

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Локтионова Оксана Геннадьевна
Должность: проректор по учебной работе
Дата подписания: 03.02.2021 18:04:28
Уникальный программный ключ:
0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра космического приборостроения и систем связи

УТВЕРЖДАЮ
Проректор по учебной работе
О.Г. Локтионова
«22» 07 2019 г.



ТЕОРИЯ ИНФОРМАЦИИ И КОДИРОВАНИЯ

Методические указания
по выполнению практических работ
для студентов, обучающихся по специальности
10.05.02 «Информационная безопасность
телекоммуникационных систем»
по дисциплине «Теория информации и кодирования»

Курск 2019

УДК 621.391

Составители: Д.С. Коптев, И.Г. Бабанин

Рецензент:

Доктор технических наук, старший научный сотрудник,
профессор кафедры космического приборостроения и систем связи
В.Г. Андронов

Теория информации и кодирования: методические указания по выполнению практических работ по дисциплине «Теория информации и кодирования» / Юго-Зап. гос. ун-т; сост.: Д.С. Коптев, И.Г. Бабанин. – Курск, 2019. – 49 с.: иллюстр. 7., табл. 12. – Библиогр.: с. 49.

Методические указания по выполнению практических работ содержат краткие теоретические сведения об о природе информации, её свойствах, методах измерения её количества и качества, общих принципах кодирования информации в системах передачи, обработки и хранения, варианты заданий для выполнения работ, а также перечень вопросов для самопроверки изучаемого материала.

Методические указания соответствуют учебному плану по специальности 10.05.02 «Информационная безопасность телекоммуникационных систем», а также рабочей программе дисциплины: «Теория информации и кодирования».

Предназначены для студентов, обучающихся по специальности 10.05.02 «Информационная безопасность телекоммуникационных систем», очной формы обучения.

Текст печатается в авторской редакции

Подписано в печать *22.07.19*. Формат 60x84 1/16.
Усл. печ. л. 2,848. Уч.-изд. л. 2,578. Тираж 100 экз. Заказ. *548* Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

Практическая работа №1 «Вероятностный подход к определению количества информации»

1 Цель работы

Ознакомиться с вероятностным подходом к определению количества информации. Проверить полученные теоретические сведения практическим путем, выполняя индивидуальные задания.

2 Краткие теоретические сведения

Понятие "количества информации" часто встречается в технической литературе, однако на практике определить количество информации очень непросто.

С 40-х годов XX века предпринимаются попытки использовать понятие информации для объяснения и описания самых разнообразных явлений и процессов. В решении этой проблемы существуют два основных подхода, которые исторически возникли одновременно. В конце 40-х годов XX века один из основоположников кибернетики американский математик Клод Шеннон развил вероятностный подход к измерению количества информации, а работы по созданию ЭВМ привели к «объемному» подходу.

Рассмотрим более подробно вероятностный подход.

В качестве примера разберем опыт, связанный с бросанием правильной игральной кости, имеющей N граней. Результатом данного опыта считаем выпадение грани с одним из следующих знаков: $1, 2, \dots, N$.

Введем в рассмотрение численную величину — **энтропию** (обозначим ее H). Энтропия - мера неопределенности некоторого опыта. В простейшем случае его исход зависит от выбора одного элемента из множества исходных.

Согласно шенноновской теории информации, в случае равновероятного выпадения каждой из граней величины N и H связаны между собой **формулой Хартли**

$$H = \log_2 N. \quad (1)$$

При введении какой-либо новой величины важным является вопрос о единице измерения этой величины.

В соответствии с формулой Хартли становится очевидным, что энтропия H будет равна единице при $N=2$. Тогда, в качестве единицы измерения принимается количество информации, связанное с проведением опыта, состоящего в получении одного из двух равновероятных исходов (примером такого опыта может служить бросание монеты, при котором возможны два исхода- «орел», «решка»). Такая единица количества информации называется «бит».

В случае, когда вероятности P_i результатов опыта (в примере, приведенном выше — бросания игральной кости) неодинаковы, имеет место формула Шеннона

$$H = -\sum_{i=1}^N P_i \cdot \log_2 P_i. \quad (2)$$

В случае равной вероятности событий $P_i = \frac{1}{N}$, формула Шеннона переходит в формулу Хартли (1).

Рассмотрим следующий пример. Необходимо определить количество информации, связанное с появлением каждого символа в сообщениях, записанные на русском языке. Считаем, что русский алфавит состоит из 33 букв и знака «пробел» для разделения слов.

По формуле Хартли

$$H = \log_2 34 \approx 5,09 \text{ бит.}$$

Однако в словах русского языка (равно как и в словах других языков) различные буквы встречаются неодинаково часто. В таблице 1 приведена вероятность частоты употребления различных знаков русского алфавита, полученная на основе анализа очень больших по объему текстов.

Таблица 1 - Частотность букв русского языка

i	Символ	$P(i)$	i	Символ	$P(i)$	i	Символ	$P(i)$
1	—	0,175	12	Л	0,035	23	Б	0,014
2	О	0,090	13	К	0,028	24	Г	0,012
3	Е	0,072	14	М	0,026	25	Ч	0,012
4	Ё	0,072	15	Д	0,025	26	Й	0,010

5	А	0,062	16	П	0,023	27	Х	0,009
6	И	0,062	17	У	0,021	28	Ж	0,007
7	Т	0,053	18	Я	0,018	29	Ю	0,006
8	Н	0,053	19	Ы	0,016	30	Ш	0,006
9	С	0,045	20	З	0,016	31	Ц	0,004
10	Р	0,040	21	Ь	0,014	32	Щ	0,003
11	В	0,038	22	Ъ	0,014	33	Э	0,003
						34	Ф	0,002

Найдем значение количества информации (энтропии) H . Для этого воспользуемся формулой Шеннона:

$$H = -\sum_{i=1}^{34} P_i \cdot \log_2 P_i \approx 4,72 \text{ бит.}$$

Полученное значение количества информации по Шеннону меньше вычисленного ранее. Это вытекает из основных свойств энтропии, как меры неопределенности сообщения. Количество информации, вычисляемое по формуле Хартли, является максимальным количеством информации, которое могло бы проходиться на один знак.

Аналогичные подсчеты количества информации можно провести и для других языков, например, использующих латинский алфавит — немецкий, французский и др. (26 различных букв и «пробел»).

По формуле Хартли получим

$$H = \log_2 27 \approx 4,76 \text{ бит.}$$

3 Пример выполнения задания

3.1 Задание

Подсчитать количество информации, приходящейся на один символ, в следующем тексте экономического содержания:

Организационно-правовые формы предприятий в своей основе определяют форму их собственности, то есть, кому принадлежит предприятие, его основные фонды, оборотные средства, материальные и денежные ресурсы. В зависимости от формы собственности в России в настоящее время различают три основные формы предпринимательской деятельности: частную, коллективную и контрактную.

3.2 Решение.

Для выполнения поставленного задания воспользуемся текстовым редактором **Microsoft Word** и табличным процессором **Microsoft Excel**.

В качестве первого шага наберем заданный текст в текстовом редакторе **Microsoft Word**.

Выполним подсчет всех символов текста. Для этого воспользуемся функцией статистики подсчета числа знаков в документе. Перейдем на вкладке **Рецензирование** в раздел **Правописание**, выделим набранный текст и нажмем кнопку «статистика» (рис. 1)

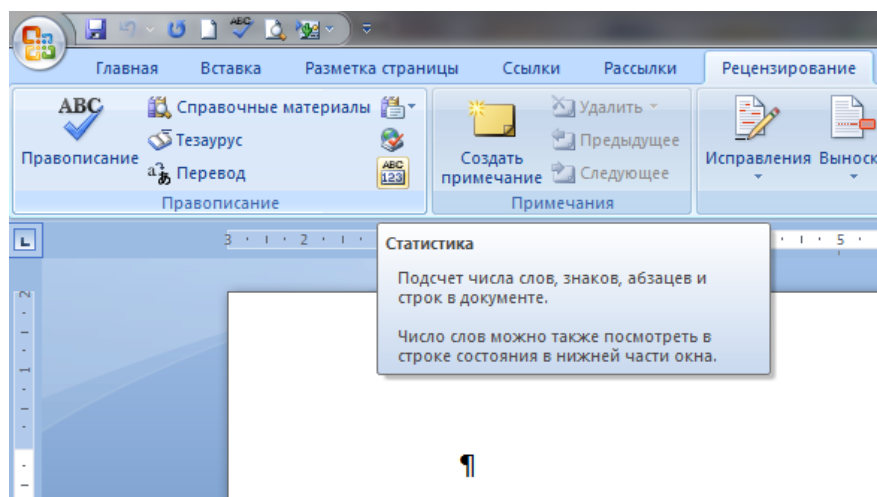


Рисунок 1 - Вид окна **Microsoft Word**

В результате появиться окно, представленное на рисунке 2.

Организационно-правовые формы предприятий в своей основе определяют форму их собственности, то есть, кому принадлежит предприятие, его основные фонды, оборотные средства, материальные и денежные ресурсы. В зависимости от формы собственности в России в настоящее время различают три основные формы предпринимательской деятельности: частную, коллективную и контрактную.

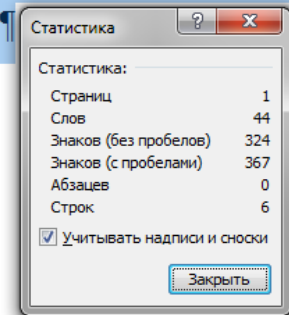


Рисунок 2 - Статистика подсчета числа знаков в документе

Таким образом, количество знаков в тексте вместе с пробелами – 367, число пробелов – 43 ($367 - 324 = 43$). Подсчитаем количество каждого символа в тексте и занесем в таблицу 2.

Определим **вероятность P_i каждого символа в тексте** как отношение количества одинаковых символов каждого значения ко всему числу символов в тексте. Вычисления в таблице будем производить в текстовом редакторе Word с помощью добавления в ячейку таблицы формулы выполнения простого расчета.

