5. ДВОИЧНЫЕ КОДЫ

Примеры решения задач

Пример 5.1. Построить код Фано для ансамбля

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6\}, \quad p(x_1) = 0{,}36,\ p(x_2) = p(x_3) = 0{,}18,$$

$$p(x_4) = 0{,}12,\ p(x_5) = 0{,}09,\ p(x_6) = 0{,}07.$$

Решение приведено в таблице 5.1.

Таблица 5.1.

X	$p(x_i)$	X^0/X^1	X^{ij}	X^{ijk}	X^{ijkl}	код	m_i
x_1	0,36	0	0			00	2
x_2	0,18	0	1			01	2
x_3	0,18	1	0			10	2
x_4	0,12	1	1	0		110	3
x_5	0,09	1	1	1	0	1110	4
x_6	0,07	1	1	1	1	1111	4

Вычислим среднюю длину кодовых слов. Воспользуемся формулой

$$\bar{m}(X) =: \sum_{x_i \in X} m_i p(x_i).$$

В нашем примере имеем:

$$\bar{m}(X) = 2(0{,}36 + 0{,}18 + 0{,}18) + 3 \cdot 0{,}12 + 4(0{,}09 + 0{,}07) = 2,44$$

Средняя длина кодовых слов для построенного кода $\bar{m}(X) = 2,44$.

Вычислим энтропию ансамбля.

$$H(X) =: -\sum_{i-1}^{6} p(x_i) \log p(x_i) = 2,37. \qquad ¤$$

Пример 5.2. Постройте код Шеннона для ансамбля

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6\}, \quad p(x_1) = 0{,}36,\ p(x_2) = p(x_3) = 0{,}18,$$

$$p(x_4) = 0{,}12,\ p(x_5) = 0{,}09,\ p(x_6) = 0{,}07.$$

Решение. Проделаем последовательно шаги алгоритма Шеннона. Вычисления приведены в таблице 5.2.

Таблица 5.2.

X	$p(x_i)$	m_i	q_i	код	ус. код	m_i ус
x_1	0,36	2	0,00	00	00	2
x_2	0,18	3	0,36 = 0,010	010	01	2
x_3	0,18	3	0,54 = 0,100	100	100	3
x_4	0,12	4	0,72 = 0,1011	1011	101	3
x_5	0,09	4	0,84 = 0,1101	1101	110	3
x_6	0,07	4	0,93 = 0,1110	1110	111	3

Находим числа m_i, $i = 1, 2, \ldots M$, исходя из неравенств:

$$\frac{1}{2^{m_i}} \le p(x_i) < \frac{1}{2^{m_i - 1}}.$$

$$\frac{1}{2^{m_1}} \le 0{,}36 = \frac{36}{100} < \frac{1}{2^{m_1 - 1}},\ m_1 = 2.$$

$$\frac{1}{2^{m_2}} \le 0{,}18 = \frac{18}{100} < \frac{1}{2^{m_2 - 1}},\ m_2 = m_3 = 2.$$

$$\frac{1}{2^{m_4}} \le 0{,}12 = \frac{12}{100} < \frac{1}{2^{m_4 - 1}},\ m_4 = 4.$$

$$\frac{1}{2^{m_5}} \le 0{,}09 = \frac{9}{100} < \frac{1}{2^{m_5 - 1}},\ m_5 = 4.$$

$$\frac{1}{2^{m_6}} \le 0{,}07 = \frac{7}{100} < \frac{1}{2^{m_6 - 1}},\ m_6 = 4.$$

Находим кумулятивные вероятности по правилу

$$q_1 = 0,\ q_2 = p(x_1),\ q_3 = p(x_1) + p(x_2),\ \ldots,\ q_M = \sum_{i=1}^{M-1} p(x_i).$$

$$q_1 = 0,\ q_2 = 0{,}36,\ q_3 = 0{,}54,\ q_4 = 0{,}72,\ q_5 = 0{,}84,\ q_6 = 0{,}93.$$

Находим первые после запятой m_i знаков в разложении числа q_i

в двоичную дробь: $i = 1, 2, \ldots M$. Цифры этого разложения, стоящие после запятой, являются кодовым словом, соответствующим сообщению x_i. Ясно, что $q_1 = 0 = 0{,}00$. Далее действуем по правилу перевода десятичной дроби в двоичную.

0,	$36 \times 2 =$
0,	$72 \times 2 =$
1,	$44 \times 2 =$
0,	$88 \times 2 =$
1,	$76 \times 2 =$
1,	$52 \times 2 =$

$0{,}36 = 0{,}010;$

0,	$54 \times 2 =$
1,	$08 \times 2 =$
0,	$16 \times 2 =$
0,	$32 \times 2 =$
0,	$64 \times 2 =$
1,	$28 \times 2 =$

$0{,}54 = 0{,}100;$

Вычислим среднюю длину построенного кода:

$$\bar{m}(X) = 2 \cdot 0{,}36 + 3(0{,}18 + 0{,}18) + 4(0{,}12 + 0{,}09 + 0{,}07) = 2,92.$$

Вычислим среднюю длину построенного усеченного кода:

$$\bar{m}(X) = 2(0{,}36 + 0{,}18) + 3(0{,}18 + 0{,}12 + 0{,}09 + 0{,}07) = 2,46.$$

Напомним, что средняя длина кодовых слов кода по алгоритму Фано равна 2,44. ¤

Пример 5.3. Постройте код Хаффмана для ансамбля из 7 сообщений, вероятности которых равны 0,3; 0,2; 0,15; 0,15; 0,1; 0,05; 0,05.

Решение. Проделаем последовательно шаги алгоритма Хаффмана. Вычисления приведены в таблице 5.3.

