

Міністерство транспорту та зв'язку України

Одеська національна академія зв'язку ім. О. С. Попова

Кафедра теорії електричного зв'язку ім. А. Г. Зюко

"ОСНОВИ ТЕОРІЇ ПЕРЕДАВАННЯ ІНФОРМАЦІЇ"

НАВЧАЛЬНИЙ ПОСІБНИК

Для студентів денної та заочної форм навчання
за напрямками "Автоматичне управління
технологічними процесами" та "Радіотехніка"

МОДУЛЬ 1

ІНФОРМАЦІЯ І КОДУВАННЯ

ЗАТВЕРДЖЕНО
методичною радою ОНАЗ
Протокол № _____
від "____" 2006 р.

Одеса 2006

Рецензенти: проф. Захарченко М.В., проф. Гофайзен О.В.,
проф. Рудий Є.М.

Укладач – доц. Пляцек О. Є.

Відп. редактор – проф. Плотніков В. М.

Редактор – Кодрул Л.А.

Комп'ютерне верстання – Гардиман Ж.А.

У модулі I посібника вводяться поняття інформації, ентропії і надмірності, що базуються на методах теорії ймовірностей, формуються питання, що стосуються оптимізації інформаційних потоків у технічних системах. Центральне місце надається дискретним марковським ланцюгам, за допомогою яких можуть бути описані джерела і канали без пам'яті.

СХВАЛЕНО
на засіданні кафедри ТЕЗ
ім. проф. Зюко А. Г.
і рекомендовано до друку
Протокол №
від “ ” 2005 р.

Навчальний посібник за змістом відповідає програмам дисциплін: "Теорія інформації", за напрямом підготовки – 0925 АВТОМАТИЗАЦІЯ ТА КОМП'ЮТЕРНО-ІНТЕГРОВАНІ ТЕХНОЛОГІЇ та "Основи теорії передавання інформації", за напрямом підготовки – 0907 РАДІОТЕХНІКА.

Основною теоретичною базою для вивчення дисципліни "Основи теорії передавання інформації" (ОТПІ) є дисципліни: "Вища математика", "Фізика", "Теорія імовірності та випадкові процеси", "Електроніка та мікросхемотехніка".

Загальний обсяг годин на вивчення дисципліни складає 166 годин, з них 79 годин на вивчення 1 модуля і 87 годин на вивчення 2 модуля. Розподіл навчального часу наступний:

Вид занять	Загальний обсяг годин	Розподіл навчального часу за модулями	
		1 МОДУЛЬ	2 МОДУЛЬ
Лекції	<i>32</i>	<i>16</i>	<i>16</i>
Практичні заняття	<i>16</i>	<i>8</i>	<i>8</i>
Лабораторні заняття	<i>16</i>	<i>8</i>	<i>8</i>
Усього аудиторних годин	<i>64</i>	<i>32</i>	<i>32</i>
Індивідуальна та самостійна робота	<i>102</i>	<i>47</i>	<i>55</i>
Разом	<i>166</i>	<i>75</i>	<i>87</i>

Програмою також передбачено виконання курсової роботи на тему: "Розрахунки й оптимізація характеристик інформаційних систем."

В курсовій роботі виконуються розрахунки параметрів кодека простого коду, інформаційних характеристик джерел повідомлень та первинних сигналів, завадостійкості демодулятора, вибір і оптимізація параметрів коректуючого коду та обчислення завадостійкості й ефективності системи передачі інформації.

Метою вивчення дисципліни є набуття студентами наступних знань та умінь.

Студент повинен знати: способи вимірювання кількості інформації, що містяться в будь-яких повідомленнях, способи кодування для економного подання повідомлень і надійного передавання їх інформаційними каналами з завадами. Для цього вивчаються такі теоретичні питання, як способи вимірювання та кількісна оцінка інформації, кодування повідомлень, їх стиснення, оцінка ефективності та завадостійкості передачі кодованих повідомлень основні закономірності теорії інформації та побудови інформаційних систем: способи описання джерел інформації, аналіз їх інформаційних характеристик, економне та завадостійке кодування та завадостійкість інформаційних систем.

Студент повинен вміти: аналізувати інформаційні потоки з метою визначення частоти збору та рівнів дискретизації сигналів для систем керування, обробки та передання інформації, в тому числі для систем реального масштабу часу, із застосуванням комп'ютерно-інтегрованих середовищ (MATLAB)

ЗМІСТ

	Стор.
Вступ	5
Розділ 1 Інформація, ентропія і надмірність	7
1.1 Інформація однієї події	7
1.2 Ентропія й надмірність	9
Розділ 2 Кодування дискретних джерел без пам'яті	14
2.1. Теорема кодування джерел 1	14
2.2. Коди Хаффмана	18
Розділ 3 Ентропія зв'язаних джерел	22
3.1 Взаємна й умовна інформація	22
3.2. Спільна й умовна ентропія	24
3.3. Висновки	26
Розділ 4 Стаціонарні дискретні джерела з пам'яттю	28
4.1. Ентропія	28
4.2. Теорема кодування джерел 2	30
4.3. Скінченні ланцюги Маркова	31
4.3.1. Дискретні за часом ланцюги Маркова	32
4.3.2. Скінченні дискретні марковські джерела з пам'яттю r	37
4.4. Ентропія стаціонарного марковського джерела	40
4.5. Кодування стаціонарних марковських джерел	43
Розділ 5 Стиснення даних	50
5.1. Вступ	50
5.2. Арифметичне кодування	52
Розділ 6 Дискретні канали без пам'яті і передавання інформації	57
6.1. Вступ	57
6.2. Двійковий симетричний канал	57
6.3. Передавання інформації	59
6.4. Висновки	66
6.5. Пропускна здатність каналу	67
6.5.1. Пропускна здатність	68
6.5.2. Пропускна здатність двійкового симетричного каналу зі стиранням	69
6.6. Теорема кодування для дискретних каналів без пам'яті	72
Розділ 7 Неперервні джерела і канали	74
7.1. Диференціальна ентропія	74
7.2. Пропускна здатність каналу і границя Шеннона	77
7.3. Приклади	81
Додатки: завдання-тести для перевірки знань і вмінь	85
Д.1. Завдання для обов'язкової частини програми модуля	85
Д.2. Завдання для поглибленого вивчення програми модуля	89
ЛІТЕРАТУРА	99

Вступ

Теорія інформації описується за допомогою імовірнісних діаграм кодування і передачі інформації у конкретних, природно – наукових застосуваннях. При цьому з'являється можливість для аналізу й оптимізації потоків інформації у технічних системах. Відокремити технічне представлення інформації від побутового можуть допомогти лінгвістичні поняття: синтаксис, семантика й прагматика. У цих поняттях синтаксис і семантика є аналогом технічних даних. Синтаксис визначає всі припустимі символи й послідовності символів. Семантика пояснює їхнє значення. Дійсний зміст і область застосування пояснюються прагматикою. Прагматика зв'язує воедино такі поняття, як дані, технічна інформація, інформація в побутовому розумінні цього слова. Остаточо, термін «інформація», який звичайно розуміють як звістку або керівництво до дії, перетворюється в повідомлення, яке має бути переданим.

- *Синтаксис + семантика* → *дані*.
- *Дані + прагматика* → *повідомлення*.

Основою подальших міркувань є розуміння інформації, як якоїсь, експериментально встановлюваної величини. Поштовхом до цього послужила робота Клода Шеннона «Математична теорія зв'язку» [1], опублікована в 1948 р. У ній К. Шеннон дав визначення понять сучасної теорії інформації і заклав основи сьогоднішньої техніки зв'язку. Шеннон, як приклад, привів широко поширені в той час перфокарти. Одна перфокарта з N можливими позиціями для отворів може містити в точності 2^N повідомлень. Якщо маємо дві перфокарти, то число повідомлень дорівнює вже 2^{2N} . Таким чином, число можливих повідомлень, що несуть дві перфокарти, дорівнює квадрату числа повідомлень, що містяться на одній перфокарті. З іншого боку, можна було б очікувати, що дві перфокарти можуть зберігати вдвічі більше інформації, ніж одна. Тут для опису міри інформації так і напрошується логарифмічна функція, яка, завдяки своїм відомим властивостям, дає очікуване подвоєння:

$$\log 2^N = N \log 2, \quad \log 2^{2N} = 2N \log 2.$$

Загальна модель зв'язку по К. Шеннону приведена на рис. В.1.

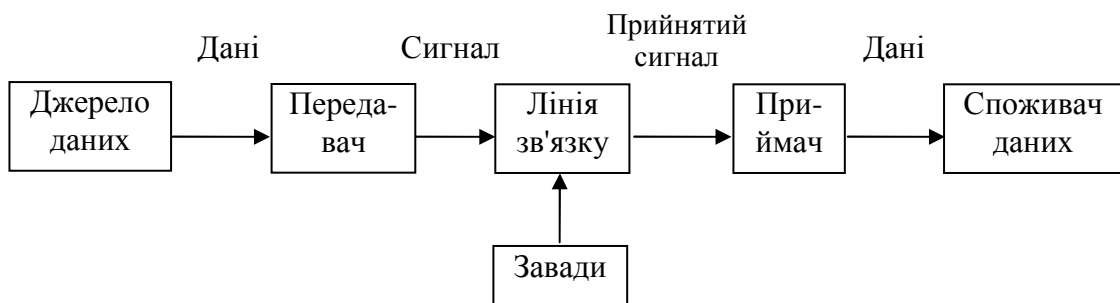


Рисунок В.1 – Модель передавання інформації каналом зв'язку за К. Шенноном

Вихідним пунктом є джерело інформації. Його повідомлення надходять на передавач. При цьому, повідомлення можуть являти собою окремі букви зв'язаного тексту, значення функцій часу, функції зміни напруги на виході мікрофо-

на, телевізійні сигнали тощо. Передавач виробляє сигнали, погоджені з фізичними властивостями каналу. Канал зображений на малюнку як лінія зв'язку з джерелом завад, які впливають на переданий сигнал і дещо його трансформують. Завада це будь-який вплив на сигнал, який викликає його випадкові відхилення від форми сигналу створеного джерелом, такий сигнал назвемо зашумленим. Зашумлений сигнал надходить у приймач, на який покладена сама складна задача.

Він повинен із зашумленого сигналу виділити передане повідомлення і відправити його споживачу.

Великою частиною цього посібника є кодування. Під кодуванням розуміють декотре відображення повідомлення за відомими правилами. При цьому, у шеннонівській моделі передачі інформації, блоки передавача і приймача потрібно розширити відповідно до цих правил. Між кодуванням джерел і кодуванням каналу існує чітка різниця. Прикладами кодування джерел можуть служити передача тексту кодом Морзе, «оцифровка» аудіо сигналу під час запису на компакт диски і т.д. Джерело повідомлень має інформаційну надмірність, яка полягає у тому, що деякі повідомлення несуть малу або навіть нульову кількість інформації. Так, наприклад при передачі телеграм вилучають окремі частини мови і сполучники, які легко поновлюються за змістом тексту. При кодуванні джерел надмірність повідомлень знижується і таке кодування часто називають стиском даних. Кодування каналів, навпаки, збільшує надмірність повідомлень. Внесення додаткових перевірочних символів дозволяє виявляти і навіть виправляти помилки, які виникають під час передавання інформації каналами. Кодування каналу надалі ми будемо називати завадостійким кодуванням. Без завадостійкого кодування було б неможливим створення накопичувачів величезної ємності, таких, як CD-ROM, DVD чи твердих дисків. Додаткові витрати на завадостійке кодування, яке забезпечує допустимі ймовірності помилок запису/зчитування, стають такими малими, що ними можна зневажати в порівнянні з виграшем від щільності запису, яка досягається при цьому. Розглянуті приклади показують, що інформація і кодування є центральними поняттями, без яких сучасна інформаційна техніка просто не існувала б.

РОЗДІЛ 1

ІНФОРМАЦІЯ, ЕНТРОПІЯ І НАДМІРНІСТЬ

1.1. Інформація однієї події

Обмін інформацією, незважаючи на свою нематеріальну природу, є невід'ємною частиною нашого життя. Норберт Вінер, один із засновників сучасної теорії інформації, охарактеризував інформацію в такий спосіб [9]:

«Інформація є інформація, а не матерія чи енергія».

Згідно з Н. Вінером, інформація є новим елементом у доповненні до матерії й енергії. Інформація для людей настільки ж важлива, наскільки важко уявити собі це поняття в природничонауковій формі. Ми говоримо, наприклад: «Ця інформація для мене важлива», маючи на увазі деяку конкретну ситуацію. Таке суб'єктивне сприйняття не підходить для технічного представлення інформації. Як строго наукове поняття, інформація повинна бути введена в техніку як вимірювана величина (аналогічно довжині в метрах, напрузі у вольтах і т.д.). Наш повсякденний досвід говорить про те, що, приймаючи інформацію до відома, ми постійно усуваємо деяку невизначеність. Це дуже нагадує експерименти з випадковими подіями. Проводячи такі експерименти, ми спостерігаємо випадкові події і це знижує невизначеність системи.

Переходимо до самої сутності теорії інформації. Насамперед дамо визначення найпростішого джерела, а потім введемо поняття кількості інформації, як вимірюваної величини для того, щоб з її допомогою охарактеризувати джерело.

Дискретне джерело без пам'яті

Найпростіше дискретне джерело без пам'яті X у кожен фіксований момент часу видає деякий символ x_i зі скінченного алфавіту $X = \{x_1, x_2, \dots, x_N\}$ з імовірністю $P(x_i) = p_i$. Вибірки символів провадяться незалежно один від одного.

Як найпростіший приклад можна привести двійкове джерело без пам'яті з алфавітом $X = \{x_1 = 0, x_2 = 1\}$ і ймовірностями $0 \leq p_1 \leq 1$ і $p_2 = 1 - p_1$. Вибір чергової цифри провадиться незалежно від попередніх і наступних вибірок. Зображення найпростішого джерела наведено на рис. 1.1. Розглянемо спочатку поодинокі події. Повсякденний досвід підказує, що події які відбуваються часто, так само як і їхні імовірності, дають нам мало інформації. Візьмемо, наприклад, повідомлення «собака вкусила людину». Це звична звістка не залучила б до себе ніякої уваги, у той час, як повідомлення «людина вкусила собаку» багато газет надрукували б великим шрифтом. З цього можна зробити висновок (і досвід нас у цьому часто переконує): часті, очікувані події несуть мало інформації і, навпаки, рідкі, тобто несподівані події, володіють високим інформаційним змістом. Отже, інформація й імовірність повинні знаходитися в обернено пропорційній залежності. Виходячи з цього, уведемо поняття кількості інформації, як вимірюваної величини на підставі наступних трьох аксіом [1].

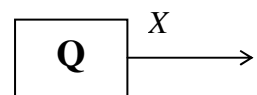


Рисунок 1.1 –
Найпростіше джерело інформації
абетки X

Аксиоми для визначення кількості інформації

1. Інформація поодинокій події $x_i \in X$, яка виникає з імовірністю p_i , має додатне значення

$$I(p_i) \geq 0. \quad (1.1)$$

2. Спільна інформація двох незалежних подій (x_i, x_j) зі спільною імовірністю $P(x_i, x_j) = p_{ij} = p_i \cdot p_j$, дорівнює сумі їхньої інформації

$$I(p_{ij}) = I(p_i) + I(p_j). \quad (1.2)$$

3. Інформація є неперервною функцією від імовірності події.

Аксиоми 1 і 2 стверджують, що інформація кількох подій не може взаємно знищуватися. Аксиома 2 уводить поняття спільності інформації подій. З аксиоми 3 випливає, що невелика зміна імовірності події приводить до невеликої зміни її інформації. Аксиома 2 визначає інформацію двох незалежних подій. З виразу (1.2) випливає, що інформація події визначається як логарифмічна функція її імовірності. Отже, інформацію можна визначити в такий спосіб:

Інформація події, яка відбувається з імовірністю p дорівнює

$$I(p) = -\log_2(p), \quad (1.3)$$

причому розмірність $[I]$ – біт.

У даній формулі використовується двійковий логарифм. Іноді використовують натуральний логарифм з одиницею виміру *нат*. Можна також переходити від однієї одиниці до іншої, застосовуючи формулу перерахування:

$$\log_a(x) = \log_b(x) / \log_b(a) = \log_b(x) \cdot \log_a(b).$$

Розмірність *біт* використовується в інформаційній техніці при двійковій системі числення. Як буде показано в наступних розділах, двійкова система дуже зручна під час опису процесу прийняття рішення, коли на будь-яке питання існує тільки дві відповіді: «так» чи «ні». У [10] приведена наочна інтерпретація поняття «біт».

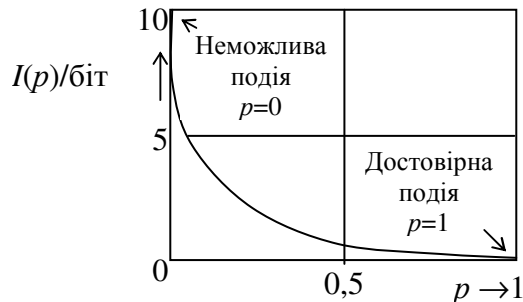


Рисунок 1.2 – Інформація символу $I(p)$ з ймовірністю її появи p

На рис. 1.2 наведено поведінку інформації як функції імовірності. Інформація події, що постійно відбувається, дорівнює нулю.

З ростом невизначеності інформація також зростає і для неможливої події прямує до нескінченності. Таким чином, інформація відповідає всім приведеним раніше міркуванням і задовольняє аксіомам 1–3. З точки зору теорії ймовірності, визначення інформації можна розглядати як деяке відображення подій. Аналогічне відображення мають стохастичні змінні.

1.2. Ентропія й надмірність

Після того як інформація окремої події визначена, розглянемо джерело подій. Для його опису будемо використовувати інформацію, яку несуть події, що містяться в ньому. За аналогією з термодинамікою введемо поняття ентропії. У термодинаміці ентропія є мірою неупорядкованості системи. У теорії інформації ентропія визначена як міра невизначеності джерела. Використовуючи інформацію окремих подій, виразимо ентропію джерела в такий спосіб: ентропія найпростішого джерела без пам'яті X з алфавітом $X = \{x_1, x_2, \dots, x_N\}$ і відповідними ймовірностями появи цих подій p_1, p_2, \dots, p_N дорівнює

$$H(x) = -\sum_{i=1}^N p_i \cdot \log_2(p_i), \quad (1.4)$$

причому розмірність $[H]$ – біт.

Уявимо собі гру, в якій деяка подія джерела може бути передбачена. Якщо джерело віддає перевагу визначеній події, сміло ставте на неї, і, в основному, ви будете вигравати. Якщо всі події рівноймовірні, то ставте на будь-яку подію: якщо невизначеність джерела максимальна, шанси на виграш мінімальні.

Приклад. Оцінка ентропії.

Пояснимо цей зв'язок на прикладі найпростішого дискретного джерела без пам'яті з табл. 1.1. Інформація джерела являє собою результат експерименту з випадковими подіями a, b, c, d . Нехай у результаті повторення цього експерименту ми одержуємо послідовність (обмежуючи кількість дослідів у визначений момент часу)

$$\{a, b, a, d, a, a, c, d, b, a, a, b, \dots\} \quad (1.5)$$

Таблиця 1.1 – Дискретне джерело без пам'яті із символами алфавіту $X = \{a, b, c, d\}$ з ймовірністю p_i й інформацією $I(p_i)$.

Символ	a	b	c	d
p_i	1/2	1/4	1/8	1/8
$I(p_i)$	1 біт	2 біт	3 біт	3 біт

Якщо підставити на місце кожної події його інформацію, одержимо типову функцію стохастичного процесу

$$\{I[n]/\text{біт}\} = \{1, 2, 1, 3, 1, 1, 3, 3, 2, 1, 1, 2, \dots\}. \quad (1.6)$$

Припустимо ергодичність (сталість поведіння) такого процесу в часі. Таку ергодичність ми, наприклад, припускаємо при киданні монети чи грального кубика. Зі зростанням числа випробувань N середнє значення інформації джерела

$$\bar{I} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} I[n] \quad (1.7)$$

прямує до математичного сподівання

$$E(I) = -\sum_{i=1}^4 p_i \cdot \log_2(p_i), \quad \text{біт}. \quad (1.8)$$

Таким чином, з огляду на збіжність ряду (1.7) до математичного сподівання, одержуємо практичну міру для інформації джерела. У розглянутому прикладі математичне сподівання $E(I)$ дорівнює 1,75 біт. З перших 12 випробувань ми також одержуємо оцінку для $\bar{I}^{(n)} = 1,75$ біт. Проводячи аналогічні міркування, Шеннон [1] заклав у визначення ентропії три наступні аксіоми.