Каждая ячейка таблицы **Word** имеет адрес, состоящий из номера столбца (обозначаемого буквами латинского алфавита) и номера строки (обозначаемого арабскими цифрами). Для вычисления P_i необходимо поместить курсор в ячейку D2, где будет помещен результат.

Таблица 2 - Частотность букв в тексте

i	Сим-вол	Ко-л-во	$P(i)$	i	Сим-вол	Ко-л-во	$P(i)$	i	Сим-вол	Ко-л-во	$P(i)$
1	_	43	0,117	14	Р	24	0,065	27	Б	3	0,008
2	,	6	0,016	15	В	16	0,044	28	Г	2	0,005
3	.	2	0,005	16	Л	8	0,022	29	Ч	2	0,005
4	-	1	0,003	17	К	6	0,016	30	Й	3	0,008
5	:	1	0,003	18	М	9	0,025	31	Х	1	0,003

6	О	35	0,095	19	Д	9	0,025	32	Ж	2	0,005
7	Е	32	0,087	20	П	9	0,025	33	Ю	5	0,014
8	Ё	0	0,000	21	У	6	0,016	34	Ш	0	0,000
9	А	14	0,038	22	Я	5	0,014	35	Ц	1	0,003
10	И	25	0,068	23	Ы	11	0,030	36	Щ	1	0,003
11	Т	25	0,068	24	З	3	0,008	37	Э	0	0,000
12	Н	25	0,068	25	Ь	4	0,011	38	Ф	5	0,014
13	С	23	0,063	26	Ъ	0	0,000				
Σ		232		Σ		110		Σ		25	
								Σ		367	

На вкладке **Макет** в разделе **Данные** выбрать кнопку формула. Окно **Формула** заполнить как показано на рисунке 3.

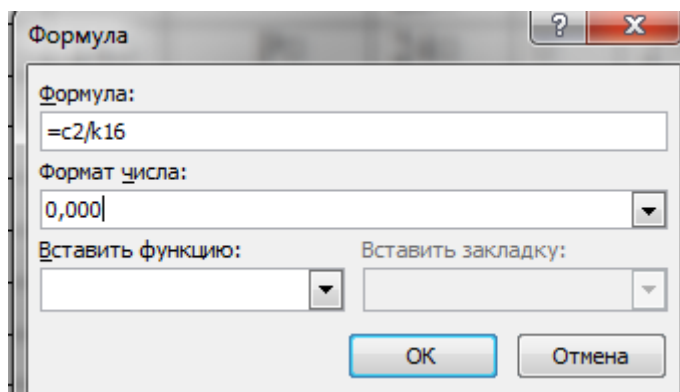


Рисунок 3 - Вставка формул в таблицу *Microsoft Word*

Аналогично вычислим остальные значения P_i каждого символа. Результаты вычислений представлены в таблице 2.

В предпоследней строке таблицы 2 знаком " Σ " обозначено суммарное количество знаков, а в последней строке знаком " Σ " обозначена общее количество всех знаков текста.

Далее по формуле Шеннона подсчитаем количество информации, приходящейся на один символ. Для этого выполним предварительные вычисления в табличном процессоре *Excel*.

Откроем лист *Excel* и заполним его как показано на рисунке 4.

	A	B	C	D	E	F	G	H	I	J	K	L
1	i	P(i)	$\log_2 P_i$	$P_i \times \log_2 P_i$	i	P(i)	$\log_2 P_i$	$P_i \times \log_2 P_i$	i	P(i)	$\log_2 P_i$	$P_i \times \log_2 P_i$
2	1	0,117	=LOG(B2;2)	=B2*C2	14	0,065	=LOG(F2;2)	=F2*G2	27	0,008	=LOG(J2;2)	=J2*K2
3	2	0,016	=LOG(B3;2)	=B3*C3	15	0,044	=LOG(F3;2)	=F3*G3	28	0,005	=LOG(J3;2)	=J3*K3
4	3	0,005	=LOG(B4;2)	=B4*C4	16	0,022	=LOG(F4;2)	=F4*G4	29	0,005	=LOG(J4;2)	=J4*K4
5	4	0,003	=LOG(B5;2)	=B5*C5	17	0,016	=LOG(F5;2)	=F5*G5	30	0,008	=LOG(J5;2)	=J5*K5
6	5	0,003	=LOG(B6;2)	=B6*C6	18	0,025	=LOG(F6;2)	=F6*G6	31	0,003	=LOG(J6;2)	=J6*K6
7	6	0,095	=LOG(B7;2)	=B7*C7	19	0,025	=LOG(F7;2)	=F7*G7	32	0,005	=LOG(J7;2)	=J7*K7
8	7	0,087	=LOG(B8;2)	=B8*C8	20	0,025	=LOG(F8;2)	=F8*G8	33	0,014	=LOG(J8;2)	=J8*K8
9	8	0		0	21	0,016	=LOG(F9;2)	=F9*G9	34	0		0
10	9	0,038	=LOG(B10;2)	=B10*C10	22	0,014	=LOG(F10;2)	=F10*G10	35	0,003	=LOG(J10;2)	=J10*K10
11	10	0,068	=LOG(B11;2)	=B11*C11	23	0,03	=LOG(F11;2)	=F11*G11	36	0,003	=LOG(J11;2)	=J11*K11
12	11	0,068	=LOG(B12;2)	=B12*C12	24	0,008	=LOG(F12;2)	=F12*G12	37	0		0
13	12	0,068	=LOG(B13;2)	=B13*C13	25	0,011	=LOG(F13;2)	=F13*G13	38	0,014	=LOG(J13;2)	=J13*K13
14	13	0,063	=LOG(B14;2)	=B14*C14	26	0		0	39			0
15	Σ	=СУММ(B2:B14)		=СУММ(D2:D14)	Σ	=СУММ(F2:F14)		=СУММ(H2:H14)	Σ	=СУММ(J2:J14)		=СУММ(L2:L14)
16									Σ	=B15+F15+J15		
17												
18												
19					H=	=-(D15+H15+L15)						
20												

Рисунок 4 – Заполнение формулами листа книги *Microsoft Excel*

При заполнении листа следует воспользоваться встроенной функцией логарифма числа по основанию 2. Для этого на вкладке **Формулы** выбираем **Математические** и щелкаем на этой кнопке. В появившемся выпадающем меню выбираем функцию LOG. Появится окно, представленное на рисунке 5. В поле число записываем адрес ячейки, в которой находится число, логарифм которого вычисляется, а в поле основание записываем "2", так как нам необходимо вычислить логарифм по основанию 2.

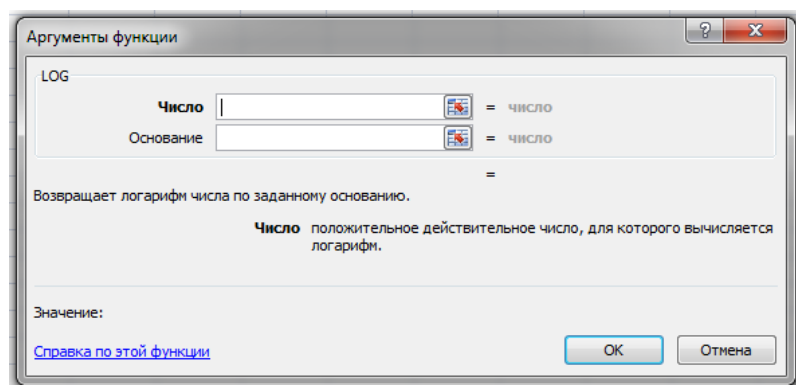


Рисунок 5 - Окно вычисления логарифма в *Microsoft Excel*

Обратим внимание на то, что в ячейках C9, G14, K9 и K12 значения логарифма не вычисляются, так как количество соответствующих символов равно нулю, а следовательно и их частота в тексте равна нулю.

В строке с номером I5 в столбцах B, F, J вычислены суммарные вероятности по столбцу. Это необходимо сделать для контроля. В строке с номером 16 в столбце J вычислено контрольное значение вероятности.

Это значение должно быть равно единице. В ячейке F19 вычисляется количество информации по формуле Шеннона (числовые значения, полученные в Excel, приведены в таблице 3).

$$P(i) \cdot \log_2 P(i)$$

Таким образом, количество информации согласно формуле Шеннона, приходящейся на один символ, в данном тексте:

$$H=4.40199 \approx 4,40 \text{ бита.}$$

Максимальное количество информации, которое могло бы приходиться на один знак в данном тексте, вычисляемое по формуле Хартли:

$$H=\log_2 367 \approx 8,5196 \text{ бит.}$$

Таблица 3

i	$P(i)$	$\log_2 P(i)$	$P(i) \cdot \log_2 P(i)$	i	$P(i)$	$\log_2 P(i)$	$P(i) \cdot \log_2 P(i)$	i	$P(i)$	$\log_2 P(i)$	$P(i) \cdot \log_2 P(i)$
1	0,117	-3,0954	-0,3622	14	0,065	-3,9434	-0,2563	27	0,008	-6,9658	-0,0557
2	0,016	-5,9658	-0,0955	15	0,044	-4,5064	-0,1983	28	0,005	-7,6439	-0,0382
3	0,005	-7,6439	-0,0382	16	0,022	-5,5064	-0,1211	29	0,005	-7,6439	-0,0382
4	0,003	-8,3808	-0,0251	17	0,016	-5,9658	-0,0955	30	0,008	-6,9658	-0,0557
5	0,003	-8,3808	-0,0251	18	0,025	-5,3219	-0,1330	31	0,003	-8,3808	-0,0251
6	0,095	-3,3959	-0,3226	19	0,025	-5,3219	-0,1330	32	0,005	-7,6439	-0,0382
7	0,087	-3,5228	-0,3065	20	0,025	-5,3219	-0,1330	33	0,014	-6,1584	-0,0862
8	0			21	0,016	-5,9658	-0,0955	34	0		
9	0,038	-4,7179	-0,1793	22	0,014	-6,1584	-0,0862	35	0,003	-8,3808	-0,0251
10	0,068	-3,8783	-0,2637	23	0,03	-5,0589	-0,1518	36	0,003	-8,3808	-0,0251
11	0,068	-3,8783	-0,2637	24	0,008	-6,9658	-0,0557	37	0		
12	0,068	-3,8783	-0,2637	25	0,011	-6,5064	-0,0716	38	0,014	-6,1584	-0,0862
13	0,063	-3,9885	-0,2513	26	0						
Σ	0,631		-2,3970	Σ	0,301		-1,5311	Σ	0,068		-0,4740
								Σ	1		-4,4020

4 Задание для выполнения

Подсчитать количество информации, находящейся на один символ по методике, изложенной в настоящих методических указаниях, в стихотворении или песне, содержащей от 320 до 380 символов. У каждого студента текст должен быть свой. Выбирать текст только на русском языке.

