

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ “ЛЬВІВСЬКА ПОЛІТЕХНІКА”
Інститут комп’ютерних наук та інформаційних технологій
Кафедра систем штучного інтелекту

Звіт до лабораторної роботи №3
з дисципліни «*Теорія інформації*»

Виконав:

ст. гр. КН-211

Головень Ростислав

Викладач:

д.т.н. Косаревич Р. Я.

Лабораторна робота №3

СТИСНЕННЯ ДАНИХ

Мета роботи: застосування методів кодування повідомлень для стиснення їх даних.

Завдання:

1. Побудувати бінарні коди Хаффмана та Шенона для:

а) розподілу випадкової величини заданої у табл.1.1 , № варіанту відповідно до журналу групи.

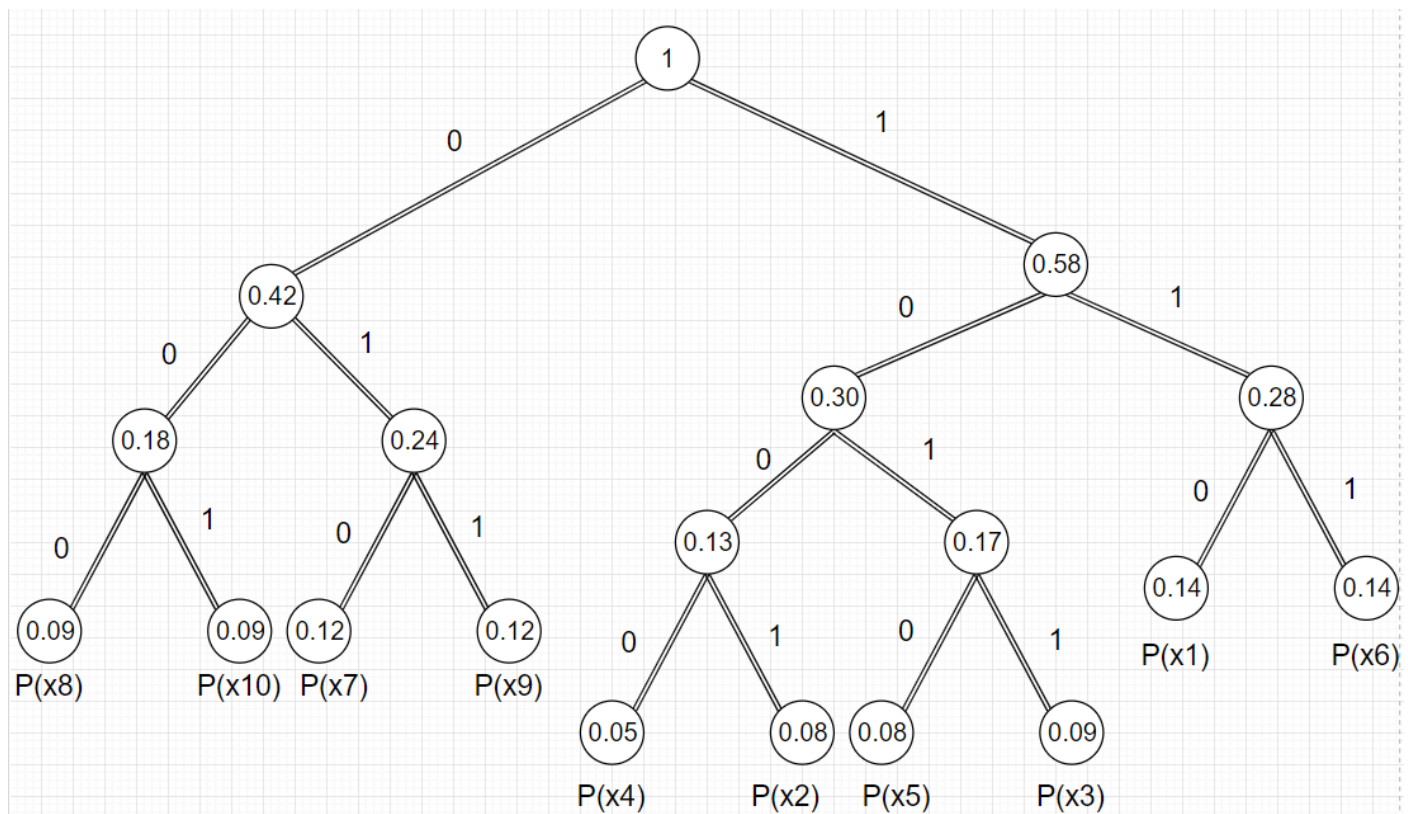
Варіант 5

P_{x_1}	P_{x_2}	P_{x_3}	P_{x_4}	P_{x_5}	P_{x_6}	P_{x_7}	P_{x_8}	P_{x_9}	$P_{x_{10}}$
0,14	0,08	0,09	0,05	0,08	0,14	0,12	0,09	0,12	0,09