Таблица 5.3.

x_i	$p(x_i)$
x_1	0,3
x_2	0,2
x_3	0,15
x_4	0,15
x_5	0,1
x_6	0,05
x_7	0,05

Код	m_i
11	2
01	2
101	3
100	3
001	3
0001	4
0000	4

$\updownarrow$ 1 0

Средняя длина кодовых слов равна 2,6:

$$\bar{m}(X) =: \sum_{x_i \in X} m_i p(x_i) =$$

$$= 2(0{,}3 + 0{,}2) + 3(0{,}15 + 0{,}15 + 0{,}1) + 4(0{,}05 + 0{,}05) = 2,6.$$

Нижняя граница согласно теореме 11.2 равна 2,354:

$$\bar{m}(X) > \frac{H(X)}{\log D} = H(X) = \sum_{i=1}^{7} p(x_i) \log p(x_i) =$$

$$= 0{,}3 \cdot \log 0{,}3 + 0{,}2 \cdot \log 0{,}2 + +2 \cdot 0{,}15 \cdot \log 0{,}15 +$$

$$+ 0{,}1 \cdot \log 0{,}1 + 2 \cdot 0{,}05 \cdot \log 0{,}05 = 2{,}354.$$

Кода со средней длиной меньшей, чем 2,6, не существует. ¤

Индивидуальные задания

1. Переведите с точностью до пяти значащих цифр после запятой в двоичную позиционную систему десятичное число m.

1. 1. $m = 31{,}75$. 1. 2. $m = 30{,}74$. 1. 3. $m = 23{,}9$
1. 4. $m = 25{,}82$. 1. 5. $m = 19{,}64$. 1. 6. $m = 26{,}93$.
1. 7. $m = 21{,}87$. 1. 8. $m = 27{,}46$. 1. 9. $m = 35{,}39$.
1. 10. $m = 28{,}45$. 1. 11. $m = 36{,}72$. 1. 12. $m = 29{,}91$.
1. 13. $m = 32{,}87$. 1. 14. $m = 34{,}66$. 1. 15. $m = 33{,}76$.

2. Элементами ансамбля $X = \{x_1, x_2, x_3, x_4, x_5\}$ являются числа. Каждому числу x_i ставится в соответствие в качестве кодового слова его двоичное разложение. Будет ли такой код однозначно декодируемым?

2. 1. $X = \{1, 2, 4, 17, 98\}$. 2. 2. $X = \{1, 4, 17, 21, 33\}$.
2. 3. $X = \{2, 24, 45, 91, 98\}$. 2. 4. $X = \{1, 2, 3, 11, 27\}$.
2. 5. $X = \{2, 13, 14, 15, 25\}$. 2. 6. $X = \{2, 5, 6, 22, 190\}$.
2. 7. $X = \{1, 2, 6, 8, 17, 38\}$. 2. 8. $X = \{0{,}5, 7, 22, 31\}$.
2. 9. $X = \{2, 3, 10{,}85, 343\}$. 2. 10. $X = \{3, 5, 9, 29, 473\}$.
2. 11. $X = \{1, 3, 12, 211, 422\}$. 2. 12. $X = \{2, 6, 7, 39, 47\}$.
2. 13. $X = \{2, 3, 4, 5, 6\}$. 2. 14. $X = \{1, 2, 5, 6, 13\}$.
2. 15. $X = \{0{,}1, 4, 9, 25\}$.

3. Является ли нижеприведенный код однозначно декодируемым? Будет ли этот код префиксным? Какова средняя длина кодового слова?

3.1. $\{0, 01\}$, $p(x_1) = 0,6$, $p(x_2) = 0,4$.
3.2. $\{0, 10, 11\}$, $p(x_1) = 0,4$, $p(x_2) = p(x_3) = 0,3$.
3.3. $\{0, 01, 11\}$, $p(x_1) = 0,2$, $p(x_2) = p(x_3) = 0,4$.
3.4. $\{0, 01, 10\}$, $p(x_1) = 0,1$, $p(x_2) = p(x_3) = 0,45$.
3.5. $\{10, 11, 110\}$, $p(x_1) = p(x_2) = 0,4$, $p(x_3) = 0,2$.
3.6. $\{00, 10, 01, 11\}$, $p(x_1) = p(x_2) = 0,3$, $p(x_3) = p(x_4) = 0,2$.
3.7. $\{00, 10, 11, 100, 110\}$, $p(x_1) = p(x_2) = p(x_3) = 0,1$, $p(x_4) = p(x_5) = 0,35$.
3.8. $\{1, 01, 001, 0001\}$, $p(x_1) = p(x_2) = 0,2$, $p(x_3) = p(x_4) = 0,3$.
3.9. $\{1, 10, 100, 1000\}$, $p(x_1) = p(x_2) = 0,1$, $p(x_3) = p(x_4) = 0,4$.
3.10. $\{10, 11, 110\}$, $p(x_1) = 0,3$, $p(x_2) = p(x_3) = 0,35$.
3.11. $\{00, 100, 101, 1111\}$, $p(x_1) = 0,1$, $p(x_2) = 0,25$, $p(x_3) = 0,35$, $p(x_4) = 0,3$.
3.12. $\{1, 110, 101, 1001\}$, $p(x_1) = p(x_2) = p(x_3) = 0,28$, $p(x_4) = 0,16$.
3.13. $\{10, 01, 101\}$, $p(x_1) = 0,31$, $p(x_2) = 0,03$, $p(x_3) = 0,66$.
3.14. $\{101, 11, 100, 10\}$, $p(x_1) = 0,15$, $p(x_2) = 0,45$, $p(x_3) = p(x_4) = 0,2$.
3.15. $\{10, 11, 100, 101, 110\}$, $p(x_1) = p(x_2) = p(x_3) = p(x_4) = p(x_5) = 0{,}2$.