5 Содержание отчета

По результатам выполненной практической работы представляется отчет, в котором должны содержаться следующие пункты:

- цель работы;
- индивидуальное задание;
- порядок выполнения работы;
- таблицы, выполненные в Excel, содержащие формулы, по которым производятся вычисления;
- таблицы, выполненные в Excel, содержащие результаты расчетов;
- выводы по результатам исследований с анализом полученных результатов.

Практическая работа №2
«Статистическое моделирование случайных событий
и дискретных случайных величин»

1 Цель работы

- 1) уяснение законов алгебры событий;
- 2) знакомство с видами единичного жребия, применяемыми в методе Монте-Карло для разыгрывания случайных событий и случайных величин;
- 3) изучение некоторых функций Excel и Mathcad.

2 Краткие теоретические сведения

2.1 Алгебра событий

Суммой событий A и B называется событие $S = A + B$, которое состоит в наступлении хотя бы одного из них.

Произведением событий A и B называется событие $D = AB$, состоящее в их совместном появлении.

Определения суммы и произведения событий распространяются на любое (конечное) число слагаемых или сомножителей.

Пример 1. Если A – появление 1 очка, B – 3-х очков, C – 5-ти очков при одном бросании игральной кости, то $S = A + B + C$ – появление нечётного числа очков. Если A – появление дамы, а B – появление пиковой масти при вытягивании одной карты из колоды, то $D = AB$ есть появление пиковой дамы.

События A и B называются несовместными, если они не могут наступить в одном и том же опыте. Ясно, что для таких событий $P(AB) = 0$.

Группа событий называется полной, если в результате опыта обязательно наступает хотя бы одно из этих событий.

Два события называются противоположными, если это несовместные события, образующие полную группу.

Пример 2. Рассмотрим в качестве опыта, приводящего к наступлению различных событий, одно бросание игральной кости. Пусть A – появление единицы, B – двойки, C – единицы, тройки или пятёрки, D – четвёрки или шестёрки. Тогда:

- событие A совместно с событием C , но несовместно с событиями B и D ;

- события A , B , C и D в совокупности образуют полную группу, но даже если событие A исключить, группа не утратит полноты;

- ни одна пара событий, выбранная из группы A , B , C , D , не является парой противоположных событий;

- события C и $B + D$ являются противоположными;

Вероятность суммы событий

Справедлива формула:

$$P(A + B) = P(A) + P(B) - P(AB). \quad (1)$$

Из этого основного утверждения вытекает целый ряд очевидных следствий:

1. Если события A и B несовместны, то $P(A + B) = P(A) + P(B)$.

2. Если события A_1, A_2, \dots, A_n образуют полную группу несовместных событий, то $P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) = 1$.

3. Для двух противоположных событий $P(A) + P(\bar{A}) = 1$.

Формулу для вероятности суммы трёх и более совместных событий мы не рассматриваем. В этом случае гораздо проще «действовать» через противоположное событие.

Пример 3. Найти вероятность того, что при бросании двух монет хотя бы на одной из них выпадет «орёл».

1^й способ. Согласно классическому определению вероятности, получаем $P(A) = 3/4$, т.к. существует 4 равновозможных исхода («орёл»-«орёл», «орёл»-«решка», «решка»-«орёл» и «решка»-«решка»), из которых 3 исхода являются благоприятными.

2^й способ. Рассматривая событие A как сумму двух событий («орёл» на первой монете, «орёл» на второй монете) по формуле (2.1) для вероятности суммы получаем $P(A) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$, где вероятность произведения события (т.е. вероятность события «орёл»-«орёл») найдена по классическому определению.

2.2 Зависимость событий. Вероятность произведения событий

Условной вероятностью $P(A|B)$ называется вероятность события A , вычисленная при условии, что событие B произошло. Событие A называется зависимым от события B , если $P(A|B) \neq P(A)$. Зависимость событий всегда взаимна, т.е. если A зависит от события B , то и B зависит от события A .

Вероятность произведения двух событий определяется формулой:

$$P(AB) = P(A)P(B|A). \quad (2)$$

Для независимых событий:

$$P(AB) = P(A)P(B). \quad (3)$$

Решим пример 6 ещё одним (уже третьим) способом. Если A - появление хотя бы одного «орла», то противоположное событие \bar{A} - появление «решек» на обеих брошенных монетах. Найдём $P(\bar{A}) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$, как вероятность произведения независимых событий (3). Тогда $P(A) = 1 - P(\bar{A}) = 3/4$.

Пример 4. Студент сдаёт два экзамена: физику и математику. Он оценивает свои шансы получить «отлично» по физике как 1 против 3^x , «отлично» по математике как 1 против 2^x . Каковы шансы студента получить хотя бы одну оценку «отлично» на двух экзаменах?

Введём вероятности отличной сдачи экзаменов по физике $P(A_1) = 1/4$ и по математике $P(A_2) = 1/3$. Как и в примере 3, здесь возможны разные способы решения.

1. По формуле (1) с учётом (3) получим:

$$P(A) = P(A_1) + P(A_2) - P(A_1)P(A_2) = 1/4 + 1/3 - 1/12 = 1/2.$$

2. Через понятие противоположного события получим:

$$P(A) = 1 - (1 - P(A_1))(1 - P(A_2)) = 1 - 3/4 * 2/3 = 1/2.$$

Есть и третий способ решения, который состоит в том, чтобы отдельно найти вероятность ровно одной оценки «отлично» и двух

оценок «отлично», и затем найденные вероятности сложить. В данном случае этот способ нерационален, однако мы обратимся к нему ниже, в пункте Дискретные случайные величины (пример 5).

2.3 Формула полной вероятности

Пусть событие A может произойти вместе с любым из несовместных друг с другом событий H_1, H_2, \dots, H_n , образующих полную группу (они называются гипотезами). Тогда вероятность события A определяется как

$$P(A) = \sum_{i=1}^n P(H_i)P(A|H_i). \quad (4)$$

Пример 4. Среди театральных зрителей женщин вдвое больше, чем мужчин. Из каждых 25 мужчин 1 является дальтоником, а среди женщин это заболевание встречается в 10 раз реже. Найти вероятность того, что выбранный наугад театральный зритель – дальтоник.

Обозначим события: H_1 - зритель - мужчина; H_2 - зритель - женщина; A - зритель - дальтоник.

Тогда $P(H_1) = 1/3$; $P(H_2) = 2/3$; $P(A|H_1) = 1/25$; $P(A|H_2) = 1/250$. Следовательно, по формуле полной вероятности:

$$P(A) = \frac{1}{3} \cdot \frac{1}{25} + \frac{2}{3} \cdot \frac{1}{250} = \frac{12}{750} = 0.016$$

2.4 Дискретные случайные величины

Величина, принимающая в результате опыта заранее неизвестное числовое значение, называется случайной величиной. Случайная величина, имеющая набор изолированных возможных значений, называется дискретной (ДСВ). Законом распределения ДСВ называется правило, по которому каждому возможному значению ставится в соответствие вероятность, с которой случайная величина может принять это значение. Закон распределения может быть задан таблично (ряд распределения), графически (многоугольник

распределения) и аналитически (формула).

Пример 5. Вернёмся к условиям примера 7, но переформулируем задачу. Студент сдаёт два экзамена: по физике и математике. Вероятность получения «пятерки» по физике $P(A_1) = 1/4$, по математике $P(A_2) = 1/3$. Найти закон распределения случайной величины X - числа полученных студентом «пятёрок» на двух экзаменах.

$$P(X = 0) = P(\bar{A}_1 \cdot \bar{A}_2) = P(\bar{A}_1) \cdot P(\bar{A}_2) = 3/4 \cdot 2/3 = 1/2;$$

$$P(X = 1) = P(A_1) \cdot P(\bar{A}_2) + P(\bar{A}_1) \cdot P(A_2) = 1/4 \cdot 2/3 + 3/4 \cdot 1/3 = 5/12;$$

$$P(X = 2) = P(A_1 \cdot A_2) = P(A_1) \cdot P(A_2) = 1/4 \cdot 1/3 = 1/12$$

События, заключающиеся в том, что ДСВ примет свои возможные значения, образуют полную группу несовместных событий. Поэтому сумма вероятностей всех возможных значений ДСВ равна 1.

Биномиальный закон распределения

Одним из наиболее известных законов распределения ДСВ является так называемый биномиальный закон.

Случайная величина X , представляющая собой число наступлений некоторого события в серии n испытаний, в каждом из которых вероятность наступления этого события одинакова и равна p , имеет распределение, называемое биномиальным. Это распределение описывается формулой Бернулли:

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, k = 0, 1, \dots, n, \quad (5)$$

где $C_n^k = \frac{n!}{(n-k)!k!}$ - число сочетаний из n по k .

Пример 6. Пусть случайная величина X - число наступлений некоторого события в серии из 3 испытаний, в каждом из которых вероятность наступления этого события одинакова и равна $1/4$. Тогда X подчинена биномиальному закону распределения:

$$P(X = x_i) = C_3^{x_i} \left(\frac{1}{4}\right)^{x_i} \left(\frac{3}{4}\right)^{3-x_i}, \text{ где } x_i = \overline{0,3},$$

или в табличном виде:

x_i	0	1	2	3
-------	---	---	---	---

p_i	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$
-------	-----------------	-----------------	----------------	----------------

2.5 Единичные жребии в методе Монте-Карло

Рассмотрим общий подход к моделированию случайных событий и величин с помощью так называемых единичных жребиев (т.е. опытов со случайным исходом). Единичный жребий может быть реализован с помощью генерации случайного числа - значения случайной величины, равномерно распределённой на интервале от 0 до 1 (работа 1). Обозначим такое случайное число через u . Нас будут интересовать три вида единичного жребия, которые мы рассмотрим последовательно.