• Код Гаффмана:

Посортуємо випадкові величини у порядку зростання:

P_{x_4}	P_{x_2}	P_{x_5}	P_{x_3}	P_{x_8}	$P_{x_{10}}$	P_{x_7}	P_{x_9}	P_{x_1}	P_{x_6}
0,05	0,08	0,08	0,09	0,09	0,09	0,12	0,12	0,14	0,14



Символ	Ймовірність	Кодове слово	Довжина кодового слова
x ₁	0,14	110	3
x ₂	0,08	1001	4
x ₃	0,09	1011	4
x ₄	0,05	1000	4
x ₅	0,08	1010	4
x ₆	0,14	111	3
x ₇	0,12	010	3
x ₈	0,09	000	3
x ₉	0,12	011	3
x ₁₀	0,09	001	3

Бінарний код Гаффмана: 1101001101110001010111010000011001

Ентропія: $\sum_{i=1}^{10} p(x_i) * \log_2(p(x_i)) = 3.265$

Очікувана довжина коду:

$0,14*3+0,08*4+0,09*4+0,05*4+0,08*4+0,14*3+0,12*3+0,09*3+0,12*3+0,09*3= 3.3$

• Код Шенона:

Символ	Ймовірність	F(x)	$\bar{F}(x)$	$\bar{F}(x)$ бінарне	Кодове слово	Довжина кодового слова
x ₁	0,14	0,14	0,07	0.00010010	0.0001	4
x ₂	0,08	0,22	0,18	0.00101110	0.00101	5
x ₃	0,09	0,31	0,265	0.01000100	0.01000	5
x ₄	0,05	0,36	0,335	0.01010110	0.010101	6
x ₅	0,08	0,44	0,4	0.01100110	0.01100	5
x ₆	0,14	0,58	0,51	0.10000011	0.1000	4
x ₇	0,12	0,70	0,64	0.10100100	0.10100	5
x ₈	0,09	0,79	0,745	0.10111111	0.10111	5
x ₉	0,12	0,91	0,85	0.11011010	0.11011	5
x ₁₀	0,09	1	0,955	0.11110100	0.11110	5

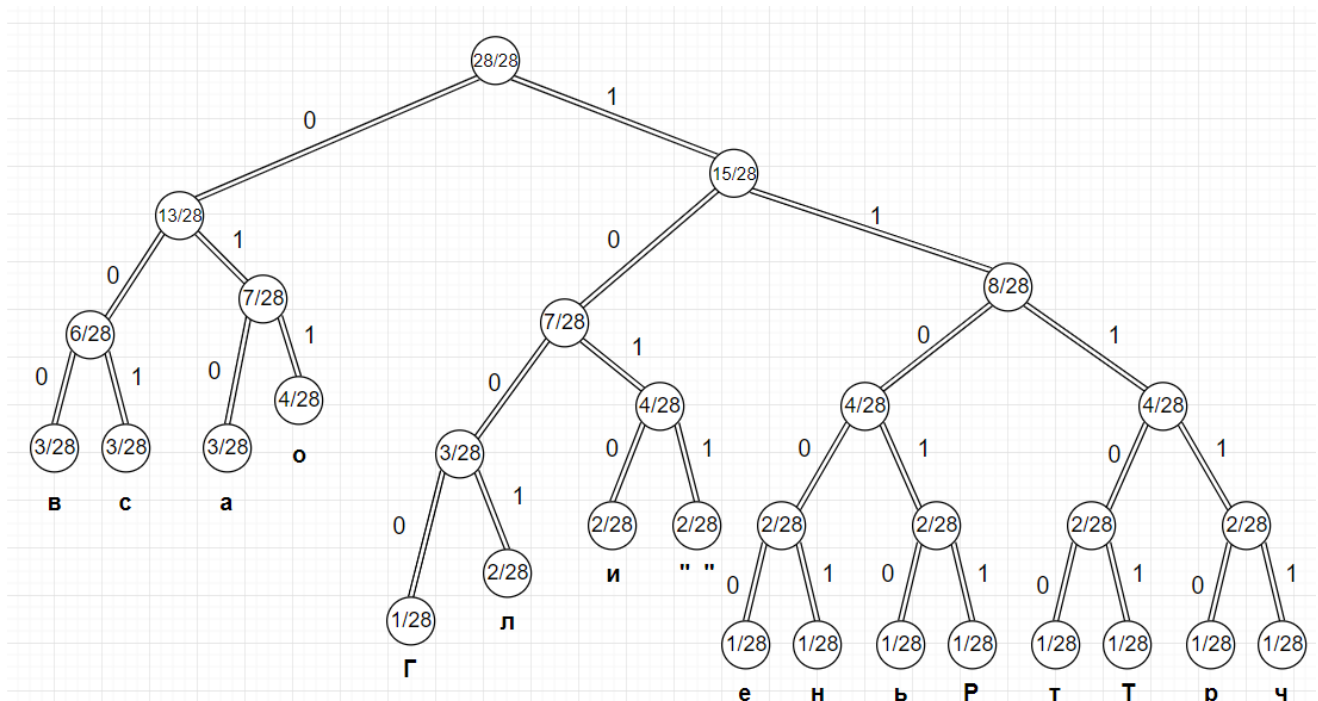
Ентропія: $\sum_{i=1}^{10} p(x_i) * \log_2(p(x_i)) = 3.265$

