Министерство образования и науки Украины
Одесский национальный университет им. И.И. Мечникова

Теория информации и кодирования

Белозёров Г.С., Варбанец П.Д., Гунявый О.А.

Одесса, 2013

Печатается по решению Ученого Совета ИМЭМ ОНУ
от 19 сентября 2013 года, протокол №1

составители: к. ф.-м. н. Г.С. Белозёров,
д. ф.-м. н. П.Д. Варбанец,
к. ф.-м. н. О.А. Гунявый

рецензенты: д. ф.-м. н. Ю.Г. Леонов,
к. ф.-м. н. С.М. Покась

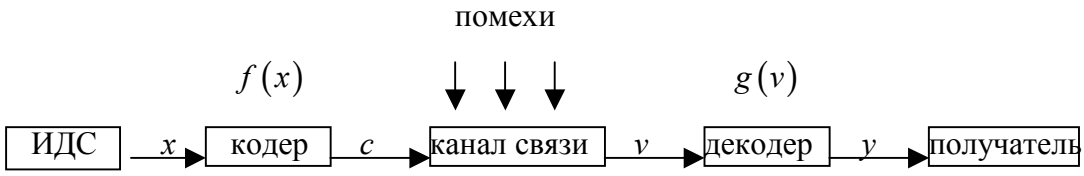
Оглавление

Введение

В настоящее время невозможно представить себе инженера-конструктора цифровых систем, который не был бы знаком с кодами, контролирующими ошибки. Необходимость в контроле ошибок сейчас настолько велика, а возможности электроники столь развиты, что научный и практический интерес к этой тематике непрерывно растёт. Умения и навыки применять кодирование стало важным для любого специалиста, создающего современные системы связи или большие цифровые системы. И это умение целится всё больше Этот курс лекций написан для студентов специальности "Компьютерная инженерия". Теоретические основы теории сопровождаются примерами, без которых трудно усваиваются некоторые положения теории. Авторы намерены подготовить учебное пособие, в котором будут изложены основные задачи теории информации и кодирования с решением тестовых задач и это составит вторую часть нашего курса "Теория информации и кодирования".

Лекция 1. Кодирование в каналах без помех.

Мы не даём определения информации. Но сможем определить количественную характеристику информации. Любую информацию приходится преобразовывать, передавать от источника информации к получателю, а также хранить информацию. В этом курсе мы излагаем математические основы теории информации. Мы различаем два типа источников информации: источник дискретных сообщений (ИДС) и источник непрерывных сообщений (ИНС). Каждое сообщение может быть передано по каналу связи, от работы которого зависит точность воспроизведения переданного сообщения получателем. Обычно каналы связи в состоянии передавать только дискретные сообщения. Поэтому непрерывные сообщения предварительно дискретизируют. (Задачу дискретизации непрерывных сообщений мы не будем рассматривать). Каждый ИДС для записи сообщений имеет свой алфавит s_1, s_2, …, s_k (это могут быть буквы, цветные шары, набор звуков и т.д.). Но канал связи может обслуживать различные ИДС, а потому пользуется своим алфавитом (обычно это элементы некоторого конечного поля $\mathbb{F}_q$). Поэтому возникает необходимость кодировать символы сообщений символами алфавита канала. Так возникает следующая принципиальная схема передачи сообщений по каналу связи:



Здесь $x-$ сообщение в алфавите ИДС, $f-$ функция, преобразующая сообщение x в кодовое слово c (записанное в алфавите канала связи), $v-$ измененное кодовое слово ($v=c+e$, где $e-$ ошибка, вызванная внешними помехами), $g-$ декодирующая функция, предназначенная для исправления возникшей ошибки, т.е. преобразующая v в c, а затем c в x, если удаётся исправить ошибку e), $y-$ полученное сообщение (y должно совпадать с x, если ошибка исправлена). Если помех нет, то всегда $y=x$. Так возникают две проблемы:

1. Если помех нет, то как найти наиболее быстрый код, позволяющий передавать любые сообщения из ИДС:
2. Если имеются помехи, то как найти достаточно скоростной код, исправляющий с высокой надёжностью допущенные ошибки в канале связи.

Методы решения этих проблем существенно различаются и в дальнейшем будут изложены.

Кодирование при отсутствии помех.

Пусть имеется ИДС, который использует для записи сообщений алфавит S из m символов $S=\{s_1,s_2,\ldots,s_m\}$. Пусть также имеется алфавит кода $A=\{a_1,a_2,\ldots,a_h\}$. В общем случае $h<m$. Каждый символ $s_i\in S$ кодируется в алфавите A словом длины n_i. Символы из S имеют различные вероятности появлений в сообщениях рассматриваемого ИДС.

Обозначим $\overline{n}=\sum_{i=1}^{m}p_in_i$, где $p_i=p(s_i)-$ вероятность появления s_i.

Следующие таблицы содержат вероятности букв русского и английского языков:

-	О	Е,Ё	А	И	Т	Н	С
0,175	0,090	0,072	0,062	0,062	0,053	0,053	0,045
Р	В	Л	К	М	Д	П	У
0,040	0,038	0,035	0,028	0,026	0,025	0,023	0,021
Я	Ы	З	Ь,Ъ	Б	Г	Ч	Й
0,018	0,016	0,016	0,014	0,014	0,013	0,012	0,010
Х	Ж	Ю	Ш	Ц	Щ	Э	Ф
0,009	0,007	0,006	0,006	0,004	0,003	0,003	0,002

A	B	C	D	E	F	G	H
0,080	0,016	0,028	0,040	0,129	0,026	0,020	0,054
I	J	K	L	M	N	O	P
0,078	0,002	0,004	0,035	0,024	0,075	0,068	0,018
Q	R	S	T	U	V	W	X
0,002	0,068	0,066	0,097	0,025	0,012	0,018	0,002
Y	Z						
0,002	0,001						

Алфавит A обычно представляет собой элементы некоторого конечного поля $\mathbb{F}_q$ (чаще всего $q=2$ или 3).

Значение $\overline{n}$ характеризует среднюю длину кода. Заметим, что $\overline{n}$ зависит не только от алфавита A, но и выбранного метода кодирования. Для данных $p_1,p_2,\ldots,p_m$ и алфавита A код с наименьшей средней длиной $\overline{n}$ называется эффективным. Основное требование, предъявляемое к кодам – это однозначность декодирования.

Пример. $S=\{s_1,s_2,s_3,s_4\}$, $A=\{0,1\}$.

a) $s_1=0$; $s_2=01$; $s_3=11$; $s_4=00$;

b) $s_1=0$; $s_2=01$; $s_3=011$; $s_4=111$;

c) $s_1=0$; $s_2=01$; $s_3=001$; $s_4=0011$.

Пусть получено сообщение: $y=0011$. Тогда декодирование по каждому из методов даёт:

a) $y\in\{s_4s_3, s_1s_1s_3\}$;

b) $y=s_1s_3$;

c) $y=s_4$.

Пусть заданы вероятности различных передаваемых символов алфавита S:

$$\begin{pmatrix} s_1 & s_2 & \dots & s_m \\ p_1 & p_2 & \dots & p_m \end{pmatrix}$$

и пусть $n_1,\dots,n_m$ – длины кодовых слов этих символов. Если $p_1\ge p_2\ge\dots\ge p_m$, то можно считать, что $n_1\le n_2\le\dots\le n_m$. Если это не так, то код заведомо неэффективен, ибо можно получить меньшую среднюю длину, меняя сопоставление кодовых слов символам $s_1,\dots,s_m$. Действительно, если $p_i>p_j$ но $n_i>n_j$, то в выражении для $\overline{n}$ содержится сумма $p_in_i+p_jn_j$, а переставляя кодовые слова символов s_i и s_j, получим новую сумму в выражении для $\overline{n}$ $p_in_j+p_jn_i$, причём остальные слагаемые в $\overline{n}$ не меняются. Но $\left(p_in_i+p_jn_j\right)-\left(p_in_j+p_jn_i\right)=$ $=p_i\left(n_i-n_j\right)-p_j\left(n_i-n_j\right)=\left(p_i-p_j\right)\left(n_i-n_j\right)>0$. Значит, новое $\overline{n}$ меньше старого $\overline{n}$. Итак, всегда можно добиться справедливости неравенств $p_1\ge p_2\ge\dots\ge p_m$ и $n_1\le n_2\le\dots\le n_m$.

Определение. Код C для совокупности символов $s_1,\dots,s_m$ называется мгновенным, если при получении знаков любого его кодового слова мы сразу об этом узнаём и не нужно исследовать последующие знаки.

Например, код c) из предыдущего примера хоть и является однозначным, но не является мгновенным. Код $s_1=0$, $s_2=10$, $s_3=110$, $s_4=111$ является мгновенным.

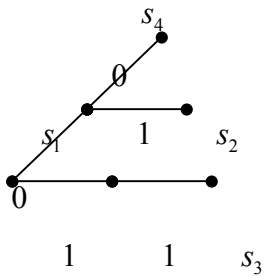
В мгновенном коде никакое кодовое слово не является префиксом другого.

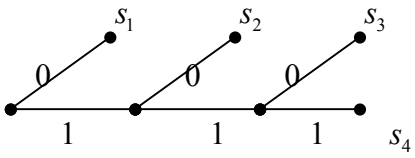
Мгновенный код всегда приводит к однозначному декодированию.

С каждым кодом можно связать граф, который называется деревом решения.

Пример.

1) $s_1 = 0,\ s_2 = 01,\ s_3 = 11,\ s_4 = 00;$ 2) $s_1 = 0,\ s_2 = 10,\ s_3 = 110,\ s_4 = 111.$





Для мгновенного кода только концевые узлы соответствуют кодовым словам.

Неравенство Крафта.

Пусть $C-$ мгновенный код в алфавите объёма h и пусть $n_1,\ldots,n_m-$ длины его кодовых слов. Тогда $$\sum_{i=1}^{m} h^{-n_i} \le 1.$$

Доказательство. Для каждого $j = 1,2,\ldots$ через ω_j обозначим количество кодовых слов длины j. Тогда имеем

$$\omega_1 \le h,\ \omega_2 \le (h-\omega_1)h,\ \omega_3 \le \big((h-\omega_1)h-\omega_2\big)h,\ \ldots,\ \omega_n \le h^n - \omega_1 h^{n-1} - \ldots - \omega_{n-1}h,$$

где $n-$ максимальная длина кодовых слов кода C. Разделив последнее неравенство на h^n, получим $$\sum_{i=1}^{n} \omega_i h^{-i} \le 1,$$

а это неравенство эквивалентно доказываемому неравенству.

Замечание. Мак-Милан доказал, что из справедливости неравенства Крафта для некоторого кода C обязательно следует, что код C однозначно декодируем. Более того, индукцией можно доказать, что из неравенства Крафта следует существование хотя бы одного мгновенного кода.

Одним из методов построения мгновенного кода является алгоритм Хаффмана (а соответствующий ему код называется кодом Хаффмана):

Итак, пусть требуется закодировать m символов $s_1,\ldots,s_m$ в алфавите объёма h и пусть $p_1, p_2,\ldots, p_m-$ вероятности этих символов.

1-ый шаг: упорядочиваем символы по убыванию их вероятностей. Т.е. можно сразу предположить, что $p_1 \ge p_2 \ge \ldots \ge p_m$;

2-й шаг: определяем число m_0, $2 \le m_0 \le h$ из сравнения $m \equiv m_0 \pmod{h-1}$, т.е. $\frac{m-m_0}{h-1} = k \in \mathbb{N}$. При $h=2$ полагаем $m_0 = 2$. Затем последние m_0 символов объединяем в группу, которой приписываем вероятность, равную сумме вероятностей этих символов;

3-й шаг: рассматриваем полученную группу как некоторый символ, и тогда оставшиеся символы вместе с этим новым символом располагаем в порядке убывания их вероятностей. Всего получится $(h-1)k+1$ символов;

4-й шаг: последние h символов объединяем в группу, т.е. переходим к шагу 2. Эту процедуру продолжаем дальше. Через $(k-1)$ шагов придём к группе из h новых символов с суммарной вероятностью равной 1;

5-й шаг: «символам» последней группы приписываем знаки из алфавита A. Каждый из этих символов либо является исходным, либо входит в группу. Для исходных символов кодирование закончено, а для групповых символов добавляем вторые знаки из алфавита A. Процесс продолжаем до тех пор, пока не исчерпаем все образованные группы.

Пример.

Пусть символы $s_1, s_2, s_3, s_4, s_5, s_6$ имеют соответственно вероятности 0,25; 0,25; 0,20; 0,15; 0,10; 0,05. Имеем

Код	Символ	p_i	$p_i^{(1)}$	$p_i^{(2)}$	$p_i^{(3)}$	$p_i^{(4)}$	$p_i^{(5)}$
10	s_1	0,25	0,25	**0,30**	**0,45**	**0,55**	**1**
11	s_2	0,25	0,25	0,25	0,30	0,45	
10	s_3	0,20	0,20	0,25	0,25		
000	s_4	0,15	0,15	0,20			
0010	s_5	0,10	**0,15**				
0011	s_6	0,05					

Здесь $m=6$, $h=2$, $m_0=2$.

Если же взять $h=3$, то получим $m_0=2$ и следующую реализацию алгоритма

Код	Символ	p_i	$p_i^{(1)}$	$p_i^{(2)}$	$p_i^{(3)}$
1	s_1	0,25	0,25	**0,50**	**1**
2	s_2	0,25	0,25	0,25	
00	s_3	0,20	0,20	0,25	
01	s_4	0,15	0,15		
020	s_5	0,10	**0,15**		
021	s_6	0,05			

Сформулируем теперь требования на дерево решения эффективного (иногда говорят оптимального) кода:

1) из всех промежуточных узлов, кроме последнего, выходят h ветвей (если это не так, то можно уменьшить среднюю длину дерева, вставив в этот промежуточный узел недостающее число ветвей из последнего узла);
2) из последнего промежуточного узла должно исходить не менее двух ветвей, так как иначе последнюю ветвь можно отрубить, уменьшив длину кода;
3) из последнего промежуточного узла должны исходить ветви с узлами, имеющими наименьшую вероятность (т.е. наименьшие по вероятности m_0 символов должны исходить из наиболее длинных промежуточных узлов, иначе можно уменьшить среднюю длину, поменяв местами ветви соответствующие наименее вероятным символам);
4) два символа с наименьшей вероятностью должны выходить из одного и того же промежуточного узла.

Мы видим, что код Хаффмана полностью удовлетворяет этим требованиям.

Лекция 2. Теорема Шеннона в каналах без помех.

Для изучения свойств кодов нам понадобятся некоторые вспомогательные утверждения.

Лемма Йенсена. Для любой случайной величины ξ и любой выпуклой функции $f(x)$ справедливо неравенство $M(f(x)) \le f(M(x))$,

где $M(\ldots)-$ математическое ожидание.

Доказательство. Так как функция $f(x)$ выпуклая, то в любой точке графика $y = f(x)$ касательная в этой точке находится над кривой.

Запишем уравнение касательной в т. x_0: $y - f(x_0) = f'(x_0)(x - x_0)$.

При $x_0 = M(\xi)$ имеем $y - f(M(\xi)) = f'(M(\xi))(x - M(\xi))$.

Отсюда, взяв $x = \xi$, $y = f(\xi)$ имеем $f(\xi) - f(M(\xi)) = f'(M(\xi))(\xi - M(\xi))$.

Переходя к математическому ожиданию, получаем

$$M(f(x)) - f(M(x)) \le 0.$$

Лемма (о двух распределениях). Пусть $p(x)$ и $q(x)-$ две функции плотности распределений случайных величин со значениями на одном и том же множестве X, тогда

$$\sum_{x \in X} p(x) \ln \frac{p(x)}{q(x)} \ge 0.$$

Доказательство. Рассмотрим уравнение касательной к графику функции $y = \ln x$ в точке $(1, 0)$, т.е. $y = x - 1$.

В силу выпуклости $y=\ln x$ $\quad\forall x>0\quad \ln x\le x-1$.

Но тогда $\sum\limits_{x\in X}p(x)\ln\dfrac{q(x)}{p(x)}\le\sum\limits_{x\in X}p(x)\left(\dfrac{q(x)}{p(x)}-1\right)=\sum\limits_{x\in X}\big(q(x)-p(x)\big)=1-1=0$.

Лемма (о средней длине мгновенного кода). Для мгновенного кода в алфавите объёма h средняя длина $\overline{n}$ не может быть меньше $\dfrac{H(s)}{\log h}$,

где $H(s)=-\sum\limits_{i=1}^{m}p_i\log p_i$, а основание логарифма больше единицы.

Доказательство. Рассмотрим мгновенный код C и пусть $n_1,\ldots,n_m-$ длины кодовых слов. В силу неравенства Крафта $\sum\limits_{i=1}^{m}h^{-n_i}\le 1$. Кроме того $\overline{n}=\sum\limits_{i=1}^{m}p_in_i$.

Обозначим $Q=\sum\limits_{i=1}^{m}h^{-n_i}$, $q_i=\dfrac{h^{-n_i}}{Q}$, так что $\sum\limits_{i=1}^{m}q_i=1$, и значит значения q_i можно рассматривать как вероятность значений некоторой случайной величины со значениями из множества C. Тогда в силу леммы о двух распределениях $\sum\limits_{i=1}^{m}p_i\log\dfrac{q_i}{p_i}\le 0$, и значит

$$-\sum_{i=1}^{m}p_i\log p_i\le-\sum_{i=1}^{m}p_i\log q_i=\sum_{i=1}^{m}p_in_i\log h+\sum_{i=1}^{m}p_i\log Q=\overline{n}\log h+\log Q\le\overline{n}\log h.$$

Тогда $\dfrac{H(s)}{\log h}\le\overline{n}$, что и требовалось доказать.

Теорема Шеннона. Пусть имеется ИДС S объёма m. Тогда для любого алфавита A объёма h существует мгновенный кода, средняя длина которого удовлетворяет неравенству

$$\frac{H(s)}{\log h}\le\overline{n}<\frac{H(s)}{\log h}+1.$$

Доказательство. Пусть $p_1,\ldots,p_m-$ вероятности символов из S. Для каждого $i=1,\ldots,m$ найдём натуральные числа n_i из неравенств

$$\log_h\frac{1}{p_i}\le n_i<\log_h\frac{1}{p_i}+1. \qquad (*)$$

Тогда $\dfrac{1}{p_i}\le h^{n_i}<\dfrac{h}{p_i}$, откуда $h^{-n_i}\le p_i$, и значит $\sum\limits_{i=1}^{m}h^{-n_i}\le 1$. Т.е. выполняется неравенство Крафта, а потому существует мгновенный код с длинами $n_1,\ldots,n_m$.

Усреднив неравенство $(*)$ по вероятностям p_i получаем окончательно

$$-\sum_{i=1}^{m}p_i\log_h p_i=\frac{H(s)}{\log h}\le\sum_{i=1}^{m}p_in_i=\overline{n}<-\sum_{i=1}^{m}p_i\log_h p_i+1=\frac{H(s)}{\log h}+1.$$

Лекция 3. Количество информации и энтропийная функция.

Рассмотрим два ИДС: X и Y и пусть $P(x,y)-$ функция плотности совместного распределения $(X,Y)=XY=\{(x,y)\,|\,x\in X, y\in Y\}$. Имеем

$$\sum_{y\in Y}P(x,y)=P(x),\qquad \sum_{x\in X}P(x,y)=P(y),\quad P(x|y)=\frac{P(x,y)}{P(y)}.$$

Мы будем трактовать x как некоторое событие на входе канала, а y – на выходе.

Для сравнения априорной вероятности $P(x)$ с апостериорной вероятностью $P(x|y)$ рассмотрим величину $I(x;y)=\log\frac{P(x|y)}{P(x)}$, которую будем называть информацией в y относительно x. Более точно, $I(x;y)-$ это количество информации, содержащейся в y относительно x. Но мы имеем

$$I(x;y)=\log\frac{P(x|y)}{P(x)}=\log\frac{P(x|y)P(y)}{P(x)P(y)}=\log\frac{P(x,y)}{P(x)P(y)}=\log\frac{P(y|x)P(x)}{P(x)P(y)}=\log\frac{P(y|x)}{P(y)}=I(y;x).$$

Т.о., количество информации, содержащейся в y относительно x равно количеству информации, содержащейся в x относительно y. Поэтому, $I(x;y)$ иногда называют взаимной информацией между x и y. Обычно основанием логарифма служат числа 2, e, 10 и тогда количество взаимной информации соответственно измеряют в битах, натах или дитах (хартли – по фамилии учёного, впервые изучавшего меру информации). $I(x;y)$ можно рассматривать как меру взаимной связи между x и y. В частности, если x и y статистически независимы, то $I(x;y)=0$. Если $I(x;y)>0$, то говорят, что событие y (соответственно x) способствует появлению события x (соответственно y).

Аналогично можно говорить о взаимной информации между x и y при условии, что произошло некоторое событие $z\in Z$ $\quad I(x;y|z)=\log\frac{P(x|yz)}{P(x|z)}$.

Свойство аддитивности количества информации.

Пусть $(x,y,z)\in XYZ$. Обозначим

$$I(x;yz)=\log\frac{P(x|yz)}{P(x)} \tag{3.1}$$

Мы интерпретируем x как сообщение на входе, а $y,z-$ символы последовательно порождаемые на выходе канала. Тогда $I(x;yz)-$ информация об x, содержащаяся в y,z.

Мы имеем

$$I(x;yz)=\log\frac{P(x|yz)}{P(x)}=\log\frac{P(x|yz)P(x|y)}{P(x)P(x|y)}=$$

$$=\log\frac{P(x|yz)}{P(x|y)}+\log\frac{P(x|y)}{P(x)}=I(x;y)+I(x;z|y). \qquad (3.2)$$

И аналогично, $$I(x;yz)=I(x;z)+I(x;y|z). \qquad (3.3)$$

Отсюда $$I(x;yz)=\frac{1}{2}\left[I(x;y)+I(x;z)+I(x;y|z)+I(x;z|y)\right]. \qquad (3.4)$$

Это соотношение симметрично относительно y и z, а потому его можно рассматривать как количество информации в x относительно y и z.