Произошло ли данное событие? Пусть нам известно, что событие A имеет вероятность p . Можно условиться считать, что если u приняло значение меньше p , то событие A произошло; при $u \geq p$ событие не произошло. Вопрос о том, почему «граничный» случай $u = p$, трактуется как не наступление события, не имеет никакого практического значения: учитывая точность компьютерного представления действительных чисел, вероятностью такого совпадения можно просто пренебречь. Во всяком случае, на результат статистического моделирования это никакого влияния не оказывает.

Какое из нескольких событий, образующих полную группу несовместных событий, произошло? Пусть события A_1, A_2, \dots, A_k наступают с вероятностями p_1, p_2, p_k образуя полную группу несовместных событий, так что:

$$\sum_{i=1}^k p_i = 1$$

На рисунке 1 показан алгоритм, который можно принять для определения наступления того или иного события в зависимости от значения случайного числа u :

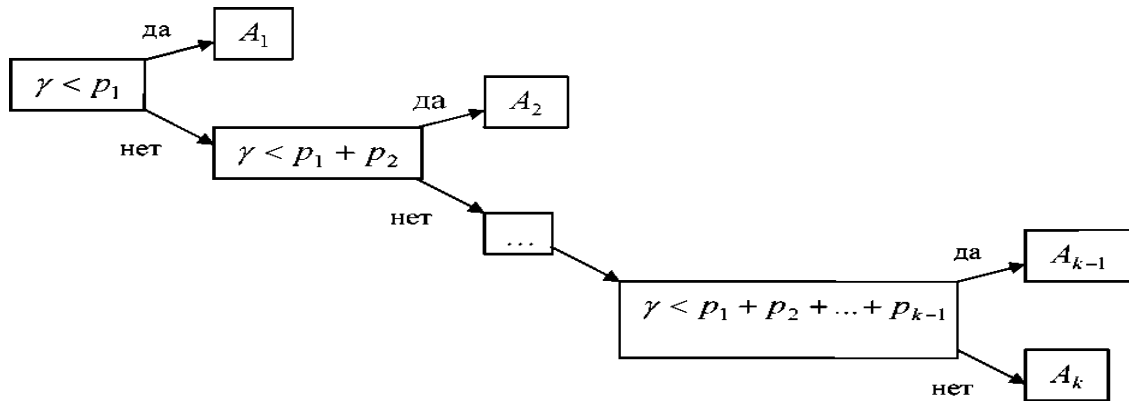


Рисунок 1 – Алгоритм определения наступления того или иного события в зависимости от значения случайного числа γ

Пусть, например, A_1, A_2, \dots, A_5 – равновозможные события, образующие полную группу несовместных событий (т.е. каждое из них имеет вероятность 0,2). Случайное число приняло значение $\gamma = 0,68$. Поскольку

$$0,2 + 0,2 + 0,2 < 0,68 < 0,2 + 0,2 + 0,2 + 0,2,$$

считаем, что произошло событие A_4 .

Какое значение приняла дискретная случайная величина? Рассмотрим дискретную случайную величину с известным законом распределения: значения x_1, x_2, \dots, x_k принимаются с вероятностями p_1, p_2, \dots, p_k соответственно. Легко понять, что этот случай сводится к предыдущей задаче, поскольку равенства $\{X = x_1\}, \{X = x_2\}, \dots, \{X = x_k\}$ можно рассматривать как события A_1, A_2, \dots, A_k . Пусть, например, случайная величина подчинена закону распределения, полученному в примере 6:

x_i	0	1	2	3
p_i	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

Допустим, что случайное число приняло значение 0,47.

Поскольку

$$\frac{27}{64} < 0,47 < \frac{54}{64},$$

считаем, что случайная величина X приняла значение 1.

3 Задания для практической работы

Задание 3.1. Смоделировать с помощью метода Монте-Карло случайные события, описанные в примере 5.

Задание 3.2. Смоделировать с помощью метода Монте-Карло ситуацию, описанную в примерах 4 и 6.

Задание 3.3. Разыграть по методу Монте-Карло дискретную случайную величину, описанную в примере 6.

Смысл настоящей работы состоит в том, чтобы, моделируя методом Монте-Карло случайные события, в которых отсутствует симметрия исходов, научиться статистически оценивать вероятности событий.

Инструкция по выполнению заданий в Excel

Для выполнения задания 3.1 создадим таблицу следующего вида:

	A	B	C	D	E
1		Доля мужчин в театре		0,333333333	
2		Вероятность дальтонизма у мужчины		0,04	
3		Вероятность дальтонизма у женщины		0,004	
4	№	Сл. число	Мужчина или женщина?	Сл.число	Дальтоник или нет?
5					

В ячейку D1 вводим число $1/3$, поскольку по условию на одного мужчину в театре приходится две женщины. В ячейки D2 и D3 вводим вероятности дальтонизма у мужчин и женщин соответственно. В ячейки B5 и D5 вводим функцию СЛЧИС(). Для заполнения ячеек C5 и E5 необходимо воспользоваться логической функцией

ЕСЛИ (лог выражение; значение если истина; значение если ложь).

Учитывая сформулированные выше правила разыгрывания случайных событий, в ячейку C5 введём `=ЕСЛИ(B5<=B1;"муж";"жен")`, а в ячейку E5 более сложное выражение `=ЕСЛИ(C5="муж";ЕСЛИ(D5<=B2;"да";"нет");ЕСЛИ(B5<=B3;"да";"нет"))`

Значки доллара в адресах некоторых ячеек поставлены для того, чтобы при последующем автозаполнении эти адреса не корректировались.

«Растягиваем» таблицу вниз до испытания №10000 и вычисляем с помощью функции СЧЁТЕСЛИ следующие величины: число мужчин, число женщин, число дальтоников и вероятность дальтонизма у наугад отобранного зрителя. Как мы видели при решении примера 8, эта вероятность равна 0,016. Статистическая оценка вероятности при числе испытаний, равном 10000, как правило, оказывается в пределах 0,014-0,018.

Переходим к заданию 3.2. Начнём с моделирования ситуации, описанной в примере 4. Создадим таблицу следующего вида:

	A	B	C	D
1	Вероятность «5» по физике		0,25	
2	Вероятность «5» по математике		0,33333333	
3	№	По физике «5»?	По математике «5»?	Хотя бы одна пятёрка?
4	1	=ЕСЛИ(СЛЧИС()<\$C\$1;"да";"нет")	=ЕСЛИ(СЛЧИС()<\$C\$2;"да";"нет")	=ЕСЛИ(В4="да";"да";ЕСЛИ(С4="да";"да";"нет"))
5

Обратите внимание на то, что в данном случае, в отличие от задания 3.1, мы не стали вводить специальных столбцов для случайных чисел, а использовали функцию СЛЧИС() как «вложение» в функцию ЕСЛИ (подумайте, почему это было невозможно в задании 3.1). «Растягиваем» таблицу и вычисляем вероятность того, что на двух экзаменах получена хотя бы одна пятёрка, как отношение числа таких испытаний, в которых это произошло, к полному числу испытаний. Статистическая оценка этой вероятности при числе испытаний, равном 10000, как правило, оказывается в пределах 0,49 - 0,51 (точное значение, как мы видели при решении примера 4, равно 0,5).

В примере 6 задача поставлена чуть более широко: найти закон распределения случайной величины X - числа полученных студентом «пятёрок» на двух экзаменах. Решение этой задачи методом Монте-Карло может быть получено с помощью следующей

таблицы в EXCEL:

	A	B	C	D
1		Вероятность «5» по физике	0,25	
2		Вероятность «5» по математике	0,33333333	
3	№	По физике «5»?	По математике «5»?	«Пятёрки» на двух экзаменах
4	1	=ЕСЛИ(СЛЧИС()<\$C\$1;1;0)	=ЕСЛИ(СЛЧИС()<\$C\$2;1;0)	=B4+C4
5

В данном случае нам удобно отображать результат экзамена в виде двоичного кода: если оценка «5», то результат равен 1, в противном случае результат равен 0. В каждом испытании число полученных «пятёрок» суммируется. После этого необходимо подсчитать число испытаний, в которых было получено 0, 1 и 2 «пятёрки» и оценить соответствующие вероятности. Эти оценки должны быть близки к точным значениям вероятностей, полученным при решении примера 6.

Переходим к заданию 3.3. Как было отмечено выше, разыгрывание дискретной случайной величины, по сути дела, сводится к разыгрыванию полной группы событий.

В диапазон ячеек B1:E2 вводим заданный ряд распределения, в ячейку C3 - формулу, которая определяет значение случайной величины в зависимости от значения случайного числа.

	A	B	C	D	E
1		0	1	2	3
2	№	0,421875	0,421875	0,140625	0,015625
3	1	=СЛЧИС()	=ЕСЛИ(B3<\$B\$2;0; ЕСЛИ(B3<\$B\$2+\$C\$2;1; ЕСЛИ(B3<\$B\$2+\$C\$2+\$D\$2;2;3)))		
4

После «растягивания» таблицы для проверки подсчитайте число испытаний, в которых случайная величина приняла значения 0, 1, 2 и 3 и оцените соответствующие вероятности.

4 Дополнительное задание

Задание 4.1 Студент сдаёт тест, который включает в себя 32

задания по 8 разделам (дидактическим единицам), по 4 задания в каждом разделе. Дидактическая единица считается освоенной, если не менее чем 2 задания из 4 решены правильно. Студент считается сдавшим тест, если он освоил все 8 дидактических единиц. Известно, что студент правильно решает каждое тестовое задание с вероятностью p . С какой вероятностью он сдаст тест? Задачу решить следующими способами.