Аксиоматичне визначення ентропії

1. Ентропія $H(X) = f(p_1, p_2, \dots, p_N)$ є неперервною функцією ймовірностей p_1, p_2, \dots, p_N .

2. Для джерел з однаковою ймовірністю подій $p_i = \frac{1}{N}$ ентропія збільшується зі зростанням числа подій N .

3. Розкладання процедури вибору подій на кілька етапів не змінює ентропію (процедуру вибору можна звести до послідовних двійкових рішень).

Приклад: Розкладання процедури вибору.

Даний приклад пояснює аксіому 3. Розглянемо три події a, b і c , що відбуваються з відповідними ймовірностями $1/2, 1/3$ і $1/6$. Для того, щоб вибрати одну з цих трьох, ми можемо поставити два незалежних питання (рис. 1.3).

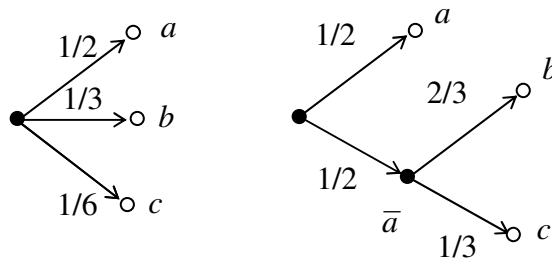


Рисунок 1.3 – Розкладення процесу вибору символів

На ці запитання можуть бути дані тільки дві відповіді: «так» чи «ні». Відповідно до аксіоми 3, до ентропії пред'являється наступна вимога:

$$H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = H_1\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} H_2\left(\frac{2}{3}, \frac{1}{3}\right), \quad (1.9)$$

причому, друге запитання задається з ймовірністю $1/2$. Покажемо в загальному вигляді (рис. 1.4), що визначення ентропії (1.8) задовольняє вимозі аксіоми 3.

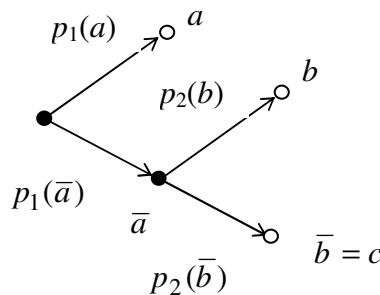


Рисунок 1.4 – Розкладення процедури прийняття рішення

Для розкладеної ентропії одержуємо

$$\frac{H(X)}{\text{бит}} = -p_1(a)\log_2(p_1(a)) - p_1(\bar{a})\log_2(p_1(\bar{a})) + p_1(\bar{a})[-p_2(b)\log_2(p_2(b)) - p_2(\bar{b})\log_2(p_2(\bar{b}))], \quad (1.10)$$

де імовірності визначаються в такий спосіб

$$p_1(\bar{a}) = p(b) + p(c), \quad (1.11)$$

$$p_2(b) + p_2(\bar{b}) = 1, \quad (1.12)$$

$$\frac{p_2(b)}{p_2(\bar{b})} = \frac{p(b)}{p(c)}. \quad (1.13)$$

Зауваження. Остання рівність підтверджується постановкою експерименту з випадковими подіями. Нехай подія a відбулася 300 разів, подія b – 200, а подія c – 100 разів. Частота кожної події у даному прикладі дорівнює її ймовірності. Якщо ми відкинемо подію a , то залишиться 300 вибірок подій b і c , частоти вибірок цих подій подвоюються, але їхнє відношення не зміниться.

З формул (1.11) – (1.13) випливає, що імовірності на другому кроці можна виразити як

$$p_2(b) = \frac{p(b)}{p(b) + p(c)} = \frac{p(b)}{p(\bar{a})}, \quad (1.14)$$

$$p_2(\bar{b}) = \frac{p(c)}{p(b) + p(c)} = \frac{p(c)}{p(\bar{a})}. \quad (1.15)$$

Підставляючи отримані вирази у формулу для ентропії, маємо

$$\frac{H(X)}{\text{бит}} = -p_1(a)\log_2(p_1(a)) - p_1(\bar{a})\log_2(p_1(\bar{a})) + p_1(\bar{a})\left[-\frac{p(b)}{p_1(\bar{a})}\log_2\left(\frac{p(b)}{p_1(\bar{a})}\right) - \frac{p(c)}{p_1(\bar{a})}\log_2\left(\frac{p(c)}{p_1(\bar{a})}\right)\right], \quad (1.16)$$

що після спрощень відповідає ентропії без розкладання процесу вибору подій

$$\frac{H(X)}{\text{бит}} = -p_1(a)\log_2(p_1(a)) - p(b)\log_2(p(b)) - p(c)\log_2(p(c)). \quad (1.17)$$

У розглянутому прикладі була використана властивість логарифмічної функції, що переводить добуток у суму. Визначення (1.4) є єдиним, при якому аксіома 3 має силу. Помітимо також, що розглянуте розкладання процесу вибору подій може бути зведене до послідовності двійкових рішень «так» і «ні». Максимальної невизначеності відповідає максимальна ентропія джерела. Сформулюємо наступну теорему.

Теорема 2.2.1. Ентропія найпростішого дискретного джерела без пам'яті максимальна, якщо всі події, що містяться в ньому, мають однакову ймовірність. У цьому випадку ентропія просто дорівнює логарифму числа подій

$$H_0 = \log_2 N [\text{бит}]. \quad (1.18)$$

Зауваження. Наступний доказ теореми є типовим у теорії інформації. У таких доказах використовуються оцінки і граничні переходи, що раніше були відомі, але не знаходили практичного застосування.

Доказ. Для доказу розглянемо два дискретних джерела без пам'яті P і Q , кожне з яких містить N подій з ймовірностями p_i і q_i відповідно. Далі скористаємося відомою верхньою оцінкою логарифмічної функції (рис. 1.5)

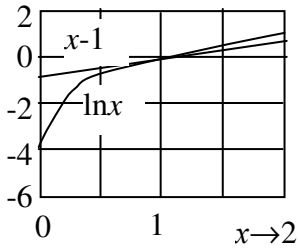


Рисунок 1.5 – Верхня оцінка логарифмічної функції

$$\ln x \leq x - 1. \quad (1.19)$$

Використовуючи цю оцінку, одержуємо

$$\ln q_i - \underbrace{\ln p_i}_{I(p_i) \text{ нат}} = \ln \left(\frac{q_i}{p_i} \right) \leq \left(\frac{q_i}{p_i} - 1 \right). \quad (1.20)$$

Помноживши обидві частини нерівності на p_i і просумувавши за всіма подіями $1 \leq i \leq N$, маємо

$$\sum_{i=1}^N p_i (\ln q_i - \ln p_i) \leq \sum_{i=1}^N p_i \left(\frac{q_i}{p_i} - 1 \right). \quad (1.21)$$

Після спрощення одержуємо

$$\frac{H(P)}{\text{нат}} + \sum_{i=1}^N p_i \ln q_i \leq \underbrace{\sum_{i=1}^N q_i}_{=1} - \underbrace{\sum_{i=1}^N p_i}_{=1} = 0 \quad (1.22)$$

і, отже

$$\frac{H(P)}{\text{нат}} \leq - \sum_{i=1}^N p_i \ln q_i. \quad (1.23)$$

Припустимо, що джерело Q містить тільки рівноймовірні події. Тоді

$$\frac{H(P)}{\text{нат}} \leq - \sum_{i=1}^N p_i \ln \left(\frac{1}{N} \right) = \ln N \underbrace{\sum_{i=1}^N p_i}_{=1} = \ln N. \quad (1.24)$$

Так як у процесі доказу на джерело P не було накладено ніяких обмежень, то дана нерівність має місце для будь-якого дискретного джерела без пам'яті, що містить N подій

$$H(X) \leq \log_2 N \text{ біт}. \quad (1.25)$$

Максимум досягається, коли всі події мають однакові імовірності.

Будь-яке джерело, що містить N подій, не всі з яких мають однакову ймовірність, має ентропію, меншу за $\log_2 N$. Розглянемо джерело ємністю $H_0 = \log_2 N$ як резервуар для збереження інформації, який ніколи не переповнюється. Різниця між максимальною ємністю H_0 й ентропією джерела X , що містить N подій, називається надмірністю джерела

$$H^* = H_0 - H(X). \quad (1.26)$$

Відносна надмірність визначається як

$$\chi = \frac{R}{H_0} = 1 - \frac{H(X)}{H_0}. \quad (1.27)$$

Приклад. Ентропія дискретного джерела без пам'яті, що містить 6 подій.

Таблиця 1.2. Дискретне джерело без пам'яті із символами x_i алфавіту $X = \{a, b, c, d, e, f\}$ з імовірністю p_i й інформацією $I(p_i)$.

x_i	a	b	c	d	e	f
p_i	0,05	0,15	0,05	0,4	0,2	0,15
$I(p_i)$ біт	4,32	2,74	4,32	1,32	2,32	2,74

Для того, щоб конкретизувати проведені вище міркування, розглянемо числовий приклад. У таблиці 1.2 задано дискретне джерело без пам'яті X із відповідними алфавітом і ймовірностями подій. Використовуючи вираз (1.3), обчислимо інформацію кожної події і доповнимо таблицю значеннями. Ентропія джерела дорівнює $H(X) = 2,25$ біт, значення H_0 для подій дорівнює $H_0 = \log_2 6 = 2,585$ біт, тоді надмірність $H^* = (\log_2 6 - 2,25)$ біт $= 0,335$ біт і, відповідно, відносна надмірність складає $\chi = 1 - \frac{2,25}{\log_2 6} = 0,130 \cong 13\%$.

Особливе значення мають дискретні двійкові джерела без пам'яті, тому що в більшості випадків з їх допомогою можна описати процес передавання даних. Нехай задано двійкове джерело без пам'яті з алфавітом $X = \{0,1\}$ і з ймовірностями для символів «0» і «1» – $p_0 = p$ і $p_1 = 1 - p_0$ відповідно. Вибір символів проводиться незалежно. Його ентропія, що називається також функцією Шеннона, залежить тільки від імовірності p .

Ентропія двійкового джерела (функція Шеннона)

$$\frac{H_b(p)}{\text{біт}} = -p \log_2 p - (1-p) \log_2 (1-p). \quad (1.28)$$

На рис. 1.6 показане поведіння функції Шеннона. Ентропія двійкового джерела всюди додатна й симетрична відносно $p = 1/2$ і має максимум при однаковій імовірності символів «0» й «1». Максимальна ентропія, яка дорівнює 1 біт відповідає двійковому рішення, тобто на запитання про значення символу - це відповідь: або «так» або «ні».

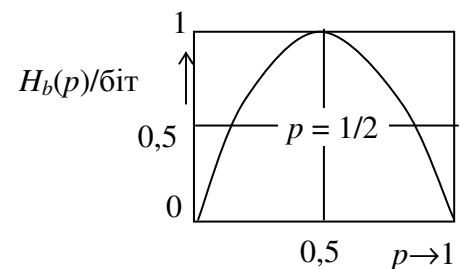


Рисунок 1.6 – Ентропія двійкового джерела

РОЗДІЛ 2

КОДУВАННЯ ДИСКРЕТНИХ ДЖЕРЕЛ БЕЗ ПАМ'ЯТІ

2.1. Теорема кодування джерел 1

У розділі 1.2 було введено поняття ентропії як середньої інформації дискретного джерела без пам'яті. Більш того, було показано, що будь-яка подія, що міститься в джерелі, може бути розкладена на послідовності двійкових рішень з результатами «так» чи «ні» без втрати інформації. Таким чином, кожній події, що міститься в алфавіті, може бути приписана деяка послідовність двійкових символів «0» чи «1» (надалі таку послідовність будемо називати кодовим словом події). При цьому не відбувається втрати інформації, тому що кожна подія може бути безпосередньо відновлена за відповідним кодовим словом. Такий процес називається *кодуванням джерела*, а сукупність кодових слів усіх подій – *кодом джерела*.

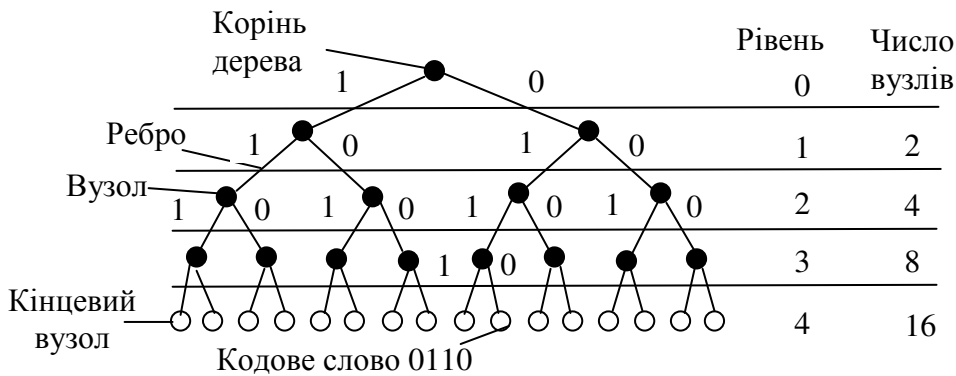


Рисунок 2.1 – Кодове дерево

Виникає питання: яке середнє число біт треба потратити для кодування джерела? Попутно виникають питання: скільки місць на платі займуть перемікальні елементи під час реалізації коду джерела і які тривалості кодових слів під час передавання інформації?

Відповіді на ці питання дає теорема Шеннона про кодування джерел. Перед тим, як приступити до розгляду цієї теореми, візьмемо як приклад двійкового кодування кодову конструкцію, зображену на рис. 2.1. Починаючи від кореня кодового дерева число ребер на кожному рівні подвоюється, причому, ліворуч розташовуються ребра, що відповідають одиницям, а праворуч – нулям кодових слів. Кодові слова утворюються при проходженні відповідних ребер, що ведуть від кореня до кінцевих вузлів. Як приклад будемо розглядати кодове слово 010.

Якщо ми пройдемо N рівнів, то одержимо 2^N кінцевих вузлів, що відповідають кодовим словам з довжиною $\log_2 2^N$ біт = N біт і можна закодувати 2^N подій. Припустимо далі, що цим кінцевим вузлам відповідають 2^N рівномірних подій із всього алфавіту джерела. Тоді, на кодування кожної події витрачається рівно N біт, що дорівнює ентропії джерела. Зв'язок між середньою довжи-

ною кодового слова й ентропією джерела узагальнює теорема кодування джерел.

Теорема 2.1.1. Теорема кодування джерел 1.

Для будь-якого дискретного джерела без пам'яті X з кінцевим алфавітом і ентропією $H(x)$ існує D -ий префіксний код, у якому середня довжина кодового слова \bar{n} задовольняє нерівності

$$\frac{H(x)}{\log D} \leq \bar{n} \leq \frac{H(x)}{\log D} + 1. \quad (2.1)$$

Термін префіксний код означає, що ніякий початок кодового слова не може бути іншим кодовим словом. Це значить, що потік подій може бути закодований без спеціального поділу цих подій. У випадку $D = 2$ використовується двійковий код. Якщо ентропія задана в бітах, то середня довжина \bar{n} також виражається в бітах, що прямо вказує на використання двійкового коду. Теорема кодування джерел указує на те, що середня довжина кодового слова не може бути менше ентропії. Однак, як буде показано в розділах 4.2 і 4.5, під час блокового кодування джерел, середня довжина кодового слова може наближатися до ентропії як завгодно близько. Ця обставина ще раз підкреслює важливість поняття ентропії. Ентропія – це міра мінімальних середніх витрат. Прикладом практичної реалізації нерівності (2.1) є код Хаффмана, розглянутий у наступному розділі.

Зауваження. Доказ теореми 2.1.1 зводиться до використання нерівності Крафта і не несе в собі яких-небудь конструктивних суджень для побудови оптимальних, з погляду витрат, кодів. Доказ має великий обсяг і дає добре уявлення про методи теорії інформації. (Проте, воно може бути поки пропущене без втрати для розуміння наступних розділів.)

Доказ.

Крок 1. Нерівність Крафта.

Для існування однозначного декодування D -го коду, що містить K кодових слів з довжинами n_1, n_2, \dots, n_k , необхідно і достатньо, щоб виконувалась нерівність Крафта

$$\sum_{k=1}^K D^{-n_k} \leq 1. \quad (2.2)$$

Для доказу цього твердження скористаємося кодовою конструкцією, зображеною на рис. 2.2.

Ми бачимо три характерні властивості цієї конструкції:

1. Існує D^i вузлів i -го порядку (i -го рівня);
2. Кожен вузол i -го порядку породжує точно D^{n-i} n -го рівня;
3. Кодова конструкція відповідає префіксному коду, тобто ніяке кодове слово не є початком іншого кодового слова, тому що кодові слова однозначно визначаються кінцевими вузлами.

Для спрощення упорядкуємо довжини кодових слів

$$n_1 \leq n_2 \leq \dots \leq n_k = n. \quad (2.3)$$

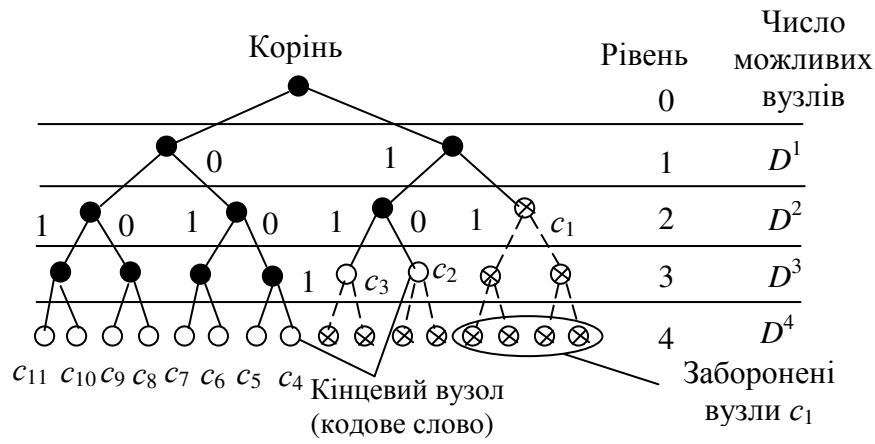


Рисунок 2.2 – Кодове дерево для $D = 2$ і $n = 4$

Тепер почнемо відлік. Кодове слово c_1 довжини n_1 забороняє в точності D^{n-n_1} можливих кінцевих вузлів на останньому n -ому рівні. Так як кодові слова префіксного коду не сполучаються, у сукупності одержимо

$$\sum_{k=1}^K D^{n-n_k} \tag{2.4}$$

заборонених вузлів на n -ому рівні. Загальне число можливих вузлів на n -ому рівні дорівнює D^n , отже

$$\sum_{k=1}^K D^{n-n_k} \leq D^n. \tag{2.5}$$

Розділивши обидві частини нерівності на D^n , одержимо нерівність Крафта.

Крок 2. Твердження Мак-Міллана.

Кожен код, що однозначно декодується, задовольняє нерівності Крафта.

Під час доказу нерівності Крафта були використані особливості префіксних кодів. Але ця умова не є необхідною. Як буде показано далі, необхідною умовою є наявність однозначного декодування коду. Піднесемо суму в нерівності Крафта до степеня L

$$\left[\sum_{k=1}^K D^{-n_k} \right]^L = \sum_{k_1=1}^K \sum_{k_2=2}^K \dots \sum_{k_L=1}^K D^{-(n_{k_1}+n_{k_2}+\dots+n_{k_L})}. \tag{2.6}$$

Позначивши через A_i число комбінацій, що містять L кодівих слів із сумарною довжиною i , запишемо (2.6) у компактній формі

$$\left[\sum_{k=1}^K D^{-n_k} \right]^L = \sum_{i=1}^{L_{n,max}} A_i D^{-i}, \tag{2.7}$$

де $L_{n,max}$ – максимальна довжина повідомлення, що містить L кодівих слів. Якщо код однозначно декодується, то всі послідовності з L кодівих слів сумарної довжини i різні. Так як є всього D^i можливих послідовностей, то $A_i \leq D^i$ і, таким чином,

$$\left[\sum_{k=1}^K D^{-n_k} \right]^L \leq \sum_{i=1}^{L_{n,max}} 1 = L_{n,max}.$$

Якщо обчислити L -ий корінь L -го степеня, одержимо оцінку зверху для суми в нерівності Крафта

$$\sum_{k=1}^K D^{-n_k} \leq (L_{n,\max})^{1/L} \quad (2.8)$$

для всіх натуральних L .