б) послідовності символів із власного П.І.Б.

• Код Гаффмана:

Головень Ростислав Тарасович

Символ	Ймовірність	Кодове слово	Довжина кодового слова
Г	1/28=3.57%	1000	4
о	4/28=14.28%	011	3
л	2/28=7.14%	1001	4
в	3/28=10.71%	000	3
е	1/28=3.57%	11000	5
н	1/28=3.57%	11001	5
ь	1/28=3.57%	11010	5
“ “	2/28=7.14%	1011	4
Р	1/28=3.57%	11011	5
с	3/28=10.71%	001	3
т	1/28=3.57%	11100	5
и	2/28=7.14%	1010	4
а	3/28=10.71%	010	3
Т	1/28=3.57%	11101	5
р	1/28=3.57%	11110	5
ч	1/28=3.57%	11111	5



Бінарний код Гаффмана:

100001110010110001100011001110101011110110110011110010100011001010000111010101111001000101100010101111

Ентропія: $\sum_{i=1}^n p(x_i) * \log_2(p(x_i)) = 3.798$

Очікувана довжина коду: 3.82

• Код Шенона:

Головень Ростислав Тарасович

Символ	Ймовірність	F(x)	$\bar{F}(x)$	$\bar{F}(x)$ бінарне	Кодове слово	Довжина кодowego слова
Г	1/28=3.57%	1/28	1/56	0.00000101	0.000001	6
о	4/28=14.28%	5/28	3/28	0.00011011	0.0001	4
л	2/28=7.14%	7/28	6/28	0.01010010	0.01010	5
в	3/28=10.71%	10/28	17/56	0.01001110	0.01001	5
е	1/28=3.57%	11/28	21/56	0.01100000	0.011000	6
н	1/28=3.57%	12/28	23/56	0.01101001	0.011010	6
ь	1/28=3.57%	13/28	25/56	0.01110010	0.011100	6
“ “	2/28=7.14%	15/28	14/28	0.10000000	0.10000	5
Р	1/28=3.57%	16/28	35/56	0.10100000	0.101000	6
с	3/28=10.71%	19/28	35/56	0.10100000	0.10100	5
т	1/28=3.57%	20/28	39/56	0.10110010	0.101100	6
и	2/28=7.14%	22/28	21/28	0.11000000	0.11000	5
а	3/28=10.71%	25/28	47/56	0.11010111	0.11010	5
Т	1/28=3.57%	26/28	51/56	0.11101001	0.111010	6
р	1/28=3.57%	27/28	53/56	0.11110010	0.111100	6
ч	1/28=3.57%	28/28	55/56	0.11111011	0.111110	6

2. Перевірити виконання нерівності Крафта для отриманої послідовності кодівих слів.