4.1.1 Аналитически найти вероятность освоения дидактической единицы как вероятность суммы несовместных событий (2, 3 или 4 правильных ответа), вычисляя отдельные слагаемые с помощью формулы Бернулли (2.5). Затем найти вероятность сдать тест как вероятность освоения всех 8 дидактических единиц. Ответ:

$$p_{test} = p^{16} (6 - 8p + 3p^2)^8.$$

4.1.2 Смоделировать результат выполнения каждого задания двоичным кодом: 0, если задание не выполнено; 1, если задание выполнено правильно. Организовать подсчёт правильно выполненных заданий в каждой дидактической единице и определение результата теста в целом. По серии испытаний оценить вероятность сдачи теста при заданном p .

4.1.3 Смоделировать результат прохождения каждой дидактической единицы как случайную величину, имеющую биномиальное распределение

$$P(X = k) = C_4^k p^k (1 - p)^{4-k}, \quad k = 0, 1, 2, 3, 4.$$

Определить результат теста в целом. По серии испытаний оценить вероятность сдачи теста при заданном p .

Сравнить результаты, получаемые всеми описанными способами.

5 Содержание отчета

По результатам выполненной практической работы представляется отчет, в котором должны содержаться следующие пункты:

- цель работы;
- индивидуальное задание;
- порядок выполнения работы;
- таблицы, выполненные в Excel, содержащие формулы,

по которым производятся вычисления;

– таблицы, выполненные в Excel, содержащие результаты расчетов;

– выводы по результатам исследований с анализом полученных результатов.

6 Контрольные вопросы

1. Что такое сумма событий, произведение событий, несовместные события, полная группа событий, противоположные события?

2. Запишите формулу для суммы двух событий, сформулируйте следствия из неё.

3. Что такое зависимые и независимые события?

4. Как вычисляется вероятность произведения событий?

5. Запишите и объясните смысл формулы полной вероятности.

6. Приведите примеры дискретных случайных величин.

7. Какая задача приводит к формуле Бернулли и биномиальному закону распределения?

8. Виды единичного жребия в методе Монте-Карло.

9. Проанализируйте результаты решения задания 3.1 - 3.3. Постройте график зависимости вероятности успешного прохождения всего теста данным студентом от вероятности правильного выполнения им тестовых заданий.

Практическая работа № 3

«Оценка обнаруживающих и корректирующих свойств кодов Хемминга»

1 Цель работы

- получение навыков составления кодов Хемминга и оценки его обнаруживающих и корректирующих свойств. Практическое усвоение методики кодирования цифровой информации «по Хеммингу» для заданных условий, определяющих свойства кода. Рассмотрение основных этапов реализации алгоритмов обнаружения и исправления ошибок на приемной стороне линии связи.

2 Теоретические сведения

Систематические коды представляют собой блочные корректирующие коды, в которых информационные и проверочные символы расположены по строго определенной системе и всегда занимают строго определенные места в кодовых словах.

Наиболее известными систематическими кодами, получившими широкое практическое применение, являются коды Хэмминга.

Код Хемминга – это блочный код, позволяющий исправлять одиночные и фиксировать двойные ошибки, разработанный Ричардом Хеммингом в сороковых годах прошлого столетия.

Идея кодов Хемминга заключается в разбиении данных на блоки фиксированной длины и вводе в эти блоки контрольных бит, дополняющих до четности несколько пересекающихся групп, охватывающих все биты блока.

Ричард Хемминг рассчитал минимальное количество проверочных бит, позволяющих однозначно исправлять однократные ошибки.

Если длина информационного блока, который требуется закодировать – m бит. Количество контрольных бит, используемых для его кодирования, – k , то закодированный блок будет иметь длину: $n = m+k$ бит. Для каждого блока такой длины возможны n различных комбинаций, содержащих ошибку. Таким образом, для каждого передаваемого информационного блока может существовать $n -$

блоков, содержащих однократную ошибку, и один блок – без ошибок. Следовательно, максимальное количество различных закодированных блоков, содержащих не больше одной ошибки, будет: $2^m(n+1)$, где $n = m+k$.

Двоичный код Хэмминга строится следующим образом:

1) Определяется число k проверочных символов из условия:

$$2^k \geq m + k + 1,$$

где m – число информационных символов.

2) Выбираются места расположения проверочных символов из условия, чтобы проверочные символы участвовали только в одной операции подсчета четности с целью упрощения процесса кодирования. Такими местами являются символы с номерами, являющиеся целыми степенями числа 2, т. е. 1, 2, 4, 8, 16 и т. д. Символы в кодовых словах Хэмминга нумеруются слева направо.

3) Определяются значения символов слова, называемого синдромом:

$$S_k S_{k-1} \dots S_{2^{s_1}},$$

Из следующих уравнений:

$$а) S_1 = x_1 \oplus x_3 \oplus x_5 \oplus x_7 \oplus \dots = 0,$$

т. е. складываются по модулю 2 значения тех символов, двоичное представление номеров которых содержит в последнем разряде 1.

$$б) S_2 = x_2 \oplus x_3 \oplus x_6 \oplus x_7 \oplus \dots = 0,$$

Складываются значения тех символов, двоичное представление номеров которых содержит 1 в предпоследнем разряде.

$$в) S_3 = x_4 \oplus x_5 \oplus x_6 \oplus x_7 \oplus \dots = 0,$$

Складываются значения символов, двоичное представление номеров которых содержит 1 в третьем от конца разряде.

$$г) S_4 = x_8 \oplus x_9 \oplus x_{10} \oplus x_{11} \oplus \dots = 0,$$

Складываются значения символов, двоичное представление номеров которых содержит 1 в четвертом от конца разряде.

$x_1, x_2, x_3, x_4, \dots$ – значения символов с номерами 1, 2, 3, 4, ...

Местоположение ошибки, то есть определение символа с ошибкой осуществляется по значению синдрома. Если синдром состоит из одних нулей, т. е.

$$S_k S_{k-1} \dots S_2 S_1 = 00 \dots 00 ,$$

то ошибка отсутствует. Если в синдроме есть символы, отличные от 0, то это говорит о наличии ошибки. Например, если $S_4 S_3 S_2 S_1 = 1000$, то это означает, что ошибка содержится в восьмом символе, так как $1000_2 = 8_{10}$.

Другими словами, синдром в коде Хэмминга определяет номер символа с ошибкой. Исправление ошибки осуществляется заменой 0 на 1 или, наоборот, 1 на 0.

Для исправления одиночной и обнаружения двойной ошибки, кроме проверок по синдрому, следует проводить еще одну проверку на четность для каждого кодового слова Хэмминга. Чтобы осуществить такую проверку, следует в конце каждого кодового слова добавить проверочный символ таким образом, чтобы сумма единиц в полученном слове всегда была четной. Тогда в случае одной ошибки проверка по синдрому укажет номер ошибочного символа, а проверка на четность укажет наличие ошибки. Если проверка по синдрому укажет на наличие ошибки, а проверка на четность не фиксирует ошибку, то в кодовом слове присутствуют две ошибки.

3 Порядок выполнения работы

Данная работа предполагает выполнение следующих этапов:

- 1) Изучить методические указания к практической работе.
- 2) Для заданных кодовых слов построить кодовые слова Хэмминга.
- 3) Осуществить проверку работоспособности кода путем изменения значения одного из символов в любом кодовом слове Хэмминга.
- 4) Оформить и защитить отчет по выполнению практической работы.

4 Варианты заданий

Номер варианта определяется как порядковый номер студента в журнале преподавателя. Пусть требуется получить кодовые векторы Хэмминга для кодовых слов, приведенных в таблице 1.

Таблица 1 – Варианты заданий для получения кодовых слов Хэмминга

Вариант	Передаваемые кодовые слова		
1	0010101	0111001	1000001
2	0001110	0100010	0001110
3	1100110	1011101	0100010
4	0101010	0100000	1101101
5	1111000	1000001	0010101
6	0000111	0001110	0001110
7	0111001	0100010	1100110
8	0100010	1101101	0101010
9	1011101	0010101	1111000
10	0100000	0001110	0100000
11	1000001	1100110	1100110
12	0001110	0101010	0101010
13	0100010	1111000	1111000
14	1101101	0100000	0100000
15	0101110	1000001	1000001

5 Содержание отчета

По результатам выполненной практической работы представляется отчет, в котором должны содержаться следующие пункты:

- цель работы;
- индивидуальное задание;
- порядок выполнения работы;
- таблицы, выполненные в Excel, содержащие формулы, по которым производятся вычисления;

- таблицы, выполненные в Excel, содержащие результаты расчетов;
- выводы по результатам исследований с анализом полученных результатов.

6 Контрольные вопросы

- 1) Какие коды называют систематическими?
- 2) Сколько ошибок способен обнаруживать и исправлять код Хэмминга?
- 3) По какому правилу определяют число проверочных символов в коде Хэмминга?
- 4) В каких местах располагаются проверочные символы в кодовых словах Хэмминга?
- 5) По какому правилу строятся уравнения для нахождения проверочных символов в коде Хэмминга?
- 6) Какую информацию при декодировании кода Хэмминга дает синдром?
- 7) Каким образом обеспечивается обнаружение двойных ошибок в кодовых словах Хэмминга?

Практическая работа №4 «Практическое ознакомление с применением корректирующих кодов»

1 Цель работы

- практическое ознакомление с применением корректирующих кодов. Исследованию подлежат: вес кода; кодовое расстояние; код с проверкой на четность.

2 Теоретическое введение

Корректирующими называются коды, позволяющие обнаруживать и исправлять ошибки. Идею представления корректирующих кодов можно представить с помощью N-мерного куба. Возьмем трехмерный куб (рисунок 1), длина ребер, в котором равна одной единице. Вершины такого куба отображают двоичные коды. Минимальное расстояние между вершинами определяется минимальным количеством ребер, находящихся между вершинами. Это расстояние называется кодовым (или хэмминговым) и обозначается буквой d .