Проаналізуємо сутність числа L . Це число незалежних кодових слів, які вибираються з $\{1, 2, \dots, K\}$ і використовуються для побудови всіх можливих послідовностей довжини, що не перевищує $L_{n,\max}$. Тому, при $L \rightarrow \infty$, ми приходимо до нерівності Крафта.

Приведені вище міркування справедливі для кожного коду, що однозначно декодується. Тому, кожен код, який однозначно декодується, задовольняє нерівності Крафта.

Крок 3.

Запишемо ліву частину нерівності (2.1) у вигляді

$$H(x) - \bar{n} \log D \leq 0.$$

Використовуючи ймовірність подій p_k і відповідні довжини кодових слів n_k , маємо

$$H(x) - \bar{n} \log D = \sum_{k=1}^K p_k \log \frac{1}{p_k} - \sum_{k=1}^K p_k n_k \log D = \sum_{k=1}^K p_k \log \frac{D^{-n_k}}{p_k}.$$

Як і під час доказу твердження (1.1), оцінимо зверху логарифмічну функцію за допомогою виразу (1.19) і одержимо

$$\sum_{k=1}^K p_k \log \frac{D^{-n_k}}{p_k} \leq \log e \left(\sum_{k=1}^K p_k \left[\frac{D^{-n_k}}{p_k} - 1 \right] \right) = \log e \left(\underbrace{\sum_{k=1}^K D^{-n_k}}_{\leq 1} - \underbrace{\sum_{k=1}^K p_k}_{=1} \right) \leq 0. \quad (2.9)$$

Тут була використана нерівність Крафта й умова нормування ймовірностей повної групи подій.

Таким чином, ліва частина нерівності (2.1) теореми про кодування джерел доведена.

Крок 4.

Права частина нерівності (2.1)

$$\bar{n} \leq \frac{H(X)}{\log D} + 1.$$

Під час доказу використовуємо нерівність Крафта, умову існування коду з відповідними довжинами кодових слів n_k і умову нормування ймовірностей подій.

Нерівність Крафта можна записати у вигляді

$$\sum_{k=1}^K D^{-n_k} \leq \sum_{k=1}^K p_k. \quad (2.10)$$

Поки має силу нерівність Крафта, ми вільні у виборі довжин кодових слів $n_k \in \mathbb{N}$. Виберемо для кожної складової таке найменше n_k , для якої

$$D^{-n_k} \leq p_k. \quad (2.11)$$

Нерівність Крафта при такому виборі буде виконуватися, отже, використовуючи конструкцію рис. 2.2, ми можемо побудувати такий префіксний код. Так

як n_k найменше ціле, при якому має місце нерівність (2.11), то для $(n_k - 1)$ справедливо $p_k < D^{-(n_k-1)}$.

Інша частина доказу формальна.

Використовуючи властивості логарифмічної функції, одержуємо

$$p_k \log p_k < p_k \log D^{-(n_k-1)} = p_k (-n_k + 1) \log D.$$

Підсумовуючи за всіма K , маємо

$$\underbrace{\sum_{k=1}^K p_k \log p_k}_{-H(X)} < \underbrace{\log D \sum_{k=1}^K p_k (-n_k)}_{-n} + \log D \underbrace{\sum_{k=1}^K p_k}_{=1}. \quad (2.12)$$

Розділивши обидві частини нерівності на $\log D$ і переставляючи члени з до множенням на (-1) (що змінює знак нерівності), одержуємо доказ, який шукали.

2.2. Коди Хаффмана

Коди Хаффмана (Давид Хаффман (1925–1999) – американський вчений) належать до сімейства кодів зі змінною довжиною кодів слів. Найвідоміший код змінної довжини – абетка Морзе (Самуель Морзе (1791–1872) – американський художник і винахідник). Основна ідея абетки Морзе – передавати букви, що часто зустрічаються, кодівими словами короткої довжини і, навпаки, букви, що рідко зустрічаються – довгими кодівими словами для того, щоб мінімізувати середні витрати часу. Такий спосіб кодування називають також кодуванням з мінімальною надмірністю чи *ентропійним кодуванням*.

Так, наприклад, в абетці Морзе літера «Е», що часто зустрічається, кодується одним символом «•», а рідка літера «Х» – чотирма символами.

У 1952 р. Хаффман показав, що запропоноване їм кодування зі змінною довжиною кодів слів є оптимальним префіксним кодом для дискретних джерел без пам'яті. Тобто, середня довжина слова коду Хаффмана мінімальна і він також є кодом без коми. Термін «код без коми» означає, що при встановленій синхронізації можлива неперервна передача потоку повідомлень з однозначним декодуванням без спеціального поділу кодів слів.

У префіксним коді ніяке кодове слово не є префіксом іншого слова.

Зауваження. Коди Хаффмана відіграють найважливішу роль у кодуванні зображень. Вони є основною частиною стандартів JPEG, MPEG і H.261 [19]. Крім цього, вони також використовуються під час аналого-цифрового перетворення аудіо сигналів. У літературі, як приклади ентропійного кодування, згадуються коди Шеннона і Фано, але вони не у всіх випадках є оптимальними і ми пропустимо їхній опис.

Кодування Хаффмана проводиться за три кроки. Ми наочно пояснимо цей процес на прикладі.

Кодування Хаффмана.

1. Упорядкування. Розташувати знаки в порядку убунання їхніх ймовірностей.

2. Редукція. Об'єднати два знаки з найменшими ймовірностями в один складений знак. Переупорядкувати список знаків відповідно до кроку 1. Продовжувати цей процес доти, поки всі знаки не будуть об'єднані.

3. Кодування. Починаємо з останнього об'єднання. Приписати першій компоненті складеного знака символ «0», а другій – символ «1». Продовжувати цей процес доти, поки всі прості знаки не будуть закодовані.

У випадку, коли кілька знаків мають однакові ймовірності, поєднуються ті два з них, що до цього мали найменше число об'єднань. Цим досягається вирівнювання довжин кодових слів, що полегшує передачу інформації.

Приклад:

Кодування Хаффмана наочно показано на прикладі джерела, заданого табл. 2.1. При цьому ми дещо спростимо кодування, відмовившись від наочних переупорядкувань знаків на другому кроці, тому що в такому маленькому прикладі це можна зробити «у розумі». Відповідно до першого кроку, розташуємо всі знаки в порядку убутання їхніх ймовірностей (рис. 2.3).

Таблиця 2.1. Ймовірності й ентропія знаків

x_i	a	b	c	d	e	f
p_i	0,05	0,15	0,05	0,4	0,2	0,15
$\frac{I(p_i)}{\text{біт}}$	4,32 біт	2,74 біт	4,32 біт	1,32 біт	2,32 біт	2,74 біт
$\frac{H(X)}{\text{біт}}$	$\approx 2,25$					

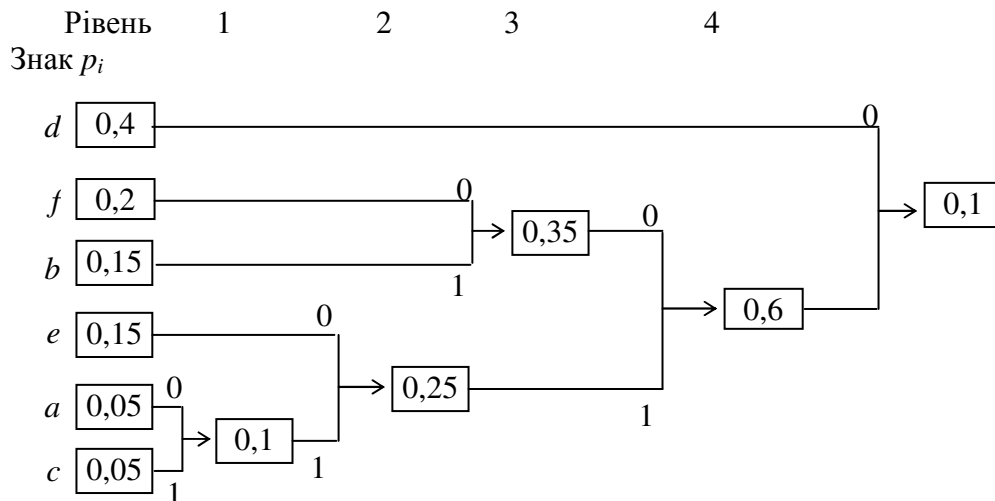


Рисунок 2.3 – Кодування Хаффмана $n = 4$.

На другому кроці об'єднаємо знаки « a » і « c », що мають найменші ймовірності, у складений знак « ac ». Далі об'єднаємо « e » з « ac » у складений знак « eac » з ймовірністю 0,25. Тепер найменші ймовірності вже мають знак « f » і « b ». Об'єднавши їх, одержимо знак « fb ». Наступна редукція дає складений знак « $fb eac$ »

з імовірністю 0,6. І, нарешті, об'єднавши всі знаки, одержимо складений знак «dfbeac», імовірність якого дорівнює 1. На третьому кроці будемо йти справа вліво, приписуючи верхнім компонентам об'єднань символ «0», а нижнім – «1». Отримані кодові слова всіх простих знаків з їхніми довжинами наведені в табл. 2.2.

Таблиця 2.2. Кодування кодом Хаффмана для рис.2.3.

x_i	d	f	b	e	a	c
p_i	0,4	0,2	0,15	0,15	0,05	0,05
Кодове слово	0	100	101	110	1110	1111
Довжина кодового слова	1	3	3	3	4	4

Даний код має мінімальну середню довжиною кодового слова. Середня довжина кодового слова визначається довжинами всіх слів n_i , «зважених» з відповідними ймовірностями p_i

$$\bar{n} = \sum_{i=1}^N p_i n_i. \quad (2.13)$$

У розглянутому вище прикладі середня довжина кодового слова $n_i = 2,3$ біт близька до ентропії $H(X) = 2,25$ біт. Найважливішою величиною, що називають *ефективністю коду* чи *фактором стиснення*, є відношення ентропії до середньої довжини кодового слова. У нашому прикладі ефективність коду дорівнює $\eta \cong 0,978$. Ефективність коду чи фактор стиснення

$$\eta = \frac{H(x)}{\bar{n}}. \quad (2.14)$$

З приклада чітко видно, що чим більше різниця між ймовірностями знаків, тим більше вираш коду Хаффмана в порівнянні з простим блоковим кодуванням.

Теорема Шеннона про кодування джерел показує, наскільки ефективним може бути таке кодування. Але теорія інформації також указує на ту обставину, що під час кодування можуть з'являтися кодові слова дуже великої довжини. Ця обставина може перешкоджати практичному використанню теореми кодування джерел.

Реалізація декодера коду Хаффмана впливає безпосередньо з рис. 2.3. На рис. 2.4 представлено дерево, що називається *кодовим деревом декодера*. Декодування кожного нового слова починається з вихідного вузла (кореня) кодового дерева. Якщо ми приймаємо «0», то йдемо ребром, що відповідає нулю

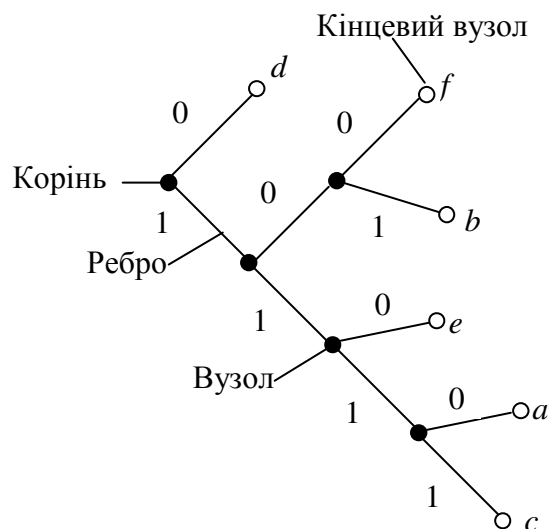


Рисунок 2.4 – Кодова конструкція для $D = 2$ і $n = 4$

(верхнім ребром). У нашому прикладі при цьому ми досягаємо кінцевого вузла d . Отже, був переданий знак d і ми починаємо декодування нового знаку з кореня.

Якщо ми приймаємо «1», то йдемо нижнім ребром і попадаємо у вузол, з якого виходять два ребра. Наступний прийнятий біт указує, яке з цих двох ребер потрібно вибрати. Ми продовжуємо цю процедуру доти, поки не досягнемо кінцевого вузла. У цьому випадку ми приймаємо рішення про переданий знак і знову переходимо до кореня кодового дерева.

При всій простоті побудови і декодування, коди Хаффмана мають три недоліки:

- Різні довжини кодових слів призводять до нерівномірних затримок декодування.
- Стиснення даних знижує надмірність і тому підвищує схильність до поширення помилок. У випадку кодування Хаффмана це означає, що один, помилково розпізнаний символ, може призвести до того, що цей і всі наступні знаки будуть декодовані невірно.
- Кодування Хаффмана припускає знання ймовірностей подій (знаків), чи, принаймні, деяких оцінок цих ймовірностей. На практиці дуже часто ймовірності подій невідомі, а їхні оцінки дуже ускладнені. Саме тому для стиснення великих масивів даних часто використовують універсальний алгоритм кодування, відомий як алгоритм Лемпеля-Зіва. Опис цього алгоритму буде приведено в розділі 5.3. Універсальний алгоритм стиснення не вимагає апріорного знання статистики джерела.

РОЗДІЛ 3

ЕНТРОПІЯ ЗВ'ЯЗАНИХ ДЖЕРЕЛ

Дотепер у своїх міркуваннях ми виходили з припущення незалежності послідовних подій. Однак, варто лише тільки відкрити, наприклад, німецький орфографічний словник, ми відразу ж знайдемо залежність між літерами, що стоять поруч, наприклад: «qu», «ch», «ck», «tz» і «sch». Читаючи німецький текст, ми бачимо, що після «q» за рідкісним винятком впливає «u» (у теорії множин говорять, що елементи деякої множини вступають у деякі відносини, тобто «зв'язуються»). У цьому випадку «u», як майже неминуха подія, практично не несе в собі ніякої інформації. Тому, під час визначення інформації подібного роду джерел, ми повинні брати до уваги існуючий взаємний зв'язок між подіями.

3.1 Взаємна й умовна інформація

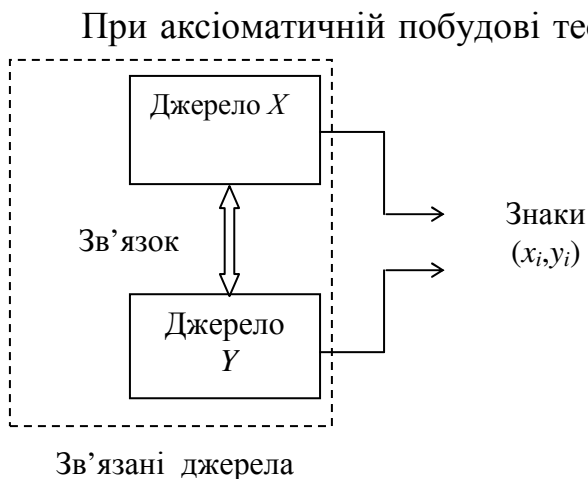


Рисунок 3.1 – Модель двох зв'язаних джерел

Якщо обидва джерела якимсь чином зв'язані між собою, то варто очікувати, що подія одного джерела дозволяє робити деяке припущення про подію іншого. У термінах теорії інформації це означає, що невизначеність другого джерела знижується, тобто джерела обмінюються взаємною інформацією. Введемо умовну імовірність $p(x|y)$ – імовірність події x за умови, що подія y відбулася. Виразимо спільну імовірність $p(x_i, y_i)$ двох подій x_i і y_i через їхню апріорну й умовну ймовірності [14].

$$p(x_i, y_i) = p(x_i | y_i) p(y_i) = p(y_i | x_i) p(x_i) \quad (3.1)$$

Використовуючи логарифмічну функцію, відразу ж одержуємо інформації подій (x_i, y_i) , (x_i) і (y_i)

$$\underbrace{\log_2 p(x_i, y_i)}_{-I(x, y) \text{ бит}} = \log_2 p(x_i | y_i) + \underbrace{\log_2 p(y_i)}_{-I(y_i) \text{ бит}} = \log_2 p(y_i | x_i) + \underbrace{\log_2 p(x_i)}_{-I(x_i) \text{ бит}} \quad (3.2)$$

тобто

$$I(x_i, y_i) = I(y_i) - \log_2 p(x_i | y_i) \text{ бит} = I(x_i) - \log_2 p(y_i | x_i) \text{ бит}. \quad (3.3)$$

Додаючи й одночасно віднімаючи $I(x_i)$ у першій частині співвідношення (3.3) і, відповідно, $I(y_i)$ у другій, одержуємо

$$\begin{aligned}
 I(x_i, y_i) &= I(x_i) + I(y_i) - \log_2 \frac{p(x_i | y_i)}{p(x_i)} \text{ бит} = \\
 &= I(x_i) + I(y_i) - \log_2 \frac{p(y_i | x_i)}{p(y_i)} \text{ бит}.
 \end{aligned}
 \tag{3.4}$$

Таким чином, інформація пари подій (x_i, y_i) визначається сумою інформації цих подій мінус деяка додатна величина, яка знижує невизначеність, тобто сама у свою чергу є інформацією. Тому, назовемо її *взаємною інформацією* пари подій.

Взаємна інформація пари подій визначається як

$$I(x_i; y_i) = \log_2 \frac{p(x_i | y_i)}{p(x_i)} = \log_2 \frac{p(y_i | x_i)}{p(y_i)} = \log_2 \left(\frac{\text{апостеріорна ймовірність}}{\text{апріорна ймовірність}} \right). \tag{3.5}$$

Варто звернути увагу на те, що взаємна інформація $I(x_i; y_i)$ завжди додатна. Важливою властивістю також є симетрія взаємної інформації відносно джерел, тому що

$$\frac{p(x_i | y_i)p(y_i)}{p(x_i)p(y_i)} = \frac{p(y_i | x_i)}{p(y_i)}.$$

Симетрія відносно джерел у виразі (3.5) дозволяє зробити висновок, що обмін інформацією між джерелами є взаємним, а не одностороннім.

Для того, щоб краще уявляти собі зміст взаємної інформації, розглянемо два граничних випадки.

1) Джерела незалежні. Тоді для пари незалежних подій маємо $p(x_i, y_i) = p(x_i)p(y_i)$, тобто джерела не обмінюються інформацією $I(x_i; y_i) = 0$.

2. Джерела жорстко зв'язані, тобто подія одного джерела однозначно визначає подію іншого $p(x_i | y_i) = p(y_i | x_i) = 1$.

У цьому випадку відбувається повний обмін інформацією $I(x_i; y_i) = I(x_i) = I(y_i)$.

З формули (3.4) випливає, що інформацію пари подій (x_i, y_i) можна інтерпретувати, як різницю між інформацією пари незалежних подій $I(x_i) + I(y_i)$ і заздалегідь передв'ященою взаємною інформацією $I(x_i; y_i)$, обумовленою зв'язаністю джерел X і Y

$$I(x_i, y_i) = I(x_i) + I(y_i) - I(x_i; y_i) \tag{3.6}$$

Розглянемо ще раз вираз (3.3) і введемо поняття умовної інформації. *Умовна інформація* (апостеріорна невизначеність)

$$I(x_i | y_i) = -\log_2 p(x_i | y_i) \text{ біт} \tag{3.7}$$

З формули (3.3) випливає

$$I(x_i; y_i) = I(y_i) + I(x_i | y_i) = I(x_i) + I(y_i | x_i), \tag{3.8}$$

тобто інформацію пари подій можна визначити як суму інформації події y_i й інформації події x_i за умови, що подія y_i вже відома, чи, навпаки, як суму інформації події x_i й інформації події y_i за умови, що подія x_i вже відома.

3.2. Спільна й умовна ентропія

Після розгляду окремих пар подій у попередньому розділі, перейдемо до середніх оцінок інформаційних характеристик джерела. На рис. 3.2 показана вихідна ситуація.