4. Можно ли построить двоичный префиксный код для ансамбля $X = \{x_1, x_2, x_3, x_4, x_5, \}$, длины кодовых слов которого будут равны $\{m_1, m_2, m_3, m_4, m_5, \}$?

4.1. $m_1 = m_2 = 1$, $m_3 = m_4 = m_5 = 3$.
4.2. $m_1 = 1$, $m_2 = m_3 = m_4 = m_5 = 2$.
4.3. $m_1 = m_2 = 2$, $m_3 = m_4 = m_5 = 3$.
4.4. $m_1 = m_2 = m_3 = 2$, $m_4 = m_5 = 3$.
4.5. $m_1 = 1$, $m_2 = m_3 = m_4 = m_5 = 3$.
4.6. $m_1 = 1$, $m_2 = m_3 = 2$, $m_4 = m_5 = 3$.
4.7. $m_1 = 1$, $m_2 = 2$, $m_3 = m_4 = m_5 = 3$.
4.8. $m_1 = 1$, $m_2 = 2$, $m_3 = m_4 = 3$, $m_5 = 4$.
4.9. $m_1 = m_2 = 2$, $m_3 = 3$, $m_4 = m_5 = 4$.
4.10. $m_1 = 2$, $m_2 = 3$, $m_3 = m_4 = m_5 = 4$.
4.11. $m_1 = 2$, $m_2 = m_3 = 3$, $m_4 = m_5 = 4$.
4.12. $m_1 = 1$, $m_2 = m_3 = 2$, $m_4 = m_5 = 4$.
4.13. $m_1 = m_2 = m_3 = m_4 = m_5 = 3$.

4.14. $m_1 = 1$, $m_2 = 2$, $m_3 = 3$, $m_4 = m_5 = 4$.
4.15. $m_1 = 1$, $m_2 = 2$, $m_3 = m_4 = m_5 = 4$.

5. Для ансамбля $X = \{x_1, \ldots, x_7\}$ постройте коды Фано, Шеннона и Хаффмана. Найдите энтропию ансамбля и средние длины кодовых слов для каждого кода.

5.1. $p(x_1) = 0{,}4$, $p(x_2) = 0{,}2$, $p(x_3) = p(x_4) = p(x_5) = 0{,}1$, $p(x_6) = p(x_7) = 0{,}05$.
5.2. $p(x_1) = 0{,}4$, $p(x_2) = p(x_3) = 0{,}2$, $p(x_4) = p(x_5) = p(x_6) = p(x_7) = 0{,}05$.
5.3. $p(x_1) = 0{,}3$, $p(x_2) = 0{,}2$, $p(x_3) = 0{,}15$, $p(x_4) = p(x_5) = p(x_6) = 0{,}1$, $p(x_7) = 0{,}05$.
5.4. $p(x_1) = p(x_2) = 0{,}25$, $p(x_3) = 0{,}15$, $p(x_4) = p(x_5) = p(x_6) = 0{,}1$, $p(x_7) = 0{,}05$.
5.5. $p(x_1) = p(x_2) = 0{,}2$, $p(x_3) = p(x_4) = 0{,}15$, $p(x_5) = p(x_6) = p(x_7) = 0{,}1$.
5.6. $p(x_1) = p(x_2) = 0{,}3$, $p(x_3) = 0{,}15$, $p(x_4) = 0{,}1$, $p(x_5) = p(x_6) = p(x_7) = 0{,}05$.
5.7. $p(x_1) = 0{,}3$, $p(x_2) = p(x_3) = 0{,}2$, $p(x_4) = p(x_5) = 0{,}1$, $p(x_6) = p(x_7) = 0{,}05$.
5.8. $p(x_1) = 0{,}5$, $p(x_2) = 0{,}2$, $p(x_3) = 0{,}1$, $p(x_4) = p(x_5) = p(x_6) = p(x_7) = 0{,}05$.
5.9. $p(x_1) = 0{,}3$, $p(x_2) = 0{,}2$, $p(x_3) = p(x_4) = p(x_5) = p(x_6) = p(x_7) = 0{,}1$.
5.10. $p(x_1) = 0{,}3$, $p(x_2) = 0{,}25$, $p(x_3) = 0{,}15$, $p(x_4) = p(x_5) = 0{,}1$, $p(x_6) = p(x_7) = 0{,}05$.
5.11. $p(x_1) = 0{,}35$, $p(x_2) = 0{,}25$, $p(x_3) = 0{,}15$, $p(x_4) = 0{,}1$, $p(x_5) = p(x_6) = p(x_7) = 0{,}05$.
5.12. $p(x_1) = 0{,}4$, $p(x_2) = p(x_3) = 0{,}15$, $p(x_4) = p(x_5) = 0{,}1$, $p(x_6) = p(x_7) = 0{,}05$.
5.13. $p(x_1) = 0{,}55$, $p(x_2) = 0{,}2$, $p(x_3) = 0{,}07$, $p(x_4) = p(x_5) = p(x_6) = 0{,}05$, $p(x_7) = 0{,}03$.
5.14. $p(x_1) = p(x_2) = 0{,}3$, $p(x_3) = p(x_4) = p(x_5) = 0{,}1$, $p(x_6) = p(x_7) = 0{,}05$.
5.15. $p(x_1) = 0{,}4$, $p(x_2) = 0{,}2$, $p(x_3) = 0{,}15$, $p(x_4) = 0{,}1$, $p(x_5) = p(x_6) = p(x_7) = 0{,}05$.