Изменяя в (3.2) порядок символов, получим

$$I(yz;x)=I(y;x)+I(z;x|y)=I(y;x)+I(x;z|y). \qquad (3.5)$$

Соотношения (3.2)-(3.5) приводят к

$$I(xu;yv)=I(xu;y)+I(xu;v|y)=I(x;y)+I(u;y|x)+I(x;v|y)+I(u;v|xy).$$

Если теперь XY и UV статистически независимы, т.е. $P((x,y);(u,v))=P(x,y)P(u,v)$, то

$$I(xu;yv)=I(x,y)+I(u,v).$$

Пример. Пусть имеется 8 сообщений занумерованных от 0 до 7. Имеем $u_3=011$, $x_0=0$, $y_1=1$, $z_2=1$.

Вероятности событий.

сообщение	код	первичная	после получения 0	после получения 1	после получения второй 1
0	000	1/4	1/3	0	0
1	001	1/4	1/3	0	0
2	010	1/8	1/6	1/2	0
3	011	1/8	1/6	1/2	1
4	100	1/16	0	0	0
5	101	1/16	0	0	0
6	110	1/16	0	0	0
7	111	1/16	0	0	0

Отсюда $I(u_3, x_0) = \log\frac{1/6}{1/8} = \log\frac{4}{3}$, $I(u_3; y_1|x_0) = \log\frac{1/2}{1/6} = \log 3$,

$$I(u_3; z_2|x_0 y_1) = \log\frac{1}{1/2} = \log 2.$$

И тогда $I(u_3; x_0 y_1 z_2) = \log\frac{4}{3} + \log 3 + \log 2 = 3(\text{бита})$,

что следует из соотношения $I(u_3; x_0 y_1 z_2) = I(u_3, x_0) + I(u_3; y_1|x_0) + I(u_3; z_2|x_0 y_1)$, которое получается аналогично соотношению (3.2).

Заметим, что $I(u_3; x_0 y_1 z_2) = \log\frac{1}{P(u_3)} = \log 8 = 3(\text{бита})$.

Количество собственной информации.

Для любых $x \in X$ и $y \in Y$

$$P(x|y) \le 1,\ P(y|x) \le 1 \tag{3.6}$$

а потому

$$I(x, y) = \log\frac{P(x|y)}{P(x)} = \log\frac{P(y|x)}{P(y)} \le \begin{cases} \log\frac{1}{P(x)} \overset{def}{=} I(x), \\ \log\frac{1}{P(y)} \overset{def}{=} I(y), \end{cases} \tag{3.7}$$

причём в (3.7) имеет место знак равенства, если в (3.6) имеет место знак равенства.

Величину $I(x)$ назовём количеством собственной информации в x.

Количеству собственной информации можно дать две интерпретации.

Если x есть сообщение на входе, то $I(x)$ измеряет количество информации, необходимое для однозначного определения x. Это следует из того, что взаимная информация между x и любым другим событием, например y, становится равной $I(x)$ только при условии $P(x|y) = 1$, т.е. y однозначно определяет x. Т.о., $I(x)-$ это максимальное количество информации, которое можно иметь об x.

Другая интерпретация: если рассматривать y как первый символ кодового слова, соответствующего сообщению x на входе, то в силу (3.6)-(3.7) информация, содержащаяся в y относительно x, равна собственной информации в y только если $P(y|x) = 1$. Иначе говоря, рассматриваемый символ y определяется однозначно сообщением x, что и следует ожидать в правильно сконструированном кодере.

Аналогично определяется собственная условная информация:

$I(x|z) = -\log P(x|z)$, $\qquad I(y|z) = -\log P(y|z)$.

Но тогда $\qquad I(x,y) = \log \dfrac{P(x|y)}{P(x)} = I(x) - I(x|y) = I(y) - I(y|x)$,

т.е. взаимная информация равна разности информации об x (соответственно y) до и после того как произошло y (соответственно x), или разности информации, требуемой для идентификации y до и после того, как становится известным x.

Иногда собственную информацию $I(x)$ называют случайной энтропией. Иначе говоря, если случайная величина задана вероятностями своих значений $P(x_k) = p_k$, то рассмотрим новую случайную величину со значениями $I(x_1),\ldots,I(x_m)$ и вероятностями $p_1,\ldots,p_m$. Математическое ожидание этой новой случайной величины называется энтропией (энтропийной функцией) случайной величины X и обозначается $H(X)$, т.е.

$$H(X) = -\sum_{i=1}^{m} p_i \log p_i = \sum_{i=1}^{m} p_i I(x_i) = I(X).$$

С этой функцией мы уже встречались, изучая среднюю длину мгновенного кода.

Из определения $H(X)$ следует, что энтропия есть среднее количество информации, которое надо иметь для вычисления отдельных значений случайной величины X. Иногда $H(X)$ называют неопределённостью при наблюдении над случайной величиной X.

Аналогично мы можем рассмотреть среднюю условную информацию:

$$I(X|Y) = \sum_{x\in X}\sum_{y\in Y} P(x,y) I(x|y) = -\sum_{x,y} P(x,y)\log P(x|y) \overset{def}{=} H(X|Y).$$

$H(X|Y)$ называется условной энтропийной функцией.

Мы можем также определить условное среднее взаимной информации:

$$I(X,y) = \sum_{x\in X} P(x|y) \log \frac{P(x|y)}{P(x)}.$$

Теорема 1. Для энтропийной функции $H(X)$ выполняется неравенство

$$0 \le H(X) \le \log m,$$

где $m-$ число различных значений, принимаемых случайной величиной X.

Доказательство. Из определения $H(X)$ и условий $0 \le p_i \le 1$ следует, что все слагаемые $-p_i \log p_i$ неотрицательны, а потому $H(X) \ge 0$

Кроме того, исследуя на максимум функцию $f(u_1,\dots u_m)=-\sum_{i=1}^{m} u_i \log u_i$ при условии $u_1+\dots+u_m=1$, получим, что он достигается при $u_1=u_2=\dots+u_m=1/m$ и при этом равен $-\sum_{i=1}^{m}\frac{1}{m}\log\frac{1}{m}=\log m$.

Замечание 1. Граничное значение $H(X)=0$ достигается только в случае вырожденной случайной величины, когда для некоторого j $p_j=1$.

Замечание 2. Для любых ИДС X и Y имеем $H(X|Y)\ge 0$. Это сразу видно из определения $H(X|Y)=\sum_{x,y} p(x,y)\log\frac{1}{p(x|y)}$.

Теорема 2. $I(X;y_i)\ge 0$, причём $I(X;y_i)=0$ если $P(x|y_i)=P(x)\ \ \forall x\in X$.

Доказательство. Имеем

$$-I(X;y_i)=\sum_{x\in X} P(x|y_i)\log\frac{P(x)}{P(x|y_i)}\le\sum_{x\in X} P(x|y_i)\left(\frac{P(x)}{P(x|y_i)}-1\right)\log e=0.$$

Далее определяем

$$I(X;Y)\overset{def}{=}\sum_{x,y} P(x,y)\log\frac{P(x|y)}{P(x)}=\sum_{y} P(y)I(X;y)\ge 0.$$

Теорема 3. Для любых ИДС X и Y имеем $H(X|Y)\le H(X)$.

Доказательство. Так как $\sum_{y} P(x,y)=P(x)$, то

$$H(X|Y)-H(X)=-\sum_{x,y} P(x,y)\log P(x|y)+\sum_{x} P(x)\log P(x)=.$$

$$=-\sum_{x,y} P(x,y)\log P(x|y)+\sum_{x,y} P(x,y)\log P(x)=\sum_{x,y} P(x,y)\log\frac{P(x)}{P(x|y)}\le$$

$$\le\sum_{x,y} P(x,y)\left(\frac{P(x)}{P(x|y)}-1\right)\log e=\left(\sum_{x} P(x)\sum_{y} P(y)-\sum_{x,y} P(x,y)\right)\log e=0,$$

откуда и следует неравенство.

Замечание. Равенство $H(X|Y)=H(X)$ достигается только в случае, когда X и Y статистически независимы.

Теорема 4.(свойство иерархической аддитивности).

$$H(X,Y)=H(X)+H(Y|X)=H(Y)+H(X|Y).$$

Доказательство.

$$H(X,Y)=-\sum_{x,y}P(x,y)\log P(x,y)=-\sum_{x,y}P(x,y)\left(\log P(x)+\log P(y|x)\right)=$$

$$=-\sum_{x}\log P(x)\sum_{y}P(x,y)-\sum_{x,y}P(x,y)\log P(y|x)=-\sum_{x}P(x)\log P(x)-\sum_{x,y}P(x,y)\log P(y|x)=$$

$$=H(X)+H(Y|X).$$

аналогично доказывается вторая часть равенства.

Следствие. Если X и Y статистически независимы, то $H(X,Y)=H(X)+H(Y)$.

Теорема 5. $H(X,Y)=H(X)+H(Y)-I(X,Y)$.

Доказательство. Имеем

$$I(X,Y)=\sum_{x,y}P(x,y)\log\frac{P(x|y)}{P(x)}=\sum_{x,y}P(x,y)\log\frac{P(x,y)}{P(x)P(y)}=I(Y,X)=$$

$$=\sum_{x,y}P(x,y)\log P(x,y)-\sum_{x,y}P(x,y)\log P(x)-\sum_{x,y}P(x,y)\log P(y)=-H(X,Y)+H(X)+H(Y).$$

Следствие. $I(X,Y)=H(X)-H(X|Y)=H(Y)-H(Y|X)$.

Замечание. Количество взаимной информации $I(X,Y)$ равно уменьшению неопределённости для одной случайной величины, если мы наблюдаем другую случайную величину. Иначе, среднее количество информации о сообщении, содержащееся в принятом сигнале равно среднему количеству информации, требуемому для определения сообщения x $\left(I(X)=H(X)\right)$ минус среднее количество информации, которое всём ещё потребуется для определения x после принятия сигнала $\left(H(X|Y)\right)$. Поэтому $H(X)$ и понимается как среднее количество переданной информации, $I(X,Y)-$ как среднее количество полученной информации, а $H(X|Y)-$ как среднее количество информации, потерянное вследствие шума (или как «ненадёжность»).

Из следствия также имеем $I(X,Y)\le\min\{H(X),H(Y)\}$.

Далее,

$$I(X,YZ)=H(X)-H(YZ|X)=H(X)-H(Y|X)-H(Z|XY)=$$

$$=H(X)-H(Y)-H(Z)+I(X,Y)+I(XY,Z)=H(X)-H(Y)-H(Z)+I(X,Z)+I(XZ,Y).$$

Теорема 6. Если Y функционально зависит от X, то $H(Y)\le H(X)$, причём равенство достигается только в случае, когда функциональная зависимость является биекцией.

Доказательство. Имеем $H(X,Y)=H(X)+H(Y|X)=H(Y)+H(X|Y)$.

Но $H(Y|X)=\sum_{a\in X}H(Y|X=a)=-\sum_{y\in Y}\sum_{a\in X}P(a,y)\log P(y|x=a)$, а случайная величина $(Y|x=a)$ является вырожденной, и поэтому $H(Y|X=a)=0$, откуда $H(Y|X)=0$. Таким образом, $H(X)=H(Y)+H(X|Y)$, откуда $H(Y)\le H(X)$, причём равенство справедливо только в случае $H(X|Y)=0$, который получается только при функциональной зависимости X от Y. Но по условию Y функционально зависит от X. Откуда X и Y биективно функционально зависимы. Что и требовалось доказать.

Теорема 7. При функциональном преобразовании $\widetilde{X}=\varphi(X)$ или $\widetilde{Y}=\psi(Y)$ количество информации не возрастает, т.е.

$$I(X,Y)\ge I(\varphi(X),Y),\qquad I(X,Y)\ge I(X,\psi(Y)).$$

Доказательство. Так как $H(Y|\varphi(X))\ge H(Y|X)$, то

$$I(\widetilde{X},Y)=H(Y)-H(Y|\widetilde{X})\le H(Y)-H(Y|X)=I(X,Y).$$

Лекция 4. Удельная энтропия n-символьных последовательностей.

Пусть зафиксирован некоторый алфавит $A=\{a^1,\dots,a^m\}$. Будем рассматривать $n-$символьные последовательности $\Xi_n=(\xi_1,\dots,\xi_n)\in A^n$. При этом полное описание ИДС задаётся моделью случайного процесса Ξ_n с дискретным временем $t\in\mathbb{N}$ и дискретным пространством состояний

$$\left\langle A^n, p_n(a_1,\dots,a_n)=P(\xi_1=a_1,\dots,\xi_n=a_n)\right\rangle,\ a_i\in A.$$

При этом мы всегда считаем, что выполняются условия самосогласованности $(1\le k_1\le\dots\le k_l\le n;1\le l\le n)$:

$$p_l(a_{k_1},\dots,a_{k_l})=\sum_{\substack{a_i\in A\\ i\in\{1,\dots,n\}\setminus\{k_1,\dots,k_l\}}}p_n(a_1,\dots,a_n).$$

Мы будем называть случайный процесс Ξ_n стационарным, если $P(\xi_{r+1}=a_1,\dots,\xi_{r+n}=a_n)$ не зависит от r, т.е. инвариантно относительно начала отсчёта.

Стационарный ИДС называется и с т о ч н и к о м б е з п а м я т и, если для любых $a_i \in A$

$$p_n(a_1,\ldots,a_n)=\prod_{i=1}^{n} p_1(a_i),$$

(т.е. ξ_i независимы в совокупности и одинаково распределены).

Нас будет интересовать поведение $H(\Xi_n)$ при $n\to\infty$. Для стационарных ИДС $H(\Xi_{n,r})=H(\Xi_n)=-\sum_{a_i\in A} p_n(a_1,\ldots,a_n)\log p_n(a_1,\ldots,a_n)$.

О п р е д е л е н и е. Удельной энтропией стационарного ИДС называется предел

$$h=\lim_{n\to\infty}\frac{1}{n}H(\Xi_n),$$

если этот предел существует.

Величину h можно рассматривать как энтропию, приходящуюся на один символ и вычисленную по бесконечно длинному случайному сообщению.

Используя символ конкатенации $\|$, получаем $\Xi_n=(\xi_n\|\Xi_{n-1})$. Положим ещё условную энтропию ξ_n относительно Ξ_{n-1}:

$$H(\xi_n|\Xi_{n-1})\overset{def}{=}-\sum_{a_i\in A} p_n(a_1,\ldots,a_n)\log p_n(a_n|a_1,\ldots,a_{n-1}),\qquad H(\xi_1|\Xi_0)\overset{def}{=}H(\xi_1).$$

Л е м м а. Для стационарного ИДС последовательность $H(\xi_n|\Xi_{n-1})$ имеет конечный предел $\lim_{n\to\infty}H(\xi_n|\Xi_{n-1})\overset{def}{=}H(\xi|\Xi_\infty)$.

Д о к а з а т е л ь с т в о. Поскольку последовательность $\{H(\xi_n|\Xi_{n-1})\}$ ограничена снизу числом 0, то наше утверждение будет доказано, если мы покажем, что эта последовательность монотонна и не возрастает.

В силу стационарности $\forall i,j\in\mathbb{N}\quad H(\xi_j|\xi_{j-i},\ldots,\xi_{j-1})=H(\xi_n|\xi_{n-i},\ldots,\xi_{n-1})$. Поэтому первые n членов нашей последовательности совпадают с последовательностью

$$H(\xi_n),H(\xi_n|\xi_{n-1}),\ldots,H(\xi_n|\xi_1,\ldots,\xi_{n-1}).$$

Но добавление условий не увеличивает энтропию, поэтому

$$H(\xi_n)\ge H(\xi_n|\xi_{n-1})\ge\ldots\ge H(\xi_n|\xi_1,\ldots,\xi_{n-1}),$$

что и требовалось доказать.

Т е о р е м а. Для произвольного стационарного ИДС удельная энтропия существует, причём $h = H\left(\xi \middle| \Xi_\infty\right)$.

Д о к а з а т е л ь с т в о. Сначала покажем, что последовательность $\{h_n\} \overset{def}{=} \left\{\frac{H\left(\Xi_n\right)}{n}\right\}$ монотонно возрастающая и ограничена снизу нулём.

В силу свойства иерархической аддитивности имеем

$$H\left(\Xi_{n+1}\right) = H\left(\Xi_n \,\|\, \xi_{n+1}\right) = H\left(\Xi_n\right) + H\left(\xi_{n+1} \middle| \Xi_n\right) \tag{4.1}$$

Но в силу свойства условной энтропии и стационарности

$$H\left(\xi_{n+1} \middle| \Xi_n\right) = H\left(\xi_{n+1} \middle| \xi_1, \ldots, \xi_n\right) \le H\left(\xi_{n+1} \middle| \xi_2, \ldots, \xi_n\right) = H\left(\xi_n \middle| \xi_1, \ldots, \xi_{n-1}\right) = H\left(\xi_n \middle| \Xi_{n-1}\right) \tag{4.2}$$

Значит $$H\left(\Xi_{n+1}\right) \le H\left(\Xi_n\right) + H\left(\xi_n \middle| \Xi_{n-1}\right). \tag{4.3}$$

Далее, по свойству иерархической аддитивности энтропии и из (4.2)

$$H\left(\Xi_n\right) = H\left(\xi_1, \ldots, \xi_n\right) = \sum_{i=1}^{n} H\left(\xi_i \middle| \Xi_{i-1}\right) \ge n H\left(\xi_n \middle| \Xi_{n-1}\right) \tag{4.4}$$

Теперь из (4.3) и (4.4) имеем:

$0 \le H\left(\Xi_{n+1}\right) \le H\left(\Xi_n\right) + \frac{1}{n} H\left(\Xi_n\right)$, откуда $0 \le \frac{H\left(\Xi_{n+1}\right)}{n+1} \le \frac{H\left(\Xi_n\right)}{n}$.

Этим показано, что $\lim\limits_{n\to\infty} h_n = \lim\limits_{n\to\infty} \frac{H\left(\Xi_n\right)}{n}$ существует.

Далее, из (4.2) для произвольного $k \le n$:

$$h_n = \frac{1}{n}\sum_{i=1}^{n} H\left(\xi_i \middle| \Xi_{i-1}\right) = \frac{1}{n}\sum_{i=1}^{k} H\left(\xi_i \middle| \Xi_{i-1}\right) + \frac{1}{n}\sum_{i=k+1}^{n} H\left(\xi_i \middle| \Xi_{i-1}\right) \le \frac{k}{n} H\left(\xi_1\right) + \frac{n-k}{n} H\left(\xi_{k+1} \middle| \Xi_k\right) \tag{4.5}$$

Теперь в силу леммы для данного $\varepsilon > 0$ найдётся $k = k\left(\varepsilon\right)$ такой, что

$$H\left(\xi_{k+1} \middle| \Xi_k\right) - H\left(\xi \middle| \Xi_\infty\right) \le \frac{\varepsilon}{2}.$$

По выбранному k, который не зависит от n, определим $n_0 = n_0\left(k, \varepsilon\right)$ так, что при $n > n_0$ выполняется неравенство $\frac{k}{n} H\left(\xi_1\right) \le \frac{\varepsilon}{2}$. Поэтому в силу (4.5) получаем оценку

$$h_n \le H\left(\xi \middle| \Xi_\infty\right) + \varepsilon \tag{4.6}$$

С другой стороны, в силу (4.4) $h_n = \frac{H\left(\Xi_n\right)}{n} \ge H\left(\xi_n \middle| \Xi_{n-1}\right) \ge H\left(\xi \middle| \Xi_\infty\right)$.

Учитывая (4.6), получаем требуемое утверждение.

С л е д с т в и е. Для стационарного ИДС справедливо асимптотическое представление при $n \to \infty$ $H\left(\Xi_n\right) = hn + \mathrm{o}\left(n\right)$.

Последнее утверждение можно уточнить: $H(\Xi_n) = hn + 2b + \mathrm{o}(1)$,

где $b = \frac{1}{2}\lim\limits_{m,n\to\infty}\left(H(\Xi_m) + H(\Xi_n) - H(\Xi_{m+n})\right) \geq 0$.

Заметим, что $b = 0$, если ИДС стационарный и без памяти, ибо тогда $H(\Xi_{m+n}) = H(\Xi_m) + H(\Xi_n)$.

Лекция 5. **Энтропийные характеристики марковских последовательностей.**

Однородные цепи Маркова (ОЦМ) часто используются как модели стохастической зависимости символов сообщений в каналах связи.

Пусть дана ОЦМ с дискретным временем и конечным пространством состояний $A = \{a_1, \ldots a_m\}$, $m < \infty$. Это означает, что вероятности будущих значений (состояний) зависят только от настоящих значений и не зависят от прошлых значений

$$P\left(\xi_{t+1} = b_{t+1} \middle| \xi_1 = b_1, \ldots, \xi_t = b_t\right) = P\left(\xi_{t+1} = b_{t+1} \middle| \xi_t = b_t\right).$$

Вероятностные характеристики ОЦМ полностью определяются вектор-столбцом начальных состояний и матрицей одношаговых переходов p_{ij}, $i, j \in \{1, \ldots, m\}$:

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix}, \quad p_{ij} = P\left(\xi_{t+1} = a_j \middle| \xi_t = a_i\right); \qquad \pi = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_m \end{pmatrix}, \qquad \pi_i = P(\xi_1 = a_i).$$

При этом выполняются условия нормировки:

$$\sum_{i=1}^{m} \pi_i = 1, \qquad \sum_{j=1}^{m} p_{ij} = 1 \ \ \forall i \in \{1, \ldots, m\}.$$

Введём ещё вектор-столбец стационарных вероятностей $\pi^* = \begin{pmatrix} \pi_1^* \\ \vdots \\ \pi_m^* \end{pmatrix}$, который является решением системы линейных алгебраических уравнений

$$\begin{cases} \sum\limits_{i=1}^{m} \pi_i^* p_{ij} = \pi_j^*; \\ \sum\limits_{i=1}^{m} \pi_i^* = 1. \end{cases}$$

Обозначим $H^*(\xi_1) = -\sum\limits_{i=1}^{m} \pi_i^* \log \pi_i^*$. $H^*(\xi_1)$ называется энтропией стационарного распределения вероятностей.