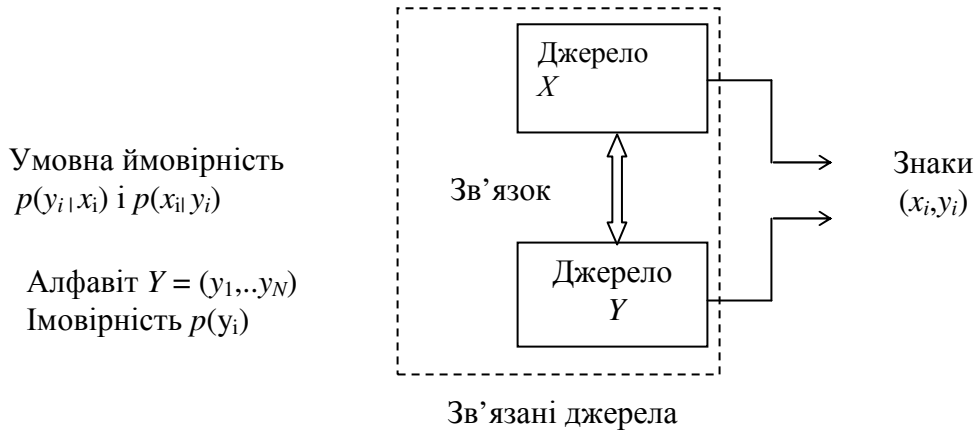


Рисунок 3.2 – Два зв'язаних дискретних джерела

Спільна ентропія двох джерел визначається як математичне сподівання інформації всіх пар подій. Спільна ентропія двох дискретних джерел без пам'яті X і Y

$$\frac{H(X, Y)}{\text{біт}} = - \sum_x \sum_y p(x, y) \log_2 p(x, y). \quad (3.9)$$

Зауваження. Тут мається на увазі, що розглядаються усі пари спільних подій, тобто

$$\sum_x \sum_y p(x, y) = \sum_{i=1}^M \sum_{j=1}^N p(x_i, y_j) = 1.$$

При усередненні умовної інформації всіх пар подій, одержимо умовну ентропію. Умовна ентропія двох дискретних джерел без пам'яті X і Y

$$\begin{aligned} \frac{H(Y|X)}{\text{біт}} &= - \sum_x \sum_y p(x, y) \log_2 p(y|x), \\ \frac{H(X|Y)}{\text{біт}} &= - \sum_x \sum_y p(x, y) \log_2 p(x|y). \end{aligned}$$

Замінивши у формулі (3.9) $\log_2 p(x, y)$ на $\log_2(p(x|y)p(y))$ і на $\log_2(p(y|x)p(y))$ одержимо

$$H(X, Y) = H(Y) + H(X|Y) = H(X) + H(Y|X). \quad (3.10)$$

Таким чином, спільна ентропія може бути представлена у вигляді суми ентропії одного джерела і деякої частини ентропії іншого джерела. Для незалежних джерел ентропія другого джерела входить у суму цілком, тому що $H(X|Y) = H(X)$ і $H(Y|X) = H(Y)$. Для зв'язаних джерел завжди $H(X|Y) < H(X)$ і $H(Y|X) < H(Y)$. Тому, у загальному випадку, завжди має місце

$$H(X, Y) \leq H(X) + H(Y) \quad (3.11)$$

Приклад: Зв'язані джерела.

Зараз самий час докладно розібрати числовий приклад, що наочно пояснює приведені вище визначення і формули. Для цієї мети було підібрано задачу, методика рішення якої може безпосередньо використовуватися на практиці.

Нехай ми маємо вибірку 100000 пар спільних подій (x_i, y_i) дискретних джерел X і Y і алфавіт кожного джерела містить чотири події. Нехай пара (x_1, y_1) зустрілася 10000 разів.

Тоді оцінка імовірності пари (x_1, y_1) дорівнює $10000/100000 = 0,1$. Оцінки інших пар подій також отримані підрахунками їхньої відносної частоти і зведені в таблицю 3.1.

Таблиця 3.1. Оцінка спільної імовірності пар знаків $p(x_i, y_i)$ і імовірність окремих знаків $p(x_i)$ і $p(y_i)$.

	y_1	y_2	y_3	y_4	$p(x_i)$
x_1	0,10	0,05	0,05	0	0,20
x_2	0,05	0,15	0,15	0	0,35
x_3	0	0,10	0,10	0,10	0,30
x_4	0	0,05	0,05	0,05	0,15
$p(y_i)$	0,15	0,35	0,35	0,15	1

Будемо вважати, що отримані оцінки близькі до ймовірностей пар подій і надалі будемо говорити вже про імовірності. Імовірності подій x_i, y_i отримані підсумовуванням рядків і стовпців. Контрольна сума $\sum_{i=1}^4 x_i = \sum_{i=1}^4 y_i = 1$ приведена в правому нижньому куті.

Тепер, коли нам відомі всі імовірності, необхідні для підрахунку ентропії, визначимо:

1. Ентропії джерел X і Y ;
2. Спільну ентропію джерел;
3. Обидві умовні ентропії;

Для контролю ми також обчислимо:

4. Умовні імовірності $p(y_i | x_i)$;
5. Визначимо умовну ентропію $H(Y|X)$.

Зауваження. Для простоти проведемо розрахунки з точністю до 4 знаків після коми.

Рішення.

1.

$$\frac{H(X)}{\text{біт}} = \sum_{i=1}^4 -p(x_i) \log_2 p(x_i) = -2 \cdot [0,15 \cdot \log_2(0,15) + 0,35 \cdot \log_2(0,35)] = 1,9261,$$

$$\frac{H(Y)}{\text{біт}} = \sum_{j=1}^4 -p(y_j) \log_2 p(y_j) = 1,8813;$$

$$2. \frac{H(X, Y)}{\text{біт}} = \sum_{i=1}^4 \sum_{j=1}^4 -p(x_i, y_j) = 3,4464;$$

3. Без довгих обчислень з виразу (3.10) одержуємо

$$H(X|Y) = H(X, Y) - H(Y) = 3,4464 - 1,8813 = 1,5651 \text{ біт},$$

$$H(Y|X) = H(X, Y) - H(X) = 3,4464 - 1,9261 = 1,5203 \text{ біт}.$$

Таблиця 3.2. Умовна ймовірність $p(y_j | x_i)$.

	y_1	y_2	y_3	y_4	$\sum_{j=1}^4 p(y_j x_i)$
x_1	1/2	1/4	1/4	0	1
x_2	1/7	3/7	3/7	0	1
x_3	0	1/3	1/3	1/3	1
x_4	0	1/3	1/3	1/3	1

4. У таблиці 3.2 приведені умовні імовірності, підраховані виходячи з таблиці 3.1. Помітимо, що при цьому ми одержали так звану стохастичну матрицю. Сума умовних ймовірностей для кожного рядка дорівнює 1.

$$5. \frac{H(Y|X)}{\text{біт}} = -\sum_{i=1}^4 \sum_{j=1}^4 p(x_i, y_j) \log_2 p(y_j | x_i) = 1,5496.$$

3.3. Висновки

Всі приведені в попередніх розділах міркування в математичній формі зведені в табл. 3.3. Нагадаємо, що основною ідеєю теорії інформації є подання інформації джерела як міри невизначеності. Ця невизначеність розкривається за допомогою експериментів з випадковими подіями з алфавіту цього джерела. Такий підхід пояснюють три стовпці таблиці. Так як інформація виходить з випадковості подій, у першому стовпці вводиться поняття ймовірності подій і спільної ймовірності пари подій як основних величин. Для пари подій вводиться також поняття умовної ймовірності. У другому стовпці дається визначення інформації події і пари подій, а також умовної і взаємної інформації. І, нарешті, у третьому стовпці, вводиться поняття ентропії як міри невизначеності джерела.

Ентропія джерела, спільна й умовна ентропії двох джерел трактуються як математичні сподівання відповідних інформації подій. Умовна ймовірність – це ймовірність однієї події за умови, що інша подія уже відбулася, тому поняття умовної інформації й умовної ентропії цілком природно виводяться з умовної ймовірності.

Взаємна інформація не має аналога в теорії ймовірності. Це зовсім нове поняття теорії інформації, що грає центральну роль в інформаційній техніці. **Взаємна інформація зв'язує поняття каналу з можливістю передачі інформації в ньому, тобто з його пропускнуою здатністю.** Це поняття буде докладно розглянуто в 6 розділі цього посібника.

Таблиця 3.3 – Дискретні джерела без пам'яті X і Y зі знаками $x \in X = (x_1, x_2, \dots, x_M)$ і $y \in Y = (y_1, y_2, \dots, y_N)$

	Інформація	Ентропія
Імовірність окремого знаку (апостеріорна ймовірність) $p(x)$	Інформація окремого знаку $I(x) = -\log_2 p(x)$ біт	Ентропія $H(x) = -\sum_x p(x) \log_2 p(x)$ біт $H(y) = -\sum_y p(y) \log_2 p(y)$ біт
Спільна інформація двох знаків $p(x, y)$	Інформація пари знаків $I(x, y) = -\log_2 p(x, y)$ біт	$H(X, Y) = -\sum_x \sum_y p(x, y) \log_2 p(x, y)$ біт
Умовна імовірність (апостеріорна ймовірність) $p(x/y) = p(x, y)/p(y)$ $p(y/x) = p(x, y)/p(x)$	Умовна інформація $I(x/y) = -\log_2 p(x/y)$ біт $I(y/x) = -\log_2 p(y/x)$ біт	$H(X/Y) = -\sum_x \sum_y p(x, y) \log_2 p(x/y)$ біт $H(Y/X) = -\sum_x \sum_y p(x, y) \log_2 p(y/x)$ біт
	Взаємна інформація $I(x, y) = \log_2 \frac{\text{апостеріорна ймовірність}}{\text{апостеріорна ймовірність}}$ біт = $\log_2 \frac{p(x/y)}{p(x)}$ біт = $\frac{p(x/y)}{p(x)}$ біт	

РОЗДІЛ 4

СТАЦІОНАРНІ ДИСКРЕТНІ ДЖЕРЕЛА З ПАМ'ЯТТЮ

4.1. Ентропія

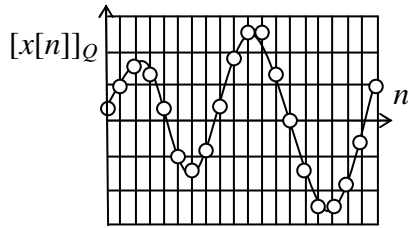


Рисунок 4.1 – Неперервний сигнал

Сигнали аналогових джерел інформації обмежені за смугою, тому корельовані за часом. Прикладом може служити аналоговий мовний сигнал у телефонній лінії. Після перетворення аналог–цифра аналогове джерело перетворюється в дискретне і, наприклад, після квантування сигналу на 256 рівнів, ми одержуємо послідовність 8-ми бітових двійкових цілих чисел від 0 до 255. Як видно з рис. 4.1, значення двох сусідніх чисел близьке один до одного, так як телефонний сигнал передається у вузькій смузі частот. Через часовий зв'язок сусідніх відліків, тобто пам'яті відліку, його невизначеність (інформація) знижується в порівнянні з аналоговим джерелом без пам'яті, тому основною задачею методів стиснення, особливо при передачі відеосигналів, є зниження надмірності.

Виникає питання про те, яким чином визначити ентропію дискретного джерела з пам'яттю. Почнемо з постановки задачі.

Визначення 4.1.1. Дискретне джерело X можна представити як дискретний за часом стохастичний процес, реалізацією, якого є послідовність подій x_n , що належать алфавіту джерела $X = \{\alpha_1, \alpha_2, \dots, \alpha_N\}$.

Зауваження. Щоб уникнути плутанини, ми позначили вміст алфавіту грецькими буквами. У цьому випадку на місці змінної x_n n -ої події може бути поставлене будь-яке число з алфавіту X .

Визначення 4.1.2. Дискретне джерело є *стаціонарним*, якщо спільні імовірності послідовностей подій не залежать від вибору початкової точки відліку часу.

Зауваження. Незалежність спостережень від точки відліку означає, що ми можемо починати вибірку з будь-якого моменту часу, тобто статистика не залежить від часу початку спостережень.

Визначення 4.1.3. Стаціонарне дискретне джерело математично цілком описане, якщо відомі всі спільні імовірності $p(x_{n_1}, x_{n_2}, \dots, x_{n_M})$ для будь-якої вибірки n_1, n_2, \dots, n_M , де $M \rightarrow \infty$.

Визначення ентропії стаціонарного дискретного джерела з пам'яттю впливає з двох підходів, що призводять до однакового результату. При першому підході ми використовуємо поняття спільної ентропії, при другому – умовної ентропії. В обох випадках ми шукаємо відповідь на наступне питання: «Якщо пам'ять джерела поширюється на кілька послідовних подій, то яку додаткову

інформацію несе окрема подія в тому випадку, якщо блок попередніх подій уже відомий?»

Підхід 1. Спільна ентропія.

Спільна ентропія двох джерел x_1 і x_2 з однаковими алфавітами й однаковими розподілами ймовірностей подій визначається як

$$H(x_1, x_2) = - \sum_{i=1}^M \sum_{j=1}^M p(x_i, x_j) \log_2 p(x_i, x_j). \quad (4.1)$$

Поширимо це визначення на L послідовних джерел X_i і знайдемо ентропію джерела X_L як

$$H_L(X) = \frac{1}{L} \cdot H(X_1, X_2, \dots, X_L) = \frac{1}{L} \cdot \sum_{x_L} p(x) \log_2 p(x), \quad (4.2)$$

де вектор $X = (x_1, x_2, \dots, x_L)$ і підсумовування провадиться за всіма можливими компонентами вектора X . Спрямувавши L до нескінченності, ми цілком охопимо пам'ять джерела й одержимо граничне значення $H_L(X)$ (якщо воно існує), що дорівнює

$$H_\infty(X) = \lim_{L \rightarrow \infty} H(X_L). \quad (4.3)$$

Підхід 2. Умовна ентропія.

Умовна ентропія L -ї події у випадку, якщо $(L-1)$ попередніх подій вже відомі, визначається як

$$H_\infty(X) = \lim_{L \rightarrow \infty} H(X_L | X_1, X_2, \dots, X_{L-1}). \quad (4.4)$$

Хоча в лівих частинах рівностей (4.3) і (4.4) ми уже використовували однакове позначення ентропії окремої події, цей факт має бути доведений. Проведемо цей доказ за 4 кроки.

Теорема 4.1.1. Для стаціонарного дискретного джерела з $H_L(x) < \infty$ має місце:

1. $H(X_L | X_1, X_2, \dots, X_{L-1})$ не зростає зі зростанням довжини блоку L ;
2. $H_L(X) \geq H(X_L | X_1, X_2, \dots, X_{L-1})$;
3. $H_L(X)$ не зростає зі зростанням довжини блоку L ;
4. Ентропія стаціонарного дискретного джерела $H_L(X)$

$$\lim_{L \rightarrow \infty} H_L(X) = \lim_{L \rightarrow \infty} H(X_L | X_1, X_2, \dots, X_{L-1}) = H_\infty(X).$$

Доказ.

1. З визначення ентропії, як міри невизначеності джерела, безпосередньо випливає, що зростання числа обмежень не може викликати зростання невизначеності, а, отже, і ентропії.

2. З «правила ланцюжка» [14] для спільної ентропії випливає

$$H_L(X) = \frac{1}{L} \cdot [H(X_1) + H(X_2 | X_1) + \dots + H(X_L | X_1, X_2, \dots, X_{L-1})]. \quad (4.5)$$

Зауваження. «Правило ланцюжка» для спільної ентропії впливає з «правила ланцюжка» для ймовірностей. Найпростіший приклад «правила ланцюжка» для ймовірностей

$$p(x, y) = p(x | y) p(y) \text{ і } p(x, y, z) = p(x | y, z) p(y | z) p(z).$$

Оскільки логарифмічна функція відображає добуток у суму, одержуємо «правило ланцюжка» для спільної ентропії.

Оскільки ентропія завжди додатна і має місце нерівність

$$H(X_1) \geq H(X_2 | X_1) \geq \dots \geq H(X_L | X_1, X_2, \dots, X_{L-1}),$$

з цього випливає нижня оцінка 2.

3. З формули (4.5) насамперед випливає розкладання

$$H_L(X) = \frac{L-1}{L} \cdot H_{L-1}(X) + \frac{1}{L} \cdot H(X_L | X_1, X_2, \dots, X_{L-1}).$$

Використовуючи вже відоме співвідношення (3.5), одержуємо нерівність

$$L \cdot H_L(X) \leq (L-1) \cdot H_{L-1}(X) + H_L(X).$$

Після підстановки одержуємо твердження 3.

$$H_L(X) \leq H_{L-1}(X).$$

4. Твердження 1., 2. і обмеження $H_1(X) < \infty$ встановлюють межі існування.

Використовуючи далі «правило ланцюжка», одержуємо

$$H_{L+j}(X) = \frac{1}{L+j} \cdot [H(X_1, X_2, \dots, X_{L-1}) + H(X_L | X_1, X_2, \dots, X_{L-1}) + \\ + H(X_{L+1} | X_1, X_2, \dots, X_L) + \dots + H(X_{L+j} | X_1, X_2, \dots, X_{L+j-1})].$$

Відповідно до твердження 1., умовна ентропія в правій частині рівності не зростає, тому справедлива оцінка

$$H_{L+j}(X) \leq \frac{1}{L+j} \cdot H(X_1, X_2, \dots, X_{L-1}) + \frac{j+1}{L+j} \cdot H(X_L | X_1, X_2, \dots, X_{L-1}).$$

Спрямувавши j до нескінченності, одержимо

$$\lim_{j \rightarrow \infty} H_{L+j}(X) \leq \lim_{j \rightarrow \infty} \frac{1}{L+j} \cdot H(X_1, X_2, \dots, X_{L-1}) + \frac{j+1}{L+j} \cdot H(X_L | X_1, X_2, \dots, X_{L-1}),$$

що дає для кожного натурального L ,

$$H_\infty(X) \leq H(X_L | X_1, X_2, \dots, X_{L-1}), \quad (4.6)$$

але тому, що для будь-якого натурального L виконується так само і 2., то (4.3) перетворюється в рівність при $L \rightarrow \infty$.

4.2. Теорема кодування джерел 2

Тепер ми можемо доповнити теорію інформації ще однією теоремою. Виявляється, що поєднуючи події джерела в блоки довжини L і кодуючи ці блоки, середня довжина кодового слова на подію може досягти ентропії джерела $H_\infty(x)$ при $L \rightarrow \infty$ як завгодно близько. При цьому пам'ять джерела цілком враховується.

Теорема 4.2.1. Теорема кодування стаціонарного дискретного джерела з ентропією $H_L(X)$.

Для блоку довжини L існує D -ий префіксний код, в якому середня довжина кодового слова на одну подію \bar{n} задовольняє нерівності

$$\frac{H(X)}{\log D} \leq \bar{n} \leq \frac{H_L(X)}{\log D} + \frac{1}{L}. \quad (4.7)$$

Теорема (4.7) не потребує спеціального доказу. Якщо ми розглядаємо блоки довжини L як нові незалежні події з ентропією, що дорівнює $L \cdot H_L(x)$ і застосуємо теорему кодування джерел 1, то ми маємо

$$\frac{LH_L(X)}{\log D} \leq L\bar{n} \leq \frac{LH_L(X)}{\log D} + 1.$$

Розділивши всі члени нерівності на L , одержуємо середню довжину кодового слова на подію

$$\frac{H_L(X)}{\log D} \leq \bar{n} \leq \frac{H_L(X)}{\log D} + \frac{1}{L}.$$

При $L \rightarrow \infty$, $H_L(x) \rightarrow H_\infty(x) = H(x)$ ми маємо

$$\frac{H_\infty(X)}{\log D} \leq \bar{n} \leq \frac{H_\infty(X)}{\log D} + \delta. \quad (4.8)$$

Таким чином, для будь-якого як завгодно малого δ , існує метод кодування блоків, що містять $L > 1/\delta$ подій, при якому для середньої довжини кодового слова на подію – \bar{n} , виконується нерівність (4.8).

Теорема кодування джерел 2 показує, що збільшуючи довжину блоку L , ми можемо як завгодно близько підійти до ентропії $H(x) = H_\infty(x)$. Однак, на практиці існують деякі обмеження, пов'язані з тим, що для того, щоб блок з L подій міг бути декодований, він повинен бути прийнятий цілком, а це, у свою чергу, може призвести до неприпустимих затримок декодування і неприпустимій ємності буфера пам'яті.