ГЛОССАРИЙ

Код Фано

Алгоритм Фано сводится к последовательному выполнению следующих шагов:

1. Сообщения ансамбля $X = \{x_1, \ldots, x_M\}$ упорядочиваем по убыванию вероятностей:

$$p(x_1) \geq p(x_2) \geq \ldots \geq p(x_M).$$

2. Разбиваем ансамбль X на два подансамбля $X^{(0)}$ и $X^{(1)}$ с помощью некоторого порогового целого числа $1 \leq k^{(1)} \leq M - 1$, так, чтобы абсолютная величина

$$K^{(1)} = \left| \sum_{i=1}^{k^{(1)}} p(x_i) - \sum_{i=k^{(1)}+1}^{M} p(x_i) \right| \tag{12.4}$$

достигала наименьшего возможного значения. Сообщениям подансамбля $X^{(0)}$ приписываем 0, сообщениям подансамбля $X^{(1)}$ приписываем 1.

3. Если подансамбли $X^{(0)}$, $X^{(1)}$ состоит более чем из двух сообщений, то разбиваем множество сообщений каждого из них на две части $X^{(00)}$, $X^{(01)}$ и $X^{(10)}$, $X^{(11)}$, соответственно, с помощью пороговых целых чисел

$$1 \leq k^{(11)} \leq k^{(1)} - 1, \; k^{(1)} \leq k^{(12)} \leq M - 1,$$

так, чтобы абсолютные величины

$$K^{(21)} = \left| \sum_{i=1}^{k^{(11)}} p(x_i) - \sum_{i=k^{(11)}+1}^{k^{(1)}} p(x_i) \right|, \tag{12.5}$$

$$K^{(22)} = \left| \sum_{i=k^{(1)}+1}^{k^{(12)}} p(x_i) - \sum_{i=k^{(12)}+1}^{M} p(x_i) \right| \tag{12.6}$$

достигали наименьших возможных значений. Сообщениям из $X^{(00)}$, $X^{(10)}$ с нулевыми последними индексами приписываем 0, сообщениям из подгрупп $X^{(01)}$, $X^{(11)}$ с единичными последними индексами приписываем 1.

Если подансамбль $X^{(ij)}$, $i, j \in \{0, 1\}$ состоит более чем из одного сообщения, то переходим к шагу 4. Если все подансамбли содержат по одному сообщению, то переходим к шагу 5.

4. Если есть подансамбли, состоящие более чем из одного сообщения, то разбиваем каждый из них на две подансамбля, исходя из соотношения, аналогичного (12.4). Сообщениям из подансамблей с нулевыми последними индексами приписываем нуль, сообщениям из подансамблей с единичными последними индексами приписываем единицу.

Если все образовавшиеся подансамбли состоят из одного сообщения, то переходим к шагу 5.

Если есть подансамбли, состоящие более чем из одного сообщения, то то повторяем шаг 4.

5. Если образовавшиеся подансамбли состоят из одного сообщения, то последовательно, начиная с первой метки, выписываем нули и единицы, относящиеся к каждому сообщению ансамбля X.

В итоге получается двоичный префиксный код для заданного ансамбля X.

Код Шеннона

Алгоритм Шеннона сводится к последовательному выполнению следующих шагов:

1. Сообщения ансамбля $X = \{x_1, \ldots, x_M\}$ упорядочиваем по убыванию вероятностей:

$$p(x_1) \geq p(x_2) \geq \ldots \geq p(x_M).$$

2. Находим числа m_i, $i = 1, 2, \ldots M$, исходя из неравенств:

$$\frac{1}{2^{m_i}} \leq p(x_i) < \frac{1}{2^{m_i - 1}}. \tag{12.7}$$

3. Сопоставим каждому сообщению кумулятивную вероятность по правилу

$$q_1 = 0, \; q_2 = p(x_1), \; q_3 = p(x_1) + p(x_2), \; q_M = \sum_{i=1}^{M-1} p(x_i). \tag{12.8}$$

4. Находим первые после запятой m_i знаков в разложении числа q_i в двоичную дробь: $i = 1, 2, \ldots M$. Цифры этого разложения, стоящие после запятой, являются кодовым словом, соответствующим сообщению x_i.

5. Если необходимо, производим операцию усечения. ???

Лемма 12.3. *Построенный по алгоритму Шеннона код является префиксным.*

Лемма 12.4. *Средняя длина кодовых слов ансамбля X кода, построенного по алгоритму Шеннона не превосходит $1 + H(X)$.*

Оптимальный побуквенный код — код Хаффмена

Предложенный Хаффменом алгоритм построения оптимальных неравномерных кодов — одно из самых важных достижений теории информации как с теоретической, так и с прикладной точек зрения. Трудно поверить, но этот алгоритм был придуман в 1952 г. студентом Дэвидом Хаффменом в процессе выполнения домашнего задания.

Пусть дан ансамбль

$$X = \{x_1, \dots x_M \mid p(x_1) \geq p(x_2) \geq \dots \geq p(x_M)\},$$

в котором M сообщений. Построим новый ансамбль X', состоящий из $M-1$ сообщений $\{x'_i, x'_2, \dots, x'_{M-1}\}$ с вероятностями

$$p(x'_i) = \begin{cases} p(x_i), \ i = 1, 2, \dots, M-2, \\ p(x_{M-1}) + p(x_M), \ i = M-1. \end{cases}$$

Любой декодируемый префиксный код для ансамбля X' можно превратить в декодируемый код для ансамбля X следующим образом. К кодовому слову, кодирующему сообщение x'_{M-1}, приписывают символы 0 и 1 для получения кодовых слов, кодирующих сообщения x_{M-1}, x_M. Остальные кодовые слова у сообщений x'_j и x_j, $j = 1, 2, \dots, M-2$, одинаковые.

Теперь нетрудно указать последовательную процедуру построения оптимального префиксного кода. Для этого необходимо обосновать, что оптимизация на каждом шаге приведет к оптимизации кода. Это обоснование выполняется с помощью следующего утверждения.