Если начальный вектор-столбец π совпадает с π^*, то мы имеем однородную стационарную цепь Маркова. Пусть ещё

$$h^* = -\sum_{i=1}^{m} \pi_i^* \sum_{j=1}^{m} p_{ij} \log p_{ij}.$$

Т е о р е м а. Для однородной стационарной цепи Маркова справедливо равенство

$$H(\Xi_n) = H^*(\xi_1) + (n-1)h^*.$$

Д о к а з а т е л ь с т в о. Из определения ОЦМ

$$p_n(a_{i_1},\ldots,a_{i_n}) = P(\xi_1 = a_{i_1}) \prod_{t=1}^{n-1} P(\xi_{t+1} = a_{i_{t+1}} \mid \xi_t = a_{i_t},\ldots,\xi_1 = a_{i_1}) = \pi_{i_1} \prod_{t=1}^{n-1} p_{i_t i_{t+1}}.$$

Заметим,

$$H(\Xi_n) = -\sum_{i_1,\ldots i_n=1}^{m} p_n(a_{i_1},\ldots,a_{i_n}) \left(\log \pi_{i_1} + \sum_{t=1}^{n-1} \log p_{i_t i_{t+1}} \right) = (\text{в силу согласованности вероятностных распределений}) = -\sum_{i_1=1}^{m} \pi^*_{i_1} \log \pi_{i_1} - \sum_{t=1}^{n-1} \sum_{i_t, i_{t+1}=1}^{m} P(\xi_t = a_{i_t}, \xi_{t+1} = a_{i_{t+1}}) \log p_{i_t i_{t+1}} =$$

$$= -\sum_{i_1=1}^{m} \pi^*_{i_1} \log \pi_{i_1} - \sum_{t=1}^{n-1} \sum_{i_t, i_{t+1}=1}^{m} P(\xi_t = a_{i_t}) p_{i_t i_{t+1}} \log p_{i_t i_{t+1}}.$$

В силу стационарности ОЦМ $P(\xi_t = a_{i_t}) = \pi^*_{i_t}$, поэтому

$$H(\Xi_n) = H^*(\xi_1) + \sum_{t=1}^{n-1} \sum_{i=1}^{m} \left(-\pi^*_i \sum_{j=1}^{m} p_{ij} \log p_{ij} \right) = H^*(\xi_1) + (n-1)h^*,$$

что и требовалось доказать.

С л е д с т в и е 1. Для стационарной ОЦМ удельная энтропия h $n-$символьной последовательности равна h^*.

Действительно, $$h = \lim_{n\to\infty} \frac{H(\Xi_n)}{n} = \lim_{n\to\infty} \frac{1}{n} \left(H^*(\xi_1) + (n-1)h^* \right) = h^*.$$

С л е д с т в и е 2. Для стационарной ОЦМ $$H(\Xi_n) = hn + 2b,$$

где $$2b = \sum_{i,j=1}^{m} \pi^*_i p_{ij} \log \frac{p_{ij}}{\pi^*_j}.$$

Действительно, из доказанной теоремы и следствия предыдущего параграфа, имеем

$H(\Xi_n) - hn = H^*(\xi_1) - h^* \overset{def}{=} 2b$. Таким образом, $$2b = -\sum_{i=1}^{m} \pi^*_i \log \pi^*_i + \sum_{i=1}^{m} \pi^*_i \sum_{j=1}^{m} p_{ij} \log p_{ij} =$$

$$= -\sum_{i=1}^{m} \pi^*_i \left(\sum_{j=1}^{m} p_{ij} = 1 \right) \log \pi^*_i + \sum_{i=1}^{m} \pi^*_i \sum_{j=1}^{m} p_{ij} \log p_{ij} = \sum_{i=1}^{m} \pi^*_i \sum_{j=1}^{m} p_{ij} \log \frac{p_{ij}}{\pi^*_i}.$$

П р и м е р. Рассмотрим бинарную ОЦМ: $\xi_1, \xi_2, \ldots \in A = \{0,1\}$, т.е. $m = 2$, $a_1 = 0$, $a_2 = 1$. Вероятностные характеристики ОЦМ

$$\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}, \; P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \; \alpha, \beta \in [0,1].$$

Если $\alpha+\beta=1$, $\pi_1=1-\alpha$, $\pi_2=\alpha$, то имеем схему независимых испытаний. Найдём стационарное распределение π^*. Для этого решим систему

$$\begin{cases}\pi_1^*(1-\alpha)+\pi_2^*\beta=\pi_1^*\\ \pi_1^*+\pi_2^*=1\end{cases}; \qquad \text{откуда} \qquad \pi^*=\begin{pmatrix}\dfrac{\beta}{\alpha+\beta}\\ \dfrac{\alpha}{\alpha+\beta}\end{pmatrix}.$$

Тогда $H^*(\xi_1)=-\dfrac{\beta}{\alpha+\beta}\log\dfrac{\beta}{\alpha+\beta}-\dfrac{\alpha}{\alpha+\beta}\log\dfrac{\alpha}{\alpha+\beta}$.

$h=-\dfrac{\beta}{\alpha+\beta}\big(\alpha\log\alpha+(1-\alpha)\log(1-\alpha)\big)-\dfrac{\alpha}{\alpha+\beta}\big(\beta\log\beta+(1-\beta)\log(1-\beta)\big)$, а потому

$$2b=2\frac{\alpha\beta}{\alpha+\beta}\log\beta+\log\left(1+\frac{\alpha}{\beta}\right)+\frac{\beta}{\alpha+\beta}(1-\alpha)\log(1-\alpha)+\frac{\alpha}{\alpha+\beta}(1-\beta)\log(1-\beta).$$

При $\alpha=\beta$ получаем $\pi_1^*=\pi_2^*=\dfrac{1}{2}$, $H^*(\xi_1)=\log 2=1$, $h=-\big((1-\alpha)\log(1-\alpha)+\alpha\log\alpha\big)$, $2b=\log 2+(1-\alpha)\log(1-\alpha)+\alpha\log=1-h$.

Заметим, что при $\alpha=\dfrac{1}{2}$ удельная энтропия максимальна, а при $\alpha\to 0$ или $\alpha\to 1$ удельная энтропия $h\to 0$, что согласуется с фактом уменьшения неопределённости.

Лекция 6. **Энтропия источника непрерывных сообщений.**

Пусть источник непрерывных сообщений задан непрерывной плотностью вероятностей $p_\xi(x)$. Положим $H(\xi)\overset{def}{=}-\int\limits_{-\infty}^{\infty}p_\xi(x)\log p_\xi(x)\,\mathrm{d}x$.

$H(\xi)$ называется дифференциальной энтропией случайного сообщения ξ с плотностью распределения $p_\xi(x)$.

Условной дифференциальной энтропией ξ относительно η называется величина

$$H(\xi|\eta)\overset{def}{=}-\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}p_{\xi,\eta}(x,y)\log p_{\xi|\eta}(x|y)\,\mathrm{d}x\,\mathrm{d}y,$$

где $p_{\xi,\eta}(x,y)-$ совместная плотность распределения случайных величин (ξ,η),

$p_{\xi|\eta}(x|y)=\dfrac{p_{\xi\eta}(x,y)}{p_\eta(y)}-$ условная плотность распределения ξ при условии $\eta=y$.

Замечание. Случайные величины ξ, η не обязательно однородны.

Лемма 1. Справедливо свойство иерархической аддитивности дифференциальной энтропии для произвольных случайных символов $(\xi,\eta)\in\mathbb{R}^2$

$$H(\xi,\eta)=H(\xi)+H(\eta|\xi)=H(\eta)+H(\xi|\eta).$$

Доказывается аналогично случая ИДС.

Лемма 2. Пусть $\xi\in\mathbb{R}^m-$ случайный символ с плотностью распределения $p_\xi(x)$, $x\in\mathbb{R}^m$, и дифференциальной энтропией $H(\xi)$, а $y=f(x):\mathbb{R}^m\to\mathbb{R}^m-$ взаимно однозначное непрерывно дифференцируемое преобразование. Тогда случайный символ $\eta=f(\xi)$ имеет дифференциальную энтропию

$$H(\eta)=H(\xi)+M\left\{\log\left|J_f(\xi)\right|\right\},$$

где $M(z)-$ математическое ожидание случайной величины z, а $J_f(x)=\left|\frac{Df(x)}{Dx}\right|-$ якобиан преобразования $y=f(x)$.

Доказательство. Обозначим $x=f^{-1}(y)-$ обратное функциональное преобразование, $J_{f^{-1}}(y)=\left|\frac{Df^{-1}(y)}{Dy}\right|-$ его якобиан. Тогда $p_\eta(y)=p_\xi\left(f^{-1}(y)\right)\left|J_{f^{-1}}(y)\right|.$

Поэтому $H(\eta)=M\left(-\log p_\eta(\xi)\right)=M\left(-\log p_\xi\left(f^{-1}(\eta)\right)\right)-M\left(\log\left|J_{f^{-1}}(y)\right|\right).$

Но известно, что $\left|J_{f^{-1}}(y)\right|=\left|J_f(x)\right|^{-1}$. Отсюда

$$H(\eta)=M\left(-\log p_\xi(\xi)\right)-M\left(\log\left|J_f(x)\right|^{-1}\right)=H(\xi)+M\left(\log\left|J_f(x)\right|\right),$$

что и требовалось доказать.

Следствие 1. При взаимно-однозначном функциональном преобразовании $y=f(x):\mathbb{R}^m\to\mathbb{R}^m$ дифференциальная энтропия может возрастать, убывать или оставаться неизменной, ибо математическое ожидание логарифма якобиана может быть любой величиной.

Следствие 2. Для любой невырожденной матрицы A порядка m и любого вектора $\mathbf{b}\in\mathbb{R}^m$ дифференциальные энтропии случайных величин $\xi\in\mathbb{R}^m$ и $\eta=A\xi+\mathbf{b}$ связаны соотношением $H(\eta)=H(\xi)+\log|A|.$

Пример 1. Если ξ имеет равномерное распределение на $[a,b]$, то $H(\xi)=\log(b-a)$.

Пример 2. Для нормальной случайной величины $\xi=N_1(\mu,\sigma^2)$ имеем

$$H(\xi)=\log\sqrt{2\pi e\sigma^2}\ .$$

Определение. Говорят, что случайная величина $\xi \in \mathbb{R}$ принадлежит классу $P_1(a,b)$, если её функция плотности $p(x)$ удовлетворяет условиям: $p(x) \geq 0$, $\int\limits_{-\infty}^{\infty} p(x)\mathrm{d}x = 1$, $p(x) = 0$ для $x \notin [a,b]$.

Определение. Случайная величина ξ принадлежит классу $P_2(a,\sigma^2)$, если $p_\xi(x) \geq 0$, $\int\limits_{-\infty}^{\infty} p_\xi(x)\mathrm{d}x = 1$, $\int\limits_{-\infty}^{\infty} x p_\xi(x)\mathrm{d}x = a$, $\int\limits_{-\infty}^{\infty} (x-a)^2 p_\xi(x)\mathrm{d}x \leq \sigma^2$.

Справедливы следующие утверждения.

Теорема 1. Для любой случайной величины $\xi \in \mathbb{R}$ с плотностью распределения $p_\xi(x) \in P_1(a,b)$ дифференциальная энтропия удовлетворяет неравенству $H(\xi) \leq \log(b-a)$,

а значит, максимум энтропийной функции для $\xi \in P_1(a,b)$ достигается для ξ равномерно распределённой на $[a,b]$.

Доказательство. Пусть $\eta -$ равномерно распределена на $[a,b]$. Тогда

$$H(\xi) - H(\eta) = -\int\limits_a^b p_\xi(x)\log p_\xi(x)\mathrm{d}x - \log(b-a) =$$

$$= -\int\limits_a^b p_\xi(x)\log p_\xi(x)\mathrm{d}x - \int\limits_a^b p_\xi(x)\log(b-a)\mathrm{d}x = -\int\limits_a^b p_\xi(x)\log\frac{p_\xi(x)}{b-a}\mathrm{d}x \leq 0$$

(по непрерывному аналогу леммы о двух распределениях).

Теорема 2. Пусть $\xi -$ случайная величина и $p_\xi(x) \in P_2(\mu,\sigma^2)$. Тогда $H(\xi) \leq \log\sqrt{2\pi e\sigma^2}$, а значит, максимум энтропийной функции для $\xi \in P_2(\mu,\sigma^2)$ достигается для $\xi = N(\mu,\sigma^2)$.

Доказательство. Рассмотрим интеграл

$$-\int\limits_{-\infty}^{\infty} p_\xi(x)\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)\mathrm{d}x = \log\sqrt{2\pi\sigma^2} + \frac{\log e}{2\sigma^2}\int\limits_{-\infty}^{\infty}(x-\mu)^2 p_\xi(x)\mathrm{d}x =$$

$$= \log\sqrt{2\pi\sigma^2} + \frac{\log e}{2\sigma^2}D_\xi \leq \log\sqrt{2\pi\sigma^2} + \frac{1}{2}\log e = \log\sqrt{2\pi e\sigma^2}.$$

Отсюда

$$H(\xi) - \log\sqrt{2\pi e\sigma^2} \leq -\int\limits_{-\infty}^{\infty} p_\xi(x)\log p_\xi(x)\mathrm{d}x + \int\limits_{-\infty}^{\infty} p_\xi(x)\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)\mathrm{d}x =$$

$$= -\int_{-\infty}^{\infty} p_\xi(x) \log \left(\frac{p_\xi(x)}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}} \right) \mathrm{d}x \le 0$$ (в силу леммы о двух распределениях).

Лекция 7. **Асимптотические свойства стационарного ИДС.**

Рассмотрим стационарный источник ИДС без памяти в алфавите $A = \{0,1\}$. Он порождает $n-$символьную последовательность $\Xi_n = (\xi_1, \ldots, \xi_n)$, где $\xi_i -$ независимые двоичные случайные символы с распределением вероятностей Бернулли

$$P(\xi_i = 1) = p, \quad P(\xi_i = 0) = 1 - p = q.$$

Для удобства будем считать, что $p < \frac{1}{2}$ (случай $p = \frac{1}{2}$ мы сейчас не рассматриваем). Обозначим через $\Xi_n(n_1) -$ сообщение, у которого n_1 единиц. Имеем $P(\Xi_n = \Xi_n(n_1)) = p^{n_1}(1-p)^{n-n_1}$. Тогда $\frac{P(\Xi_n = \Xi_n(0))}{P(\Xi_n = \Xi_n(n))} = \frac{(1-p)^n}{p^n} \to \infty$ при $n \to \infty$, т.е. сообщения $\Xi_n(n)$ и $\Xi_n(0)$ не равномощны.

Пусть $\varepsilon_n -$ случайная величина, значения которой совпадают с числом единиц в сообщении Ξ_n. Очевидно, $\varepsilon_n = \xi_1 + \ldots + \xi_n$. И, значит, $M(\varepsilon_n) = np$, $D(\varepsilon_n) = npq$.

Т е о р е м а. Множество всех реализаций случайной величины ε_n можно разбить на два непересекающихся множества A_n и B_n, так что при $n \to \infty$:

1) $P(\Xi_n \in A_n) \to 0$, а значит, $P(\Xi_n \in B_n) \to 1$;
2) реализации из множества B_n относительно равновероятны в следующем смысле: $\frac{\log P(\Xi_n') - \log P(\Xi_n'')}{\log P(\Xi_n')} \to 0$ для $\Xi_n', \Xi_n'' \in B_n$ ($\Xi_n', \Xi_n'' -$ реализации Ξ_n со значениями в B_n);
3) если $N(B_n) -$ число элементов множества B_n, то $\log N(B_n) \sim nH(\xi)$, где $H(\xi) = -p \log p - q \log q$.

Д о к а з а т е л ь с т в о. К случайной величине ε_n применим неравенство Чебышёва

$$P(|\varepsilon_n - M(\varepsilon_n)| \ge \delta) = P(|\varepsilon_n - np| \ge \delta) \le \frac{D(\varepsilon_n)}{\delta^2} = \frac{npq}{\delta^2}.$$

Возьмём $\delta = n^{3/4}$. Тогда $P\left(\left|\varepsilon_n - np\right| \ge n^{3/4}\right) \le \dfrac{D(\varepsilon_n)}{\delta^2} = \dfrac{pq}{n^{1/2}}$.

$P\left(\left|\varepsilon_n - np\right| \ge n^{3/4}\right)$ – это вероятность суммы несовместных событий:

$\varepsilon_n = n_1, \varepsilon_n = n_2, \ldots, \varepsilon_n = n_m$, где $n_1, \ldots n_m$ – все натуральные решения неравенства $\left|x - np\right| \ge n^{3/4}$, $1 \le x \le n$.

Теперь определим класс A_n: $\Xi_n \in A_n \Leftrightarrow$ соответствующее значение $\varepsilon_n = n'$ удовлетворяет неравенству $\left|n' - np\right| \ge n^{3/4}$.

Тогда $P\left(\Xi_n \in A_n\right) = P\left(\left|\varepsilon_n - np\right| \ge n^{3/4}\right) \le \dfrac{pq}{n^{1/2}} \to 0$ при $n \to \infty$.

В класс B_n отнесём все остальные Ξ_n: $B_n = \left\{E_n \middle| \left|\varepsilon_n - np\right| < n^{3/4}\right\}$.

Пусть теперь Ξ_n' и Ξ_n'' – реализации случайных векторов из множества B_n, и пусть им соответствуют значения ε_n, равные соответственно n' и n''. Тогда имеем неравенства

$$n' = np + \mathrm{O}\left(n^{3/4}\right),\ n'' = np + \mathrm{O}\left(n^{3/4}\right). \tag{5.1}$$

Кроме того,

$$P\left(\varepsilon_n = n'\right) = p^{n'} q^{n-n'},\ P\left(\varepsilon_n = n''\right) = p^{n''} q^{n-n''}. \tag{5.2}$$

Таким образом,

$$\log P\left(\Xi_n'\right) = \log P\left(\varepsilon_n = n'\right) = n' \log p + \left(n - n'\right)\log q = np \log p + nq \log q + O\left(n^{3/4}\right);$$

$$\log P\left(\Xi_n''\right) = \log P\left(\varepsilon_n = n''\right) = n'' \log p + \left(n - n''\right)\log q = np \log p + nq \log q + O\left(n^{3/4}\right) \tag{5.3}$$

$$\left|\frac{\log P\left(\Xi_n'\right) - \log P\left(\Xi_n''\right)}{\log P\left(\Xi_n'\right)}\right| = \left|\frac{\mathrm{O}\left(n^{3/4}\right)}{np \log p + nq \log q + \mathrm{O}\left(n^{3/4}\right)}\right| = \mathrm{O}\left(n^{-1/4}\right) \to 0 \text{ при } n \to \infty.$$

Далее мы видели, что $1 - \dfrac{pq}{n^{1/2}} \le \sum P\left(\Xi_n \middle| \Xi_n \in B_n\right) \le 1$. Поэтому, если $K = N\left(B_n\right)$ – количество элементов множества B_n, то в силу (5.3)

$$1 - \frac{pq}{n^{1/2}} \le \sum_{\Xi_n \in B_n} P\left(\Xi_n'\right) = P\left(\Xi_n \in B_n\right) \le K \max_{\Xi_n \in B_n} P\left(\Xi_n'\right) = K a^{-nH(\xi) + \mathrm{O}\left(n^{3/4}\right)},$$

где a – основание логарифма в определении $H\left(\xi\right)$.

Следовательно, $$K \ge \left(1 - \frac{pq}{n^{1/2}}\right) a^{nH(\xi) + \mathrm{O}\left(n^{3/4}\right)} \tag{5.4}$$

Аналогично, $$K \max_{\Xi_n \in B_n} P\left(\Xi_n'\right) = K a^{-nH(\xi) + \mathrm{O}\left(n^{3/4}\right)} \le 1 \tag{5.5}$$

Из соотношений (5.4) и (5.5) следует третье утверждение теоремы.

Замечание. Доказанная теорема легко обобщается и на случай произвольного алфавита A, содержащего m элементов. В этом случае, в силу теоремы, из всех m^n реализаций заслуживают внимание только $K \approx a^{nH(\xi)}$ реализаций, которые можно считать равновероятными. Таким образом, доля реализаций, заслуживающих внимание, равна

$$\frac{a^{nH(\xi)}}{m^n} = a^{n\left(H(\xi)-\log m\right)}.$$

А поскольку $H(\xi) < \log m$, если ξ не есть равномерно распределённой случайной величиной, то эта доля неограниченно убывает при $n \to \infty$. Заметим также, что в полученных оценках постоянные в символах «О» зависят от p и q.

Например, если $p = 0.2$, $n = 100$, $a = 2$, $m = 2$, то доля «заслуживающих внимание» реализаций равна $2^{100(-0.2\log 0.2 - 0.8\log 0.8)} \approx 2^{-27.8} \approx 10^{-8}$.

Установленный в теореме результат для стационарных ИДС без памяти может быть обобщён на

Теорема (Стратонович). Пусть $X-$ $n-$ мерная случайная величина со значениями в A^n, где $A-$ алфавит длины m, причём $H(X)$ не убывает с ростом n и $\frac{-\log P(\Xi_n)}{H(X)} \overset{\text{по вероятности}}{\to} 1$,

т.е. $$P\left\{\left|\frac{-\log P(\Xi_n)}{H(X)} - 1\right| \geq \varepsilon\right\} < \varepsilon.$$

Тогда множество реализаций X можно разбить на два непересекающихся множества A_n и B_n, так что при $n \to \infty$:

1) $\sum P(\Xi_n \in A_n) \to 0$, $\quad \sum P(\Xi_n \in B_n) \to 1$;

2) $\frac{\log P(\Xi_n') - \log P(\Xi_n'')}{\log P(\Xi_n')} \to 0$ для $\Xi_n', \Xi_n'' \in B_n$;

3) если $N(B_n)-$ число элементов множества B_n, то $\log N(B_n) \sim nH(X)$.

Эту теорему мы доказывать не будем, но рассмотрим случай марковских последовательностей.

Лекция 8. **Асимптотические свойства стационарной ОЦМ.**

Рассмотрим стационарную ОЦМ и пусть $H = H(X|X) = -\sum_{i,j=1}^{m} \pi_i p_{ij} \log p_{ij}$ – её средняя условная энтропия.