4.3. Скінченні ланцюги Маркова (матеріал підвищеної складності)

У цьому і наступному параграфі буде розглядатися спеціальна форма дискретних джерел з пам'яттю – марковські джерела. Їхній опис подібний до марковських ланцюгів, що знайшли різноманітне застосування в інших областях науки і техніки. Так, на основі марковських ланцюгів будуються моделі розпізнавання мови, моделі передачі телефонними каналами, що комутуються. Ланцюги Маркова (А. А. Марков (1856–1922) – видатний російський математик) використовуються під час дослідження моделей мереж зв'язку (канали Гільберта-Еліота) і в теорії керування транспортними потоками. Значення ланцюгів Маркова ґрунтується на тім факті, що з їхньою допомогою можна скласти майже адекватну математичну модель багатьох процесів, які спостерігаються на практиці.

4.3.1. Дискретні за часом ланцюги Маркова

У цьому розділі вводиться поняття скінченних дискретних за часом марковських ланцюгів. Почнемо розгляд цього поняття на найпростішому прикладі «випадкових блукань».

Приклад: Випадкові блукання студента.

Нас цікавить питання про те, чи добереться студент, (який сумнівається «піти – чи не піти на лекцію») до аудиторії. Поставимо питання по-іншому

(рис. 4.2): яка P – ймовірність того, що випадкові блукання на 7 часовому кроці приведуть студента в просторовий стан S_1 ? чи, іншими словами, («випадкові блукання» приведуть до стану S_1 у момент часу $n = 7$).

Ситуація, зображена на рис. 4.2 уже містить у собі найважливіші ознаки ланцюга Маркова. Під марковським ланцюгом розуміється дискретний за часом і за станами марковський процес – $S(n)$. Його реалізацією є безліч шляхів, що ведуть зі стану S_1 у стан S_i .

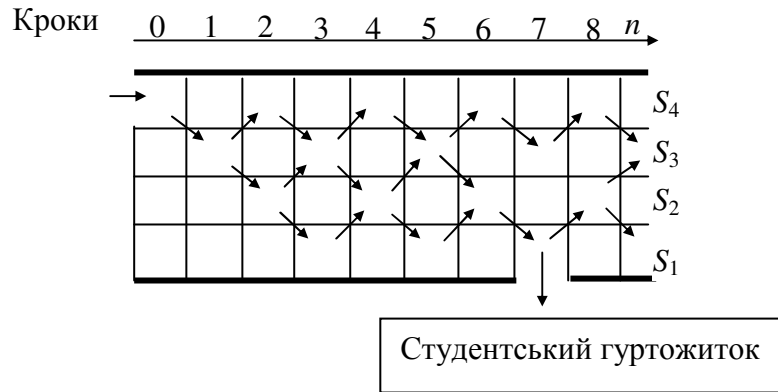


Рисунок 4.2 – Випадкові блукання студента

Вихідним пунктом для опису марковського ланцюга є множина станів

$$S = \{S_1, S_2, \dots, S_N\} \quad (4.9)$$

де N – натуральні числа і стохастичний вектор розподілу ймовірностей станів у момент часу n

$$p_n = (p_n(1), p_n(2), \dots, p_n(N)) \quad (4.10)$$

Для того, щоб повністю визначити ланцюг Маркова, нам залишається задати метод підрахунку ймовірностей $p_n(i)$ для будь-якого моменту часу n . З визначення імовірності маємо

$$0 \leq p_n(i) \leq 1 \forall i = 1, 2, \dots, N \text{ и } \sum_{i=1}^N p_n(i) = 1 \quad (4.11)$$

Особливе значення має розподіл ймовірностей на початку спостереження, тобто початкові умови

$$p_0 = (p_0(1), p_0(2), \dots, p_0(N)) \quad (4.12)$$

Зміна станів описується перехідними ймовірностями

$$\pi_{n_2, n_1}(j/i) = P(s[n_1] = S_i \cap s[n_2] = S_j) \quad (4.13)$$

Ці перехідні, імовірності мають наступні властивості:

1. Один зі станів у момент часу n завжди досягається при будь-якому S_1 у момент n_1

$$\sum_{j=1}^N \pi_{n_2, n_1}(j/i) = 1 \quad (4.14)$$

2. Граничний розподіл, тобто імовірності j -го стану в момент часу n_2

$$\sum_{j=1}^N \pi_{n_2, n_1}(j/i) = p_{n_2}(j). \quad (4.15)$$

3. Рекурсивна формула для перехідних ймовірностей (окремий випадок рівняння Колмогорова–Чепмена.)

$$\pi_{n_3, n_1}(j/i) \sum_{i=1}^N \pi_{n_3, n_2}(j/i) \pi_{n_3, n_1}(j/i). \quad (4.16)$$

Дискретна рівність Колмогорова–Чепмена виходить з трансформації «правила ланцюжка» для ймовірностей

$$p(x_1, x_2) = p(x_2/x_1) p(x_1), \quad (4.17)$$

$$p(x_1, x_2, x_3) = p(x_3/x_1, x_2) p(x_2/x_1) p(x_1) \quad (4.18)$$

Помноживши обидві частини рівності на $1/p(x_1)$, одержимо

$$p(x_2, x_3/x_3) = p(x_3/x_1, x_2) p(x_2/x_1). \quad (4.19)$$

Підсумовуючи за всіма x_2 , приходимо до дискретної форми рівняння Колмогорова–Чепмена

$$p(x_3/x_1) = \sum_{x_2} p(x_3/x_1, x_2) p(x_2/x_1). \quad (4.20)$$

При перетворенні (4.20) у (4.16), ми припускаємо

$$p(x_3, x_1, x_2) = p(x_3/x_2). \quad (4.21)$$

Остання рівність характеризує марковський процес.

Визначення 4.3.1. Марковським процесом називається процес, в якому минуле не впливає на майбутнє, якщо сьогоднішнє відоме.

Зауваження. У розглянутому прикладі відоме минуле – це стохастична змінна X_1 , відоме сьогоднішнє – X_2 , невідоме майбутнє – X_3 . Застосування ланцюгів Маркова сильно спрощується, коли вони гомогенні, стаціонарні і регулярні. Розглянемо ці три поняття.

Визначення 4.3.2. Ланцюг Маркова гомогенний, якщо перехідні ймовірності між станами не залежать від вибору часової точки відліку. Таким чином, перехідні ймовірності залежать тільки від різниці часів відліків l

$$\pi(j/i) = p(s[n]s_1 \cup s[n+l] = s_1) \quad n = 0, 1, 2, 3, \dots \quad (4.22)$$

У цьому випадку для $l = n_2 - n_1$ і $m = n_3 - n_2$ рівність Колмогорова–Чепмена (4.16) можна записати у вигляді

$$\pi_{l+m}(j/i) = \sum_r \pi_m(j/r) \pi_l(r/i). \quad (4.23)$$

Ця рівність може бути представлена у вигляді суми покомпонентних добутків векторів-рядків на вектори-стовпці. Записавши перехідні ймовірності у вигляді матриці, одержимо рівняння (4.23) у матричній формі

$$\begin{pmatrix} \pi_{l+m}(1/1) & \pi_{l+m}(2/1) & \dots & \pi_{l+m}(N/1) \\ \pi_{l+m}(1/2) & \pi_{l+m}(2/2) & \dots & \pi_{l+m}(N/2) \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{l+m}(1/N) & \pi_{l+m}(2/N) & \dots & \pi_{l+m}(N/N) \end{pmatrix} = \begin{pmatrix} \pi_{l+m}(1/1) & \pi_{l+m}(2/1) & \dots & \pi_{l+m}(N/1) \\ \pi_{l+m}(1/2) & \pi_{l+m}(2/2) & \dots & \pi_{l+m}(N/2) \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{l+m}(1/N) & \pi_{l+m}(2/N) & \dots & \pi_{l+m}(N/N) \end{pmatrix} \begin{pmatrix} \pi_{l+m}(1/1) & \pi_{l+m}(2/1) & \dots & \pi_{l+m}(N/1) \\ \pi_{l+m}(1/2) & \pi_{l+m}(2/2) & \dots & \pi_{l+m}(N/2) \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{l+m}(1/N) & \pi_{l+m}(2/N) & \dots & \pi_{l+m}(N/N) \end{pmatrix} \quad (4.24)$$

Цей процес можна починати з першого кроку за допомогою матриці перехідних ймовірностей

$$\Pi = \begin{pmatrix} \pi(1/1) & \pi(2/1) & \dots & \pi(N/1) \\ \pi(1/2) & \pi(2/2) & \dots & \pi(N/2) \\ \vdots & \vdots & \ddots & \vdots \\ \pi(1/N) & \pi(2/N) & \dots & \pi(N/N) \end{pmatrix}. \quad (4.25)$$

Помітимо, що матриця Π – стохастична матриця із сумою ймовірностей рядків, що дорівнює 1. Багаторазово застосовуючи таке матричне перетворення до вихідного розподілу n станів p_0 ми можемо одержати розподіл ймовірностей p_n у будь-який момент часу

$$p_n = p_0 \Pi^n \quad (4.26)$$

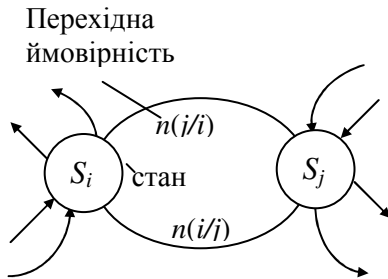


Рисунок 4.3 – Граф станів гомогенного ланцюга Маркова

Теорема 4.3.1. Гомогенний ланцюг Маркова цілком характеризується матрицею перехідних ймовірностей і вихідним розподілом станів. Наведені вище міркування можна наочно узагальнити за допомогою графа станів (рис. 4.3). Тут вузли відповідають станам, а шляхи між ними – переходам між станами. Кожному шляху приписано вагу, яка дорівнює перехідній ймовірності. Таким чином, граф станів подає повну інформацію про матрицю перехідних ймовірностей.

Приклад: Випадкові блукання студента (продовження).

Випадкові блукання (див. рис. 4.2) надано на рис. 4.4 у вигляді графа станів.

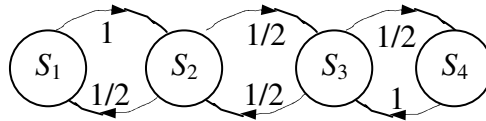


Рисунок 4.4 – Граф станів для випадкових блукань студента

Графу станів відповідає матриця перехідних ймовірностей

$$\Pi = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (4.27)$$

З рис. 4.2 випливає, що вихідний розподіл станів

$$p_0 = (0,0,0,1) \quad (4.28)$$

Ймовірність того, що в результаті випадкових блукань студент на 7-му часовому кроці з'явиться в дверях гуртожитку, визначається першим компонентом вектора станів на 7 кроці. З p_7 (4.26) випливає

$$p_7 = p_0 \Pi^7 \quad (4.29)$$

Розрахунки за допомогою комп'ютера дають

$$\Pi_7 = \begin{pmatrix} 0 & 0,619 & 0 & 0,3281 \\ 0,3359 & 0 & 0,6641 & 0 \\ 0 & 0,6641 & 0 & 0,3359 \\ 0,3281 & 0 & 0,6719 & 0 \end{pmatrix}. \quad (4.30).$$

Звідси

$$p_7 = (0,0,0,1) \cdot \Pi^7 = (0,3281, 0, 0,6719, 0) \quad (4.31)$$

і ймовірність, яку шукали дорівнює

$$p_7(1) = 0,3281 \quad (4.32)$$

Найважливішим окремим випадком марковського ланцюга є випадок, коли розподіл станів не залежить від часу спостереження.

Визначення 4.3.3. Гомогенний ланцюг Маркова стаціонарний, якщо розподіл станів постійний за часом. У цьому випадку початковий розподіл є власним вектором матриці перехідних ймовірностей, тобто

$$p_0 \Pi = p_0. \quad (4.33)$$

Зауваження. Якщо вектор розподілу станів є стохастичним вектором із сумою компонентів, що дорівнює 1, то сума компонентів власного вектора також дорівнює 1.

Для того, щоб ланцюг Маркова був стаціонарним, має виконуватися (4.33). Тоді

$$p_0 \Pi = (p_1, p_2, p_3, p_4) \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{pmatrix} = (p_2/2, p_1 + p_3/2, p_2/2 + p_4, p_3/2), \quad (4.34)$$

$$p_1 + p_2 + p_3 + p_4 = 1. \quad (4.35)$$

Використовуючи умови (4.33) і (4.35), знаходимо стаціонарний розподіл станів

$$p_0 = (1/6, 1/3, 1/3, 1/6). \quad (4.36)$$

З рекурсивного співвідношення (4.26) постають наступні найважливіші запитання: Що відбувається на «протязі довгого часу», тобто при $n \rightarrow \infty$? Чи встановлюється стаціонарний розподіл станів? Чи є щось подібне до стаціонарного розподілу, наприклад, два сталих розподіли станів?

Визначення 4.3.4. Гомогенний ланцюг Маркова називається регулярним, якщо:

- Гранична матриця

$$\Pi_\infty = \lim_{n \rightarrow \infty} \Pi^n \quad (4.37)$$

існує, причому, усі N рядків граничної матриці являють собою граничний розподіл p_∞ ;

- Граничний розподіл є єдиним стаціонарним розподілом ймовірностей станів будь-якого регулярного ланцюга Маркова;
- Ланцюг Маркова завжди регулярний, якщо існує деяке натуральне n , при якому всі компоненти деякого стовпця матриці Π^n відмінні від нуля.

Останнє з тверджень визначення 4.3.4 рівносильне наступному: ланцюг Маркова є регулярним, якщо на деякому кроці n існує щонайменше один стан, який може бути досягнутий з будь-якого початкового стану.

Приклад: Випадкові блукання (продовження).

Розглянемо приклад випадкових блукань студента і з'ясуємо, чи є відповідним цим блуканням регулярний ланцюг Маркова.

Матриця перехідних ймовірностей (4.27) має два граничних стани залежно від того, чи є число часових кроків n парним чи непарним

$$\lim_{N \rightarrow \infty} \Pi^{2N} = \begin{pmatrix} 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \end{pmatrix}, \quad (4.38)$$

$$\lim_{N \rightarrow \infty} \Pi^{2N+1} = \begin{pmatrix} 0 & 2/3 & 0 & 1/3 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 1/3 & 0 & 2/3 & 0 \end{pmatrix}, \quad (4.39)$$

тому, у даному випадку ланцюг Маркова не є регулярним. З іншого боку, не виконується й остання умова з визначення 4.3.4, тому що на кожному кроці всі парні стани переходять у непарні і навпаки (див. рис. 4.2).

Зауваження. Якщо ми виберемо початковий розподіл, наприклад рівний p_0 з (4.4), то на будь-якому часовому кроці кожний зі станів досяжний.

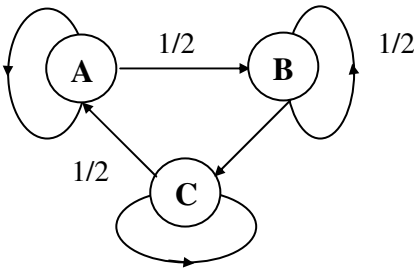


Рисунок 4.5 – Граф станів

Приклад: Марковський ланцюг із трьома станами.

Нехай марковський ланцюг заданий графом станів (рис. 4.5)

1. Побудувати матрицю перехідних ймовірностей.

2. Показати, що ланцюг Маркова стаціонарний. При цьому виходити з рівномірного початкового розподілу станів.

3. Показати, що ланцюг Маркова регулярний.

4. Побудувати граничну перехідну матрицю.

Рішення.

1. Матриця перехідних ймовірностей

$$\Pi = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}; \quad (4.40)$$

2. Стаціонарність.

Так як початковий розподіл станів рівномірний, то

$$p_0 = 1/3 (1, 1, 1), \quad (4.41)$$

при цьому виконується умова (4.4)

$$p_0 \Pi = 1/3 (1, 1, 1) \begin{pmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix} = 1/3 (1, 1, 1) = p_0; \quad (4.42)$$

3. Ланцюг Маркова регулярний, тому що

$$\Pi^2 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} = 1/4 \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{pmatrix} \quad (4.43)$$

і для Π^2 виконується остання з умов визначення 4.3.4;

4. Гранична перехідна матриця.

Оскільки ланцюг Маркова регулярний, скористаємося визначенням 4.3.4, відповідно до якого $p_\infty = p_0$ і з (4.41) маємо

$$\Pi_{\infty} = \begin{pmatrix} p_0 \\ p_0 \\ p_0 \end{pmatrix} = 1/3 \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \quad (4.44)$$

4.3.2. Скінченні дискретні марковські джерела з пам'яттю r

Марковські ланцюги можна з успіхом використовувати для моделювання скінчених дискретних джерел з пам'яттю. Припускаючи, що стохастичні параметри джерел з пам'яттю можуть бути підраховані як середні за часом величини, (тобто джерела мають властивість ергодичності), намітимо шлях подальших міркувань.

Нехай задана довільна послідовність $\{x[n]\} = \{a, b, a, c, b, b, f, d, b, b, a, c, \dots\}$ джерела з алфавітом $X = \{a, b, c, d\}$. Ми вже раніше визначили частоти подій, як оцінки для ймовірностей подій $p(a)$, $p(b)$, $p(c)$ і $p(d)$ і знайшли ентропію джерела, вважаючи події незалежними. Якщо джерело має пам'ять, то його ентропія може бути тільки менше, тобто раніше ми знаходили оцінку зверху. Виникає питання, яким чином можна включити до аналізу пам'ять. Для цього необхідно врахувати залежність між подіями. Оцінимо умовні ймовірності джерела $p(ala)$, $p(bla)$, $p(c/a)$, $p(d/a)$, $p(a/b)$, ..., $p(d/d)$ двох послідовних подій шляхом підрахунку частот парних подій. Після цього джерело може бути розкладено на чотири підджерела, що відповідають першим символам у парних подіях.

На рис. 4.6 цей перший крок міркувань наочно продемонстрований. Тут символ a визначає один з чотирьох підджерел.

Подіям, що відбуваються за символом b (шляхам на графі), приписуються ваги, рівні ймовірностям, наприклад $p^{(a)}(b) = p(b/a)$. Таким чином, кожне таке підджерело вже можна розглядати як деяке самостійне джерело без пам'яті. Ентропія такого джерела може бути обчислена відомими методами. Вихідне джерело з пам'яттю являє собою стохастичну сукупність чотирьох підджерел без пам'яті, а його ентропія визначається середніми значеннями ентропії цих підджерел. Ми можемо продовжити міркування, розглядаючи все більш тривалі стани підджерел (наприклад, вектори \mathbf{a} , \mathbf{a} чи $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$) доти, поки всю пам'ять вихідного джерела не буде охоплено.

Ці евристичні міркування узагальнені в наступному визначенні.

Визначення 4.3.5. Скінченне дискретне марковське джерело з r цілком визначається наступними умовами:

1. Задано непорожню множину станів $S = \{S_1, S_2, \dots, S_N\}$, причому, S містить вектори довжини r .