Лемма 12.5. *Если оптимален однозначно декодируемый префиксный код для ансамбля X', то оптимален полученный из него префиксный код для ансамбля X.*

Таким образом, задача построения оптимального префиксного кода сводится к задаче построения оптимального префиксного кода для ансамбля, содержащего на одно сообщение меньше. В этом ансамбле снова

можно выделить два наименее вероятных сообщения и, объединяя их, получить новый ансамбль, содержащий теперь уже на два сообщения меньше, чем исходный. Очевидно, что таким образом можно дойти до ансамбля, содержащего всего два слова, оптимальным кодом для которого является просто 0 для одного сообщения и 1 для другого. Описанный метод построения оптимального префиксного кода называется *методом Хаффмена.*

Алгоритм Хаффмана. 1. Сообщения ансамбля $X = \{x_1, \dots, x_M\}$ упорядочиваем по убыванию вероятностей:

$$p(x_1) \geq p(x_2) \geq \dots \geq p(x_M).$$

2. Два наименее вероятные сообщения склеиваем в одно и приписываем ему суммарную вероятность склеенных сообщений.

3. Если полученный ансамбль состоит из двух или более сообщений, то переходим к первому пункту, т. е. в полученном ансамбле вероятности упорядочиваем по убыванию. Затем два наименее вероятные сообщения склеиваем в одно и приписываем ему суммарную вероятность склеенных сообщений.

4. Через конечное число шагов получим ансамбль, состоящий из одного сообщения.

5. Двигаясь от заключительного ансамбля к начальному и расставляя кодовые символы 0 и 1 по принципу «вверх — 1», «вниз —0» получаем оптимальный код.

6. ТРОИЧНЫЕ КОДЫ

Примеры решения задач

Пример 6.1. Перевести с точностью до пяти значащих цифр после запятой в троичную систему счисления десятичное число 29,38.

Решение. Для начала переведем целую часть. Для перевода целое десятичное число делят нацело с остатком (целочисленное деление) на 3 до тех пор, пока частное больше нуля. Остатки, записанные слева направо от последнего к первому являются целым несимметричным троичным эквивалентом целого десятичного числа.

$$29 = 3 \cdot 9 + 2,\ 9 = 3 \cdot 3 + 0,\ 3 = 3 \cdot 1 + 0,\ 1 = 3 \cdot 0 + 1.$$

Теперь, записав все остатки от последнего к первому справа налево, получим результат $29_{10} = 1002_3$.

Переведем дробную часть десятичного числа в троичную систему счисления.

$$\begin{array}{r|l} 0, & 38 \\ & 3 \\ \hline \mathbf{1}, & 14 \\ & 3 \\ \hline \mathbf{0}, & 42 \\ & 3 \\ \hline \mathbf{1}, & 26 \\ & 3 \\ \hline \mathbf{0}, & 78 \\ & 3 \\ \hline \mathbf{2}, & 34 \end{array}$$

Перевод дробного числа осуществляется по следующему алгоритму:

— дробная часть десятичной дроби умножается на основание 3;

— в полученном произведении выделяется целая часть;

— алгоритм завершается, если дробная часть полученного произведения равна нулю или если достигнута требуемая точность вычислений. Иначе вычисления продолжаются с первого шага.

Сложим вместе дробную и целую части полученные ранее и получим:

ОТВЕТ: $29{,}38_{10} = 1002{,}10102_3$. ¤

Пример 6.2. Перевести с точностью до пяти значащих цифр после запятой в двоичную систему счисления троичное число $m = 2101{,}12$.

Решение. Переведем число из троичной системы счисления в десятичную: $m = 2 \cdot 3^3 + 1 \cdot 3^2 + 1 \cdot 3^0 + 1 \cdot 3^{-1} + 2 \cdot 3^{-2} = 62{,}556_{10}$

Дальше переводим число из десятичной в двоичную:

$m = 62{,}556_{10} = 111110{,}10001_2$

ОТВЕТ: $2101{,}12_3 = 111110{,}10001_2$. ¤

Пример 6.3. Элементами ансамбля $X = \{7, 10, 30, 56, 270\}$ являются числа. Каждому числу x_i, ставится в соответствие в качестве кодового слова его троичное разложение. Будет ли такой код однозначно декодируемым?

Решение. На первом шаге получим кодовые слова для каждого сообщения ансамбля, для этого преобразуем числа из десятичной системы счисления в троичную:

$7_{10} = 21_3$, $10_{10} = 101_3$, $30_{10} = 1010_3$ $56_{10} = 2002_3$, $270_{10} = 101021_3$.

Мы получили кодовые слова для ансамбля X:

$$21,\ 101,\ 1010,\ 2002,\ 101021.$$

Теперь проверим является ли этот код однозначно декодируемым. Кодовое слово 101021 можно декодировать по разному: как последовательность сообщений 30 и 7, или как сообщение 270. Следовательно, код является неоднозначно декодируемым.

ОТВЕТ: Нет, не будет. ¤

Пример 6.4. Для ансамбля X постройте троичные коды Фано, Шеннона и Хаффмана. Найдите энтропию ансамбля и средние длины кодовых слов.

$$X = \{x_1, \ldots, x_7\},\ p(x_1) = 0{,}4;\ p(x_2) = 0{,}2;$$

$$p(x_3) = p(x_4) = p(x_5) = 0{,}1;\ p(x_6) = p(x_7) = 0{,}05.$$

Решение. Вычислим энтропию ансамбля:

$$H(X) = -0{,}4 \log_3 0{,}4 - 0{,}2 \log_3 0{,}2 - 3 \cdot 0{,}1 \log_3 0{,}1 - 2 \cdot 0{,}05 \log_3 0{,}05 =$$

$$0{,}334 + 0{,}293 + 0{,}629 + 0{,}273 = 1{,}529.$$

Строим троичные коды по алгоритму Фано.