Выясним свойства цепи Маркова, определяемые её средней энтропией H. Пусть проведено n опытов, т.е. получена последовательность длины n. Пусть $M_i -$ число появлений значения $a_i \in A$ в этих опытах, а $M_{ij} -$ число переходов от a_i к a_j.Тогда по закону больших чисел для $\forall \varepsilon, \delta > 0$ и достаточно большом n: $P\left\{\left|\frac{M_i}{n} - \pi_i\right| < \varepsilon\right\} > 1 - \delta$, $P\left\{\left|\frac{M_{ij}}{M_i} - \pi_{ij}\right| < \varepsilon\right\} > 1 - \delta$.

Заметим, что из неравенств

$$\left|\frac{M_i}{n} - \pi_i\right| < \varepsilon, \qquad \left|\frac{M_{ij}}{M_i} - p_{ij}\right| < \varepsilon \tag{6.1}$$

и $p_{ij} \le 1$, $M_i \le n$ следуют неравенства

$$\left|M_i p_{ij} - n\pi_i p_{ij}\right| < np_{ij}\varepsilon \le n\varepsilon, \qquad \left|M_{ij} - M_i p_{ij}\right| < M_i\varepsilon \le n\varepsilon. \tag{6.2}$$

Но тогда

$$\left|M_{ij} - n\pi_i p_{ij}\right| < 2n\varepsilon, \tag{6.3}$$

откуда $M_{ij} = n\left(\pi_i p_{ij} + s_{ij}\right)$, $\left|s_{ij}\right| < 2\varepsilon$.

Нарушение этого неравенства возможно только в том случае, когда нарушается хотя бы одно из неравенств (6.2), а значит , хотя бы одно из неравенств (6.1). Но вероятность нарушения каждого из неравенств в (6.1) меньше δ (при больших n). Значит, вероятность нарушения (6.3) не превосходит 2δ: $P\left(\left|M_{ij} - n\pi_i p_{ij}\right| \ge 2n\varepsilon\right) < 2\delta$.

Отсюда заключаем, что вероятность того, что нарушится неравенство (6.3) хотя бы для одной пары (i, j) не больше суммы вероятностей нарушения хотя бы одного из неравенств (6.3) т.е. не превосходит $2m^2\delta$. Но при большом n можно взять δ как угодно малым, а значит, неравенства (6.3) выполняются для любой пары (i, j) с вероятностью как угодно близкой к 1, таким образом, $\lim\limits_{n\to\infty} P\left(\left|M_{ij} - n\pi_i p_{ij}\right| < 2n\varepsilon\right) = 1$.

Зафиксируем теперь $n-$последовательность X_n значений $a_i \in A$ в n опытах. Тогда вероятность появления этого события (т.е. появления заданного X_n) равна $Q = \pi_k p_{ki_1} p_{i_1 i_2} \dots p_{i_{m-2} i_{m-1}}$. С учётом принятых обозначений

$$Q = \pi_k \prod_{i,j=1}^{m} p_{ij}^{M_{ij}}. \tag{6.4}$$

Так как в n опытах могут получаться различные $n-$последовательности X_n, то вероятность Q можно рассматривать как случайную величину, т.е. в равенстве (6.4) номер k и числа M_{ij} являются случайными величинами. Положим $M_{ij} = n\left(\pi_i p_{ij} + s_{ij}\right)$. Тогда из (6.4) имеем

$$\log\frac{1}{Q}=\log\frac{1}{\pi_k}-\sum_{i,j=1}^{m}\log p_{ij}^{M_{ij}}=$$

$$=\log\frac{1}{\pi_k}-n\sum_{i,j=1}^{m}\left(\pi_i p_{ij}+s_{ij}\right)\log p_{ij}=\log\frac{1}{\pi_k}+nH-n\sum_{i,j=1}^{m}s_{ij}\log p_{ij} \qquad (6.5)$$

А из (6.3) имеем $\left|s_{ij}\right|<2\varepsilon$, откуда $\log\frac{1}{Q}=\log\frac{1}{\pi_k}+nH+\mathrm{O}\left(n\varepsilon\sum_{i,j=1}^{m}\log\frac{1}{p_{ij}}\right)$, или

$$\left|\frac{1}{n}\log\frac{1}{Q}-H\right|<\frac{1}{n}\log\frac{1}{\pi_k}+\mathrm{O}\left(\varepsilon\sum_{i,j=1}^{m}\log\frac{1}{p_{ij}}\right). \qquad (6.6)$$

При достаточно большом n и достаточно малом ε правая часть в (6.6) будет меньше произвольно малого $\eta>0$. (Мы считаем, что все $p_{ij}\neq 0$, так как иначе предполагаемая последовательность X_n имеет вероятность 0, а такие последовательности нас не интересуют). Но так как вероятность выполнения неравенства (6.3) как угодно близка к 1 при больших n, то, значит, вероятность того, что правая часть (6.6) меньше произвольно малого $\eta>0$ также близка к 1, а потому $\lim_{n\to\infty}P\left(\left|\frac{1}{n}\log\frac{1}{Q}-H\right|<\eta\right)=1$.

Таким образом, с вероятностью, близкой к 1, появляется одна из таких последовательностей X_n, для которой вероятность Q удовлетворяет неравенству

$$a^{-(H+\eta)n}<Q<a^{-(H-\eta)n}, \qquad (6.7)$$

где $a-$ основание логарифма в определении H. Последнее неравенство выделяет «наиболее вероятное» множество B_n последовательностей X_n при неограниченном возрастании n. Действительно, расположим последовательности в порядке возрастания их вероятностей. Найдём наименьшее число наиболее вероятных последовательностей, сумма вероятностей которых $\geq q$, где $q\notin\{0,1\}$. Обозначим это число через $\nu(q)$. Так как для последовательностей, которые не удовлетворяют неравенству (6.7), сумма вероятностей которых как угодно мала, то в число $\nu(q)$ наиболее вероятных последовательностей они не войдут. Поэтому

$$\nu(q)=\nu_1(q)+\nu_2(q),$$

где $\nu_2(q)-$ число последовательностей, для которых вероятности $Q\geq a^{-(H-\eta)n}$, а $\nu_1(q)-$ число последовательностей, которые удовлетворяют (6.7). (Мы учитываем, что $q<1$). Таким образом, вероятность каждой из $\nu(q)$ наиболее вероятных последовательностей превышает $a^{-(H+\eta)n}$, и $\nu(q)a^{-(H+\eta)n}<q<1$, а потому $\log\nu(q)-n(H+\eta)<0$, откуда $\frac{\log\nu(q)}{n}-H\leq\eta$.

Так как сумма вероятностей последовательностей, которые не удовлетворяют (6.7), меньше δ, то сумма вероятностей $\nu_1(q)$ последовательностей больше $q-\delta$. Сумма вероятностей каждой из этих последовательностей меньше $a^{-(H-\eta)n}$, а потому

$$q-\delta<\nu_1(q)a^{-(H-\eta)n}\le\nu(q)a^{-(H-\eta)n}.$$

Отсюда $\log(q-\delta)<\log\nu(q)-n(H-\eta)$, так что $\dfrac{\log\nu(q)}{n}>\dfrac{1}{n}\log(q-\delta)+(H-\eta)$.

Значит, $\lim\limits_{n\to\infty}\dfrac{\log\nu(q)}{n}=H$.

Таким образом, при больших n наименьшее число наиболее вероятных последовательностей, сумма вероятностей которых $\ge q$, удовлетворяет неравенству

$$a^{n(H-\eta)}<\nu(q)\le a^{n(H+\eta)}. \tag{6.8}$$

Заметим, что оценка (6.8) величины $\nu(q)$ не зависит от q. Это есть следствие того, что при $q_1<q_2$ величины $\nu(q_1)$, $\nu(q_2)$, $\nu(q_2)-\nu(q_1)$ растут с ростом n, но $\nu(q_1)$, $\nu(q_2)$ растут быстрее, чем $\nu(q_2)-\nu(q_1)$, т.е. $\nu(q_1)$, $\nu(q_2)$ – величины одного и того же порядка роста.

Полученная оценка (6.8) показывает, какое число возможных $n-$последовательностей X_n нужно учитывать, т.е. практически рассматривать при исследованиях.

Например, если $a=2$, $m=8$, $H=2.75$ бит, $n=100$, то

$m^n=8^{100}=2^{300}$, $a^{nH}=m^{\frac{nH}{\log m}}=2^{3\cdot\frac{275}{3}}=2^{275}$, откуда $\dfrac{m^n}{\nu(q)}\approx\dfrac{m^n}{a^{nH}}=2^{25}>30\cdot10^6$,

т.е. число наиболее вероятных последовательностей составляет одну тридцатимиллионную долю числа всех последовательностей.

Пример. Определить коэффициент сжатия текста передаваемых по каналу связи сообщений, рассматривая последовательность знаков текста как стационарную ОЦМ с энтропией H.

Мы видим, что для текстов длины n имеется $\approx a^{nH}$ практически возможных текстов, но $a^{nH}=m^{\frac{nH}{\log m}}$, где $m-$ число символов алфавита, а потому число знаков для передачи a^{nH} сообщений оценивается как $\approx\dfrac{nH}{\log m}$, т.е. сообщение из n знаков текста можно передавать в среднем при помощи $\dfrac{nH}{\log m}$ знаков того же текста, т.е. коэффициент сжатия равен $\dfrac{H}{\log m}$.

Лекция 9. **Каналы связи с помехами.**

Мы видели, что целью кодирования в каналах связи без помех является уменьшение избыточности, т.е. уменьшение средней длины кода. Но в каналах связи с помехами целью кодирования является минимизация вероятности ошибки. И она может быть достигнута за счёт избыточности. Так возникает задача: как, не очень увеличивая избыточность минимизировать вероятность ошибки декодирования. Мы будем рассматривать информационные слова одной и той же длины k и будем им сопоставлять кодовые слова длины n, $n > k$. Основное внимание мы уделяем линейным кодам.

Определение. Код C называется линейным, если его кодовые слова образуют подпространство пространства $\mathbb{F}_q^n$, где $q-$ число элементов конечного поля, элементы которого образуют алфавит кода. (Ясно, что q есть степень простого числа).

Линейный код длины n и числом символов k будем называть $(n,k)-$ кодом. Кодовые слова будем обозначать строками длины n. Пусть $\overline{c}_1,\ldots,\overline{c}_k-$ базис подпространства кодовых слов и пусть $G-$ матрица, составленная из этих строк. Ясно, что размер G равен $k\times n$. Так как ранг G равен k, то система линейных однородных уравнений (СЛОУ) с матрицей G имеет подпространство решений размерности $m = n-k$. Из базиса пространства решений образуем матрицу H, так что $GH^T = 0$, где $0-$ нулевая матрица размерности $k\times m$. Матрицу H называют проверочной матрицей кода. При построении матрицы H можно выбрать фундаментальное решение так, что последние m столбцов матрицы H образуют единичную матрицу $I_m = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$, так что $H = (AI_m)$, где $A-$ матрица размера $m\times k$.

Учитывая, что $GH^T = 0 \Rightarrow HG^T = 0$, мы получаем, что тот же код C можно получить как пространство решений СЛОУ с матрицей H. А потому в качестве порождающей матрицы кода можно взять матрицу $\left(I_k\left(-A^T\right)\right)$. Поэтому проверочная и порождающая матрицы кода имеют вид: $H = (AI_m)$, $G = \left(I_k\left(-A^T\right)\right)$.

Такой вид матриц H и G называется стандартным. В этом случае каждое кодовое слово $\overline{c} = (c_1,\ldots,c_k,\ldots c_n)$ обладает тем свойством, что первые k элементов строки $\overline{c}$ совпадают с информационными, т.е. $c_1 = a_1,\ldots c_k = a_k$, остальные называются проверочными. А сам код называется систематическим.

Пусть $\overline{x},\overline{y}\in\mathbb{F}_q^n$. Обозначим через $w(\overline{x})-$ число ненулевых компонентов вектора $\overline{x}$ (вес $\overline{x}$), а через $d(\overline{x},\overline{y})=w(\overline{x}-\overline{y})-$ расстояние Хэмминга между векторами $\overline{x}$ и $\overline{y}$.

Лемма. Пространство $\mathbb{F}_q^n$ с расстоянием Хэмминга $d(\overline{x},\overline{y})$ является метрическим пространством, т.е.

1. $d(\overline{x},\overline{y})=d(\overline{y},\overline{x})$;
2. $d(\overline{x},\overline{x})\geq 0$, причём $d(\overline{x},\overline{y})=0\Leftrightarrow\overline{x}=\overline{y}$;
3. (неравенство треугольника): $\forall\overline{x},\overline{y},\overline{z}\in\mathbb{F}_q^n$

$$d(\overline{x},\overline{y})=d(\overline{x},\overline{z})+d(\overline{y},\overline{z}).$$

Определение. Пусть $C-$ линейный код. Минимальным кодовым расстоянием кода C называется $d_C=\min\limits_{\substack{\overline{c}_1,\overline{c}_2\in C\\ \overline{c}_1\neq\overline{c}_2}} d(\overline{c}_1,\overline{c}_2)$.

Определение. Сферой радиуса t с центром в точке $\overline{x}_0\in\mathbb{F}_q^n$ называется множество точек $\overline{x}\in\mathbb{F}_q^n$, для которых $d(\overline{x},\overline{x}_0)\leq t$.

Определение. Говорят, что код C исправляет t ошибок, если $\forall\overline{y}\in\mathbb{F}_q^n$ имеется не более одного кодового слова $\overline{c}$, такого, что $d(\overline{y},\overline{c})\leq t$.

Принципом декодирования является правило, по которому полученное сообщение (т.е. вектор $\overline{y}\in\mathbb{F}_q^n$) декодируется в ближайшее (в смысле расстояния Хэмминга) кодовое слово.

Теорема. Код C с минимальным расстоянием d_C исправляет t ошибок $\Leftrightarrow$ $d_C\geq 2t+1$.

Доказательство. Для кодового слова $\overline{c}\in C$ обозначим через $B_t(\overline{c})$ сферу радиуса t с центром в т. $\overline{c}$. Из условия $d_C\geq 2t+1$ следует, что при $\overline{c}_1\neq\overline{c}_2$ $B_t(\overline{c}_1)\cap B_t(\overline{c}_2)=\varnothing$, и значит, каждое сообщение $\overline{y}$, содержащее не более t ошибок, а потому попадает не более, чем в одну сферу $B_t(\overline{c})$. И наоборот, если код исправляет t ошибок и менее, то при $\overline{c}_1\neq\overline{c}_2$ $B_t(\overline{c}_1)\cap B_t(\overline{c}_2)=\varnothing$, а значит, $d(\overline{c}_1,\overline{c}_2)>2t$, откуда $d_C\geq 2t+1$.

Теорема. Линейный код C с проверочной матрицей H будет иметь минимальное расстояние $d_C\geq s+1$, где $s\in\mathbb{N}$ $\Leftrightarrow$ любые s столбцов матрицы H линейно независимы.

Доказательство.

Необходимость. Пусть $d_C\geq s+1$. Предположим, что в матрице H имеется s столбцов с номерами $i_1,\ldots i_s$, которые линейно зависимы: $\alpha_1\overline{h}_{i_1}+\ldots+\alpha_s\overline{h}_{i_s}=0$, причём не все $\alpha_j\in\mathbb{F}_q$ равны нулю. Рассмотрим вектор $\overline{c}=(0,\ldots,\alpha_1,\ldots,\alpha_s,\ldots 0)$, где α_j стоит на i_j-ом месте.

Тогда $H\overline{c}^T = 0 \Rightarrow \overline{c} \in C,\ \overline{c} \neq \overline{0}$. Но $d\left(\overline{c}, \overline{0}\right) \leq s$, что невозможно, ибо $d_C \geq s+1$. Значит, в H любые s столбцов линейно независимы.

Достаточность. Пусть в H любые s столбцов линейно независимы, но $d_C \leq s$. В этом случае найдутся $\overline{c}_1, \overline{c}_2 \in C$ такие, что $\overline{c}_1 \neq \overline{c}_2$ и $d\left(\overline{c}_1, \overline{c}_2\right) \leq s$. Мы имеем $\overline{c} = \overline{c}_1 - \overline{c}_2 = \left(0,\ldots,\beta_1,0,\ldots,\beta_k,0,\ldots 0\right)$, $\beta_i \neq 0$, $i = 1,\ldots,k$, причём $k \leq s$. Так как $0 \neq \overline{c} \in C$, то $H\overline{c}^T = \overline{0}$. Отсюда следует, что в матрице H имеется k линейно зависимых столбцов, причём $k \leq s$, что противоречит предположению.

Пример. Зафиксируем натуральное m и положим $n = 2^m - 1$. Обозначим через H матрицу, столбцами которой являются все $m-$значные числа от 1 до 2^{m-1} в системе счисления с основанием 2: $H = \begin{pmatrix} 0 & 0 & 0 & \ldots & 1 & 1 \\ 0 & 0 & 0 & \ldots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 0 & 1 & \ldots & 0 & 1 \end{pmatrix}$.

Линейный код с проверочной матрицей H называется бинарным кодом Хэмминга. В матрице H любые два столбца линейно независимы, но имеются три линейно зависимых столбца (например, $\overline{h}_3 = \overline{h}_1 + \overline{h}_2$). Поэтому $s = 2$, и значит, $d_C \geq 3$. Этот код исправляет одиночные ошибки.

Лекция 10. **Синдром вектора. Декодирование по методу смежного класса.**

Рассмотрим $(n,k)-$код C над полем $\mathbb{F}_q$. Ясно, что $|C| = q^k$. Рассмотрим фактор-группу $\mathbb{F}_q^n / C$, порядок которой равен $\left|\mathbb{F}_q^n\right| / |C| = q^{n-k}$.

Определение. Лидером смежного класса $A \in \mathbb{F}_q^n / C$ называется вектор $\overline{y}^0 \in A$ с наименьшим весом. Если в A имеется несколько векторов с наименьшим весом, то фиксируем один из них и называем его лидером.

Определение. Синдромом вектора $\overline{y} \in \mathbb{F}_q^n$ называется вектор $s\left(\overline{y}\right) = \overline{s} = H\overline{y}^T$.

Очевидно, что $\overline{s} \in \mathbb{F}_q^m$. Ясно, что вектора одного и того же смежного класса имеют равные синдромы, а разных – разные.

Предполагая, что канал связи имеет малую вероятность искажения символов, положим в основу декодирования принцип: при прохождении сообщения через канал связи возможными векторами ошибок могут быть только лидеры смежных классов. Отсюда следует следующий

метод декодирования по методу смежного класса: получаемое сообщение $\overline{y}$ равно сумме $\overline{c}+\overline{e}$, где $\overline{c}-$ посланное кодовое слово, $\overline{e}-$ вектор ошибки. Мы имеем

$$s\left(\overline{y}\right)=\overline{s}=H\overline{y}^T=H\left(\overline{c}^T+\overline{e}^T\right)=H\overline{c}^T+H\overline{e}^T=H\overline{e}^T.$$

Вычислив $\overline{s}$ определяем смежный класс, в котором лежит вектор $\overline{e}$ (он лидер этого смежного класса), а значит, и вектор $\overline{y}$. И тогда $\overline{c}=\overline{y}-\overline{e}$.

Для практического использования этого метода необходимо иметь таблицу лидеров и их синдромов. Однако при большом числе смежных классов просмотр таких таблиц затруднителен.

Теорема. Пусть $C-$ линейный бинарный код с проверочной матрицей H. Тогда синдром всякого вектора $\overline{y}\in\mathbb{F}_2^n$ равен сумме столбцов матрицы, номера которых совпадают с позициями ненулевых компонентов вектора ошибок.

Доказательство. Следует из определения синдрома.

Замечание. Из теоремы следует, что в случае только одной ошибки, синдром совпадает с тем столбцом, на позиции которого находится эта ошибка. В частности, для кода Хэмминга, в случае одиночной ошибки синдромом есть двоичная запись номера позиции, где произошла ошибка.

Верхние оценки минимального кодового расстояния.

Мы выше уже видели как минимальное кодовое расстояние характеризует возможности кода исправлять ошибки. Поскольку ранг матрицы H не превосходит m, то $d_C\le m+1$. В случае $d_C=m+1=n-k+1$ код C называется MDS-кодом. Рассмотрим $(n,k)-$код C с минимальным расстоянием d. Через $M=|C|$ будем обозначать мощность кода. Нас интересует связь между n, d и M. Обозначим

$$A(n,d)=\max_C\left\{M\middle|\text{длина } C=n,\ d_C=d\right\}.$$

Определение. Код (n,d,M) называется максимальным, если он не содержится ни в каком коде $(n,d,M+1)$. (Для такого кода $|C|=A(n,d)$).

Теорема (неравенство Чернова). Пусть $q>1-$ натуральное, $n\in\mathbb{N}$. Тогда для $\forall t<n\cdot\dfrac{q-1}{q}$ справедливо неравенство

$$\frac{1}{n+1}2^{n\varphi\left(\frac{t}{n}\right)}\le C_n^t(q-1)^t\le\sum_{i=1}^{t}C_n^i(q-1)^i\le 2^{n\varphi\left(\frac{t}{n}\right)},$$

где $\varphi(x)=x\log_q(q-1)-x\log_q x-(1-x)\log_q(1-x),\ 0\le x\le 1$.

Доказательство. Мы рассмотрим только случай $q=2$. Тогда $\varphi(x)=-x\log x-(1-x)\log(1-x)=H(x)$, $\lambda=\frac{t}{n}<\frac{1}{2}$. Доказываемое неравенство приобретает вид: $$\frac{1}{n+1}2^{nH(\lambda)}\le C_n^t\le\sum_{i=1}^{t}C_n^i\le 2^{nH(\lambda)}.$$

Далее, $1=\left(\lambda+(1-\lambda)\right)^n=\sum_{i=0}^{n}C_n^i\lambda^i(1-\lambda)^{n-i}\ge\sum_{i=1}^{t}C_n^i(1-\lambda)^n\left(\frac{\lambda}{1-\lambda}\right)^{n-i}\ge(1-\lambda)^n\left(\frac{\lambda}{1-\lambda}\right)^t\sum_{i=1}^{t}C_n^i$,

откуда $$\sum_{i=1}^{t}C_n^i\le(1-\lambda)^{-(n-t)}\lambda^{-t}=2^{-n(1-\lambda)\log(1-\lambda)-n\frac{t}{n}\log\lambda}=2^{n\varphi\left(\frac{t}{n}\right)}.$$

Также из формулы Стирлинга $n^{-1}\log\left(\sum_{i=0}^{t}C_n^i\right)\ge n^{-1}\log C_n^t\ge H(\lambda)+\mathrm{o}(1)$, откуда $\frac{1}{n+1}2^{nH(\lambda)}\le C_n^t$.