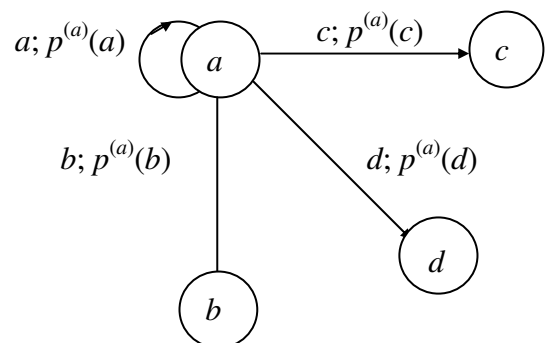


Рисунок 4.6 – Подання джерела у вигляді ланцюга Маркова (перший крок)

2. Кожен стан S_1 відповідає дискретному джерелу без пам'яті з алфавітом $X_1 = \{x_1, x_2, \dots, x_M\}$ і ймовірностями j -их символів алфавіту $p^{(1)}(j)$;

3. Стан $S[n] = (x[n-r], x[n-r+1], \dots, x[n-1])$ з $r-1$ послідовних символів і черговий символ $x[n]$ створюють новий стан $S[n+1] = (x[n-r+1], x[n-r+2], \dots, x[n])$;

4. Задано початковий розподіл станів $p_0 = (p_0(1), p_0(2), \dots, p_0(N)) \dots$

Ми бачимо, що пам'ять r охоплює r послідовних символів, так як на ймовірність чергового символу впливає в точності r попередніх символів. Пояснимо це більш докладно на прикладі

Приклад: Марковське джерело з пам'яттю $r = 2$

Розглянемо двійкове джерело з алфавітом $X = \{0,1\}$. Комбінації двох символів дають чотири стани

$$S = \{S_1 \cong (0, 0), S_2 \cong (1, 1), S_3 \cong (0, 1), S_4 \cong (1, 0)\}. \quad (4.45)$$

Перехідні $p_{S_1}(p(x[n] = 0), p(x[n] = 1))$ ймовірності між станами задаються величинами

$$p_{S_1} = (0,1), p_{S_2} = (1/2, 1/2), p_{S_3} = (2/3, 1/3), p_{S_4} = (3/4, 1/4) \quad (4.46)$$

Якщо задати ще і початковий розподіл станів

$$p_0 = \{p_0(1), p_0(2), p_0(3), p_0(4)\}, \quad (4.47)$$

то всі вимоги з 4.3.5 виконані і скінчене марковське джерело визначене. Умови (4.45) і (4.46) є достатніми для побудови графа станів, який надано на рис. 4.7.

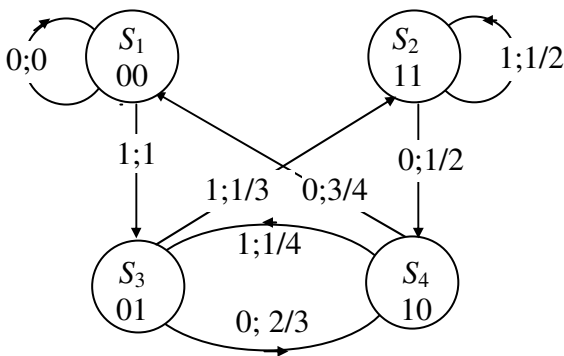


Рисунок 4.7 – Граф станів марковського джерела з пам'яттю r

Проаналізуємо матрицю перехідних ймовірностей і дослідимо її на регулярність. Матриця перехідних ймовірностей будується за графом станів рис. 4.7 і має вигляд

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/3 & 0 & 2/3 \\ 3/4 & 0 & 1/4 & 0 \end{pmatrix} \quad (4.48)$$

Регулярність перевіряється за допомогою граничної матриці. Відповідно до визначення 4.3.4

$$P_\infty = \frac{1}{4} \begin{pmatrix} 9 & 8 & 12 & 12 \\ 9 & 8 & 12 & 12 \\ 9 & 8 & 12 & 12 \\ 9 & 8 & 12 & 12 \end{pmatrix}. \quad (4.49)$$

Зауваження. Граничну матрицю було знайдено за допомогою програмної системи MatLab (<http://www.mathworks.com>).

Усі рядки граничної матриці однакові, отже вона є регулярною. Відповідний граничний розподіл має вигляд

$$p_\infty = \frac{1}{41}(9,8,12,12). \quad (4.50)$$

Принципи покрокової апроксимації джерела з пам'яттю узагальнює наступне твердження.

Теорема 4.3.2. Стационарне марковське джерело з пам'яттю r може бути апроксимоване стаціонарним марковським джерелом з пам'яттю l , де $0 \leq l < r$.

Якщо величина r заздалегідь відома, то на першому кроці апроксимації розглядається джерело без пам'яті.

Модель джерела без пам'яті цілком описується розподілом ймовірностей символів. Середня ймовірність символів – це ймовірність, що оцінює спостерігач, не знаючи, в якому стані знаходиться джерело, тому, воно визначається стаціонарним розподілом ймовірностей станів p_∞ і ймовірностями символів a_1, \dots, a_M у станах S_1, \dots, S_N

$$(p(a_1), \dots, p(a_M)) = p_\infty \cdot \begin{pmatrix} p_{s_1}(a_1) & p_{s_1}(a_2) & \dots & p_{s_1}(a_M) \\ p_{s_2}(a_1) & p_{s_2}(a_2) & \dots & p_{s_2}(a_M) \\ \vdots & \vdots & \ddots & \vdots \\ p_{s_N}(a_1) & p_{s_N}(a_2) & \dots & p_{s_N}(a_M) \end{pmatrix}. \quad (4.51)$$

Приклад: Марковське джерело з пам'яттю $r=2$ (продовження).

• Джерело без пам'яті $l=0$. У числовому прикладі для $a_1=0$ і $a_2=1$ одержуємо

$$\{p(0), p(1)\} = \frac{1}{41}(9,8,12,12) \cdot \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \\ 2/3 & 1/3 \\ 3/4 & 1/4 \end{pmatrix} = \frac{1}{41}(21,20). \quad (4.52)$$

• Стаціонарне марковське джерело з пам'яттю $l=1$, у цьому випадку модель джерела має два стани. Відповідний граф станів зображено на рис. 4.8.

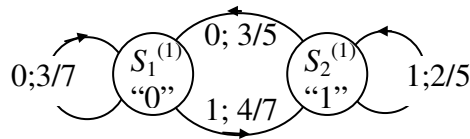


Рисунок 4.8 – Граф станів марковського джерела, яке апроксимується

Визначимо відповідні графу ймовірності. Ймовірності станів дорівнюють

$$P(S_1^{(1)}) = p(0) = \frac{21}{41} \text{ и } P(S_2^{(1)}) = p(1) = \frac{20}{41} \quad (4.53)$$

Спільні ймовірності пар символів можна безпосередньо оцінити за первинним джерелом. Вони будуть дорівнювати ймовірностям станів (4.47)

$$\begin{aligned} p(0,0) = p_\infty(1) = \frac{9}{41}, \quad p(1,1) = p_\infty(2) = \frac{8}{41} \\ p(0,1) = p_\infty(3) = \frac{12}{41}, \quad p(1,0) = p_\infty(4) = \frac{12}{41}. \end{aligned} \quad (4.54)$$

Тепер можна визначити перехідні ймовірності для джерела, яке апроксимується. Відповідно до їхнього визначення

$$\pi^{(1)}(j/i) = \frac{p(i,j)}{p(i)}, \quad (4.55)$$

одержуємо матрицю перехідних ймовірностей

$$\Pi^{(1)} = \begin{pmatrix} 3/7 & 4/7 \\ 3/5 & 2/5 \end{pmatrix}. \quad (4.56)$$

Регулярність перевіряємо шляхом пошуку граничної матриці

$$\Pi_\infty^{(1)} = \begin{pmatrix} 0,5122 & 0,4878 \\ 0,5122 & 0,4878 \end{pmatrix}. \quad (4.57)$$

Так як рядки граничної матриці рівні, ми маємо регулярний марковський ланцюг із граничним розподілом

$$p_{\infty}^{(1)} \approx (0,5122, 0,4878) \quad (4.58)$$

Так як стани відповідають символам 0 і 1, повинне виконуватися

$$p_{\infty}^{(1)} \approx (21/41, 20/41). \quad (4.59)$$

4.4. Ентропія стаціонарного марковського джерела

Ентропія стаціонарного марковського джерела обчислюється, виходячи з того, що кожний стан джерела є підджерелом без пам'яті, яке має визначену ентропію. Таким чином, ентропія первинного джерела дорівнює математичному сподіванню ентропій підджерел.

Теорема 4.4.1. *Стаціонарне марковське джерело з алфавітом із M символів, яке має N станів, тобто N підджерел, ентропія кожного з яких дорівнює*

$$H(X/S_1) = -\sum_{m=1}^M p_{s_1}(x_m) \log_2(p_{s_1}(x_m)) \text{ біт}, \quad (4.60)$$

має ентропію, рівну математичному сподіванню ентропій підджерел

$$H_{\infty}(x) = \sum_{i=1}^N p_{\infty} H(X/S_i). \quad (4.61)$$

Надалі буде показано, що евристичний підхід (4.61) відповідає загальній властивості ентропії стаціонарного дискретного джерела у твердженні 4.1.1.

Зауваження. *Доказ твердження досить складний, тому його можна опустити без втрати розуміння наступних розділів.*

Доказ. Ентропія марковського джерела [10]. Доказ проводиться в три кроки.

Крок 1.

На першому кроці покажемо, що при відомому стані $z_0 = S_j$ умовна ентропія марковського джерела визначається як

$$H(X_l | X_{l-1}, \dots, X_0, Z_0 = S_j) = \sum_{r=1}^N \pi_l(r/j) H(X/Z = S_1), \quad (4.62)$$

де через X_l позначено l -е підджерело, а через Z_l – стан на кроці l

Для того, щоб довести (4.62), розглянемо і перетворимо умовні імовірності $P(x_l | x_{l-1}, \dots, x_0, Z_0 = S_j)$, які є підґрунтям ентропій. Як подальше обмеження введемо стан Z_i

$$P(x_l | x_{l-1}, \dots, x_0, Z_0 = S_j) = P(x_l | Z_l, x_{l-1}, \dots, x_0, Z_0 = S_j) \quad (4.63)$$

Так як Z_l цілком визначається початковим станом Z_0 і символами x_0, \dots, x_{l-1} додаткове обмеження не впливає на умовну ймовірність, тобто рівність справедлива. Друге перетворення впливає з передбачуваної властивості марковського джерела, відповідно до якого l -ий символ залежить від l -го стану

$$P(x_l | Z_l, x_{l-1}, \dots, x_0, Z_0 = S_j) = P(x_l | Z_l). \quad (4.64)$$

Для того, щоб знайти умовну ентропію (4.62), потрібно помножити умовні імовірності, що лежать в основі ентропії, на відповідні імовірності і знайти математичне сподівання

$$\frac{N(X_l | X_{l-1}, \dots, X_0, Z_0 = S_j)}{\text{біт}} = \quad (4.65)$$

$$= \sum_{x_0, x_1, \dots, x_l, Z_l} P(x_l, x_{l-1}, \dots, x_0, Z_l | Z_0 = S_j) \times \log_2 P(x_l, x_{l-1}, \dots, x_0, Z_0 = S_j)$$

Величини Z_l не чинять впливу на спільні імовірності, тому що розподіл суми за всіма станами Z_l є граничним. З огляду на праву частину (4.64), остаточно одержуємо

$$\frac{N(X_l | X_{l-1}, \dots, X_0, Z_0 = S_j)}{\text{біт}} = \sum_{x_l, Z_l} P(x_l, Z_l | Z_0 = S_j) \log_2 P(x_l | Z_l) \quad (4.66)$$

Ймовірність символу x_l на l -ому часовому кроці при заданих Z_0 і S_l дорівнює імовірності S_1 у S_2 за умови, що в стані S_1 заданий символ x_l

$$P(x_l Z_l | Z_0 = S_j) = P(x_l | S_j) = P(x_l | S_j) \pi_l(i/j) \quad (4.67)$$

тому, для (4.66) ми можемо написати

$$\sum_{x_l, Z_l} P(x_l, Z_l | Z_0 = S_j) \log_2 P(x_l | Z_l) = \sum_{x_l, S_l} \pi_l(i/j) P(x_l | S_l) \log_2 P(x_l | S_l). \quad (4.68)$$

Сума за всіма символами X_l при заданому стані S_l дорівнює ентропії i -го підджерела

$$\sum_{S_l} \pi_l(i/j) \sum_{x_l} P(x_l | S_l) \log P(x_l | S_l) \sum_{i=1}^N \pi(i-j) H(X | S_l) \quad (4.69)$$

і твердження (4.62) доведене.

Крок 2.

Дотепер ми виходили з фіксованого початкового стану. Для того, щоб визначити ентропію марковського джерела через математичне сподівання, будемо вважати початковий стан випадковим

$$H(X_l | X_{l-1}, \dots, X_0, Z_0) = \sum_{j=1}^N \sum_{i=1}^N P_{Z_0}(j) \pi_l(i/j) H(X | S_l) \quad (4.70)$$

Сума по j містить у собі всі переходи в i -й стан, що дає в результаті ймовірність i -го стану на l -му кроці. З огляду на стаціонарність джерела, одержуємо

$$\sum_{j=1}^N P_{Z_0}(j) \pi_l(i/j) = p_\infty(i). \quad (4.71)$$

Остаточно маємо

$$H(X_l | X_{l-1}, \dots, X_0, Z_0) = \sum_{i=1}^N p_\infty(i) H(X | S_l). \quad (4.72)$$

Помітимо, що умовна ентропія стаціонарного марковського джерела не залежить від числа символів l . Праві частини рівностей (4.72) і (4.61) однакові. Залишається довести рівність лівих частин (4.72) і (4.61) при $\rightarrow \infty$.

Крок 3.

Розглянемо вираз для ентропії при відомому початковому стані і перетворимо його за допомогою «правила ланцюжка»

$$\frac{1}{L} H(X_L, \dots, X_0 | Z_0) = \frac{1}{L} [H(X_0 | Z_0) + H(X_1 | X_0) + \dots + H(X_L | X_{L-1}, \dots, X_0, Z_0)] \quad (4.83)$$

Так як марковське джерело стаціонарне, усі доданки правої частини рівні і визначаються відповідно до (4.72). Якщо є L доданків, то для будь-якого натурального L одержуємо

$$\lim_{L \rightarrow \infty} \frac{1}{L} H(X_L, \dots, X_0 | Z_0) = \sum_{i=1}^N p_{\infty}(i) H(X | S_i) \quad (4.74)$$

Умова в лівій частині ще не порушена. Використовуючи універсальне співвідношення

$$H(X) = I(X; Y) + H(X|Y) \quad (4.75)$$

Одержимо

$$H(X_L, \dots, X_0 | Z_0) = H(X_L, \dots, X_0) - I(X_L, \dots, X_0; Z_0). \quad (4.76)$$

З (4.74) випливає

$$\lim_{L \rightarrow \infty} [H(X_L, \dots, X_0) - I(X_L, \dots, X_0; Z_0)] = \sum_{i=1}^N p_{\infty}(i) H(X | S_i). \quad (4.77)$$

З іншого боку, використовуючи універсальне співвідношення

$$I(X; Y) = H(Y) - H(Y|X) \quad (4.78)$$

одержимо

$$I(X_L, \dots, X_0; Z_0) = H(Z_0) - H(Z_0 | X_L, \dots, X_0) \dots \quad (4.79)$$

Для ентропії має місце оцінка

$$0 \leq H(Z_0 | X_L, \dots, X_0) \leq H(Z_0) \leq \log N_t \quad (4.80)$$

отже

$$0 \leq I(X_L, \dots, X_0; Z_0) \leq H(Z_0) \leq \log N \quad (4.81)$$

і в граничному випадку одержимо

$$\lim_{L \rightarrow \infty} \frac{1}{L} I(X_L, \dots, X_0; Z_0) \leq 0. \quad (4.82)$$

Використовуючи (4.82), з (4.77) для ентропії стаціонарного марковського джерела остаточно одержуємо (4.61).

Як приклад розглянемо знову марковське джерело з $r = 2$ і його апроксимації з попередніх розділів.

1. Джерело без пам'яті.

Насамперед визначимо ентропію джерела без урахування пам'яті. З розподілу ймовірностей двох символів одержимо ентропію, яка дорівнює

$$\frac{H^0(X)}{\text{біт}} = -\frac{21}{41} = \log_2 \frac{21}{41} - \frac{20}{41} \log_2 \frac{20}{41} \approx 0,999 \quad (4.83)$$

Ентропія близька до одиниці, тому що символи майже рівноймовірні.

2. Марковське джерело з пам'яттю $r = 1$

Розглянемо апроксимацію первинного джерела джерелом з $r = 1$, що складається з двох підджерел. З урахуванням ймовірностей (див. рис. 4.8), одержуємо

$$\begin{aligned} \frac{H_1^{(1)}(X/Z = S_1^{(1)})}{\text{біт}} &= -\pi^{(1)}(1/1) \log_2 \pi^{(1)}(1/1) - \\ &- \pi^{(1)}(2/1) \log_2 \pi^{(1)}(2/1) = -3/7 \log_2 3/7 - 4/7 \log_2 4/7 \approx 0,985 \end{aligned} \quad (4.84)$$

$$\begin{aligned} \frac{H_2^{(1)}(X/Z = S_2^{(1)})}{\text{біт}} &= -\pi^{(1)}(1/2) \log_2 \pi^{(1)}(1/2) - \\ &- \pi^{(1)}(2/2) \log_2 \pi^{(1)}(2/2) = -3/5 \log_2 3/5 - 2/5 \log_2 2/5 \approx 0,971. \end{aligned} \quad (4.85)$$

Таким чином, ентропія марковського джерела з $r = 1$ дорівнює

$$\begin{aligned} \frac{H^{(1)}(X)}{\text{біт}} &= P(S_1^{(1)})H_1^{(1)}(X) + P(S_2^{(1)})H_2^{(1)}(X) = \\ &= \frac{21}{41}0,985 + \frac{20}{41}0,971 \approx 0,979 \end{aligned} \quad (4.86)$$

у порівнянні з (1), ентропія дещо зменшилась.

3. Марковське джерело з пам'яттю $r = 2$.

У цьому випадку ми повинні брати до уваги 4 підджерела (див. рис. 4.7). Для стану S_1 маємо

$$H_L(X - Z = S_1) = 0 \text{ біт}. \quad (4.87)$$

Так як підджерело 2 постійно виробляє символ «1», ентропія в стані S_2 дорівнює

$$H_2(X - Z = S_2) = 1 \text{ біт}. \quad (4.88)$$

Підджерела 3 і 4 мають ентропію відповідно

$$\frac{H_3(X/Z) = S_3}{\text{біт}} = -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} \approx 0,918 \quad (4.89)$$

$$\frac{H_4(X/Z) = S_4}{\text{біт}} = -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} \approx 0,811. \quad (4.90)$$

В результаті одержуємо ентропію джерела з $r = 2$, що дорівнює

$$\begin{aligned} H_\infty(X) \sum_{r=1}^4 P(S_r) H_r(X) &= \\ &= \frac{1}{41} (8 + 12 \cdot 0,918 + 12 \cdot 0,811) \text{ біт} \approx 0,701 \text{ біт} \end{aligned} \quad (4.91)$$

Отже, ми бачимо, що в порівнянні з марковським джерелом з пам'яттю $r = 1$, ентропія зменшилася.

4.5. Кодування стаціонарних марковських джерел

Насамперед згадаємо теорему кодування джерел 4.2.1. У цій теоремі розглядаються блоки, що містять L символів. Теорема стверджує, що у випадку, коли L прямує до нескінченності, для блоків довжини L існує префіксний код, у якому середня довжина кодового слова на один символ як завжди близька до спільної ентропії $H_L(X)$. У випадку марковського джерела з пам'яттю r , спільна ентропія для блоку довжини $L = r$ є граничним випадком і теорема кодування може бути сформульована в такий спосіб:

Теорема 4.5.1. *Теорема кодування стаціонарних марковських джерел з пам'яттю r і ентропією $H_\infty(x)$.*

Для блоку довжини $L \geq r$ існує D -ий префіксний код, в якого середня довжина кодового слова \bar{n} задовольняє нерівності

$$\frac{H_\infty(x)}{\log D} \leq \bar{n} \leq \frac{H_\infty(x)}{\log D} + \frac{1}{L}. \quad (4.92)$$

З розкладання марковського джерела на підджерела без пам'яті безпосередньо впливає стратегія оптимального кодування.

Якщо початкові стани відомі, то під час кодування джерела всі наступні стани однозначно визначені. При цьому в передавачі і приймачі для кожної множини підджерел можливо провести кодування і декодування Хаффмана, враховуючи розподіл ймовірностей символів і станів.

Практична реалізація кодів Хаффмана показує, що для досягнення істотного кодового виграшу, необхідно кодувати блоки досить великої довжини. Крім цього, базові стани завжди повинні визначатися r символами для того, щоб перехід до блоків більшої довжини був відносно нескладним.

Запропонована проста стратегія цілком враховує пам'ять джерела і, отже, у граничному випадку, дозволяє одержати оптимальний префіксний код.

Кодування стаціонарного марковського джерела X з пам'яттю r і ентропією, що дорівнює $H_\infty(X)$.

1. Об'єднати в блоки $l = r + 1$ символів джерела.

2. Провести кодування Хаффмана для блоків.