X	$p(x_i)$	$X^0/X^1/X^2$	X^{ij}	X^{ijk}	код	m_i
x_1	0,4	0			0	1
x_2	0,2	1	0		10	2
x_3	0,1	1	1		11	2
x_4	0,1	2	0		12	2
x_5	0,1	2	1		20	2
x_6	0,05	2	2	0	21	2
x_7	0,05	2	2	1	22	2

Находим среднюю длину кодовых слов:
$\bar{m}(x) = 1 \cdot 0{,}4 + 2 \cdot 0{,}6 = 1{,}6$.

Строим троичные коды по алгоритму Шеннона.

Теперь построим код Шеннона. Находим числа m_i, исходя из неравенств:

$$\frac{1}{3^{m_i}} \leq p(x_i) < \frac{1}{3^{m_i - 1}},$$

$p(x_i)$	0,4	0,2	0,1	0,05
m_i	1	2	3	4

.

Находим кумулятивные вероятности:

$$q_1 = 0;\ q_2 = 0{,}4;\ q_3 = 0{,}6;\ q_4 = 0{,}7;\ q_5 = 0{,}8;\ q_6 = 0{,}9;\ q_7 = 0{,}95.$$

Находим первые после запятой m_i знаков в разложении числа i в троичную дробь. Цифры этого разложения, стоящие после запятой, являются кодовым словом, соответствующим сообщению x_i.

X	$p(x_i)$	m_i	q_i	Код	Ус. код	m_i ус.
x_1	0,4	1	0,0	0	0	1
x_2	0,2	2	$0{,}4 = 0{,}01$	01	01	2
x_3	0,1	3	$0{,}6 = 0{,}121$	121	121	3
x_4	0,1	3	$0{,}7 = 0{,}200$	200	200	3
x_5	0,1	3	$0{,}8 = 0{,}210$	210	210	3
x_6	0,05	4	$0{,}9 = 0{,}2200$	2200	220	3
x_7	0,05	4	$0{,}95 = 0{,}2211$	2211	221	3

.

Находим среднюю длину кодовых слов:
Для обычного кода $\bar{m} = 1 \cdot 0{,}4 + 0{,}2 \cdot 2 + 0{,}3 \cdot 3 + 0{,}1 \cdot 4 = 2{,}1$.
Для усеченного $\bar{m} = 2{,}0$.

Строим троичные коды по алгоритму Хаффмана.

Проделаем последовательно шаги алгоритма Хаффмана, смотри глоссарий. Вычисления приведены в таблице:

X	$p(x_i)$	Код	m_i
x_1	0,4	0	1
x_2	0,2	1	1
x_3	0,1	20	2
x_4	0,1	21	2
x_5	0,1	220	3
x_6	0,05	221	3
x_7	0,05	222	3

Средняя длина слова равна $\bar{m} = 0{,}6 \cdot 1 + 0{,}2 \cdot 2 + 0{,}2 \cdot 3 = 1{,}6$

Пример 6.5. Для ансамбля $X = \{x_1, \ldots, x_6\}$ даны двоичные кодовые слова. Преобразовать слова из двоичной в троичную системы счисления. Сравнить средние длины кодовых слов, построенные в двоичных и троичных систем счисления. Считать что все сообщения равновероятны.

$x_1 = 01$, $x_2 = 100$, $x_3 = 1111$, $x_4 = 0110$, $x_5 = 11110$, $x_6 = 10011$.

Решение. Переведем коды ансамбля в десятеричную систему счисления: $x_1 = 01_2 = 1_{10}$, $x_2 = 100_2 = 4_{10}$, $x_3 = 1111_2 = 15_{10}$, $x_4 = 0110_2 = 6_{10}$, $x_5 = 11110_2 = 30_{10}$, $x_6 = 10011_2 = 19_{10}$.
Переведем из десятичной системы счисления в троичную как в примере 1:

$x_1 = 1_{10} = 1_3$, $x_2 = 4_{10} = 10_3$, $x_3 = 15_{10} = 120_3$, $x_4 = 6_{10} = 20_3$, $x_5 = 30_{10} = 1010_3$, $x_6 = 19_{10} = 201_3$.
Найдем средние длины кодовых слов:
для двоичной $\bar{m}_2 = (2 + 3 + 4 + 4 + 5 + 5)/6 = 23/6 =\approx 3{,}83$.
для троичной $\bar{m}_3 = (1 + 2 + 3 + 2 + 4 + 3)/6 = 15/6 = 2{,}5$.

Индивидуальные задания

1. Переведите с точностью до пяти значащих цифр после запятой в троичную систему счисления десятичное число m.

1.1. $m = 27{,}58$. 1.6. $m = 31{,}59$. 1.11. $m = 32{,}78$
1.2. $m = 34{,}89$. 1.7. $m = 35{,}12$. 1.12. $m = 37{,}21$.
1.3. $m = 38{,}37$. 1.8. $m = 28{,}56$. 1.13. $m = 39{,}62$.
1.4. $m = 40{,}55$. 1.9. $m = 34{,}47$. 1.14. $m = 37{,}46$.
1.5. $m = 29{,}86$. 1.10. $m = 31{,}69$. 1.15. $m = 34{,}84$.

2. Переведите с точностью до пяти значащих цифр после запятой в двоичную систему счисления троичное число m.

2.1. $m = 2101{,}12$. 2.6. $m = 1022{,}021$. 2.11. $m = 1022{,}022$
2.2. $m = 1211{,}2111$. 2.7. $m = 2211{,}212$. 2.12. $m = 22011{,}01$.
2.3. $m = 2012{,}102$. 2.8. $m = 1222{,}202$. 2.13. $m = 12121{,}11$.
2.4. $m = 1121{,}201$. 2.9. $m = 1111{,}221$. 2.14. $m = 20021{,}21$.
2.5. $m = 2122{,}1102$. 2.10. $m = 2021{,}122$. 2.15. $m = 21221{,}22$.