Нерівність Крафта виконується для всіх послідовностей кодівих символів

- Для Гаффмана (дискр): $\sum D^{-1} \leq 1 = 1$
- Для Гаффмана (ПБ): $\sum D^{-1} \leq 1 = 1$
- Для Шенона (дискр): $\sum D^{-1} \leq 1 = 0,359$
- Для Шенона (ПБ): $\sum D^{-1} \leq 1 = 0.39$

Розрахунки у файлі Excel прикладені до звіту.

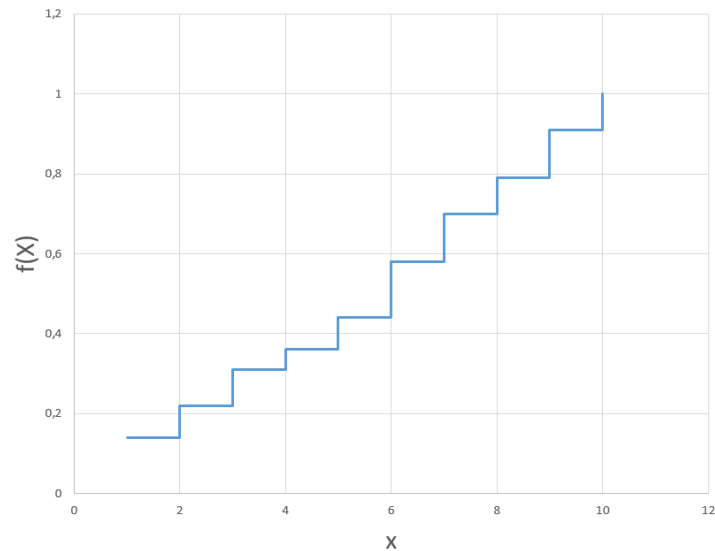
3. Обчислити ентропію вихідної послідовності і порівняти із очікуваною довжиною коду.

Ентропія вихідної послідовності практично співпадає з очікуваною довжиною коду

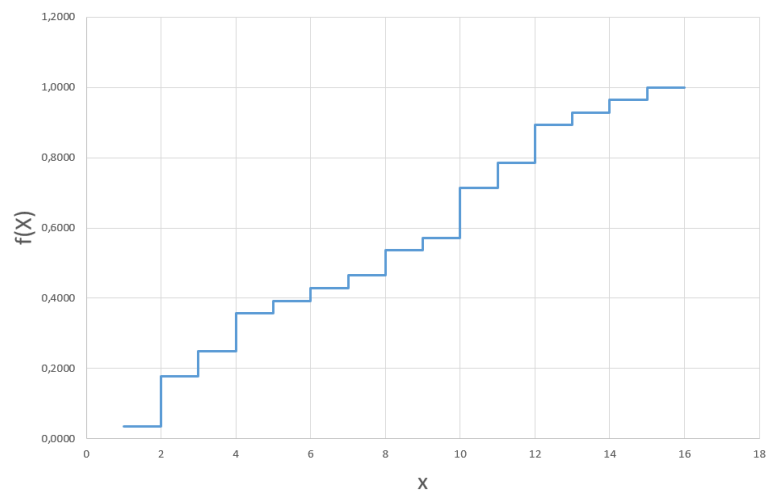
- Для випадкової величини:
Ентропія = 3.265, а очікувана довжина = 3.3
- Для ПБ:
Ентропія = 3.798, а очікувана довжина = 3.82

Графіки кумулятивної функції розподілу:

- для випадкової величини:



- для ПІБ:



Висновок: на даній лабораторній роботі я ознайомився з алгоритмами кодування та навчився їх застосовувати для кодування повідомлень. У кожного є, як свої плюси, так і мінуси.

Алгоритм Гаффмана на вході отримує таблицю частот, кодує символи у дерево від листків до кореня, а тоді будує символ-код на основі побудованого дерева. Символам з більшою ймовірністю ставляться у відповідність коротші коди. Але як мінус - для декодування потрібно мати таблицю частот, якою користувався «кодер».

Алгоритм Шенона-Фано-Еліаса, який став основою арифметичного кодування, використовує кумулятивну функцію розподілу випадкової величини для побудови коду.

Обидва алгоритми володіють властивістю префіксності, що дозволяє їх однозначно декодувати.

Я перевіряв нерівність Крафта для моїх послідовностей кодових символів і умова виконується, а отже префіксний код існує. Також можна спостерігати, що очікувана довжина коду практично співпадає з вихідною ентропією вихідної послідовності.