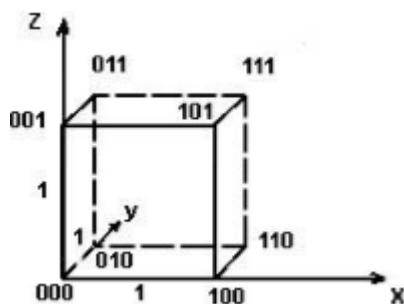


Рисунок 1 – Представление двоичных кодов с помощью куба

Иначе, кодовое расстояние – это то минимальное число элементов, в которых одна кодовая комбинация отличается от другой. Для определения кодового расстояния достаточно сравнить две кодовые комбинации по модулю 2. Так, сложив две комбинации:

$$\begin{array}{r} 10110101101 \\ + \end{array}$$

11001010101

01111111000

определим, что расстояние между ними $d=7$.

Заметим, что кодовое расстояние d между комбинацией X_i и нулевой комбинацией $X_0 = 00..0$ называют весом W комбинации X_i , т.е. вес X_i равен числу "1" в ней. Для кода с $N=3$ восемь кодовых комбинаций размещаются на вершинах трехмерного куба. Такой код имеет кодовое расстояние $d=1$, и для передачи используются все восемь кодовых комбинаций 000,001,...,111. Такой код является не помехоустойчивым, он не в состоянии обнаружить ошибку.

Если выберем комбинации с кодовым расстоянием $d=2$, например, 000,110,101,011, то такой код позволит обнаруживать однократные ошибки. Назовем эти комбинации разрешенными, предназначенными для передачи информации. Все остальные 001,010,100,111 - запрещенные.

Любая одиночная ошибка приводит к тому, что разрешенная комбинация переходит в ближайшую, запрещенную комбинацию (см. рисунок 2). Получив запрещенную комбинацию, мы обнаружим ошибку. Выберем далее вершины с кодовым расстоянием $d=3$

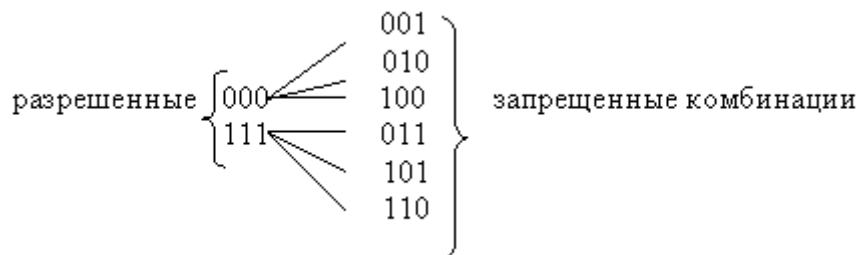


Рисунок 2 – Запрещенные комбинации

Такой код может исправить одну одиночную ошибку или обнаружить две ошибки. Таким образом, увеличивая кодовое расстояние можно увеличить помехоустойчивость кода. В общем случае кодовое расстояние определяется по формуле $d=t+1+1$ где t - число

исправляемых ошибок, l - число обнаруживаемых ошибок. Обычно $l > t$.

Большинство корректирующих кодов образуются путем добавления к исходной k - комбинации r - контрольных символов. В итоге в линию передаются $n = k + r$ символов. При этом корректирующие коды называются (n, k) кодами. Как можно определить необходимое число контрольных символов?

Для построения кода способного обнаруживать и исправлять одиночную ошибку необходимое число контрольных разрядов будет составлять $n - k \geq \log_2(n + 1)$. Это равносильно известной задаче о минимуме числа контрольных вопросов, на которые могут быть даны ответы вида "да" или "нет", для однозначного определения одного из элементов конечного множества.

Если необходимо исправить две ошибки, то число различных исходов будет составлять C_n^2 . Тогда $n - k \geq \log_2(1 + C_n^1 + C_n^2)$, в этом случае обнаруживаются однократные и двукратные ошибки. В общем случае, число контрольных символов должно быть не меньше

$$n - k \geq \log(1 + C_n^1 + C_n^2 + \dots + C_n^t) = \log \sum_{i=0}^t C_n^i \quad (1)$$

Формула 1 называется неравенством Хэмминга, или нижней границей Хэмминга для числа контрольных символов.

К основным характеристикам корректирующих кодов относятся:

- число разрешённых и запрещённых кодовых комбинаций;
- избыточность кода;
- число обнаруживаемых или исправляемых ошибок;
- минимальное кодовое расстояние;
- корректирующие возможности кодов.

Для блочных двоичных кодов, с числом символов в блоках равным n , общее число возможных кодовых комбинаций определяется значением

$$N_o = 2^n \quad (2)$$

Число разрешённых кодовых комбинаций при наличии k информационных разрядов в первичном коде равно

$$N_k = 2^k \quad (3)$$

Очевидно, что число запрещённых комбинаций равно:

$$N_z = N_o - N_k = 2^n - 2^k \quad (4)$$

а с учётом $r=n-k$ отношение будет:

$$N_o / N_k = 2^{n-k} = 2^r \quad (5)$$

где r – число избыточных (проверочных) разрядов в блочном коде.

Избыточностью корректирующего кода называют величину

$$\chi = \frac{r}{n} = \frac{n-k}{n} = 1 - \frac{k}{n} \quad (6)$$

откуда следует:

$$B_k = \frac{k}{n} = 1 - \chi \quad (7)$$

Эта величина показывает, какую часть общего числа символов кодовой комбинации составляют информационные символы. В теории кодирования величину V_k называют относительной скоростью кода. Если производительность источника информации равна H символов в секунду, то скорость передачи после кодирования этой информации окажется равной,

$$V = H \cdot \frac{k}{n}, \quad (8)$$

поскольку в закодированной последовательности из каждых n символов только k – символов являются информационными.

Если число ошибок, которые нужно обнаружить или исправить, значительно, то необходимо иметь код с большим числом проверочных символов. Чтобы при этом скорость передачи оставалась достаточно высокой, необходимо в каждом кодовом блоке одновременно увеличивать как общее число символов, так и число информационных символов. При этом длительность кодовых блоков будет существенно возрастать, что приведёт к задержке информации при передаче и приёме. Чем сложнее кодирование, тем длительнее временная задержка информации.

При применении двоичных кодов учитывают только дискретные искажения, при которых единица переходит в нуль ($1 \rightarrow 0$) или нуль переходит в единицу ($0 \rightarrow 1$). Переход $1 \rightarrow 0$ или $0 \rightarrow 1$ только в одном элементе кодовой комбинации называют единичной ошибкой (единичным искажением). В общем случае под кратностью ошибки подразумевают число позиций кодовой комбинации, на которых под действием помехи одни символы оказались заменёнными на другие. Возможны двукратные ($g = 2$) и многократные ($g > 2$) искажения элементов в кодовой комбинации в пределах $0 \leq g \leq n$.

Минимальное кодовое расстояние является основным параметром, характеризующим корректирующие способности данного кода. Если код используется только для обнаружения ошибок кратностью g_0 , то необходимо и достаточно, чтобы минимальное кодовое расстояние было равно

$$d_{min} \geq g_0 + 1. \quad (9)$$

В этом случае никакая комбинация из g_0 ошибок не может превести одну разрешённую кодовую комбинацию в другую разрешённую. Таким образом, условие обнаружения всех ошибок кратностью g_0 можно записать в виде:

$$g_0 \leq d_{min} - 1. \quad (10)$$

Чтобы можно было исправить все ошибки кратностью g_i и менее, необходимо иметь минимальное расстояние, удовлетворяющее условию

$$d_{min} \geq 2 \cdot g_i + 1. \quad (11)$$

В этом случае любая кодовая комбинация с числом ошибок g_i отличается от каждой разрешённой комбинации не менее чем в g_{i+1} позициях. Если условие (11) не выполнено, возможен случай, когда ошибки кратности g исказят переданную комбинацию так, что она станет ближе к одной из разрешённых комбинаций, чем к переданной или даже перейдёт в другую разрешённую комбинацию. В соответствии с этим, условие исправления всех ошибок кратностью не более g_i можно записать в виде:

$$g_i \leq (d_{min} - 1) / 2. \quad (12)$$

Из (9) и (11) следует, что если код исправляет все ошибки кратностью g_i , то число ошибок, которые он может обнаружить, равно $g_0 = 2 \cdot g_i$. Следует отметить, что соотношения (9) и (11) устанавливают лишь гарантированное минимальное число обнаруживаемых или исправляемых ошибок при заданном d_{\min} и не ограничивают возможность обнаружения ошибок большей кратности. Например, простейший код с проверкой на чётность с $d_{\min} = 2$ позволяет обнаруживать не только одиночные ошибки, но и любое нечётное число ошибок в пределах $g_0 < n$.

Вопрос о минимально необходимой избыточности, при которой код обладает нужными корректирующими свойствами, является одним из важнейших в теории кодирования. Этот вопрос до сих пор не получил полного решения. В настоящее время получен лишь ряд верхних и нижних оценок (границ), которые устанавливают связь между максимально возможным минимальным расстоянием корректирующего кода и его избыточностью.

Так, граница Плоткина даёт верхнюю границу кодового расстояния d_{\min} при заданном числе разрядов n в кодовой комбинации и числе информационных разрядов k , и для двоичных кодов:

$$d_{\min} \leq \frac{n \cdot 2^{k-1}}{2^k - 1} \quad (13)$$

или

$$r \geq 2 \cdot (d_{\min} - 1) - \log_2 d_{\min} \quad (14)$$

при

$$n \geq 2 \cdot d_{\min} - 1.$$

Верхняя граница Хемминга устанавливает максимально возможное число разрешённых кодовых комбинаций (2^k) любого помехоустойчивого кода при заданных значениях n и d_{\min} :

$$2^k \leq 2^n / \sum_{i=0}^{\frac{d_{\min}-1}{2}} C_n^i \quad (15)$$

где C_i^n - число сочетаний из n элементов по i элементам.