Теорема (о верхней границе Хэмминга). Пусть $C-$ линейный $(n,k)-$код, $|C|=M$. Если этот код исправляет t ошибок, то $M\sum_{i=0}^{t}C_n^i(q-1)^i\le q^n$, где $q-$ объём алфавита кода.

Доказательство. Подсчитаем число элементов сферы $B_t(\overline{c})$. Имеем $\left|B_t(\overline{c})\right|=\sum_{i=0}^{t}C_n^i(q-1)^i$, где мы учли, что в сфере содержится точка $\overline{c}$, а число точек $\overline{x}$ сферы, для которых $w(\overline{c}-\overline{x})=i$, равно $C_n^i(q-1)^i$. Во всех сферах с центрами в кодовых словах содержится $M\sum_{i=0}^{t}C_n^i(q-1)^i$ точек, причём эти сферы не пересекаются. Всего точек в $\mathbb{F}_q^n$ в точности q^n, откуда и следует неравенство.

Определение. Если в неравенстве Хэмминга достигается равенство, то линейный $(n,k)-$код называется совершенным.

Совершенными кодами являются:

1) код с повторениями;
2) обобщённый код Хэмминга;
3) двоичный код Галлея (23,12);
4) троичный код Галлея (11,6).

Имеются ещё нелинейные коды Васильева и Шонгейна, которые также удовлетворяют равенству Хэмминга. Для кода с повторениями это утверждение есть следствием свойств биноминальных коэффициентов (с учётом того, что $M=|C|=q$, $t=\frac{n-1}{2}$).

Рассмотрим обобщённый код Хэммнга. Пусть имеется поле $\mathbb{F}_q$ и пусть $n=\frac{q^m-1}{q-1}$, $m\in\mathbb{N}$. Обозначим через $\alpha-$ примитивный элемент поля $\mathbb{F}_{q^m}$ (т.е. $\alpha-$ порождающий элемент мультипликативной группы поля $\mathbb{F}_{q^m}$). Базисом поля $\mathbb{F}_{q^m}$ являются $1,\alpha,\ldots,\alpha^{m-1}$. Тогда при некоторых $a_{ij}\in\mathbb{F}_q$ для $i=1,2,\ldots,n$ $\alpha^i=\sum_{j=0}^{m-1}a_{ij}\alpha^j$. Образуем матрицу H размера $m\times n$, из столбцов $A_i=\begin{pmatrix}a_{i0} & a_{i1} & \ldots & a_{im-1}\end{pmatrix}^T$.

Обобщённым кодом Хэмминга называется линейный код с проверочной матрицей H.

Лемма. Обобщённый код Хэмминга является совершенным кодом, исправляющим одиночные ошибки.

Доказательство. Заметим, что любые два столбца матрицы H не пропорциональны. Иначе мы имели бы $\alpha^{i_1}=a\alpha^{i_2}$, где $i_1\neq i_2$ и $a\in\mathbb{F}_q^*$, откуда $\alpha^{(i_1-i_2)(q-1)}=1$ и, значит, $\left(q^m-1\right)\left|(i_1-i_2)(q-1)\right.$, что невозможно, так как $|i_1-i_2|(q-1)\le(n-1)(q-1)<n(q-1)=q^m-1$. Значит любые два столбца матрицы H линейно независимы, а потому для этого кода $d_C\ge 3$, т.е. код исправляет одиночные ошибки. Построим теперь сферы радиуса 1 с центрами в кодовых словах. Имеем $B_1(\bar{c})=1+n(q-1)=q^m$. Но $k=n-m$, а потому имеется q^k кодовых слов, т.е. $M=|C|=q^k$. А значит, все сферы содержат $q^k\cdot q^m$ векторов из $\mathbb{F}_q^n$, т.е. эти сферы не пересекаясь покроют всё пространство $\mathbb{F}_q^n$. Что и требовалось показать.

Совершенность указанных выше кодов Галлея мы покажем позже.

Теорема (о границе Плоткина). Для минимального расстояния линейного $(n,k)-$кода справедливо неравенство $d_C\le\frac{nq^{k-1}(q-1)}{q^k-1}$.

Доказательство. Пусть $i-$номер столбца порождающей матрицы кода, отличный от нулевого. Обозначим через D_i подмножество кодовых слов, у которых $i-$я компонента нулевая. Ясно, что D_i- подпространство в C, причём порядок фактор-группы $|C/D_i|=\left|\frac{C}{D_i}\right|=q$, а значит, $|D_i|=q^{k-1}$. Поскольку в порождающей матрице нет нулевых столбцов, то имеется n подпространств $D_1,\ldots,D_n$, у каждого из которых векторы содержат хотя бы по одной нулевой

компоненте. Значит, сумма весов всех кодовых слов $\le nq^k - nq^{k-1} = nq^{k-1}(q-1)$. В то же время, в коде имеется $q^k - 1$ ненулевых слов и вес каждого не меньше d_C. А потому

$d_C(q^k-1) \le$ сумма весов кодовых слов $= \sum_{\bar{c}\in C} w(\bar{c}) \le nq^{k-1}(q-1)$, откуда $\qquad d_C = \dfrac{nq^{k-1}(q-1)}{q^k-1}$.

Теорема (Варшамова-Гильберта). Пусть для $n,k\in\mathbb{N}$, таких, что $1\le k<n$, существует $d\in\mathbb{N}$, для которого выполняется неравенство $q^{n-k} \ge \sum_{i=0}^{d-2} C_n^i (q-1)^i$. Тогда существует, по крайней мере, один линейный $(n,k)-$код, минимальное расстояние которого $d_C \ge d$.

Доказательство. Мы покажем, что существует матрица размера $(n-k)\times n$, у которой любые $d-1$ столбцов линейно независимы. В качестве 1-го столбца возьмём любой ненулевой столбец из $\mathbb{F}_q^{n-k}$, в качестве 2-го столбца – любой вектор из $\mathbb{F}_q^{n-k}$, ему не пропорциональный. Пусть уже построены $j-1$ столбцов, $d\le j<n$, причём любые $d-1$ из них линейно независимы. Рассмотрим всевозможные линейные комбинации этих столбцов в количестве не более $d-2$ векторов. Всех таких векторов (т.е. линейных комбинаций) $\sum_{i=0}^{d-2} C_{j-1}^i (q-1)^i$. Но тогда, в силу условий теоремы, найдётся вектор (а значит, и $j-$й столбец), который будет отличен от этих линейных комбинаций, а потому и линейно независим от любых $d-2$ столбцов из числа уже построенных $j-1$ столбцов. Так последовательно мы построим n столбцов матрицы H, у которой любые $d-1$ столбцов линейно независимы, а потому соответствующий $(n,k)-$код имеет минимальное расстояние $d_C \ge d$.

Лекция 11. **Порождение новых кодов из заданных кодов.**

1. Дуальные коды. Рассмотрим линейный $(n,k)-$код в алфавите $\mathbb{F}_q$. Тогда $C\in\mathbb{F}_q^n$. Пусть H и $G-$ его проверочная и порождающая матрицы, тогда $HG^T=0$. Определим на $\mathbb{F}_q^n$ скалярное произведение как сумму попарных произведений одноимённых координат. Обозначим через $C^\perp-$ ортогональное дополнение к C в $\mathbb{F}_q^n$. Поскольку $GH^T=0$, то учитывая, что $\operatorname{rank} H = m$, сразу получаем, что $C^\perp$ порождено строками матрицы H. Таким образом, мы получаем линейный $(n,m)-$код $C^\perp$ с проверочной матрицей G и порождающей матрицей H. Этот код называется дуальным к коду C.

Определение. Линейный код называется самодуальным, если $C = C^{\perp}$, а значит, $k = m$, $n = 2k$.

Определение. Пунктированием бинарного кода C называется инвертирование всех символов кодовых слов. Получаемое множество C^* является подпространством в $\mathbb{F}_q^n$, причём коды C и C^* ортогональны.

2. Простые преобразования линейного кода. Пусть имеется линейный $(n,k)-$код. Нас интересуют незначительные изменения данного кода, которые приводят к новым кодам (вообще говоря не линейным):

Расширение кода – увеличение длины кодовых слов путём добавлением новых проверочных символов, что приводит к возрастанию большего размера порождающей матрицы.

Удлинение кода – увеличение длины кода путём добавления новых информационных символов, что приводит к увеличению обоих размеров порождающей матрицы.

Выкалывание кодовых координат – уменьшение длины кода удалением проверочных символов, что приводит к уменьшению большего размера порождающей матрицы.

Укорочение кода - уменьшение длины кода удалением информационных символов, что приводит к уменьшению обоих размеров порождающей матрицы.

Пополнение кода – увеличение числа информационных символов, что приводит к увеличению меньшего размера порождающей матрицы.

Код с выбрасыванием – уменьшение числа информационных символов без изменения длины кода, что приводит к уменьшению меньшего размера порождающей матрицы.

Эти преобразования используются для модификации известных кодов с целью получения новых хороших кодов.

Любой двоичный $(n,k,d^*)-$ код с нечётным минимальным расстоянием можно расширить до $(n+1,k,d^*+1)-$ кода добавлением к каждому кодовому слову суммы всех его компонент в качестве проверки на чётность. (Ибо слова с нечётным весом получат в качестве дополнительного символа единицу). Значит, все кодовые слова веса d^* становятся кодовыми словами веса d^*+1. В этом случае вместо матрицы H получаем матрицу

$$H' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & & & \\ \vdots & & H & \\ 0 & & & \end{pmatrix}.$$

Так $\left(2^m-1, 2^m-1-m\right)=(n,k)-$ код Хэмминга может быть расширен до $\left(2^m, 2^m-m\right)-$кода, исправляющего одиночные ошибки и обнаруживающего двойные ошибки. Этот код также называется кодом Хэмминга.

Из кода в алфавите $q=p^r$, $p-$простое, можно получить новый код выбирая в исходном коде только слова в алфавите $q_1=p^l$, где $l \mid r$. Полученный таким образом код называется подкодом кода C.

3. Коды Рида-Маллера.

На векторах пространства $\mathbb{F}_2^n$ введём покомпонентное умножение: если $\overline{a}=\left(a_0, a_1, \ldots a_{n-1}\right)$, $\overline{b}=\left(b_0, b_1, \ldots b_{n-1}\right)$, то $\overline{a}\overline{b}=\left(a_0b_0, a_1b_1, \ldots a_{n-1}b_{n-1}\right)$. Зафиксируем целые $m>r\geq 0$. Положим $n=2^m$. Порождающую матрицу линейного кода будем строить из блоков $G_0, G_1, \ldots, G_r$, причём $G_0=(1,1,\ldots,1)\in \mathbb{F}_2^n$. Далее, G_1 есть матрица размера $m\times n$, столбцы которой представляют двоичную запись $m-$значных чисел от 0 до $n-1$. Для $l>1$ строки матрицы G_l являются всевозможными произведениями различных l строк матрицы G_1. Отсюда следует, что матрица G_l состоит из C_m^l строк. Таким образом, $G=\begin{pmatrix} G_0 \\ G_1 \\ \vdots \\ G_r \end{pmatrix}$, причём размер матрицы G равен $k\times n$, где $k=\sum\limits_{j=0}^{r} C_m^j$. Очевидно, что $n-k=\sum\limits_{j=0}^{m-r-1} C_m^j$.

Построенный $(n,k)-$код называется кодом Рида-Маллера порядка r.

Например, при $m=4$, откуда $n=16$, возьмём $r=3$. Тогда

$G_0=(1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 11\ \ 11\ \ 11\ \ 11\ \ 11)$,

$$G_1=\begin{pmatrix} 0 & 0 & 0 & 0 & \ldots & 1 & 1 \\ 0 & 0 & 0 & 0 & \ldots & 1 & 1 \\ 0 & 0 & 1 & 1 & \ldots & 1 & 1 \\ 0 & 1 & 0 & 1 & \ldots & 0 & 1 \end{pmatrix}=\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix},$$

$$G_2=\begin{pmatrix} a_1a_2 \\ a_1a_3 \\ a_1a_4 \\ a_2a_3 \\ a_2a_4 \\ a_3a_4 \end{pmatrix}=\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$G_3 = \begin{pmatrix} a_1a_2a_3 \\ a_1a_2a_4 \\ a_1a_3a_4 \\ a_2a_3a_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & 1 & 0 & 1 \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Поэтому G есть матрица размера $(1+4+6+4\times16)=(15\times16)$. Легко видеть, что этот код есть код с проверкой на чётность. Это линейный $(16,15)-$код.

Если взять $r=2$, то будем иметь $G=\begin{pmatrix} G_0 \\ G_1 \\ G_2 \end{pmatrix}$, что приводит к $(16,11)-$коду, который является расширением кода Хэмминга, получаемым добавлением проверки на чётность.

Заметим, что код Рида-Маллера $r-$порядка может быть получен из кода $(r-1)-$порядка, а код $(r-1)-$порядка получается из кода $r-$порядка с помощью выбрасывания.

Мы сейчас построим алгоритм декодирования Рида, который позволяет исправлять $2^{m-r-1}-1$ ошибок и восстанавливать k информационных символов. Предположим, что мы имеем декодер для кода Рида-Маллера $(r-1)-$порядка, исправляющего $2^{m-r}-1$ ошибок. Мы построим декодер для кода Рида-Маллера $r-$порядка, исправляющего $2^{m-r-1}-1$ ошибок, сведя этот случай к предыдущему. Но мы знаем, что код Рида-Маллера $0-$го порядка может быть декодирован мажоритарным методом, а тогда по индукции получим метод декодирования кодов высших порядков.

Разобьём информационный вектор длины k на $r+1$ частей: $\overline{a}=[I_0,I_1,\dots,I_r]$, где I_l содержит C_m^l информационных битов. Вспомним также, что $k=\sum_{l=0}^{r}C_m^l$. Каждая часть вектора $\overline{a}$ будет умножатся на свой блок матрицы G:

$\overline{c}=\overline{a}G=[I_0,I_1,\dots,I_r]\begin{pmatrix} G_0 \\ G_1 \\ \vdots \\ G_r \end{pmatrix}-$ это метод кодирования.

Если мы сможем восстановить информационные биты в $r-$ой части, то сможем восстановить их вклад в принятое слово а затем вычесть этот вклад, то задача сведётся к декодированию кода Рида-Маллера $(r-1)-$порядка, и т.д., пока не придём к декодированию кода Рида-Маллера $0-$го, что осуществляется мажоритарным методом.

Полученное слово в виде $\overline{v} = [I_0, I_1, \ldots, I_r] \begin{pmatrix} G_0 \\ G_1 \\ \vdots \\ G_r \end{pmatrix} + \overline{e}$.

Рассмотрим декодирование последнего информационного символа i_{k-1}. При построении кодового слова он умножается на последнюю строку матрицы G_r. Вычислим 2^{m-r} проверочных сумм по 2^r бит в каждой из 2^m бит принятого слова, так что принятый бит входит только в одну сумму. Проверочные суммы строятся так, что символ i_{k-1} вносит вклад только в один бит, а все другие информационные символы вносят вклад в чётное число битов в каждой проверочной сумме. Таким образом, при отсутствии ошибок все проверочные суммы равны i_{k-1}. Но даже если имеется не более $2^{m-r-1}-1$ ошибок, большинство проверочных сумм будет равняться i_{k-1}. Это следует из того, что в последней строке матрицы G_r единицы разделяет 2^r-1 нулей.

Первая проверочная сумма представляет собой сумму по $\bmod 2$ первых 2^r битов принятого слова, вторая – сумму по $\bmod 2$ следующих 2^r битов и т.д. Всего получается 2^{m-r} проверочных сумм, а по предположению имеется не более $2^{m-r-1}-1$ ошибок. Значит, мажоритарное голосование проверочных сумм даёт правильное значение i_{k-1}. Например, для $(16,11)-$кода Рида-Маллера имеем 4 проверочных суммы $(m=4, r=2)$:

$$i_{10}^{(1)} = v_0 + v_1 + v_2 + v_3;\ i_{10}^{(2)} = v_4 + v_5 + v_6 + v_7;\ i_{10}^{(3)} = v_8 + v_9 + v_{10} + v_{11};\ i_{10}^{(4)} = v_{12} + v_{13} + v_{14} + v_{15}.$$

Если произошла одна ошибка, то одна из оценок, даваемых проверочными суммами, неверна и мажоритарное голосование даёт правильное значение i_{10}. Если произошло две ошибки, то ни одно значение проверочных сумм не встречается чаще другого, и поэтому обнаруживается двойная ошибка, которая не исправляется.

Аналогично может быть продекодирован любой из информационных символов, который умножается на одну из строк матрицы G_r. Для этого надо перестановкой столбцов матрицы G_r добиться того, чтобы соответствующая строка матрицы G_r выглядела так, как в предыдущем случае, когда единицы отделяются 2^r-1 нулями.

После того, как информационные символы найдены (это символы, входящие в I_{r-1}), их вклад в кодовое слово вычитается из принятого слова. Это эквивалентно тому, что мы получаем слово Рида-Маллера $(r-1)-$го порядка. И всё начинаем сначала. Этот процесс повторяется до тех пор, пока не будут найдены все информационные символы (биты).

Лекция 12. **Циклические коды.**

Определение. Линейный $(n,k)-$код C называется циклическим, если с каждым кодовым словом $\overline{c}=(c_0,c_1,\ldots,c_{n-1})$ он содержит и кодовое слово $\overline{c}_1=(c_{n-1},c_0,c_1,\ldots,c_{n-2})$, где $c_i\in\mathbb{F}_q$.

Пусть $(n,q)=1$ и пусть $I=\langle x^n-1\rangle-$ главный идеал в кольце $\mathbb{F}_q[x]$. Фактор-кольцо $\mathbb{F}_q[x]/I$ является кольцом главных идеалов, состоящим из q^n классов, представителями которых являются многочлены степени, меньше n. Между элементами из $\mathbb{F}_q[x]/I$ и векторами из $\mathbb{F}_q^n$ можно установить взаимнооднозначное соответствие $v_0+v_1x+\ldots+v_{n-1}x^{n-1}\rightleftharpoons(v_0,v_1,\ldots,v_{n-1})$.

Лемма (критерий цикличности кода). Линейный $(n,k)-$код C будет циклическим $\Leftrightarrow$ многочлены, соответствующие кодовым словам, образуют идеал в $\mathbb{F}_q[x]/I$.

Доказательство.

Необходимость. Пусть $C-$ циклический код. Обозначим через $C[x]-$ множество кодовых многочленов. Ясно, что для $\overline{c}_1,\overline{c}_2\in C$ имеем $\overline{c}_1-\overline{c}_2\in C\Rightarrow c_1(x)-c_2(x)\in C[x]$. Так что первое требование в определении идеала выполнено. Далее,

$$x\cdot c(x)=x\left(c_0+c_1x+\ldots+c_{n-1}x^{n-1}\right)=c_0x+c_1x^2+\ldots+c_{n-1}x^n=c_{n-1}+c_0x+\ldots+c_{n-2}x^{n-1}$$

т.к. $x^n=x^n-1+1\equiv 1$. И так как $(c_{n-1},c_0,c_1,\ldots,c_{n-2})\in C$, то $x\cdot c(x)\in C[x]$, т.е. выполнено второе требование идеала. Необходимость доказана.

Достаточность. Так как $C[x]-$ идеал в $\mathbb{F}_q[x]/I$, то $\forall c(x)\in C[x]$ имеем $x\cdot c(x)\in C[x]$, а потому для каждого $(c_0,c_1,\ldots,c_{n-1})\in C$ имеем $(c_{n-1},c_0,c_1,\ldots,c_{n-2})\in C$, ч.т.д.

Пусть $C-$ циклический код. Тогда $C[x]-$ главный идеал. Выберем порождающий многочлен этого идеала. Обозначим его через $g(x)$, $\deg g(x)\le n-1$. Следовательно, $\forall c(x)\in C[x]$ имеем $c(x)\vdots g(x)\left(\mathrm{mod}\left(x^n-1\right)\right)$. В частности, многочлен x^n-1, который играет роль нуля в фактор-кольце $\mathbb{F}_q[x]/I$, делится на $g(x)$. Многочлен $g(x)$ будем называть порождающим многочленом циклического кода C.

Итак, для $c(x)\in C[x]$ $\exists a(x)$ так что $c(x)=a(x)g(x)$. Если $\deg g(x)=m$, то $\deg a(x)\le n-1-m$, т.е. $\deg a(x)\le k-1$, где $k=n-m$. Т.о., $a(x)$ определяется коэффициентами

$a_0, a_1, \ldots, a_{k-1}$. Это даёт возможность строить кодовые слова $(c_0, c_1, \ldots, c_{n-1})$ по информационным $(a_0, a_1, \ldots, a_{k-1})$ из равенства $c(x) = a(x)g(x) = \left(a_0 + a_1 x + \ldots + a_{k-1}x^{k-1}\right)\left(g_0 + g_1 x + \ldots + g_m x^m\right)$.

Обозначим $h(x) = \dfrac{x^n - 1}{g(x)}$ и назовём его проверочным многочленом. $h(x) = h_0 + h_1 x + \ldots + h_k x^k$, т.к. $n = m + k$. Теперь очевидно, что $c(x) \in C[x] \Leftrightarrow c(x)h(x) \vdots \left(x^n - 1\right)$, или другими словами, $c(x) \in C[x] \Leftrightarrow c(x)h(x)$ является нулём в $\mathbb{F}_q[x]/I$.