3. Якщо середня довжина кодового слова на символ істотно відрізняється від ентропії $H_\infty(X)$, то збільшити довжину блоку за рахунок наступних символів. Провести кодування Хаффмана для блоків більшої довжини. Продовжити цей процес до задовільного наближення середньої довжини кодового слова до ентропії.

Приклад: Кодування марковського джерела з пам'яттю $r = 2$ (продовження).

Перевіримо ефективність запропонованого алгоритму на числовому прикладі з попереднього розділу.

Відповідно до пам'яті джерела $r = 2$, об'єднаємо в блоки кожні три символи джерела і проведемо кодування Хаффмана. Необхідні імовірності станів для блоків визначаються стаціонарним розподілом ймовірностей і матрицею перехідних ймовірностей (4.48) чи графом станів (рис. 4.7). Ймовірність блоку 001, наприклад, дорівнює

$$P_{001} = P(S_1)\pi(3/1) = P_\infty(1)\pi(3/1) = 9/41 \cdot 1 = \frac{9}{41}. \quad (4.93)$$

Можна так само помітити, що блок 000 ніколи не з'являється і, отже, не повинен кодуватися.

З табл. 4.1. знаходимо, що середня довжина кодового слова на символ дорівнює 0,911. Тому ефективність кодування відносно мала

$$\eta = \frac{H_\infty(x)}{\bar{n}} = \frac{0,701}{0,911} \approx 0,77. \quad (4.94)$$

Шляхом побудови блоків більшої довжини, ефективність кодування може бути істотно підвищена. Повторимо кодування Хаффмана для блоків, довжина яких дорівнює 4 (табл. 4.2). З табл.4.2. випливає, що середня довжина кодового слова на символ дорівнює 0,744. Ефективність кодування вже дорівнює

$$\eta = \frac{H_\infty(x)}{\bar{n}} \approx \frac{0,701}{0,744} \approx 0,94 \quad (4.95)$$

і ми, в основному, реалізували велику частину можливостей кодування, однак, подальшим збільшенням довжини блоку можна збільшити виграш від кодування.

Приклад: Марковське джерело першого порядку.

На рис. 4.9 заданий граф станів марковського джерела першого порядку.

1. Доповнити відсутні ймовірності рис. 4.9 і знайти стаціонарний розподіл ймовірностей станів.

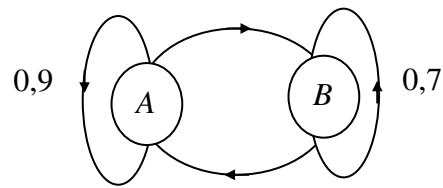


Рисунок 4.9. Граф станів марковського джерела першого порядку

Таблиця 4.1 – Кодування Хаффмана для марковського джерела з пам'яттю $r = 2$ і довжиною блоку 3.

Символи p_i	Кодова конструкція	Кодове слово		
		n_i	$n_i p_i$	
001 9/41		00	2	18/41
100 9/41		10	2	18/41
010 8/41		010	3	24/41
110 4/41		110	3	24/41
111 4/41		111	3	12/41
011 4/41		0110	4	16/41
101 3/41		0111	4	12/41
000 0		—	—	—
		$\frac{\pi}{\text{біт}} = \frac{1}{3} \cdot \frac{112}{41} = 0,911$		

2. Знайти ентропію джерела.

3. Провести кодування Хаффмана для блоків, що складаються з трьох двійкових символів.

4. Яку ефективність має кодування?

Рішення.

1. Перехідні імовірності рис. 4.9 дорівнюють

$$\begin{aligned} \pi(A/A) &= 0,9 \quad \pi(B/A) = 0,1 \\ \pi(A/B) &= 0,3 \quad \pi(B/B) = 0,7. \end{aligned} \tag{4.96}$$

Стаціонарне значення ймовірності стану A визначається, виходячи з наступних рівностей

$$\begin{aligned} P(A) &= 1 - P(B) \\ P(A) &= 0,9 \cdot P(A) + 0,3 \cdot P(B) \end{aligned} \tag{4.97}$$

тому,

$$\begin{aligned} P(A) &= 0,9 \cdot P(A) + 0,3 \cdot (1 - P(A)) \\ P(A) \cdot [1 - 0,9 + 0,3] &= 0,3 \\ P(A) &= 3/4. \end{aligned} \tag{4.98}$$

З цього випливає, що

$$P(B) = 1/4. \tag{4.99}$$

Таблиця 4.2 – Кодування Хаффмана для марковського джерела з пам'яттю $r = 2$ і довжиною блоку 4

Символи p_i		Кодова конструкція	Кодове слово		
			n_i	$n_i p_i$	
1001	9/41		00	2	18/41
0010	6/41		100	3	18/41
0100	6/41		110	3	18/41
0011	3/41		1010	4	12/41
1100	3/41		1110	4	12/41
0101	2/41		0100	4	8/41
1110	2/41		0101	4	8/41
1111	2/41		0110	4	8/41
0110	2/41		0111	4	8/41
0111	2/41		10110	5	10/41
1010	2/41		10111	5	10/41
1101	1/41		11110	5	5/41
1011	1/41		11111	5	5/41
1000	0		–	–	–
0000	0		–	–	–
0001	0		–	–	–

$\frac{\pi}{\text{біт}} = \frac{1}{4} \cdot \frac{122}{41} \approx 0,744$

2. Ентропія джерела визначається як

$$\begin{aligned} \frac{H(A)}{\text{біт}} &= -0,9 \log_2 0,9 - 0,1 \log_2 0,1 = 0,469 \\ \frac{H(B)}{\text{біт}} &= -0,7 \log_2 0,7 - 0,3 \log_2 0,3 = 0,811 \\ H_\infty(X) &= P(A) \cdot H(A) + P(B) \cdot H(B) = 0,572 \text{ біт} \end{aligned} \tag{4.100}$$

3 Імовірності блоків, що складаються з трьох двійкових символів, дорівнюють

Блок	Ймовірність станів	Кодова слово
AAA	0,6075	
BBB	0,1225	100
AAV	0,0675	110
VAA	0,0675	1010
AVV	0,0525	1011
VVA	0,0525	1110
VAV	0,0225	11110
VVV	0,0075	11111

Рисунок 4.10 – Кодування Хаффмана

3 Середня довжина кодового слова визначається як

$$\bar{n} = 1/3(0,6075 + 3(0,1225 + 0,0675) + 4(0,0675 + 0,0525 + 0,0525) + 5(0,0225 + 0,0075)) = 0,6725, \quad (4.101)$$

отже, ефективність кодування дорівнює

$$\eta = \frac{H_{\infty}(x)}{\bar{n}} = \frac{0,572}{0,6725} \approx 0,85. \quad (4.102)$$

4.6. Висновки

У приведених нижче висновках читач може знайти найважливіші твердження і зв'язок марковських ланцюгів і марковських джерел.

Марковські ланцюги.

• Марковським процесом називається стохастичний процес, в якому сьогоднішнє відоме, а майбутнє не залежить від минулого.

• Дискретний за часом і станами марковський процес називається марковським ланцюгом. Його реалізацією є послідовність станів

$$S_t \in S = \{S_1, S_2, \dots, S_N\}.$$

Ланцюг Маркова є гомогенний, якщо перехідні імовірності між станами не залежать від вибору часової точки відліку, отже,

$$\pi(j/t) = P(S_j/S_r) \text{ для } j, r = 0, 1, 2, 3, \dots, N.$$

Гомогенний марковський ланцюг цілком визначається матрицею перехідних ймовірностей Π

$$\Pi = (\pi(j/r))_{N \times N} = \begin{pmatrix} \pi(1/2) & \pi(2/1) & \dots & \pi(N/1) \\ \pi(1/2) & \pi(2/2) & \dots & \pi(N/2) \\ \vdots & \vdots & \ddots & \vdots \\ \pi(1/N) & \pi(2/N) & \dots & \pi(N/N) \end{pmatrix}$$

і початковим розподілом $p_0 = (p_0(S_1), p_0(S_2), \dots, p_0(S_N)) \dots$

Еквівалентом матриці перехідних ймовірностей є граф станів (рис. 4.11) з вузлами, шляхами і вагою шляхів, які відповідають станам, переходам між станами і перехідним ймовірностям.

Розподіл станів гомогенного марковського ланцюга на n -му кроці визначається як

$$p_0 = p_0 \Pi.$$

Гомогенний ланцюг Маркова стаціонарний, якщо розподіл станів постійний за часом. У цьому випадку початковий розподіл є власним вектором перехідної матриці, тобто гомогенний ланцюг Маркова регулярний, якщо існує межа

$$\Pi_{\infty} = \lim_{k \rightarrow \infty} \Pi^k,$$

причому всі рядки Π_{∞} дорівнюють граничному розподілу p_{∞} .

• Граничний розподіл p_{∞} є єдиним стаціонарним розподілом регулярного ланцюга Маркова.

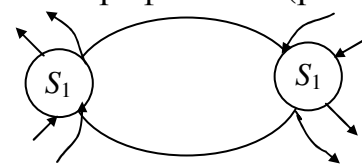


Рисунок 4.11. Граф станів

Стаціонарні Марковські джерела.

Скінчене дискретне марковське джерело з пам'яттю r цілком визначається наступними умовами.

Задано непорожню множину станів $s = (s_1, s_2, \dots, s_N)$, причому S містить усі вектори довжини r . Кожен стан $S_t \in S$ відповідає дискретному джерелу без пам'яті з алфавітом $X_1 = \{x_1, x_2, \dots, x_M\}$ і з ймовірностями j -их символів алфавіту $p^{(1)}(j)$.

Стан $S(n) = (x[n-r], x[n-r+1], \dots, x[n-1])$ з r послідовних символів і черговий символ $x[n]$ утворюють новий стан $s[n+1] = (x[n-r],$

Задано початковий розподіл станів $p_0 = (p_0(1), p_0(2), \dots, p_0(N))$...

Ентропія стаціонарного марковського джерела з пам'яттю r визначається як математичне сподівання умовних ентропій підджерел

$H_\infty(X) = \sum_{r=1}^N p_\infty(r) H(X/S_r)$, при цьому умовна ентропія i -го підджерела, яка відповідає стану i дорівнює

$$H_\infty(X/S_i) = -\sum_{m=1}^M p_{s_i}(x_m) \log_2(p_{s_i}(x_m)), \text{ біт.}$$

РОЗДІЛ 5

СТИСНЕННЯ ДАНИХ

5.1. Вступ

Задачею *стиснення даних* є мінімізація технічних витрат на збереження чи передачу інформації шляхом оптимального кодування джерел. При цьому розрізняють два поняття:

1 *Несуттєва інформація* – це інформація, яку можна ігнорувати під час передачі. Прикладом може служити традиційна телефонія. У телефонних каналах передача інформації здійснюється у смузі 0,3–3,4 кГц. Всі інші спектральні складові відкидаються, при цьому суттєва частина переданої інформації втрачається. Ясно, що первинний мовний сигнал не може бути цілком відновлений на прийомному кінці. У цьому випадку говорять про кодування з втратами.

2 Під *надмірністю* розуміють кількаразове повторення в повідомленні необхідної для споживача інформації. Надмірність може бути усунута без втрат інформації. Прикладом є кодування Хаффмана. Таке кодування називають кодуванням без втрат.

Найважливішими прикладами стиснення даних є цифрове радіомовлення (Digital Audio Broadcasting, DAB) і цифрове телебачення (Digital Video Broadcasting, DVB) [19]. Обидві системи працюють на основі кодування аудіо і відео сигналів, яке використовує стандарти MPEG (Motion Pictures Experts Group). Кодування аудіосигналу засноване на психологічній моделі сприйняття мови, яка використовує приховані спектральні і часові ефекти, при цьому в сигнальному блоці відбувається видалення несприйнятої на слух частини аудіосигналу (несуттєвої інформації). Аналогічні ефекти використовуються також під час кодування відеосигналу (зокрема, психологічний ефект руху). Кодування зображень дозволяє досягти ще більшого ступеня стиснення.

Ступінь стиснення визначається затратами для передачі чи збереження інформації без стиснення k_0 і затратами з використанням деякого методу стиснення k_m

$$G_k = \frac{k_0 - k_m}{k_0}. \quad (5.1)$$

Ступінь стиснення залежить від алгоритму, який використовується і властивостей джерела. Приведемо деякі числові приклади ступенів стиснення, які досягаються на практиці:

– до 80 % для текстових даних (у форматі редактора Word 97 за допомогою програми стиснення ZIP);

– 87,5 % при переході від РСМ–телефонії зі швидкістю 64 Кбіт/с до передачі інформації за рекомендацією ІТУ G.725 зі швидкістю 8 Кбіт/с;

– 90 % при кодуванні інформації стереофонічних аудіо компакт дисків зі швидкістю $216 \times \text{біт} \times 44 \text{ кГц} = 1408 \text{ Кбіт/с}$ методом, який використовує стан-

дарт стиснення MPEG (Advanced Audio Coding) зі швидкістю 112 Кбіт/с і майже рівнозначною якістю мови.

Наступним прикладом є ентропія німецького літературного тексту. Результати частотного аналізу надані на рис. 5.1. Якщо розглядати літери ізольовано, то одержимо ентропію, яка приблизно дорівнює 4,7 біт/літеру. Якщо об'єднати літери в блоки, то ми вже використовуємо такі безумовні зв'язки, як склади, слова і т.д., тому, для блоків дуже великої довжини асимптотично досяжна границя дорівнює 1,6 біт/літеру.

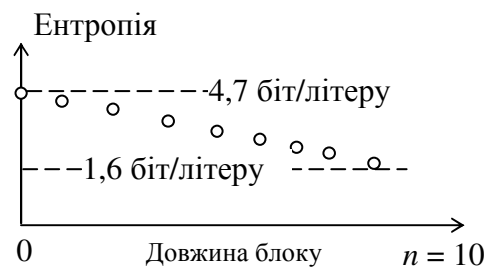


Рисунок 5.1 – Ентропія німецької літературної мови як функція довжини блоку

Алгоритми стиснення даних можна розділити на три групи:

1. *Статичні алгоритми*, наприклад, *кодування Хаффмана*.

Стиснення німецького літературного тексту методом Хаффмана, у порівнянні зі стисненням інформації, яке складається з довільних стандартних знаків ASCII, дозволяє досягти приблизно 50 % вигащу.

2. *Адаптивні алгоритми*, наприклад, модифіковане кодування – Хаффмана. Тут розподіл ймовірностей знаків спочатку приймається рівномірним, а потім змінюється з часом в міру накопичення статистики.

3. *Динамічні алгоритми*, наприклад кодування, що використовується в рекомендації ITU V42. bis.

Основна проблема *ентропійного кодування* полягає в тому, що воно потребує знання розподілу ймовірностей знаків. Дуже часто статистика знаків заздалегідь невідома й ефективному кодуванню повинен передувати частотний аналіз. Тут на допомогу приходять *універсальні алгоритми*.

1 Універсальні алгоритми стиснення, які є за своєю суттю адаптивними, не мають потреби в апріорній статистиці. Таке ефективне кодування починається відразу ж після надходження інформації на вхід кодера.

2 Крім цього, існують «швидкі» алгоритми з відносно простою технічною складністю.

3 Кожний з передбачуваних алгоритмів допомагає досягти високого ступеня стиснення.

Як приклад методів стиснення, розглянемо два найважливіших алгоритми: арифметичне кодування, при якому здійснюється динамічний частотний аналіз і універсальний алгоритм Лемпеля-Зіва. Алгоритм Лемпеля-Зіва LZ77 був запропонований у 1977 р. і модифікований у 1984 р. Він використовується в рекомендації ITU V.42.bis і називається LZW алгоритмом.

5.2. Арифметичне кодування

Під час *арифметичного кодування* ми виходимо з того факту, що при нормалізованому розподілі сума ймовірностей знаків (і відповідних їм відносних частот) джерела завжди дорівнює одиниці. Якщо відносні частоти знаків невідомі передавачу і приймачу, то:

- вони можуть визначатися, наприклад, шляхом поточних статистичних вимірювань переданої інформації у фіксовані моменти часу;
- приймач і передавач разом, виходячи з відносних частот, встановлюють жорсткі правила кодування.

Особливістю арифметичного кодування є те, що для відображення послідовності знаків у потоки натуральних чисел на інтервалі $[0, 1]$ використовуються відносні частоти.

Результатом такого відображення є стиснення знаків (познакове стиснення) відповідно до їхніх ймовірностей. Пояснимо ідею арифметичного кодування за допомогою наступного прикладу.

Розглянемо арифметичне кодування послідовності літер «GELEEESER». Відносні частоти літер у цьому потоці наведені в табл. 5.1.

Таблиця 5.1. Літери і їхні відносні частоти

Літера	E	S	G	L	R
Відносна частота	0,5	0,2	0,1	0,1	0,1

Процедура кодування представлена на рис. 5.2.

Першій літері «G», відповідно до її відносної частоти, відповідає інтервал $[0.7, 0.8]$. Відповідно до алгоритму, кожен ланцюжок літер, що починаються з G, завжди буде відображатися в число, яке належить цьому інтервалу. Таким чином, у розглянутому прикладі перша десяткова цифра після коми вже визначена.

Кодування наступних літер провадиться аналогічно з тією важливою різницею, що тепер розподілу щораз буде піддаватися інтервал, вибраний на попередньому кроці. З рис. 5.2 випливає, що літері «E» на другому кроці відповідає інтервал $[0.7, 0.75]$.

Табл. 5.3, в якій алгоритм кодування простежений по кроках, показує, що послідовність «GELEEESER» відображається в число 740387 (0 і кома не мають потреб у відображенні). Відзначимо, що:

1. Літерам, які часто зустрічаються, ставляться у відповідність великі інтервали. На їхнє відображення витрачається менше десяткових цифр, чим на відображення літер, які рідко зустрічаються.
2. Довгі повідомлення відображаються в «довгі» числа. Подання цих чисел у двійковій формі, необхідне для передачі повідомлень, призводить до появи кодових слів великої довжини.

Практична реалізація ковзного алгоритму арифметичного кодування вимагає великої точності, яка обмежується довжиною кодових слів. Для скорочення необхідної довжини регістра під час реалізації арифметичного кодування вико-

ристовується цілочислова арифметика з видачею вже готових проміжних результатів.

Алгоритм подано в табл. 5.2. Механізм його дії розкривається за допомогою рис. 5.3. У нашому прикладі для реалізації кодування досить регістра для збереження шести поточних десяткових цифр.

Відповідно до табл. 5.2, на першому кроці відбувається ініціалізація змінних *LO* і *HI*. Для першої літери «*G*», що кодується, ширина інтервалу дорівнює $B = 1'000'000$. Нижня і верхня границі дорівнюють відповідно

$LO = 0 + 1'000'000 \cdot 0,7 = 700'000$ і $HI = 0 + 1'000'000 \cdot 0,8 - 1 = 799'999$. Перша десяткова цифра уже визначена і дорівнює 7, а вміст регістрів *LO* і *HI* зсувається на одну позицію вліво. Помітимо, що в регістрі *LO* місце, яке звільнилося, займає 0, а в регістрі *HI* – 9.