Отсюда можно получить выражение для оценки числа проверочных символов:

$$r \geq \log_2 \left(\sum_{i=0}^{\frac{d_{\min}-1}{2}} C_n^i \right) \quad (16)$$

Для значений $(d_{\min}/n) \leq 0.3$ разница между границей Хемминга и границей Плоткина сравнительно невелика. Граница Варшамова-Гильберта для больших значений n определяет нижнюю границу для числа проверочных разрядов, необходимого для обеспечения заданного кодового расстояния:

$$r \geq \log_2 \left(\sum_{i=0}^{d-2} C_{n-1}^i \right) \quad (17)$$

Отметим, что для некоторых частных случаев Хемминг получил простые соотношения, позволяющие определить необходимое число проверочных символов:

$$\begin{aligned} r &\geq \log_2(n+1) \text{ для } d_{\min}=3, \\ r &\geq \log_2(2n) \text{ для } d_{\min}=4. \end{aligned}$$

Все приведённые выше оценки дают представление о верхней границе числа d_{\min} при фиксированных значениях n и k или оценку снизу числа проверочных символов r при заданных k и d_{\min} .

Существующие методы построения избыточных кодов решают в основном задачу нахождения такого алгоритма кодирования и декодирования, который позволял бы наиболее просто построить и реализовать код с заданным значением d_{\min} . Поэтому различные корректирующие коды при одинаковых d_{\min} сравниваются по сложности кодирующего и декодирующего устройств. Этот критерий является в ряде случаев определяющим при выборе того или иного кода.

Поясним идею проверки на чётность на примере простейшего корректирующего кода, который так и называется кодом с проверкой на чётность или кодом с проверкой по паритету (равенству).

В таком коде к кодовым комбинациям безизбыточного первичного двоичного k - разрядного кода добавляется один дополнительный разряд (символ проверки на чётность, называемый проверочным, или контрольным). Если число символов "1" сходной кодовой комбинации чётное, то в дополнительном разряде формируют контрольный символ 0, а если число символов "1" нечётное, то в дополнительном разряде формируют символ 1. В результате общее число символов "1" в любой передаваемой кодовой комбинации всегда будет чётным.

Таким образом, правило формирования проверочного символа сводится к следующему:

$$r_1 = i_1 \oplus i_2 \oplus \dots \oplus i_k,$$

где i - соответствующий информационный символ (0 или 1), k - общее их число, а под операцией " \oplus " здесь и далее понимается сложение по mod 2. Очевидно, что добавление дополнительного разряда увеличивает общее число возможных комбинаций вдвое по сравнению с числом комбинаций исходного первичного кода, а условие чётности разделяет все комбинации на разрешённые и не-

разрешённые. Код с проверкой на чётность позволяет обнаруживать одиночную ошибку при приёме кодовой комбинации, так как такая ошибка нарушает условие чётности, переводя разрешённую комбинацию в запрещённую.

Критерием правильности принятой комбинации является равенство нулю результата S суммирования по mod 2 всех n символов кода, включая проверочный символ r_1 . При наличии одиночной ошибки S принимает значение 1:

$$S = r_1 \oplus i_1 \oplus i_2 \oplus \dots \oplus i_k = \{0 - \text{ошибки нет}$$

$$(\text{_____} n \text{_____}) = \{1 - \text{единичная ошибка}$$

Этот код является $(k+1, k)$ – кодом, или $(n, n-1)$ – кодом. Минимальное расстояние кода равно двум ($d_{\min}=2$), и, следовательно, никакие ошибки не могут быть исправлены. Простой код с проверкой на чётность может использоваться только для обнаружения (но не исправления) однократных ошибок.

Увеличивая число дополнительных проверочных разрядов и формируя по определённым правилам проверочные символы r , равные 0 или 1, можно усилить корректирующие свойства кода так, чтобы он позволял не только обнаруживать, но и исправлять ошибки.

3 Порядок выполнения работы

3.1 Исследование веса кодовой комбинации

1) Для перевода чисел из десятичной системы счисления в двоичную необходимо в среде Mathcad задать функцию $\text{binary}(x)$, где x – число в десятичной системе счисления. Пример реализации подобной функции представлен на рисунке 3;

$$\text{Binary}(N) := \left| \begin{array}{l} p \leftarrow \text{floor}\left(\frac{\ln(N)}{\ln(2)}\right) \\ \text{for } i \in 0..p \\ \quad b_{p-i} \leftarrow \text{mod}\left(\text{floor}\left(\frac{N}{2^i}\right), 2\right) \\ b^T \end{array} \right.$$

Рисунок 3 – Пример реализации функции binary

- 2) Проверьте работоспособность заданной функции;
- 3) В качестве исследуемой кодовой комбинации используйте дату рождения вашей мамы (например, 01.01.1970);
- 4) Переведите полученное число в двоичную систему счисления при помощи заданной ранее функции binary;
- 5) Посчитайте количество единиц в получившемся числе при помощи встроенной функции Mathcad “Сумма вектора”, которая находится в блоке “Векторы и матрицы” на Панели инструментов.

3.2 Исследование кодового расстояния:

- 1) В качестве исследуемых кодовых комбинаций используйте вашу дату рождения (например, 01.01.1995) и дату через 400 дней от вашей даты рождения;
- 2) Переведите полученные числа в двоичную систему счисления с помощью функции binary. Присвойте двум переменным получившиеся значения;
- 3) Для поразрядного сложения полученных чисел необходимо задать функцию XOR(x,y), где x – первое число, y – второе. Пример реализации подобной функции представлен на рисунке 4.

$$\text{XOR}(x,y) := \left| \begin{array}{l} z \leftarrow 0 \\ \text{for } i \in 0.. \text{rows}(x) - 1 \\ \quad z_i \leftarrow x_i \oplus y_i \\ z \end{array} \right.$$

Рисунок 4 – Пример реализации функции xor

4) Посчитайте количество единиц в получившемся числе при помощи встроенной функции Mathcad “Сумма вектора”, которая находится в блоке “Векторы и матрицы” на Панели инструментов.

3.3 Исследование кода с проверкой на четность

1) В качестве исследуемой кодовой комбинации используйте ваш номер телефона без 8 (например, 9031112233);

2) Переведите полученное число в двоичную систему счисления с помощью функции binary. Присвойте переменной получившееся значение;

3) Посчитайте количество единиц в получившемся числе при помощи встроенной функции Mathcad “Сумма вектора”, которая находится в блоке “Векторы и матрицы” на Панели инструментов;

4) Транспонируйте полученную матрицу при помощи встроенной функции Mathcad “Транспонирование матрицы”, которая находится в блоке “Векторы и матрицы” на Панели инструментов;

5) Для транспонированной матрицы примените функцию reverse(T), где T – транспонированная матрица;

6) Необходимо добавить контрольный разряд. Для этого необходимо задать элемент T с индексом (n-1,0), где n – число разрядов в двоичном числе. Значения этого элемента зависит от того результата, который вы получили в шаге 3.3: если число четное, то контрольному разряду присваивается значение 0, нечетное – единица;

7) Для вывода скорректированного кода, необходимо выполнить обратные преобразования: для полученной матрицы используем функцию reverse(T), где T - транспонированная матрица, а затем транспонируем ее. Если все выполнено верно, то контрольный разряд окажется у вас старшим.

4 Содержание отчета

По результатам выполненной практической работы представляется отчет, в котором должны содержаться следующие пункты:

- цель работы;
- индивидуальное задание;
- порядок выполнения работы;
- таблицы, выполненные в Excel, содержащие формулы, по которым производятся вычисления;
- таблицы, выполненные в Excel, содержащие результаты расчетов;
- выводы по результатам исследований с анализом полученных результатов.

5 Контрольные вопросы

- 1) Что называют корректирующими кодами?
- 2) Что такое кодовое расстояние и как его вычислить для двух кодовых комбинаций?
- 3) Что называют весом кодовой комбинации?
- 4) Назовите основные характеристики корректирующих кодов?
- 5) Что такое число разрешенных и запрещенных комбинаций?
- 6) Что такое избыточность корректирующего кода и относительная скорость кода?
- 7) Какой смысл имеют верхние границы Плоткина и Хэмминга?
- 8) Что такое граница Варшамова-Гилберта и чему она равна $d_{\min}=3$ и 4?
- 9) В чем смысл кода с проверкой на четность? Охарактеризуйте его на примере трехразрядного кода с числом разрешенных комбинаций равным четырем.

Практическая работа №5
«Расчёт пропускной способности
дискретного и непрерывного каналов связи»

1 Цель работы

- 1) Провести анализ пропускной способности каналов связи.
- 2) Вычислить пропускную способность дискретного и непрерывного канала по заданным вариантам.

2 Теоретические сведения

В современных телекоммуникационных системах имеются два вида каналов связи: дискретный и непрерывный. Они определяются типом сигналов, идущим по ним. Под сигналом понимается физический процесс (например, изменяющееся во времени напряжение), отображающий некоторую информацию или сообщение. Математически сигнал описывается функцией определенного типа. Аналоговые сигналы описываются непрерывной (или кусочно-непрерывной) функцией $x_a(t)$, причем сама функция и аргумент t могут принимать любые значения на некоторых интервалах $x_a^t \leq x \leq x_a^{t'}$, $t^t \leq t \leq t^{t'}$. На рисунке 1 а представлен пример аналогового сигнала, изменяющегося во времени по закону $x_a(t) = A \cdot e^{-\alpha(t)}$, где $A=1, \alpha > 0$. Другой пример аналогового сигнала, показанный на рисунке 1 б, изменяется во времени по закону $x_a(t) = U_m \cdot \sin(2 \cdot \pi \cdot f \cdot t)$.