3. Элементами ансамбля $X = \{x_1, x_2, x_3, x_4, x_5\}$ являются числа. Каждому числу x_i ставится в соответствие в качестве кодового слова его троичное разложение. Будет ли такой код однозначно декодируемым?

3.1. $X = \{2, 5, 21, 46, 194\}$. 3.2. $X = \{1, 8, 11, 59, 201\}$.
3.3. $X = \{1, 12, 34, 71, 115\}$. 3.4. $X = \{4, 7, 13, 68, 171\}$.
3.5. $X = \{5, 10, 19, 91, 289\}$. 3.6. $X = \{3, 8, 21, 71, 182\}$.
3.7. $X = \{6, 16, 24, 67, 178\}$. 3.8. $X = \{2, 9, 22, 78, 136\}$.
3.9. $X = \{4, 11, 26, 82, 134\}$. 3.10. $X = \{5, 7, 29, 51, 143\}$.
3.11. $X = \{4, 11, 32, 65, 103\}$. 3.12. $X = \{2, 14, 37, 83, 126\}$.
3.13. $X = \{8, 12, 16, 28, 340\}$. 3.14. $X = \{7, 13, 27, 43, 110\}$.
3.15. $X = \{4, 16, 21, 11, 102\}$.

4. Для ансамбля $X = \{x_1, \ldots, x_7\}$ постройте троичные коды Фано, Шеннона и Хаффмана. Найдите энтропию ансамбля и средние длины кодовых слов для каждого кода.

4.1. $p(x_1) = p(x_2) = 0{,}3$, $p(x_3) = p(x_4) = p(x_5) = 0{,}1$, $p(x_6) = p(x_7) = 0{,}05$.
4.2. $p(x_1) = p(x_2) = 0{,}25$, $p(x_3) = p(x_4) = p(x_5) = p(x_6) = p(x_7) = 0{,}1$.
4.3. $p(x_1) = 0{,}3$, $p(x_2) = 0{,}2$, $p(x_3) = 0{,}15$, $p(x_4) = p(x_5) = p(x_6) = 0{,}1$, $p(x_7) = 0{,}05$.
4.4. $p(x_1) = 0{,}3$, $p(x_2) = 0{,}25$, $p(x_3) = 0{,}15$, $p(x_4) = p(x_5) = 0{,}1$, $p(x_6) = p(x_7) = 0{,}05$.
4.5. $p(x_1) = 0{,}5$, $p(x_2) = 0{,}2$, $p(x_3) = 0{,}1$, $p(x_4) = p(x_5) = p(x_6) = p(x_7) = 0{,}05$.

4.6. $p(x_1) = 0{,}3$, $p(x_2) = 0{,}2$, $p(x_3) = p(x_4) = p(x_5) = p(x_6) = p(x_7) = 0{,}1$.
4.7. $p(x_1) = 0{,}4$, $p(x_2) = p(x_3) = 0{,}15$, $p(x_4) = p(x_5) = 0{,}1$, $p(x_6) = p(x_7) = 0{,}05$.
4.8. $p(x_1) = p(x_2) = 0{,}2$, $p(x_3) = 0$, $p(x_4) = 0{,}3$, $p(x_5) = p(x_6) = p(x_7) = 0{,}1$.
4.9. $p(x_1) = 0{,}55$, $p(x_2) = 0{,}2$, $p(x_3) = 0{,}07$, $p(x_4) = p(x_5) = p(x_6) = 0{,}05$, $p(x_7) = 0{,}03$.
4.10. $p(x_1) = 0{,}3$, $p(x_2) = p(x_3) = 0{,}2$, $p(x_4) = p(x_5) = 0{,}1$, $p(x_6) = p(x_7) = 0{,}05$.
4.11. $p(x_1) = p(x_2) = 0{,}3$, $p(x_3) = p(x_4) = p(x_5) = 0{,}1$, $p(x_6) = p(x_7) = 0{,}05$.
4.12. $p(x_1) = p(x_2) = 0{,}25$, $p(x_3) = 0{,}15$, $p(x_4) = p(x_5) = p(x_6) = 0{,}1$, $p(x_7) = 0{,}05$.
4.13. $p(x_1) = 0{,}4$, $p(x_2) = 0{,}2$, $p(x_3) = 0{,}15$, $p(x_4) = 0{,}1$, $p(x_5) = p(x_6) = p(x_7) = 0{,}05$.
4.14. $p(x_1) = 0{,}35$, $p(x_2) = 0{,}25$, $p(x_3) = 0{,}15$, $p(x_4) = 0{,}1$, $p(x_5) = p(x_6) = p(x_7) = 0{,}05$.
4.15. $p(x_1) = p(x_2) = 0{,}3$, $p(x_3) = 0{,}15$, $p(x_4) = 0{,}1$, $p(x_5) = p(x_6) = p(x_7) = 0{,}05$.

5. Для ансамбля $X = \{x_1, \ldots, x_6\}$ даны двоичные кодовые слова. Преобразуйте слова из двоичной в троичную системы счисления. Сравните средние длины кодовых слов построенные в двоичных и троичных систем счисления. Считать что все сообщения равновероятны.

