Румянцев, Ромащенко, Шень. Заметки о теории кодирования, 2011.
<http://www.mscme.ru/~anromash/courses/coding-theory.ps>
- [9] А.М. Яглом, И.М. Яглом. Вероятность и информация, 1973.
- [10] Ф.Дж.А. Мак-Вильямс, Н.Дж.А. Слоэн. Теория кодов, исправляющих ошибки, 1979.
- [11] E. Nisan, N. Kushilevitz. Communication complexity, 1997.
Дополнительная литература.
- [12] Р. Галлагер. Теория информации и надежная связь, 1974.
- [13] В.М. Сидельников. Теория кодирования, 2008.