Рассмотрим две матрицы соответственно размерности $k \times n$ и $m \times n$:

$$G = \begin{pmatrix} g_0 & g_1 & \ldots & g_m & 0 & 0 & \ldots & 0 \\ 0 & g_0 & \ldots & g_{m-1} & g_m & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & g_0 & \ldots & g_m \end{pmatrix}, \quad H = \begin{pmatrix} 0 & 0 & \ldots & 0 & h_k & h_{k-1} & \ldots & h_0 \\ 0 & 0 & \ldots & h_k & h_{k-1} & h_{k-2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h_k & h_{k-1} & \ldots & h_0 & 0 & 0 & \ldots & 0 \end{pmatrix}.$$

Лемма. Матрицы G и H являются соответственно порождающей и проверочной матрицей матрицами циклического кода с порождающим многочленом $g(x)$.

Доказательство. Многочлены $g(x), xg(x), \ldots, x^{k-1}g(x)$ являются кодовыми, а соответствующие им кодовые слова линейно независимы в пространстве $\mathbb{F}_q^n$, их количество равно размерности пространства кодовых слов, значит они образуют базис этого пространства, а потому $G-$ порождающая матрица. Покажем, что $HG^{\dot{O}} = 0$. Рассмотрим элементы d_{ij}, расположенные на пересечении $i-$й строки и $j-$го столбца матрицы $HG^{\dot{O}}$. Имеем

$d_{11} = g_m h_{k-1} + g_{m-1}h_k$, $d_{12} = g_m h_{k-2} + g_{m-1}h_{k-1} + g_{m-2}h_k$ и т.д.

Все значения d_{ij} представляют собой коэффициенты многочлена $g(x)h(x)$ за исключением старшего коэффициента и свободного члена. Но $g(x)h(x) = x^n - 1$, откуда $d_{ij} = 0$. Т.е. $HG^{\dot{O}} = 0$. А это означает, что H является проверочной матрицей кода.

Замечание. Построенные матрицы G и H соответствуют, вообще говоря, несистематической форме линейного кода. Напомним, что в систематической (стандартной) форме $H = \left(A \quad I_m\right)$, $G = \left(I_k \quad -A^{\dot{O}}\right)$. Покажем, как можно получить циклический код в стандартной форме.

Для любого натурального $j < n$ обозначим $x^j = b_j(x)g(x) + r_j(x)$, где $\deg r_j(x) < m$. Таким образом, $x^j - r_j(x) \in C[x]$. Обозначим $g_j(x) = x^k\left(x^j - r_j(x)\right) \in C[x]$. Тогда учитывая, что $x^n = x^n - 1 + 1 \equiv 1\left(\mathrm{mod}\left(x^n - 1\right)\right)$, имеем

$g_{n-k}(x)=1-x^k r_{n-k}(x),$

$g_{n-k+1}(x)=x-x^k r_{n-k+1}(x),$

……

$g_{n-1}(x)=x^{k-1}-x^k r_{n-1}(x).$

Эти многочлены линейно независимы над $\mathbb{F}_q$ и являются кодовыми. Так что порождающая матрица этого кода имеем вид $G=\left(I_k \quad -A^{\dot{O}}\right)$ и тогда $H=\left(A \quad I_m\right)$.

Теорема. Пусть $n=2^m-1$, и пусть $g(x)-$ минимальный многочлен примитивного элемента поля $\mathbb{F}_{2^m}$, откуда $\deg g(x)=m$. Тогда циклический код, порождённый многочленом $g(x)$, эквивалентен коду Хэмминга. (Два кода называются эквивалентными, если их проверочные матрицы получаются одна из другой перестановкой столбцов).

Доказательство. Пусть $\alpha-$ примитивный элемент поля $\mathbb{F}_{2^m}$, тогда его порядок в группе $\mathbb{F}_{2^m}^*$ равен 2^m-1. Элементы $1,\alpha,\ldots,\alpha^{m-1}$ образуют базис пространства $\mathbb{F}_{2^m}$ над $\mathbb{F}_2$. Пусть $\alpha^i=\sum_{j=0}^{m-1}a_{ij}\alpha^j$, где $i=0,1,\ldots,n-1$. Рассмотрим матрицу $H=\begin{pmatrix} a_{00} & \ldots & a_{n-1,0} \\ \vdots & \ddots & \vdots \\ a_{0,m-1} & \ldots & a_{n-1,m-1} \end{pmatrix}$.

Так как $g(\alpha)=0$, то $c(\alpha)=a(\alpha)g(\alpha)=0$ для каждого кодового многочлена $c(x)$.

Пусть $\overline{c}=(c_0,\ldots c_{n-1})\in C$, тогда $c(\alpha)=c_0+c_1\alpha+\ldots+c_{n-1}\alpha^{n-1}=0$. Из способа построения матрицы H заключаем, что $c(\alpha)=0\Leftrightarrow H\overline{c}^{\dot{O}}=0$.

Все столбцы матрицы H отличны от нулевого и различны между собой. Так что они являются координатными столбцами различных элементов поля $\mathbb{F}_{2^m}$. Эти координатные столбцы представляют собой запись $m-$значных чисел в алфавите $\mathbb{F}_2$, а значит, матрица H отличается от проверочной матрицы бинарного кода Хэмминга только расположением столбцов, ч.т.д.

Следствие. Бинарный код Хэмминга является циклическим кодом.

Из приведённых выше рассуждений видно, что каждый нетривиальный делитель многочлена x^n-1 над полем $\mathbb{F}_q$ порождает циклический код. Так как по предположению $(n,q)=1$, то x^n-1 не имеет кратных корней. Разложим x^n-1 в произведение неприводимых: $x^n-1=f_1(x)\ldots f_r(x)$, тогда при $i\neq j \qquad (f_i(x),f_j(x))=1$. В этом случае многочлен x^n-1

имеет $\sum_{i=0}^{r} C_r^i$ различных делителей, из которых два многочлена: 1 и x^n-1, являются тривиальными. Таким образом, для данного n имеется в точности $\sum_{i=0}^{r} C_r^i - 2 = 2^r - 2$ различных циклических кодов в алфавите $\mathbb{F}_q$, $(n,q)=1$.

Пусть $n \mid q^m-1$ и $\alpha-$ примитивный элемент поля $\mathbb{F}_{q^m}$. Через $H\left(\alpha^{i_1},\alpha^{i_2},\ldots,\alpha^{i_n}\right)$ будем обозначать матрицу, столбцами которой служат коэффициенты представлений элементов $\alpha^{i_1},\alpha^{i_2},\ldots,\alpha^{i_n}$ в базисе $1,\alpha,\ldots,\alpha^{m-1}$. Мы уже видели, что при $q=2$ и $n=2^m-1$ матрица $H\left(1,\alpha,\ldots,\alpha^{n-1}\right)$ является проверочной матрицей линейного кода (эквивалентного коду Хэмминга), исправляющего одиночные ошибки.

Т е о р е м а. Пусть $\alpha-$ примитивный элемент поля $\mathbb{F}_{q^m}$, и пусть $\beta=\alpha^{q-1}, n=\dfrac{q^m-1}{q-1}$. Тогда минимальное расстояние кода с проверочной матрицей $H\left(1,\beta,\ldots,\beta^{n-1}\right)$ не меньше трёх $\Leftrightarrow$ $(n,q-1)=1 \Leftrightarrow (m,q)=1$.

Д о к а з а т е л ь с т в о. Пусть два столбца матрицы H линейно зависимы: $\beta^i=\gamma\beta^j, \gamma\in\mathbb{F}_q, i\neq j$. Тогда $\beta^{i-j}=\gamma \Rightarrow \beta^{(i-j)(q-1)}=1$. Элемент $\alpha^{\frac{q^m-1}{q-1}}$ является примитивным элементом поля $\mathbb{F}_q$, откуда $\beta^{i-j}=\alpha^{\frac{q^m-1}{q-1}k}=\alpha^{nk}, 0\le k<q-1$. И т.к. $\beta=\alpha^{q-1}$, то $(q-1)(i-j)=nk$. Но $(i-j)<n$, $k<q-1$, а значит, $q-1$ и n не взаимно простые. Таким образом, H содержит два линейно зависимых столбца $\Leftrightarrow (n,q-1)>1$.

Далее, $n=\dfrac{q^m-1}{q-1}=q^{m-1}+q^{m-2}+\ldots+q+1$, $q^{m-j}-1=(q-1)s_j$, откуда $q^{m-j}=(q-1)s_j+1$, а значит, $n=(q-1)\sum_{j=1}^{m} s_j + m \Rightarrow (n,q-1)=(m,q-1)$, ч.т.д.

П р и м е р. Пусть $q=4, m=4$. Тогда $n=\dfrac{4^4-1}{4-1}=85$. Пусть $\alpha-$ примитивный элемент поля $\mathbb{F}_{4^4}$ над $\mathbb{F}_2$. Положим $\beta=\alpha^3$. Мультипликативная группа поля $\mathbb{F}_4^*$ имеет порядок 3, а т.к. в $\mathbb{F}_{4^4}$ имеется только два элемента порядка 3: α^{85} и α^{170}, то $\mathbb{F}_4=\left\{0,1,\alpha^{85},\alpha^{170}\right\}=\left\{0,1,\tilde{2},\tilde{3}\right\}$, где мы переобозначили $\alpha^{85}=\tilde{2}$, $\alpha^{170}=\tilde{3}$.

Рассматривая неприводимые многочлены 8-й степени над $\mathbb{F}_2$ (ибо $4^4 = 256 = 2^8$), мы получаем, что многочлен $p(x) = x^8 + x^4 + x^3 + x^2 + 1$ имеет корнем примитивный элемент поля $\mathbb{F}_{4^4}$ над $\mathbb{F}_2$, т.е. $\alpha^8 = \alpha^4 + \alpha^3 + \alpha^2 + 1$. Используя это равенство и множество сопряжённых с β над $\mathbb{F}_4$: $\left\{\beta, \beta^4, \beta^{4^2} = \beta^{16}, \beta^{4^3} = \beta^{64}\right\}$; получаем порождающий многочлен для β над $\mathbb{F}_4$:

$$g(x) = (x-\beta)\left(x-\beta^4\right)\left(x-\beta^{16}\right)\left(x-\beta^{64}\right) =$$

$$= \left(x-\alpha^3\right)\left(x-\alpha^{12}\right)\left(x-\alpha^{48}\right)\left(x-\alpha^{142}\right) = x^4 + x^3 + \alpha^{170}x + 1 = x^4 + x^3 + \tilde{3}x + 1.$$

Этот многочлен порождает циклический код (85,81) над $\mathbb{F}_4$.

Лекция 13. **Циклические коды, исправляющие две ошибки.**

Мы здесь рассматриваем коды над полем $\mathbb{F}_2$. Ранее мы показали, что для $n = 2^m - 1$ и для примитивного элемента α поля $\mathbb{F}_{2^m}$ существует циклический код (порождённый минимальным многочленом для α), исправляющий одиночные ошибки. Пусть $g(x)-$ многочлен наименьшей степени, корнями которого являются α и α^3. Мы опишем процедуру декодирования этого циклического кода, которая приводит к исправлению двойных ошибок, а значит минимальное расстояние этого кода не меньше пяти.

Принятое слово запишем в виде $v(x) = a(x)g(x) + e(x), \quad \deg v(x) < n,$

Многочлен $e(x)$ содержит не более двух ненулевых коэффициентов. Т.о. $e(x) = 0$, или $e(x) = x^i$, или $e(x) = x^i + x^j$, где i и j указывают номера позиций, где произошли ошибки. Т.к. $v(\alpha) = e(\alpha)$, то $e(\alpha) = 0$, либо $e(\alpha) = \alpha^i$, либо $e(\alpha) = \alpha^i + \alpha^j$. Поэтому $x_1 = \alpha^i$, $x_2 = \alpha^j$ называют л о к а т о р а м и ошибок. Если произошла одна ошибка, то полагаем $x_2 = 0$, а если нет ошибок, то $x_1 = x_2 = 0$.

Обозначим $s_1 = v(\alpha)$, $s_3 = v\left(\alpha^3\right)$. Значения s_1 и s_3 называют компонентами синдрома и их легко вычислить по принятому слову. Если произошло две ошибки, то $\begin{cases} s_1 = x_1 + x_2 \\ s_3 = x_1^3 + x_2^3 \end{cases}$.

Поскольку $i, j < n$, то $s_1 = 0$ только когда не произошло ни одной ошибки. Если же $s_1 \neq 0$, то из предыдущей системы $s_1^3 + s_3 = x_1^2 x_2 + x_1 x_2^2 = (x_1 + x_2) x_1 x_2 = s_1 x_1 x_2$.

Отсюда следует, что x_1, x_2 являются корнями уравнения

$$x^2 + s_1 x + \frac{s_1^3 + s_3}{s_1} = 0 \quad (\text{если } s_1 \neq 0).$$

Из этого уравнения определяем локаторы ошибок (двух или одной) и значит мы знаем, как исправить двойные ошибки этого кода.

Заметим, что строить циклический код по корням α и α^2 нецелесообразно, т.к. в силу свойства автоморфизма Фробениуса α и α^2 являются корнями одного и того же минимального многочлена и тогда мы приходим к циклическому коду, эквивалентному коду Хэмминга, который не может исправлять двойные ошибки (его минимальное расстояние меньше пяти).

Лекция 14. **Циклические коды, исправляющие пакеты ошибок.**

Хорошо иметь коды, исправляющие t ошибок любой конфигурации. Однако бывает, что каналы приводят к пакетам ошибок. Значит, целесообразно разрабатывать коды, исправляющие пакеты ошибок с достаточно высокой скоростью.

Определение. Циклическим пакетом длины t называется вектор, все ненулевые компоненты которого расположены среди t последовательных (по циклу) компонент, первая и последняя из которых отличны от нуля.

Пакет ошибок можно задавать многочленом вида $e(x)=x^i b(x)$, где $\deg b(x)<t$ и у многочлена $b(x)$ отличны от нуля коэффициенты b_0 и b_{t-1}. Здесь x^i указывает начальный локатор ошибок.

Синдромным многочленом циклического кода, порождённого многочленом $g(x)$, является остаток от деления $v(x)$ (а значит и $e(x)$) на $g(x)$. Таким образом,

$$s(x)=R_{g(x)}\left[e(x)\right]=R_{g(x)}\left[v(x)\right]\equiv v(x)\left(\operatorname{mod} g(x)\right).$$

Ясно, что синдромные многочлены циклического кода, исправляющего пакеты ошибок, должны быть различными.

Например, многочлен $g(x)=x^6+x^3+x^2+x+1$ над полем $\mathbb{F}_2$ является делителем многочлена $x^{15}-1$, а потому порождает циклический код длины 15. Перечислим все циклические пакеты длины $t\le 3$:

$$e(x)=x^i,$$

$$e(x)=x^i(1+x),$$

$$e(x)=x^i\left(1+x^2\right),$$

$$e(x)=x^i\left(1+x+x^2\right),$$

где $i = 0,\ldots,14$. Непосредственной проверкой можно убедиться, что все синдромы этих ошибок различны, значит, циклический код, порождённый $g(x)$, исправляет пакеты ошибок длины, не превосходящей трёх.

Замечание. Сумма кодового слова и исправляющего пакета не может равняться сумме другого кодового слова и исправляющего пакета (так как иначе синдромные многочлены этих пакетов совпадут, и значит, они не могут быть однозначно определены и исправлены). В частности, в данном примере, кодовое слово не может быть пакетом длины 6. В общем случае, если линейный код исправляет пакеты длины t и меньше, то он не может иметь ненулевых кодовых слов, являющихся пакетами длины $2t$ или меньше.

Теорема (граница Рейгера). Каждый линейный код, исправляющий все пакеты длины t и меньше, должен содержать не менее $2t$ проверочных символов.

Доказательство. Пусть код исправляет пакеты длины t и менее. Тогда сдери его кодовых слов нет пакетов длины $2t$ или менее. Рассмотрим множество D векторов из $\mathbb{F}_q^n$, у которых все компоненты, за исключением первых $2t$, равны нулю. Так как разность векторов из одного и того же смежного класса фактор-группы $\mathbb{F}_q^n / C$ принадлежит C, то векторы из D лежат в различных смежных классах. Но в D имеется q^{2t} векторов, значит $\left[\mathbb{F}_q^n : C\right] \geq q^{2t}$. Поскольку $\left[\mathbb{F}_q^n : C\right] = q^n : q^k = q^{n-k} = q^m$, то $n - k \geq 2t$, т.е. число проверочных символов в коде C не менее $2t$, что и требовалось доказать.

Рассмотренный выше циклический код достигает границы Рейгера. В самой теореме цикличность кода не предполагается. С помощью ЭВМ было найдено много хороших циклических кодов, исправляющих пакеты ошибок. Укажем некоторые из них.

Порождающий многочлен	Параметры кода	Длина исправляемого пакета
$x^4 + x^3 + x^2 + 1$	$(7,3)$	2
$x^5 + x^4 + x^2 + 1$	$(15,10)$	2
$x^6 + x^5 + x^4 + x^3 + 1$	$(15,9)$	3
$x^8 + x^7 + x^6 + x^3 + 1$	$(63,55)$	3
$x^{12} + x^8 + x^5 + x^3 + 1$	$(511,499)$	4

Из приведённых циклических кодов можно построить более длинные коды методом перемежения. Чтобы из $(n,k)-$кода получить $(jn, jk)-$код, выберем из исходного кода

j кодовых слов и укрупним кодовое слово, чередуя их символы: пусть $\overline{c}_1 = \left(c_1^{(1)},\ldots,c_n^{(1)}\right),\ldots,\overline{c}_j = \left(c_1^{(j)},\ldots,c_n^{(j)}\right)-$ выделенные кодовые слова. Строим новое кодовое слово

$$\tilde{c} = \left(c_1^{(1)},\ldots,c_1^{(j)},c_2^{(1)},\ldots,c_2^{(j)},\ldots,c_n^{(1)},\ldots,c_n^{(j)}\right).$$

Если исходный код исправлял пакеты ошибок длины, не превосходящей t, то построенный код будет исправлять пакеты ошибок длины, не превосходящей jt. Так, например, выбирая в коде $(15,9)$ четыре кодовых слова, получим $(60,36)-$код, который может исправлять пакеты ошибок длины, не превосходящей 12, так как исходный код исправлял пакеты длины, не превосходящей 3. Покажем, что для циклического кода новый код перемежения также будет циклическим. Действительно, пусть $g(x)-$ порождающий многочлен данного циклического кода. Покажем, что $g\left(x^j\right)-$ порождающий многочлен кода перемежения. Для этого покажем, что перемежение символов нескольких информационных многочленов с последующим умножением на $g\left(x^j\right)$ даёт тоже самое кодовое слово, что и умножение каждого из исходных информационных многочленов на $g(x)$ с последующим перемежением этих слов $(n,k)-$кода. Точнее, пусть $c_i(x) = a_i(x)g(x)$, $i = 1,\ldots,j-$ выбранные кодовые слова (многочлены). Для формирования слова кода перемежения каждое из этих кодовых слов растягивается вставкой $(j-1)$ нулей между соседними символами. Кодовое слово из кода перемежения получается теперь задержкой $l-$го слова на $(l-1)$ тактов, $l = 1,\ldots,j$ с последующим сложением

$$\tilde{c}(x) = c_1\left(x^j\right) + xc_2\left(x^j\right) + \ldots + x^{j-1}c_j\left(x^j\right) = \left[a_1\left(x^j\right) + xa_2\left(x^j\right) + \ldots + x^{j-1}a_j\left(x^j\right)\right]g\left(x^j\right).$$

Стоящий в квадратных скобках многочлен соответствует информационному слову $\overline{a}$, которое является словом перемежения информационных слов $\overline{a}_1,\ldots,\overline{a}_j$. Значит, $\tilde{c}(x) = a(x)g\left(x^j\right)$.

Кроме кодов, найденных на ЭВМ и исправляющих пакеты ошибок, известны коды, построенные аналитическими методами, которые также исправляют пакеты ошибок. К таким кодам относятся коды Файра.

Коды Файра.

Определение. Кодом Файра в алфавите q называется циклический $(n,k)-$код над $\mathbb{F}_q$, с порождающим многочленом $g(x)=\left(x^{2t-1}-1\right)p(x)$, где $p(x)-$ примитивный многочлен степени m над $\mathbb{F}_q$, причём $m\geq t$.

Теорема. Для кода Файра $n=\left(q^m-1\right)(2t-1)$, $k=n-m-(2t-1)$.

Доказательство. Пусть $n-$ наименьшее натуральное, для которого $\left(x^n-1\right)\vdots g(x)$. В силу примитивности $p(x)$ имеем $\left(x^{q^m-1}-1\right)\vdots p(x)$, а следовательно, $n\vdots\left(q^m-1\right)$. Но $\left(x^n-1\right)\vdots\left(x^{2t-1}-1\right)$, откуда $n\vdots(2t-1)$. Многочлены $x^{q^m-1}-1$ и $x^{2t-1}-1$ не имеют общих корней, за исключением единицы, а потому $\left(2t-1,q^m-1\right)=1$, откуда $n=\left(q^m-1\right)(2t-1)l$. Но в виду минимальности n имеем $n=\left(q^m-1\right)(2t-1)$.

Теорема. Код Файра исправляет все пакеты ошибок длины t и менее.