5.1. $X = \{10, 100, 101, 110, 10110, 10010\}$.
5.2. $X = \{111, 110, 100, 101, 10100, 11010\}$.
5.3. $X = \{100, 101, 1100, 1010, 10110, 10101\}$.
5.4. $X = \{111, 110, 1101, 110, 10111, 10000\}$.
5.5. $X = \{001, 000, 0010, 0101, 01011, 01001\}$.
5.6. $X = \{010, 011, 0101, 0110, 01011, 01001\}$.
5.7. $X = \{11, 101, 1001, 0011, 10011, 11101\}$.
5.8. $X = \{00, 101, 1011, 0100, 10000, 10001\}$.
5.9. $X = \{01, 110, 1000, 1010, 00000, 11010\}$.
5.10. $X = \{101, 111, 0000, 1110, 01011, 11111\}$.
5.11. $X = \{11, 111, 1110, 0100, 11011, 01011\}$.
5.12. $X = \{010, 110, 0110, 1101, 01111, 11001\}$.

5. 13. $X = \{101, 011, 1011, 1100, 10011, 01101\}$.
5. 14. $X = \{111, 100, 1011, 0110, 11001, 01001\}$.
5. 15. $X = \{10, 101, 1010, 1010, 10101, 10100\}$.

ГЛОССАРИЙ

Код Фано

Алгоритм Фано сводится к последовательному выполнению следующих шагов:

1. Сообщения ансамбля $X = \{x_1, \ldots, x_M\}$ упорядочиваем по убыванию вероятностей:

$$p(x_1) \geq p(x_2) \geq \ldots \geq p(x_M).$$

2. Разбиваем ансамбль X на три подансамбля $X^{(0)}$, $X^{(1)}$ и $X^{(2)}$ так, чтобы абсолютная величина

$$K_1 = \mid \sum_{x_i \in X^{(0)}} p(x_i) - \sum_{x_i \in X^{(1)}} p(x_i) \mid + \mid \sum_{x_i \in X^{(0)}} p(x_i) - \sum_{x_i \in X^{(2)}} p(x_i) \mid + \\ + \mid \sum_{x_i \in X^{(1)}} p(x_i) - \sum_{x_i \in X^{(2)}} p(x_i) \mid \tag{6.1}$$

достигала наименьшего возможного значения. Сообщениям подансамбля $X^{(0)}$ приписываем 0, сообщениям подансамбля $X^{(1)}$ приписываем 1, сообщениям подансамбля $X^{(2)}$ приписываем 2.

3. Если подансамбль $X^{(0)}$ состоит из четырех и более сообщений, то разбиваем множество его сообщений на три части $X^{(00)}$, $X^{(01)}$ и $X^{(02)}$ таким образом, чтобы соответствующая абсолютная величина K_2, вычисленная по формуле

$$K_2 = \mid \sum_{x_i \in X^{(00)}} p(x_i) - \sum_{x_i \in X^{(01)}} p(x_i) \mid + \mid \sum_{x_i \in X^{(00)}} p(x_i) - \sum_{x_i \in X^{(02)}} p(x_i) \mid + \\ + \mid \sum_{x_i \in X^{(01)}} p(x_i) - \sum_{x_i \in X^{(02)}} p(x_i) \mid, \tag{6.1}$$

достигала наименьшего возможного значения. Сообщениям из ансамбля $X^{(00)}$ приписываем 0, сообщениям из ансамбля $X^{(01)}$ приписываем 1, а сообщениям из ансамбля $X^{(02)}$ приписываем 2.

Аналогично поступаем с ансамблями $X^{(1)}$ и $X^{(2)}$.

4. Если на каком-то этапе получаем подансамбль из четырех и более сообщений, то к нему применяем шаг 3.

5. Если на каком-то этапе получили подансамбль из одного, двух или трех сообщений, то для них дописываем в кодовые слова 0, 1 и 2. Для этих подансамблей процесс кодирования прекращается.

В итоге получается троичный префиксный код для заданного ансамбля X.

Код Шеннона

Алгоритм Шеннона сводится к последовательному выполнению следующих шагов:

1. Сообщения ансамбля $X = \{x_1, \ldots, x_M\}$ упорядочиваем по убыванию вероятностей:

$$p(x_1) \geq p(x_2) \geq \ldots \geq p(x_M).$$

2. Находим числа m_i, $i = 1, 2, \ldots M$, исходя из неравенств:

$$\frac{1}{3^{m_i}} \leq p(x_i) < \frac{1}{3^{m_i - 1}}. \tag{1.2}$$

3. Сопоставим каждому сообщению кумулятивную вероятность по правилу:

$$q_1 = 0,\ q_2 = p(x_1),\ q_3 = p(x_1) + p(x_2),\ q_M = \sum_{i=1}^{M-1} p(x_i). \tag{1.3}$$

4. Находим первые после запятой m_i знаков в разложении числа q_i в троичную дробь: $i = 1, 2, \ldots M$. Цифры этого разложения, стоящие после запятой, являются кодовым словом, соответствующим сообщению x_i.

5. Если необходимо, производим операцию усечения.

Оптимальный код Хаффмена

1. Сообщения ансамбля $X = \{x_1, \ldots, x_M\}$ упорядочиваем по убыванию вероятностей:

$$p(x_1) \geq p(x_2) \geq \ldots \geq p(x_M).$$

2. Три наименее вероятные сообщения склеиваем в одно и приписываем ему суммарную вероятность склеенных сообщений.

3. Если полученный ансамбль состоит из трех или более сообщений, то переходим к первому и второму пунктам, т. е. в полученном ансамбле вероятности упорядочиваем по убыванию. Затем три наименее вероятные сообщения склеиваем в одно и приписываем ему суммарную вероятность склеенных сообщений.

4. Через конечное число шагов получим ансамбль, состоящий из одного или двух сообщения.

5. Двигаясь от заключительного ансамбля к начальному и расставляя кодовые символы 0, 1 и 2 по принципу «вверх — 1», «вниз — 0» и «центр — 2» получаем оптимальный код.