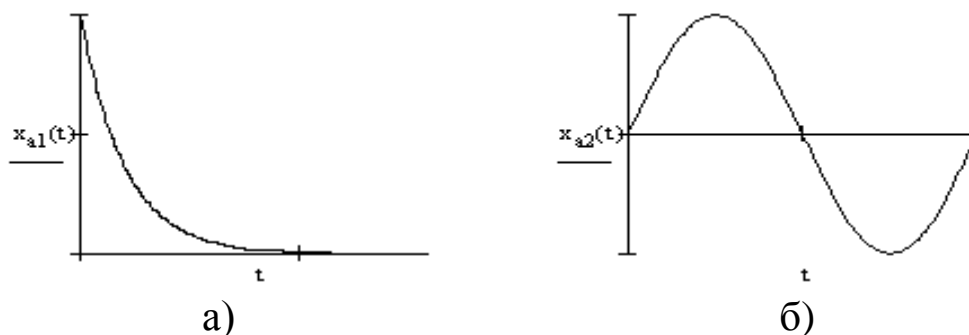


Рисунок 1 – Аналоговые сигналы

Дискретные сигналы отличаются от аналоговых тем, что их значения известны лишь в дискретные моменты времени. Дискретные сигналы описываются решетчатыми функциями – последовательностями – $x_o(n \cdot t)$, где $t = \text{const}$ – интервал (период) дискретизации, $n=0,1,2,\dots$. Сама функция $x_o(n \cdot t)$ может в дискретные моменты принимать произвольные значения на некотором интервале. Эти значения функции называются выборками или отсчетами функции. Другим обозначением решетчатой функции $x_o(n \cdot t)$ является $x(n)$ или x_n . На рисунке 2 (а и б) представлены примеры решетчатых функций $x_o(t) = A \cdot e^{-a(t)}$ и $x_o(t) = U_m \cdot \sin(2 \cdot \pi \cdot f \cdot t)$. Последовательность $x(n)$ может быть конечной или бесконечной, в зависимости от интервала определения функции.

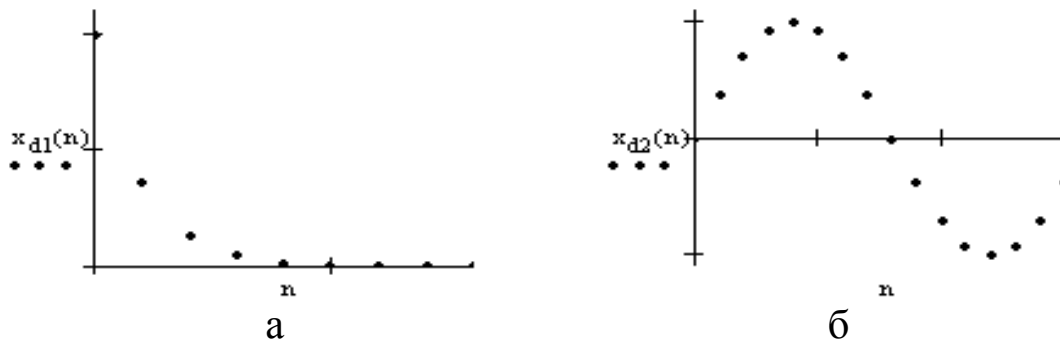


Рисунок 2 – Дискретные сигналы

Важной характеристикой канала связи, как дискретного, так и непрерывного является его пропускная способность. Чем выше скорость передачи данных и ниже вероятность битовой ошибки, тем эффективнее система связи. Наибольшая возможная в данном канале скорость передачи информации называется его пропускной способностью. Пропускная способность канала есть скорость передачи информации при использовании «наилучших» (оптимальных) для данного канала источника, кодера и декодера, поэтому она характеризует только канал. С передачей одного элементарного сигнала связано некоторое количество информации $I(s)$. Если общее число различных элементарных сигналов n , а вероятности их появления $p(a_i)$ ($i = 1 \dots n$), то согласно формуле Шеннона:

$$I_s = -\sum_{i=1}^n p(a_i) \cdot \log_2 p(a_i) \quad (1)$$

Однако, как обсуждалось выше, оптимальным будет такой вариант кодирования, при котором появление всех элементарных сигналов (знаков вторичного алфавита) оказывается равновероятным – в таком случае:

$$I_s = I_s \max = \log_2 n \quad (2)$$

Это значение является предельным (наибольшим) для информационного содержания элементарного сигнала выбранного вторичного алфавита. Поскольку такое количество информации передается за время t , можно ввести величину, характеризующую предельную интенсивность информационного потока через канал – пропускную способность канала C :

$$C = \frac{I_s^{\max}}{t} = L \cdot I_s^{\max} \quad (3)$$

Данная величина является характеристикой канала связи, поскольку зависит только от его особенностей. Это выражение служит определением пропускной способности как идеального канала (без помех), так и реального канала с помехами – просто, как мы увидим далее, информационное содержание элементарного сигнала в реальном канале оказывается меньше $\log_2 n$.

Если $I_s \max$ выражено в битах, а t – в секундах, то единицей измерения C будет бит/с. Раньше такая единица называлась *бод*, однако, название не прижилось, и по этой причине пропускная способность канала связи измеряется в *бит/с*. Производными единицами являются:

$$1 \text{ Кбит/с} = 10^3 \text{ бит/с};$$

$$1 \text{ Мбит/с} = 10^6 \text{ бит/с};$$

$$1 \text{ Гбит/с} = 10^9 \text{ бит/с}.$$

При отсутствии в канале связи помех $I_s \max = \log_2 n$; тогда $C = L \cdot \log_2 n = \frac{\log_2 n}{t} = v_m \cdot \log_2 n$ – максимально возможное значение пропускной способности (это обстоятельство отражено индексом "0"); в реальном канале $I_s \max \leq \log_2 n$ и, следовательно, $C \leq C_0$.

Пропускная способность дискретного m -ичного канала определяется выражением:

$$C_{\text{дк}} = V \cdot \left[\log_2 n + p \cdot \log_2 n \frac{p}{n-1} + (1-p) \log_2 (1-p) \right] \quad (4)$$

где: V - скорость модуляции, [Бод]

p - вероятность ошибки сигналов в канале

n - число вариантов кодовых символов (основание кода, например $m=2, 4, 8, 16, \dots$)

Пропускная способность двоичного канала (т.е. при $n=2$) определяется выражением:

$$C_{\text{ддк}} = V \cdot [1 + p \cdot \log_2 p + (1-p) \log_2 (1-p)] \quad (5)$$

При расчетах необходимо учитывать, что микрокалькуляторы и программа Mathcad производят вычисления только с десятичными и натуральными логарифмами.

Для перехода от десятичного к двоичному логарифму необходимо воспользоваться выражением:

$$\log_2 p = \frac{\log_{10} p}{\log_{10} 2} \text{ [Бит]} \quad (6)$$

Пропускная способность непрерывного канала определяется выражением:

$$C_{нк} = F_k \cdot \log_2 \left[1 + \frac{P_c}{P_{ш}} \right] \quad (7)$$

где: F – полоса пропускания канала, [Гц], $\frac{P_c}{P_{ш}} = h$ – отношение мощности сигнала к мощности шума в канале (ОСШ).

3 Задание на практическую работу

Номер варианта студента определяется как порядковый номер студента в журнале преподавателя.

1) Вычислить пропускную способность дискретного m -ичного канала по заданным вариантам.

2) Вычислить пропускную способность непрерывного канала по заданным вариантам.

Таблица 1 – Варианты заданий

Вариант по списку (N)	Вероятность ошибки сигналов в канале (p)	Основание кода	Скорость модуляции в Битах	Полоса пропускания канала	Мощность сигнала	Мощность шума
1	0.(10modN)	2	50	3100	10^7	7500
2	0.(10modN)	4	100	6000	10^8	7700
3	0.(10modN)	8	200	12000	10^9	8500
4	0.(10modN)	16	50	3100	10^7	9500
5	0.(10modN)	32	100	6000	10^8	7500
6	0.(10modN)	2	200	12000	10^9	7700
7	0.(10modN)	4	50	3100	10^9	8500
8	0.(10modN)	8	100	6000	10^8	9500
9	0.(10modN)	16	200	12000	10^7	7500
10	0.(10modN)	32	50	3100	10^7	7700
11	0.(10modN)	2	100	6000	10^8	8500
12	0.(10modN)	4	200	12000	10^9	9500
13	0.(10modN)	8	50	3100	10^9	7500
14	0.(10modN)	16	100	6000	10^8	7700
15	0.(10modN)	32	200	12000	10^7	8500
16	0.(10modN)	2	50	3100	10^7	9500
17	0.(10modN)	4	100	6000	10^8	7500

18	0.(10modN)	8	200	12000	10^9	7700
19	0.(10modN)	16	50	3100	10^9	8500

4 Работа с MathCad

Следует познакомить MathCad с величинами, которыми мы будем оперировать. Для этого введем в рабочем поле наши величины, и присвоим им численные значения (взяты значения варианта №8): $p := 0.2$ $n := 8$ $V := 100$, p вычисляется как ноль целых, а его десятичная часть – остаток от деления 10 на номер варианта.

Затем нужно посчитать пропускную способность дискретного m -ичного канала, воспользовавшись формулой:

$$C_{\text{дк}} = V \cdot \left[\log_2 n + p \cdot \log_2 n \frac{p}{n-1} + (1-p) \log_2 (1-p) \right].$$

Необходимо правильно ввести ее в среду MathCad, не забывая о том, что данная программа не может считать логарифмы по основанию 2. Учитывая сей факт, формула в среде MathCad будет иметь следующий вид:

$$C_{\text{дк}} := V \cdot \left[\left[\frac{\log(n)}{\log(2)} + p \cdot \frac{\log\left[\frac{(n \cdot p)}{n-1}\right]}{\log(2)} + (1-p) \cdot \frac{\log(1-p)}{\log(2)} \right] \right]$$

Для расчета пропускной способности непрерывного канала связи воспользуемся формулой:

$$C_{\text{НК}} = F_k \cdot \log_2 \left[1 + \frac{P_c}{P_u} \right]$$

Для вывода на экран результата вычисления нужно ввести “C=”.

5 Контрольные вопросы

- 1) Дайте определение сигнала.
- 2) Чем отличаются дискретные и двоичные сигналы?

- 3) Дайте определение пропускной способности канала.
- 4) Формула максимального значения пропускной способности канала.
- 5) Формула пропускной способности непрерывного канала.
- 6) Формула пропускной способности дискретного канала.