Доказательство. Каждый код будет исправлять пакеты длины t и менее, для различных пакетов длины не более t их синдромы различны. Следовательно, эти пакеты лежат в различных смежных классах группы $\mathbb{F}_q^n/C$. Рассмотрим два пакета $x^ib_1(x)$ и $x^jb_2(x)$, где $b_1(x),b_2(x)-$ многочлены степени меньше t, причём их свободные члены отличны от нуля. Предположим, что для кода Файра эти пакеты имеют одинаковые синдромы. Тогда $x^i\left(b_1(x)-x^{j-i}b_2(x)\right)-$ кодовое слово. В силу цикличности кода можно считать, что $i=0$. Тогда

$$b_1(x)-x^jb_2(x)=\left(x^{2t-1}-1\right)p(x)a(x)=g(x)a(x)\left(\mathrm{mod}\left(x^n-1\right)\right) \qquad (1)$$

Учтём, что для каждого ν, $0\leq\nu<q^m-1$ $\left(x^{\nu(2t-1)}-1\right)\vdots\left(x^{2t-1}-1\right)$. Значит,

$$\left(x^{\nu(2t-1)}-1\right)b_1(x)=\left(x^{2t-1}-1\right)Q_1(x)\left(\mathrm{mod}\left(x^n-1\right)\right). \qquad (2)$$

Путём сложения (1) и (2) получаем

$$x^{\nu(2t-1)}b_1(x)-x^jb_2(x)=\left(x^{2t-1}-1\right)Q_2(x)\left(\mathrm{mod}\left(x^n-1\right)\right)$$

и значит,

$$x^{\nu(2t-1)}\left[b_1(x)-x^{j-\nu(2t-1)}b_2(x)\right]=\left(x^{2t-1}-1\right)Q_3(x)\left(\mathrm{mod}\left(x^n-1\right)\right). \qquad (3)$$

Отсюда

$$x^{\nu(2t-1)-(t-1)}\left[x^{t-1}b_1(x)-x^{j+t-1-\nu(2t-1)}b_2(x)\right]=\left(x^{2t-1}-1\right)Q_3'(x)\left(\mathrm{mod}\left(x^n-1\right)\right). \qquad (4)$$

Мы можем выбрать неотрицательное ν таким образом, чтобы в (3) или (4) множителем при $b_2(x)$ будет x^l, причём $0 \le l < t$. Таким образом, за счёт выбора ν получаем:

$$x^{\nu(2t-1)}\left[b_1(x) - x^l b_2(x)\right] = \left(x^{2t-1}-1\right)Q_3(x)\left(\mathrm{mod}\left(x^n-1\right)\right)$$

или

$$x^{\nu(2t-1)-(t-1)}\left[x^{t-1}b_1(x) - x^l b_2(x)\right] = \left(x^{2t-1}-1\right)Q_3'(x)\left(\mathrm{mod}\left(x^n-1\right)\right),$$

где $\deg b_1(x), \deg b_2(x) < t$, $k < t$. Так что степень многочлена в квадратных скобках не превосходит $2t-2$. Но $x^{2t-1}-1$ должно делить $x^{2t-1}-1$. Значит,

$b_1(x) - x^l b_2(x) = 0$ или $x^{t-1}b_1(x) - x^l b_2(x) = 0$.

Так как свободные члены b_{10} и b_{20} отличны от нуля, то $l = 0$ или, соответственно, $l = t-1$. В любом случае $j = \nu(2t-1)$ и $b_1(x) = b_2(x) = b(x)$. Покажем, что $b(x) = 0$.

Из соотношения (1) имеем

$$-\left(x^{\nu(2t-1)}-1\right)b(x) = a(x)\left(x^{2t-1}-1\right)p(x). \qquad (5)$$

При $\nu \ne 0$ многочлен $p(x)$ не может делить $\left(x^{\nu(2t-1)}-1\right)$, так как $\left(2t-1, q^m-1\right)=1$ и $\nu < q^m-1$. Поэтому в равенстве (5) многочлен может делить только $b(x)$. Но $\deg b(x) < \deg p(x)$. Следовательно, при $\nu \ne 0$ $b(x) = 0$, т.е. ошибок нет. Если $\nu = 0$, то из того, что $j = \nu(2t-1)$ следует $j = 0$, а тогда равенство $b_1(x) = b_2(x)$ противоречит выбору двух разных пакетов. Таким образом, два различных пакета $b_1(x)$ и $x^j b_2(x)$ принадлежат разным смежным классам. Следовательно, код способен исправлять пакеты длины t и менее, что и требовалось доказать.

Если $m = t$, то число проверочных символов равно $3t-1$, что всего лишь на $t-1$ превышает границу Рейгера. Например, при $m = t = 10$, $q = 2$, получаем $n = 1029 \cdot 19$, $k = n - 19 = 1028 \cdot 19$. Т.е. получаем $(19437, 19408)-$код Файра, исправляющий пакеты длины 10 и меньше.

Лекция 15. **Коды Боуоза-Чоудхури-Хоквингема (БЧХ коды).**

Определение. Пусть задано поле $\mathbb{F}_q$, натуральное m и элемент $\beta \in \mathbb{F}_{q^m}$ порядка n. Тогда для любых натуральных $d < n$ и $b < n$ определяем БЧХ код как циклический код, порождённый многочленом $g(x) = \mathrm{НОК}\left[f_b(x), f_{b+1}(x), \ldots, f_{b+d-2}(x)\right]$,

где $f_i(x)-$ минимальный многочлен элемента β^i.

Если $b=1$, то получаем БЧХ код в узком смысле, а если $n=q^m-1$, т.е. $\beta-$ примитивный элемент поля $\mathbb{F}_{q^m}$, то код называется примитивным БЧХ кодом. Если $n=q-1$, то получаем код Рида-Соломона. Число d обычно выбирается равным $2t+1$ и называется конструктивным расстоянием БЧХ кода.

Лемма 1. Минимальное расстояние БЧХ кода не меньше его конструктивного расстояния, т.е. $d_C \geq d$.

Доказательство. Из определения БЧХ кода, построенного на корнях $\beta^b, \beta^{b+1}, \ldots, \beta^{b+d-2}$, следует, что проверочная матрица этого кода имеет вид

$$H=\begin{pmatrix} 1 & \beta^b & \beta^{2b} & \ldots & \beta^{b(n-1)} \\ 1 & \beta^{b+1} & \beta^{2(b+1)} & \ldots & \beta^{(b+1)(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \beta^{b+d-2} & \beta^{2(b+d-2)} & \ldots & \beta^{(b+d-2)(n-1)} \end{pmatrix}.$$

Здесь каждая строка заменяется матрицей порядка m, столбцы которой являются коэффициентами представлений соответствующей степени β в базисе $\{1, \beta, \beta^2, \ldots, \beta^{n-1}\}$. Матрица H имеет размер $m(b+d-1)\times n$. Но мы будем рассматривать матрицу H именно в записанном выше виде.

Теперь, чтобы доказать наше утверждение достаточно показать, что любые $(d-1)$ столбцов этой матрицы линейно независимы. Но для этого достаточно показать, что определитель, составленный из любых $(d-1)$ столбцов этой матрицы, отличен от нуля. Имеем

$$\Delta=\begin{vmatrix} \beta^{i_1 b} & \beta^{i_2 b} & \ldots & \beta^{i_{d-1} b} \\ \beta^{i_1 (b+1)} & \beta^{i_2 (b+1)} & \ldots & \beta^{i_{d-1} (b+1)} \\ \vdots & \vdots & \ddots & \vdots \\ \beta^{i_1 (b+d-2)} & \beta^{i_2 (b+d-2)} & \ldots & \beta^{i_{d-1} (b+d-2)} \end{vmatrix} = \beta^{(i_1+i_2+\ldots+i_{d-1})b} \begin{vmatrix} 1 & 1 & \ldots & 1 \\ \beta^{i_1} & \beta^{i_2} & \ldots & \beta^{i_{d-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \beta^{i_1 (d-2)} & \beta^{i_2 (d-2)} & \ldots & \beta^{i_{d-1} (d-2)} \end{vmatrix}.$$

Справа стоит определитель Ван дер Монда, порождённый элементами $\beta^{i_1}, \beta^{i_2}, \ldots, \beta^{i_{d-1}}$. Но т.к. β имеет порядок n, то все эти степени β различны, а потому $\Delta \neq 0$.

Следствие. БЧХ код с конструктивным расстоянием $d=2t+1$ исправляет по крайней мере t ошибок.

Пример. Пусть $q=2$, $m=4$. Примитивный многочлен четвёртой степени над $\mathbb{F}_2$ равен $p(x)=x^4+x+1$. Его корень обозначим через α. Поле $\mathbb{F}_{2^4}=\mathbb{F}_{16}$ состоит из нуля и $\alpha^i, i=0,1,\ldots,14$. Порождающий многочлен кода, исправляющего двойные ошибки, равен

$$g(x)=\text{НОК}\left[f_1(x), f_2(x), f_3(x), f_4(x)\right]=$$

$$= \text{НОК}\left[x^4+x+1, x^4+x+1, x^4+x^3+x^2+x+1, x^4+x+1\right] = x^8+x^7+x^6+x^4+1.$$

Таблица степеней $\alpha^i, i=0,1,\ldots,14$, и их минимальных многочленов над $\mathbb{F}_2$ имеет вид

	В базисе $\{1,\alpha,\alpha^2,\alpha^3\}$	В двоичном виде	Минимальные многочлены
1	1	0000	$x-1$
α	α	0001	x^4+x+1
α^2	α^2	0010	x^4+x+1
α^3	α^3	0100	$x^4+x^3+x^2+x+1$
α^4	$\alpha+1$	1000	x^4+x+1
α^5	$\alpha^2+\alpha$	0011	x^2+x+1
α^6	$\alpha^3+\alpha^2$	0110	$x^4+x^3+x^2+x+1$
α^7	$\alpha^3+\alpha+1$	1011	x^4+x^3+1
α^8	α^2+1	0101	x^4+x+1
α^9	$\alpha^3+\alpha$	1010	$x^4+x^3+x^2+x+1$
α^{10}	$\alpha^2+\alpha+1$	0111	x^2+x+1
α^{11}	$\alpha^3+\alpha^2+\alpha$	1110	x^4+x^3+1
α^{12}	$\alpha^3+\alpha^2+\alpha+1$	1111	$x^4+x^3+x^2+x+1$
α^{13}	$\alpha^3+\alpha^2+1$	1101	x^4+x^3+1
α^{14}	α^3+1	1001	x^4+x^3+1

Порождающий многочлен кода, исправляющего тройные ошибки, равен

$$g(x) = \text{НОК}\left[f_1(x), f_2(x), f_3(x), f_4(x), f_5(x), f_6(x)\right] = x^{10}+x^8+x^5+x^4+x^2+x+1.$$

Это $(15,5)-$код.

Для $t=4$

$$g(x) = \text{НОК}\left[f_1(x),\ldots,f_8(x)\right] = x^{14}+x^{13}+\ldots+x+1 = \frac{x^{15}-1}{x-1}.$$

Это $(15,1)-$код с повторением, который исправляет даже 7 ошибок.

Но поле $\mathbb{F}_{16}$ можно рассматривать как расширение поля $\mathbb{F}_4$, т.е. $\mathbb{F}_{16} = \mathbb{F}_{4^2}$. Сначала построим поле $\mathbb{F}_4$. Обозначим через β корень неприводимого над $\mathbb{F}_4$ многочлена x^2+x+1. Заметим, что β^2 тоже корень этого многочлена, так что $\mathbb{F}_4 = \{0,1,\beta,\beta^2\}$. Обозначим $\beta = 2$, $\beta^2 = 3$. К тому же заметим, что $\beta^2 = \beta+1$.

Чтобы получить поле $\mathbb{F}_{4^2}$ надо рассмотреть неприводимый над $\mathbb{F}_4$ многочлен второй степени. Таким многочленом не может быть x^2+x+1, т.к. его корнем является $\beta \in \mathbb{F}_4$. Возьмём

многочлен x^2+x+2 и обозначим через γ его корень. Так как $\beta,\gamma\in\mathbb{F}_{16}$, то их можно выразить через α. Из приведённой выше таблицы видно, что $\beta=\alpha^5$, $\beta^2=\alpha^{10}$. Но так как $\alpha^5=\alpha^2+\alpha$, то легко убедиться, что корнем многочлена x^2+x+2 является α.

И теперь мы можем построить следующую таблицу для элементов поля $\mathbb{F}_{16}$ и их минимальных многочленов над $\mathbb{F}_4$.

	В базисе $\{1,\gamma\}$	В четверичном виде	Минимальные многочлены
1	1	01	$x+1$
α	γ	10	x^2+x+2
α^2	$\gamma+2$	12	x^2+x+3
α^3	$3\gamma+2$	32	x^2+3x+1
α^4	$\gamma+1$	11	x^2+x+2
α^5	2	02	$x+2$
α^6	2γ	20	x^2+2x+1
α^7	$2\gamma+3$	23	x^2+2x+2
α^8	$\gamma+3$	13	x^2+x+3
α^9	$2\gamma+2$	22	x^2+x+1
α^{10}	3	03	x^2+3x+3
α^{11}	3γ	30	x^2+3x+3
α^{12}	$3\gamma+1$	31	x^2+3x+1
α^{13}	$2\gamma+1$	21	x^2+2x+2
α^{14}	$3\gamma+3$	33	x^2+3x+3

Из этой таблицы видно, что порождающий многочлен для исправляющего одиночные ошибки БЧХ над $\mathbb{F}_4$ длины 15 равен

$$g(x)=\text{НОК}\left[f_1(x),f_2(x)\right]=\left(x^2+x+2\right)\left(x^2+x+3\right)=x^4+x+1.$$

Значит, над $\mathbb{F}_4$ мы получили $(15,11)-$код БЧХ. Этим кодом последовательность 11 четверичных символов, что эквивалентно 22 битам, кодируется в последовательность 15 четверичных символов. Этот код не является кодом Хэмминга, т.к. для кода Хэмминга $n=\dfrac{q^m-1}{q-1}$.

Для кода БЧХ над $\mathbb{F}_4$ длины 15 и $t=2$

$$g(x)=\text{НОК}\left[f_1(x),f_2(x),f_3(x),f_4(x)\right]=$$

$$=\left(x^2+x+2\right)\left(x^2+x+3\right)\left(x^2+3x+1\right)=x^6+3x^5+x^4+x^3+2x^2+2x+1.$$

Получили $(15,9)-$код БЧХ над $\mathbb{F}_4$, исправляющий 2 ошибки.

Для $t=3$

$$g(x)=x^9+3x^8+3x^7+2x^6+x^5+2x^4+x+2.$$

Имеем $(15,6)-$код, исправляющий тройные ошибки.

Лекция 16. **Декодирование БЧХ кодов. Декодер Питерсона-Горенстейна-Цирлера.**

БЧХ коды могут быть декодированы по методу декодирования циклических кодов. Однако, для них разработаны и свои специфические декодеры. Мы рассмотрим декодер, разработанный Питерсоном для $q=2$, а в общем случае Горенстейном и Цирлером. Мы для простоты выкладок будем считать $b=1$, а элемент α порядка n не обязательно примитивный в поле $\mathbb{F}_{q^m}$.

Пусть $g(x)-$ минимальный многочлен, построенный на корнях $\alpha,\alpha^2,\ldots,\alpha^{2t}$, т.е. наш код исправляет t ошибок. Поэтому если $e(x)=e_{n-1}x^{n-1}+\ldots+e_1x+e_0-$ многочлен ошибок, то у него не более t коэффициентов отличны от нуля, т.е. $e(x)=e_{i_1}x^{i_1}+\ldots+e_{i_r}x^{i_r}$, где $e_{i_1},\ldots,e_{i_r}\in\mathbb{F}_2$, $0\le r\le t$, причём ни e_i ни r нам не известны.

Для удобства обозначим значения ошибок e_{i_l} через Y_l, а локаторы ошибок α^{i_l} через X_l. Поскольку порядок α равен n, то все X_l различны.

Для $j=1,2,\ldots,2t$ вычислим компоненты $s_1,\ldots,s_{2t}$ синдрома:

$$s_j=v(\alpha^j)=c(\alpha^j)+e(\alpha^j)=e(\alpha^j).$$

Тогда получаем следующую систему уравнений относительно X_l, Y_l:

$$\begin{cases} s_1=Y_1X_1+\ldots Y_rX_r \\ s_2=Y_1X_1^2+\ldots Y_rX_r^2 \\ \qquad\ldots \\ s_{2t}=Y_1X_1^{2t}+\ldots Y_rX_r^{2t} \end{cases}.$$

Решать эту систему непосредственно очень трудно. Рассмотрим многочлен локаторов ошибок $\Lambda(x)=\prod_{l=1}^{r}(1-xX_l)=\Lambda_rx^r+\ldots+\Lambda_1x+1$.

Положим $x=X_l^{-1}$ и умножим на $Y_lX_l^{j+r}$. Тогда получим

$$Y_lX_l^{j+r}\left(1+\Lambda_1X_l^{-1}+\ldots+\Lambda_rX_l^{-r}\right)=0 \quad \text{или} \quad Y_l\left(X_l^{j+r}+\Lambda_1X_l^{j+r-1}+\ldots+\Lambda_rX_l^{j}\right)=0.$$

Теперь суммируя по $l=1,\ldots r$, получим

$$s_{j+r}+\Lambda_1s_{j+r-1}+\ldots+\Lambda_rs_j=0,\ j=1,\ldots r. \qquad (*)$$

Последнее равенство рассматриваем как систему линейных уравнений относительно неизвестных $\Lambda_1,\ldots,\Lambda_r$.

Лемма 1. Система линейных уравнений (*) однозначно разрешима $\Leftrightarrow$ в полученном сообщении сочно r ошибок.

Доказательство. Система r линейных уравнений с r неизвестными разрешима однозначно $\Leftrightarrow$ матрица системы невырожденная. Матрица системы (*) имеет вид

$$A=\begin{pmatrix} s_1 & s_2 & \ldots & s_r \\ s_2 & s_3 & \ldots & s_{r+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_r & s_{r+1} & \ldots & s_{2r-1} \end{pmatrix}.$$

Но легко убедиться, что $A=BCB^T$, где

$$B=\begin{pmatrix} 1 & 1 & \ldots & 1 \\ X_1 & X_2 & \ldots & X_r \\ \vdots & \vdots & \ddots & \vdots \\ X_1^{r-1} & X_2^{r-1} & \ldots & X_r^{r-1} \end{pmatrix}, \qquad C=\begin{pmatrix} Y_1X_1 & 0 & \ldots & 0 \\ 0 & Y_2X_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & Y_rX_r \end{pmatrix}.$$

Действительно, элемент, стоящий на (i,j) месте в матрице BCB^T равен

$$\sum_{l=1}^{r} X_l^{i-1}\sum_{k=1}^{r} Y_lX_l\delta_{lk}X_k^{j-1}=\sum_{l=1}^{r} Y_lX_l^{i+j-1}=s_{i+j-1}=a_{ij},$$

где $\delta_{lk}-$ символ Кронекера.

Поэтому имеем $|A|=|B|^2|C|$, откуда видно, что $|A|\neq 0\Leftrightarrow\left(|C|\neq 0\right)\wedge\left(|B|\neq 0\right)$ $\Leftrightarrow$ имеется точно r ошибок, т.е. $X_1\neq 0,\ldots,X_r\neq 0$ и $Y_1\neq 0,\ldots,Y_r\neq 0$, что и требовалось доказать.

Если с помощью предыдущей леммы мы нашли коэффициенты многочлена локатора ошибок, то находим корни этого многочлена (они все различны), а затем для определения значений ошибок решаем систему уравнений относительно Y_j:

$$\begin{cases} s_1=Y_1X_1+\ldots Y_rX_r \\ s_2=Y_1X_1^2+\ldots Y_rX_r^2 \\ \qquad\ldots \\ s_r=Y_1X_1^r+\ldots Y_rX_r^r \end{cases}. \qquad (**)$$

Лемма. Система уравнений (**) разрешима однозначно, если в сообщении имеется точно r ошибок.

Доказательство. Определитель системы (**) имеет вид

$$\Delta = \begin{vmatrix} X_1 & X_2 & \dots & X_r \\ X_1^2 & X_2^2 & \dots & X_r^2 \\ \vdots & \vdots & \ddots & \vdots \\ X_1^r & X_2^r & \dots & X_r^r \end{vmatrix} = X_1 \cdot \ldots \cdot X_r \begin{vmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_r \\ \vdots & \vdots & \ddots & \vdots \\ X_1^{r-1} & X_2^{r-1} & \dots & X_r^{r-1} \end{vmatrix} \neq 0,$$

если $X_1, \dots, X_r$ различны между собой.

Теперь мы можем построить алгоритм декодирования БЧХ кода:

Шаг 1. Находим правильное значение r. В качестве пробного значения r берём $r = t$ и вычисляем матрицу A. Если $|A| \neq 0$, то $r = t$. Иначе, испытываем значение $r = t - 1$, повторяя процедуру.

Шаг 2. Обращаем матрицу A для найденного значения r и находим коэффициенты многочлена локаторов ошибок: $\begin{pmatrix} \Lambda_r & \dots & \Lambda_1 \end{pmatrix}^T = A^{-1} \begin{pmatrix} -s_{r+1} & \dots & -s_{2r} \end{pmatrix}^T$.

Шаг 3. Находим корни многочлена локаторов ошибок, т.е. находим локаторы ошибок $X_1, \dots, X_r$, используя процедуру Ченя.

Шаг 4. Решаем систему уравнений (**), т.е. находим значения ошибок e_l. В случае $q = 2$ все значения $e_l = 1$: $\begin{pmatrix} Y_1 & \dots & Y_r \end{pmatrix}^T = \begin{pmatrix} X_1 & \dots & X_r \\ \vdots & \ddots & \vdots \\ X_1^r & \dots & X_r^r \end{pmatrix} \begin{pmatrix} s_1 & \dots & s_r \end{pmatrix}^T$.

Замечание. Процедура Ченя состоит в последовательном вычислении значений $\Lambda(\alpha^j)$ для каждого $j = 1, \dots, n$, и проверки равенства этих значений нулю. Для вычисления значения $\Lambda(x)$ в точке $x = \beta$ можно использовать схему Горнера:

$$\Lambda(\beta) = \left(\dots \left(\left(\left(\Lambda_r \beta + \Lambda_{r-1} \right) \beta + \Lambda_{r-2} \right) \beta + \Lambda_{r-3} \right) \beta + \dots + \Lambda_0 \right).$$

В этом случае для нахождения $\Lambda(\beta)$ понадобится r умножений и r сложений.

Пример. Рассмотрим БЧХ код длины 15 с $q = 2$, исправляющий тройные ошибки. Это $(15,5)$ – код с порождающим многочленом $g(x) = x^{10} + x^8 + x^5 + x^4 + x^3 + x^2 + x + 1$.

Пусть $v(x) = x^{11} + x^{10} + x^9 + x^6 + 1$. Сначала вычислим компоненты синдрома.

$$s_1=\alpha^{11}+\alpha^{10}+\alpha^9+\alpha^6+1=\alpha^3+\alpha^2+\alpha=\alpha^{11},$$
$$s_2=\alpha^7+\alpha^5+\alpha^3+\alpha^{12}+1=\alpha^3+\alpha+\alpha=\alpha^7,$$
$$s_3=\alpha^3+1+\alpha^{12}+\alpha^3+1=\alpha^{12},$$
$$s_4=\alpha^{14}+\alpha^{10}+\alpha^6+\alpha^9+1=\alpha^3+1=\alpha^{14},$$
$$s_5=\alpha^{10}+\alpha^5+1+1+1=0,$$
$$s_6=\alpha^6+1+\alpha^9+\alpha^6+1=\alpha^9.$$

При $r=3$ $\begin{vmatrix} s_1 & s_2 & s_3 \\ s_2 & s_3 & s_4 \\ s_3 & s_4 & s_5 \end{vmatrix}=\begin{vmatrix} \alpha^{11} & \alpha^7 & \alpha^{12} \\ \alpha^7 & \alpha^{12} & \alpha^{14} \\ \alpha^{12} & \alpha^{14} & 0 \end{vmatrix}=\alpha^{7+7+12}\begin{vmatrix} \alpha^4 & 1 & \alpha^5 \\ 1 & \alpha^5 & \alpha^7 \\ \alpha^5 & \alpha^7 & 0 \end{vmatrix}=$

$$=\alpha^{26}\left(\alpha^7+\alpha^7+\alpha^{10}+\alpha^{13}\right)=\alpha^{26}\left(\alpha^3+\alpha\right)=\alpha^{28}\alpha^9=\alpha^5.$$

Следовательно, $r=3$, $|A|=\alpha^5$.

$$A^{-1}=\alpha^{10}\begin{pmatrix} \alpha^{13} & \alpha^{11} & \alpha^5 \\ \alpha^{11} & \alpha^9 & \alpha^2 \\ \alpha^5 & \alpha^2 & \alpha^6 \end{pmatrix}=\begin{pmatrix} \alpha^8 & \alpha^6 & 1 \\ \alpha^6 & \alpha^4 & \alpha^{12} \\ 1 & \alpha^{12} & \alpha \end{pmatrix}.$$

Поэтому $\begin{pmatrix} \Lambda_3 \\ \Lambda_2 \\ \Lambda_1 \end{pmatrix}=\begin{pmatrix} \alpha^8 & \alpha^6 & 1 \\ \alpha^6 & \alpha^4 & \alpha^{12} \\ 1 & \alpha^{12} & \alpha \end{pmatrix}\begin{pmatrix} \alpha^{14} \\ 0 \\ \alpha^9 \end{pmatrix}=\begin{pmatrix} 1 \\ \alpha^9 \\ \alpha^{11} \end{pmatrix}.$

Многочлен локаторов ошибок: $\Lambda(x)=x^3+\alpha^9x^2+\alpha^{11}x+1$.

$$\Lambda(\alpha)=\left(\left(\left(\alpha+\alpha^9\right)\alpha+\alpha^{11}\right)\alpha+1\right)=\alpha^{12}+\alpha^{11}+\alpha^3+1\neq 0,$$
$$\Lambda(\alpha^2)=\left(\left(\left(\alpha^2+\alpha^9\right)\alpha^2+\alpha^{11}\right)\alpha^2+1\right)=\alpha^{13}+\alpha^{13}+\alpha^6+1\neq 0,$$
$$\Lambda(\alpha^3)=\left(\left(\left(\alpha^3+\alpha^9\right)\alpha^3+\alpha^{11}\right)\alpha^3+1\right)=\alpha^0+\alpha^{14}+\alpha^9+1\neq 0,$$
$$\Lambda(\alpha^4)=\left(\left(\left(\alpha^4+\alpha^9\right)\alpha^4+\alpha^{11}\right)\alpha^4+1\right)=\alpha^2+\alpha^{15}+\alpha^{12}+1\neq 0,$$
$$\Lambda(\alpha^5)=\left(\left(\left(\alpha^5+\alpha^9\right)\alpha^5+\alpha^{11}\right)\alpha^5+1\right)=\alpha^4+\alpha+1+1\neq 0,$$
$$\Lambda(\alpha^6)=\left(\left(\left(\alpha^6+\alpha^9\right)\alpha^6+\alpha^{11}\right)\alpha^6+1\right)=\alpha^6+\alpha^2+\alpha^3+1\neq 0,$$
$$\Lambda(\alpha^7)=\left(\left(\left(\alpha^7+\alpha^9\right)\alpha^7+\alpha^{11}\right)\alpha^7+1\right)=\alpha^8+\alpha^3+\alpha^6+1=0,$$
$$\Lambda(\alpha^8)=\left(\left(\left(\alpha^8+\alpha^9\right)\alpha^8+\alpha^{11}\right)\alpha^8+1\right)=\alpha^{10}+\alpha^4+\alpha^9+1\neq 0,$$
$$\Lambda(\alpha^9)=\left(\left(\left(\alpha^9+\alpha^9\right)\alpha^9+\alpha^{11}\right)\alpha^9+1\right)=\alpha^{12}+\alpha^{12}+\alpha^5+1\neq 0,$$
$$\Lambda(\alpha^{10})=\left(\left(\left(\alpha^{10}+\alpha^9\right)\alpha^{10}+\alpha^{11}\right)\alpha^{10}+1\right)=\alpha^{14}+\alpha^6+1+1\neq 0,$$
$$\Lambda(\alpha^{11})=\left(\left(\left(\alpha^{11}+\alpha^9\right)\alpha^{11}+\alpha^{11}\right)\alpha^{11}+1\right)=\alpha^3+\alpha+\alpha^7+1=0,$$
$$\Lambda(\alpha^{12})=\left(\left(\left(\alpha^{12}+\alpha^9\right)\alpha^{12}+\alpha^{11}\right)\alpha^{12}+1\right)=\alpha^3+\alpha^8+\alpha^6+1\neq 0,$$
$$\Lambda(\alpha^{13})=\left(\left(\left(\alpha^{13}+\alpha^9\right)\alpha^{13}+\alpha^{11}\right)\alpha^{13}+1\right)=\alpha^9+\alpha^5+\alpha^9+1\neq 0,$$
$$\Lambda(\alpha^{14})=\left(\left(\left(\alpha^{14}+\alpha^9\right)\alpha^{14}+\alpha^{11}\right)\alpha^{14}+1\right)=\alpha^{12}+\alpha^7+\alpha^{10}+1\neq 0.$$

Лекция 17. **Коды Рида-Соломона.**

Рассмотрим БЧХ код, у которого $m=1$ и значит, $n=q-1$, если $\alpha-$ примитивный элемент поля $\mathbb{F}_q$. Минимальный многочлен каждого элемента $\beta\in\mathbb{F}_q$ равен $x-\beta$. В коде Рида-Соломона, исправляющем t ошибок, обычно берут $b=1$, и тогда порождающий многочлен кода имеем вид $g(x)=(x-\alpha)(x-\alpha^2)\ldots(x-\alpha^{2t})$.

Так как $\deg g(x)=2t$, то $k=n-2t$.

Иногда с помощью разумного выбора b удаётся упростить кодер. Таким образом,

$$g(x)=(x-\alpha^b)(x-\alpha^{b+1})\ldots(x-\alpha^{2t+b-1}).$$

Пример. Возьмём $q=16$, $t=2$, $b=1$, $\alpha-$ примитивный элемент поля $\mathbb{F}_{16}$. Тогда

$$g(x)=(x-\alpha)(x-\alpha^2)(x-\alpha^3)(x-\alpha^4)=x^4+\alpha^{13}x^3+\alpha^6x^2+\alpha^3x+\alpha^{10}.$$

Поскольку $\deg g(x)=4$, то $k=15-4=11$, т.е. мы имеем $(15,11)-$код, исправляющий двойные ошибки. Заметим, что информационное слово представляет собой 11 символов из $\mathbb{F}_{16}$, что эквивалентно 44 битам.

Теорема. Код Рида-Соломона имеет минимальное расстояние $n-k+1$ и является кодом с максимальным расстоянием среди всех $(n,k)-$кодов.

Доказательство. У матрицы ранга $n-k$ любые $n-k+1$ столбцов линейно зависимы, а потому $d_C\le n-k+1$. Далее, пусть $d=2t+1-$ конструктивное расстояние кода. Тогда $d_C\ge d=2t+1=n-k+1$. Следовательно, $d_C=n-k+1=d$.

Замечание. Значение $n-k+1$ называется границей Синглтона для минимального расстояния $(n,k)-$кода. Из доказанной теоремы следует, что если для пары (n,k) существует код Рида-Соломона, то он наиболее оптимален (исправляет больше ошибок). Но не для всякой (n,k) пары существуют коды Рида-Соломона.

Лекция 18. **Двоичный код Галлея.**

Рассмотрим совершенный код длины n над $\mathbb{F}_q$ с минимальным расстоянием $d_C=2t+1$. Мы уже видели, что при $t=1$ обобщённый код Хэмминга является совершенным и исправляет одиночные ошибки. В 1949 г. Галлей показал, что при $q=2$ существует совершенный код, исправляющий тройные ошибки. Будем строить этот код. Для совершенного кода над $\mathbb{F}_q$,

который исправляет t ошибок, сферы радиуса t с центрами в кодовых словах покрывают всё пространство $\mathbb{F}_q^n$ без пересечений. Поэтому $\left|C\right|\sum_{i=1}^{t} C_n^i (q-1)^i = q^n$.

Так как $\left|C\right| = q^k$, где $k-$ число информационных символов кода, то $\sum_{i=1}^{t} C_n^i (q-1)^i = q^{n-k}$.

В бинарном случае (т.е. при $q=2$) для некоторого $l = n-k$ мы должны иметь

$$\sum_{i=0}^{t} C_n^i = 2^l. \tag{1}$$

Попытаемся описать все бинарные совершенные коды с $t=3$. Имеем

$$\sum_{i=0}^{3} C_n^i = 1+n+\frac{n(n-1)}{2!}+\frac{n(n-1)(n-2)}{3!} = \frac{6+5n+n^3}{6} = \frac{(n+1)\left(n^2-n+6\right)}{6}.$$

Поэтому из (1) мы должны иметь

$$(n+1)\left[(n+1)^2-3(n+1)+8\right] = 3\cdot 2^{l+1} \tag{2}$$

Если $(n+1)\vdots 16$, то содержимое в квадратных скобках делится на 8, но не делится на 16, а значит, это содержимое может равняться либо 8, либо 24.

Если $\left(n^2-n+6\right) = (n+1)^2-3(n+1)+8 = 8$, то $n=-1$ или $n=2$, что невозможно, так как $n \geq 2t+1 = 7$. Если $\left(n^2-n+6\right) = 24$, то $n \notin \mathbb{N}$. Поэтому заключаем, что $n+1$ является делителем 24. А потому $n \in \{7,11,23\}$.

Но $n=11$ не удовлетворяет равенству (2). При $n=7$ мы имеем код $C = \{(0,0,0,0,0,0,0),(1,1,1,1,1,1,1)\}$ (это единственные кодовые слова длины 7 с весом 0 или 7).

Покажем, что существует совершенный код длины 23 (пока у нас выполнено только необходимое условие (1)).

Рассмотрим $(7,4)-$ код Хэмминга с матрицей $H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$.

Ему эквивалентен код C с матрицей $\tilde{H} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$.

Коду C принадлежат $0 = (0,0,0,0,0,0,0)$ и семь циклических сдвигов слова $(1,1,0,1,0,0,0)$. Кроме того, $1 = (1,1,1,1,1,1,1) \in C$ а также семь циклических сдвигов слова $(1,1,1,0,0,1,0)$. Этими словами исчерпывается весь код C, ибо $\left|C\right| = 2^{7-3} = 16$.

Обозначим через $\overline{C}$ код, полученный из C добавлением знака проверки на чётность $\left(c_n = \sum_{i=0}^{n-1} c_i\right)$. Теперь легко проверить, что $\overline{C}$ и $\overline{C}^*-$ коды, и они обладают свойством $\overline{C} \cap \overline{C}^* = \{0,1\}$. Минимальное расстояние этих кодов равно 4 и оба кода самодуальны (проверяется непосредственно).

Образуем новый код $\tilde{C}$ длины 24: $\tilde{C} \overset{def}{=} \left\{(a+x, b+x, a+b+x) \middle| a,b \in \overline{C}, x \in \overline{C}^*\right\}$.

То, что $\tilde{C}-$ линейное пространство над $\mathbb{F}_2$, следует из того, что сумма двух слов из $\tilde{C}$ есть слово из $\tilde{C}$. А остальные свойства линейного пространства очевидны.

Далее, если a,b пробегают базис $\overline{C}$, $x-$ базис $\overline{C}^*$, то слова $(a,0,a)$, $(0,b,b)$, (x,x,x) образуют базис $\tilde{C}$. Таким образом, мы показали, что $\tilde{C}$ есть $(24,12)-$код.

Любые два вектора из $\tilde{C}$ ортогональны, значит $\tilde{C}-$ самодуальный код. Кроме того, все базисные слова кодов $\overline{C}$ и $\overline{C}^*$ имеют вес, делящийся на 4. Отсюда все слова кода $\tilde{C}$ имеют вес, делящийся на 4. Предположим, что минимальный вес кода $\tilde{C}$ меньше восьми, и пусть $w(c) < 8$. Пусть $c = (a+x, b+x, a+b+x)$. Так как векторы $a+x$, $b+x$, $a+b+x$ имеют чётный вес, то хотя бы один из этих векторов равен нулю. Если это $a+x$ или $b+x$, то $w(c) < 8$ только если $x = 0$ или $x = 1$. Не ограничивая общности можем считать, что $x = 0$. Но так как вес векторов a, b, $a+b$ кратен 4, то равенство $w(c) = 4$ невозможно, значит $w(c) = 8$, т.е. $d_{\tilde{C}} = 8$. Если теперь в коде $\tilde{C}$ выбросить последнюю координату (а она является проверкой на чётность), то получаем $(23,12)-$код с минимальным расстоянием 7. Так как для построенного кода выполняется равенство (1), то мы имеем совершенный бинарный $(23,12)-$код, исправляющий тройные ошибки. Он называется бинарным кодом Галлея. Ниже мы покажем, что этот код является циклическим.

Лекция 19. **Квадратично-вычетные коды.**

Определение. Бинарный циклический код называется квадратично-вычетным, если его длина равна простому числу p, причём $p \mid (2^m - 1)$ для некоторого натурального m, а корнями порождающего многочлена являются все элементы $\alpha^j \in \mathbb{F}_{2^m}$, где $\alpha-$ примитивный элемент поля $\mathbb{F}_{2^m}$, а j пробегает все квадратичные вычеты по модулю p, $0 < j < p$.

Рассмотрим многочлены

$$g(x) = x^{11} + x^{10} + x^6 + x^5 + x^4 + x^2 + 1,$$
$$\tilde{g}(x) = x^{11} + x^9 + x^7 + x^6 + x^5 + x + 1.$$

Поскольку $\tilde{g}(x) = x^{11} g\left(\frac{1}{x}\right)$, то $\tilde{g}(x)-$ возвратный многочлен для $g(x)$. Непосредственной проверкой убеждаемся, что $x^{23} - 1 = (x-1) g(x) \tilde{g}(x)$.

Эти многочлены порождают $(23,12)-$коды, которые очевидно эквивалентны. На самом деле эти коды совпадают с построенным кодом Галлея.

Покажем, что код Галлея является квадратично-вычетным. Действительно, построим неприводимый многочлен, порождающий поле $\mathbb{F}_{2^{11}}$, в котором содержится $2^{11} = 2048$ элементов. Мы имеем $2047 = 23 \cdot 89$. Пусть $\alpha-$ примитивный элемент поля $\mathbb{F}_{2^{11}}$, т.е. порядок α равен 2047. Обозначим $\beta = \alpha^{89}$, $\beta^{-1} = \alpha^{-89}$. Тогда $\beta \in \mathbb{F}_{2^{11}}$ имеет порядок 23.

Пусть B и B^{-1} означают соответственно множества элементов поля $\mathbb{F}_{2^{11}}$, сопряжённые над $\mathbb{F}_2$ соответственно с β и β^{-1}:

$$B = \left\{\beta, \beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32} = \beta^9, \beta^{64} = \beta^{18}, \beta^{128} = \beta^{13}, \beta^{256} = \beta^3, \beta^{512} = \beta^6, \beta^{1024} = \beta^{12}\right\},$$
$$B^{-1} = \left\{\beta^{-1}, \beta^{-2}, \beta^{-4}, \beta^{-8}, \beta^{-16}, \beta^{-9}, \beta^{-18}, \beta^{-13}, \beta^{-3}, \beta^{-6}, \beta^{-12}\right\}.$$

Тогда минимальные многочлены для β и β^{-1} соответственно равны

$$f_\beta(x) = \prod_{k=0}^{10} \left(x - \beta^{2^k}\right), \qquad f_{\beta^{-1}}(x) = \prod_{k=0}^{10} \left(x - \beta^{-2^k}\right).$$

Множества B и B^{-1} не пересекаются. Их элементы имеют порядок 23. Значит,

$$x^{23} - 1 = (x-1) f_\beta(x) f_{\beta^{-1}}(x).$$

Так как $f_\beta(x)$ и $f_{\beta^{-1}}(x)$ неприводимы, то в силу единственности разложения на неприводимые, получаем – $f_\beta(x) = g(x)$, $f_{\beta^{-1}}(x) = \tilde{g}(x)$ или наоборот.

Таким образом, мы доказали, что корнями $\tilde{g}(x)$ (или $g(x)$) являются элементы $\alpha^{-89 \cdot 2^k}$, $k = 0,1,\ldots,10$. Однако,

$$\left(\frac{-89 \cdot 2^k}{23}\right) = -\left(\frac{89}{23}\right) \cdot \left(\frac{2}{23}\right)^k = -(-1)^{\frac{23^2-1}{8}} \left(\frac{20}{23}\right) = -\left(\frac{5}{23}\right) = -(-1)^{\frac{5-1}{2} \cdot \frac{23-1}{2}} \left(\frac{3}{5}\right) = 1,$$

т.е. циклический код, порожденный $f_{\beta^{-1}}(x)$ является квадратично-вычетным. Значит, код Галлея является квадратично-вычетным.

Замечание 1. Для многочлена $f_\beta(x)$ его корни $\beta^{2^k}=\alpha^{89\cdot 2^k}=\alpha^{j_k}$ таковы, что j_k- квадратичные невычеты по модулю p. Но порождаемый им код эквивалентен коду Голея, а потому также является квадратично-вычетным.

Замечание 2. Пусть $p-$ простое, $p\left|\left(2^m-1\right)\right.$ и $\alpha-$ примитивный элемент поля $\mathbb{F}_{2^m}$. Положим $g(x)=\prod_{j-\text{êâ. âû÷åò}(\bmod p)}\left(x-\alpha^j\right)$. Но многочлен $g(x)$ порождает квадратично-выченый код только в том случае, когда коэффициенты $g(x)$ принадлежат $\mathbb{F}_2$.

Замечание 3. Если даже многочлен $g(x)$ из предыдущего замечания имеет коэффициенты из $\mathbb{F}_2$, то важно знать минимальное расстояние соответствующего кода. Сделать это бывает трудно. Часто квадратично-выченые $(n,k)-$коды обладают большим минимальным расстоянием, что является его достоинством.

В следующей таблице указаны квадратично-выченые коды, для которых вычислено d_C:

n	k	d_C
7	4	3
17	9	5
23	12	7
31	16	7
41	21	9
47	24	11
71	36	11

n	k	d_C
73	37	13
79	40	15
89	45	17
97	49	15
103	52	19
127	64	19
151	76	19

(Подчёркнутым выделены $(n,k)-$коды с максимально возможным значением d_C).

Теорема. Минимальное расстояние квадратично-вычетного кода длины p не меньше $\sqrt{p}$.

Доказательство. Пусть $s-$ квадратичный невычет по модулю p. Тогда каждый элемент поля $\mathbb{F}_p$ можно записать в виде $j\cdot s$, где $j\in\mathbb{F}_p$, причём $j\cdot s-$ квадратичный вычет $\Leftrightarrow$ $j-$ квадратичный невычет.

Пусть $c(x)-$ кодовый многочлен, соответствующий кодовому слову с минимальным весом d_C. Тогда $c(x)=a(x)g(x)$.

Определим $\tilde{c}(x)=c\left(x^s\right)\left(\bmod\left(x^p-1\right)\right)$. Имеем $\tilde{c}(x)=a\left(x^s\right)g\left(x^s\right)\left(\bmod\left(x^p-1\right)\right)$, причём вес $\tilde{c}(x)$ самое большее d_C.

Но если $j-$ квадратичный вычет, то $g\left(\alpha^{js}\right)\neq 0$, т.к. $j\cdot s-$ квадратичный невычет. И наоборот, $g\left(\alpha^{js}\right)=0$ если $j-$ квадратичный невычет. Таким образом, $g\left(x^{s}\right)=\tilde{g}\left(x\right)\left(\mathrm{mod}\left(x^{p}-1\right)\right)$. Следовательно, $c\left(x\right)\tilde{c}\left(x\right)\left(\mathrm{mod}\left(x^{p}-1\right)\right)$ делится как на $g\left(x\right)$, так и на $\tilde{g}\left(x\right)$, а потому кратен $x^{p-1}+x^{p-2}+\ldots+x+1$. Но его степень не превосходит p, откуда $c\left(x\right)\tilde{c}\left(x\right)=x^{p-1}+x^{p-2}+\ldots+x+1\left(\mathrm{mod}\left(x^{p}-1\right)\right)$. Вес многочлена справа равен $p-1$, а вес каждого многочлена справа не превосходит d_{C}. Значит, вес правой части не превосходит d_{C}^{2}. Следовательно, $d_{C}\geq\sqrt{p}$.

Замечание. Если 2– квадратичный вычет $\mathrm{mod}\,p$, то для всех $k=0,1,2,\ldots$ имеем $2^{k}-$ квадратичный вычет $\mathrm{mod}\,p$. Поэтому, если $j-$ квадратичный вычет $\mathrm{mod}\,p$, то $2^{k}j-$ квадратичные вычеты $\mathrm{mod}\,p$. Так что α^{j} и все сопряжённые с ним являются корнями порождающего многочлена квадратично-вычетного кода. Таким образом, квадратично-вычетные коды существуют только для тех p, для которых 2 является квадратичным вычетом $\mathrm{mod}\,p$, т.е. для простых вида $8k\pm 1$.