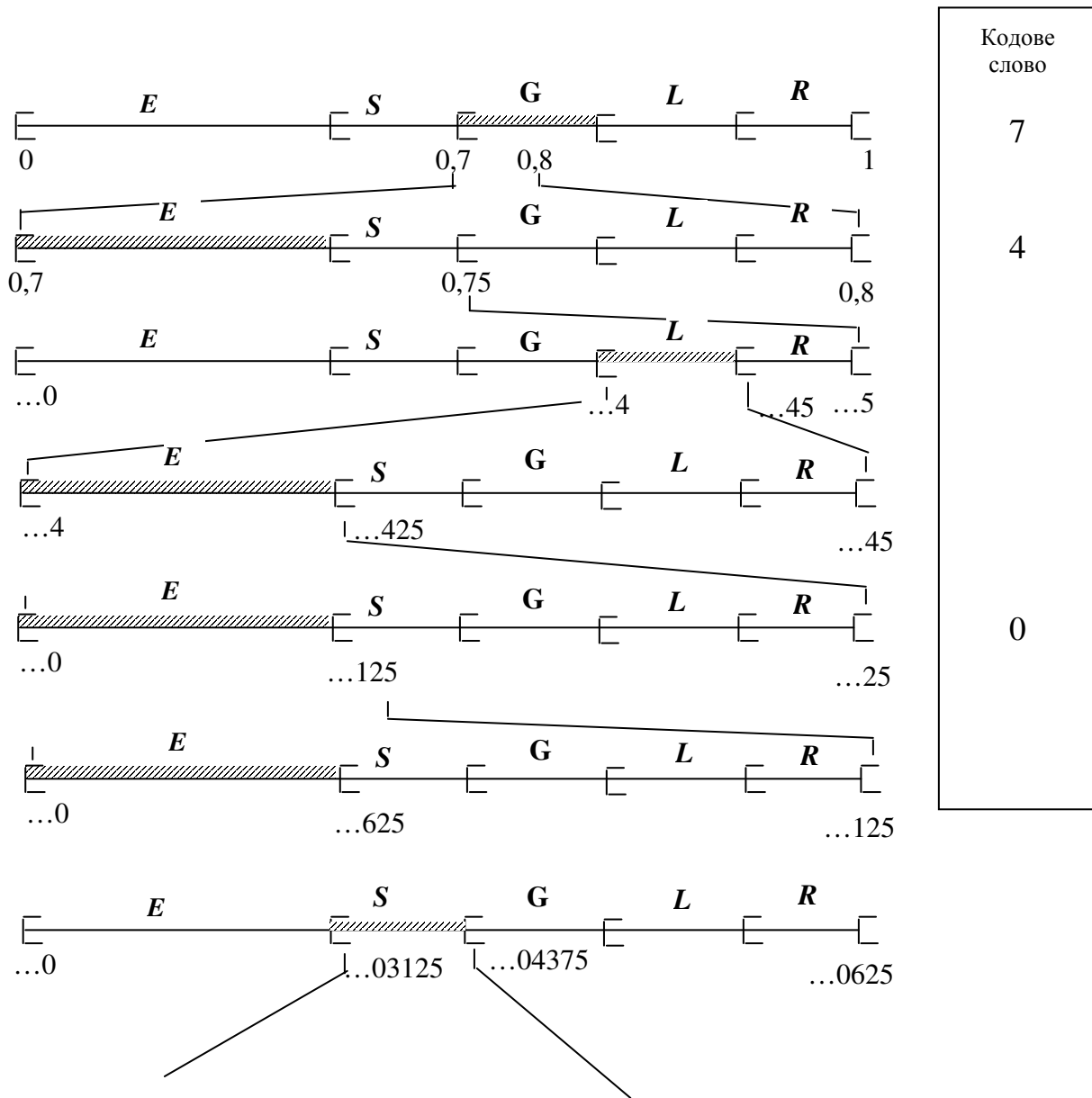


Рисунок 5.2. Арифметичне кодування

Таблиця 5.2 – Арифметичне кодування фрази «GELEEESER».

Знак	LO	HI	Вихід	Знак	LO	HI	Вихід
Старт	000'000	999'999		E	000'000	062'499	
G	700'000	799'999		Вихід/зсув	000'000	624'999	0
Вихід/зсув	000'000	999'999	7	S	312'500	437'499	
E	000'000	499'999		S	375'000	399'999	
L	400'000	449'999		Вихід/зсув	750'000	999'999	3
Вихід/зсув	000'000	499'999	4	E	750'000	874'999	
E	000'000	249'999		R	862'500	874'999	87
E	000'000	124'999		Кінець			

3. Кодування є динамічною процедурою, яка орієнтується на блоки. Сам процес кодування може бути доповнений ковзними вікнами, які містять поточний словник фраз і Look-ahead буфером.

4.

Фраза															Буфер Look-ahead													
F	A	C	H	H	O	C	H	S	C	H	U	L	E		F	U	L	D	A		F	A	C	H	B	E	R	E
21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1								

Рисунок 5.3 – Ковзне вікно алгоритму LZ77, яке відповідає фразі FACH

В процесі кодування оброблюваний текст відображається в послідовність покажчиків чи прапорців. Структура закодованого тексту показана на рис. 5.4. У прикладі, наведеному на рис. 5.3, ланцюжок літер «FACH» замінюється послідовністю [21, 4, B].

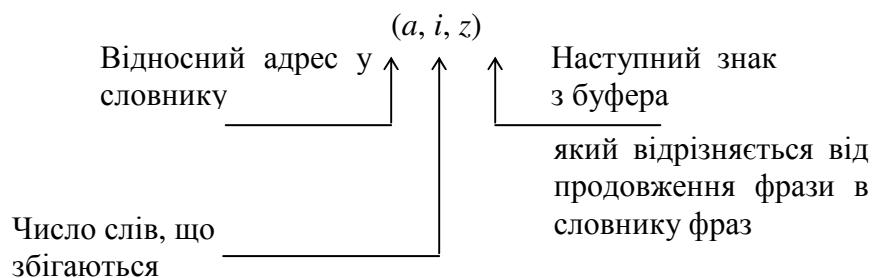


Рисунок 5.4 – Структура покажчиків

На рис. 5.5 представлено декотрий окремий випадок і наведено його алгоритмічне рішення. На шостому кроці кодування черговим є знак «E», який не міститься в словнику фраз. У цьому випадку він кодується за допомогою так званої «нульової фрази». «Нульова фраза» має вигляд [0, 0, символ] і при декодуванні розпізнається за двома нулями, які стоять поруч.

Ще одним цікавим випадком є повторення знаків, так зване Character Rans. У цьому випадку, для заміни знаку використовуються вже два прапорці. Перший «0» служить ознакою появи деякої особливості тексту. Наступні «1» і символ вказують на те, що знак повторюється. У другому прапорці вказується число повторень і наступний знак.

Фраза															Буфер Look-ahead								
I	N	F	O	R	M	F	T	I	O	N		I	S		I	N	F	O	R	M	A	T	ION NOT-MATTER NOR ENERGY
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	1	2	3	4	5	6	7		
															[17,7,T]								

Фраза															Буфер Look-ahead								
I	O	N		I	S		I	N	F	O	R	M	A	T	I	O	N		N	O	T	-MATTER NOR ENERGY	
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	1	2	3	4	5	6	7		
															[15,4,N]								

Фраза															Буфер Look-ahead								
S		I	N	F	O	R	M	A	T	I	O	N		N	O	T		M	A	T	T	E R NOR ENERGY	
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	1	2	3	4	5	6	7		
															[10,1,T]								

Фраза															Буфер Look-ahead								
I	N	F	O	R	M	F	T	I	O	N			N	O	T		M	A	T	T	E	R	NOR ENERGY
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	1	2	3	4	5	6	7		
															[4,1,M]								

Фраза															Буфер Look-ahead								
F	O	R	M	A	T	I	O	N			N	O	T		M	A	T	T	E	R		N	O R ENERGY
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	1	2	3	4	5	6	7		
															[11,2,T]								

Фраза															Буфер Look-ahead									
M	A	T	I	O	N			N	O	T			M	A	T	T	E	R		N	O	R		E NERGY
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	1	2	3	4	5	6	7			
															[0,0,E] Нульова фраза									

Фраза															Буфер Look-ahead								
A	T	I	O	N			N	O	T			M	A	T	T	E	R		N	O	R		E N ENERGY
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	1	2	3	4	5	6	7		
															[0,0,R] Нульова фраза								

Рисунок 5.5 – Приклад кодування за алгоритмом LZ77 з нульовою фразою і повторенням знаків

Витрати на кодування визначаються довжиною вікна, яке містить словник фраз ω_p довжиною Look ahead буфера ω_L і витратами на двійкове подання покажчика

$$K_c/\text{біт} = \log_2 \omega_p + \log_2 \omega_L + 8. \quad (5.2)$$

Кодування Лемпеля–Зіва приводить до стиснення даних у тому випадку, якщо витрати на кодування, тобто довжина покажчика в двійковому численні в

середньому виявляється менше, ніж для безпосереднього кодування, наприклад, кодом ASCII, що відповідає 8 бітам на один знак.

У типовому випадку $\omega_p = 2^{12} = 4096$ і $\omega_L = 2^4 = 16$ і витрати на двійкове подання покажчика складають 24 біта. Для фрази, яка складається з чотирьох знаків, і вже міститься в словнику фраз, економія, у порівнянні з прямим кодуванням кодом ASCII (32 біта), складає 25 %.

Для кодування Лемпеля–Зіва встановлено, що:

- ланцюжки знаків, які часто з'являються, кодуються дуже ефективно;
- знаки і послідовності знаків, які рідко з'являються, з часом видаляються зі словника фраз;
- знаки, які повторюються, також кодуються ефективно;
- на кодування нульових фраз витрачається відносно велике число біт;
- методи теорії інформації дозволяють довести, що кодування методом

Лемпеля–Зіва асимптотично оптимальне. Це означає, що для дуже довгого тексту надмірність зникає, тобто середнє число біт, необхідне для кодування одного знаку, наближається до ентропії тексту;

– практично досяжна ступінь стиснення для довгих текстів складає 50–60%.

РОЗДІЛ 6

ДИСКРЕТНІ КАНАЛИ БЕЗ ПАМ'ЯТІ І ПЕРЕДАВАННЯ ІНФОРМАЦІЇ

6.1. Вступ

У розділі 3 були розглянуті два зв'язаних джерела інформації. Були введені



Рисунок 6.1 – Модель передавання

такі ключові поняття як спільна, взаємна й умовна інформації пар подій (символів) для зв'язаних джерел. На їхній основі ми прийшли до фундаментальних понять інформації – спільної, взаємної й умовної ентропії. (див. табл. 3.3).

Там же було відзначено, що спільна й умовна ентропії мають аналоги в теорії ймовірностей і визначаються як математичні сподівання спільних і умовних ймовірностей усіх пар подій двох джерел.

Ми продовжимо ці міркування, приділивши основну увагу взаємній ентропії. Для опису каналів передавання інформації використовуємо концепцію двох зв'язаних джерел. Виявляється, що за допомогою понять, введених у розділі 3, вдається цілком описати процес передавання інформації каналами без пам'яті. У результаті, ми оцінимо можливість передавання інформації каналом, тобто пропускну здатність каналу.

У шеннонівській моделі каналу зв'язку інформація одного джерела (передавача) передається каналом приймачу і видається споживачу. Для споживача має значення тільки вихід приймача, тобто приймач для нього сам є джерелом інформації, тому, модель зв'язаних джерел може бути застосована до ланцюжка "Передавач – Канал – Приймач" (рис. 6.1). Якщо відбувається передавання інформації, то символи одного джерела повинні впливати на символи іншого джерела. Як приклад розглянемо двійкові симетричні канали без пам'яті.

6.2. Двійковий симетричний канал

Двійковий симетричний канал (ДСК) являє найпростіший приклад взаємодії двох дискретних джерел без пам'яті. Він є дискретною двійковою моделлю передавання інформації по каналу з адитивним білим гауссовим шумом (АБГШ).

Зауваження. При перевірці ефективності алгоритмів завадостійкого кодування, для розрахунків і моделювання каналів зв'язку методом Монте-Карло успішно застосовуються дискретні моделі каналів.

Двійковий симетричний канал описується за допомогою діаграми переходів (рис. 6.2). На діаграмі представлені можливі переходи двійкових символів (0,1) джерела X у двійкові символи джерела Y. Кожному переходу приписана перехідна ймовірність.

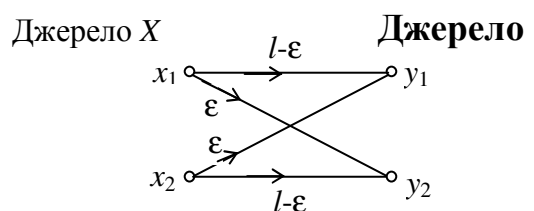


Рисунок 6.2 – Діаграма передавання даних двійковим симетричним каналом

З рис. 6.2 видно, що помилковим переходам відповідає імовірність ϵ , тому, звичайно говорять, що під час передавання двійкової інформації по ДСК, помилка відбувається з імовірністю ϵ . Еквівалентом діаграми переходів є матриця каналу. Вона містить перехідні імовірності і є стохастичною матрицею, у якій сума всіх елементів кожного рядка дорівнює одиниці.

Матриця каналу з вхідним алфавітом, що складається з M символів x_i і вихідним алфавітом, що складається з N символів y_j , містить усі перехідні імовірності $P(y_j | x_i)$ і має вигляд

$$P_{Y|X} = \begin{pmatrix} p(y_1 | x_1) & p(y_2 | x_1) & \dots & p(y_N | x_1) \\ p(y_1 | x_2) & p(y_2 | x_2) & \dots & p(y_N | x_2) \\ \vdots & \vdots & \ddots & \vdots \\ p(y_1 | x_M) & p(y_2 | x_M) & \dots & p(y_N | x_M) \end{pmatrix}. \quad (6.1)$$

У випадку ДСК маємо

$$P_{Y|X}^{\text{ДСК}} = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}. \quad (6.2)$$

Із симетрії переходів випливає, що рівномірний розподіл символів на вході каналу викликає рівномірний розподіл вихідних символів.

Випишемо умовні і взаємні інформації всіх можливих пар подій, припускаючи рівномірний розподіл вхідних символів. Для ДСК маємо

$$\begin{aligned} I(y_1 | x_1) &= -\log_2 p(y_1 | x_1) \text{ біт} = -\log_2 (1-\epsilon) \text{ біт}; \\ I(y_2 | x_1) &= -\log_2 p(y_2 | x_1) \text{ біт} = -\log_2 \epsilon \text{ біт}; \\ I(y_1 | x_2) &= -\log_2 p(y_1 | x_2) \text{ біт} = -\log_2 \epsilon \text{ біт}; \\ I(y_2 | x_2) &= -\log_2 p(y_2 | x_2) \text{ біт} = -\log_2 (1-\epsilon) \text{ біт}. \end{aligned}$$

Звідси випливає

$$\begin{aligned} I(x_1; y_1) &= \log_2 \frac{p(y_1 | x_1)}{p(y_1)} \text{ біт} = \log_2 \frac{(1-\epsilon)}{1/2} = (1 + \log_2 (1-\epsilon)) \text{ біт}; \\ I(x_1; y_2) &= \log_2 \frac{p(y_2 | x_1)}{p(y_2)} \text{ біт} = \log_2 \frac{\epsilon}{1/2} = (1 + \log_2 \epsilon) \text{ біт}; \\ I(x_2; y_1) &= \log_2 \frac{p(y_1 | x_2)}{p(y_1)} \text{ біт} = \log_2 \frac{\epsilon}{1/2} = (1 + \log_2 \epsilon) \text{ біт}; \\ I(x_2; y_2) &= \log_2 \frac{p(y_2 | x_2)}{p(y_2)} \text{ біт} = \log_2 \frac{(1-\epsilon)}{1/2} = (1 + \log_2 (1-\epsilon)) \text{ біт}. \end{aligned}$$

Розглянемо три особливих випадки

1. $\epsilon = 0$ (передача без помилок)

$$I(x_1; y_1) = I(x_2; y_2) = 1 \text{ біт}.$$

Іншої взаємної інформації не існує, тому що пари взаємних символів (x_1, y_2) і (x_2, y_1) ніколи не можуть з'явитись. Інформація передається від джерела X до джерела Y без втрат.

2. $\epsilon = 1/2$. Для всіх пар символів (x_i, y_j) маємо

$$I(x_i, y_j) = \log_2 \frac{1/2}{1/2} \text{ біт} = 0.$$

Джерела X і Y незалежні. Передачі інформації не відбувається.

3. $\epsilon = 1$. У цьому випадку або якісь імовірності переплутані, чи ми десь цілком помиляємося. Виявивши цей факт і проінвертувавши прийняті символи y_i , ми прийдемо до першого випадку.

На закінчення розглянемо поведження умовної $I(y_i | x_i) = I(y_1 | x_1)$ і взаємної $I(y_i; x_i) = I(y_1; x_1)$ інформації в ДСК як функції ймовірності помилки ϵ .

Умовну інформацію $I(y_i | x_i)$ можна розглядати як невизначеність, внесена

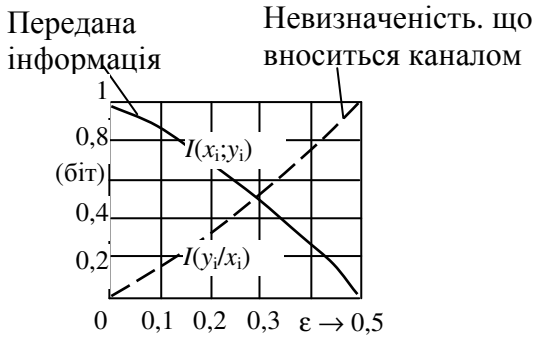


Рисунок 6.3 – $I(y_i | x_i)$ – умовна і $I(y_i; x_i)$ – взаємна інформації як функції від ϵ

каналом, а взаємну інформацію $I(y_i; x_i)$ – як інформацію, передану каналом. Під час передавання одного двійкового символу $\epsilon = 0$ інформація передається без втрат, тому $I(y_i | x_i) = 0$, а $I(y_i; x_i) = 1$ біт. Зі зростанням імовірності помилки ϵ , невизначеність, внесена каналом, зростає, а передана інформація, навпаки, зменшується. При $\epsilon = 1/2$ передавання інформації відсутнє, тому $I(y_i | x_i) = 1$ біт, а $I(y_i; x_i) = 0$. Сума ж умовної $I(y_i | x_i)$ і взаємної $I(y_i; x_i)$ інформації не залежить від ϵ і завжди дорівнює одному біту.

6.3. Передавання інформації

Після розгляду окремих пар подій у попередньому розділі, повернемося знову до моделі передавання інформації. На рис. 6.4 показана вихідна ситуація.

Опис каналу за допомогою перехідних ймовірностей зводиться до спільних ймовірностей пар подій. З цього погляду, обидва джерела в моделі передачі інформації рівнозначні, тому підрозділяти джерела на передавач і приймач, маючи на увазі напрямок передачі інформації, тут і надалі не завжди має сенс.

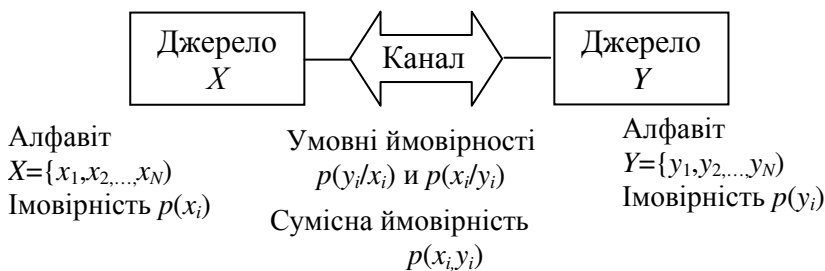


Рисунок 6.4 – Два дискретних джерела без пам'яті, зв'язані каналом

У розділі 3 (див. табл. 3.3) спільна ентропія двох джерел визначена як математичне сподівання інформації всіх можливих пар подій

$$\frac{H(X, Y)}{\text{біт}} = - \sum_x \sum_y p(x, y) \log_2 p(x, y).$$

Точно так само визначається умовна ентропія

$$\begin{aligned} \frac{H(Y|X)}{\text{біт}} &= -\sum_x \sum_y p(x,y) \log_2 p(y|x), \\ \frac{H(X|Y)}{\text{біт}} &= -\sum_x \sum_y p(x,y) \log_2 p(x|y) \end{aligned} \quad (6.3)$$

З цього випливає

$$H(X,Y) = H(Y) + H(X|Y) + H(X) + H(Y|X) \quad (6.4)$$

і

$$H(X,Y) \leq H(Y) + H(X) \quad (6.5)$$

причому, знак рівності має місце тільки для незалежних джерел.

У випадку двох джерел, зв'язаних каналом, спільна невизначеність знижується, тому що подія одного джерела дозволяє заздалегідь припускати подію іншого джерела. З точки зору теорії інформації, зниження невизначеності означає обмін інформацією між джерелами. Міркуючи аналогічно, дійдемо висновку, що середнє значення інформації, переданої каналом, визначається як математичне сподівання взаємних інформацій усіх пар подій.

Середнє значення інформації, яким обмінюються два дискретних джерела без пам'яті X і Y , дорівнює

$$\frac{I(X;Y)}{\text{біт}} = \sum_x \sum_y p(x,y) \log_2 \frac{p(y|x)}{p(y)} = \sum_x \sum_y p(x,y) \log_2 \frac{p(x|y)}{p(x)} \quad (6.6)$$

З визначення переданої інформації випливає

$$\frac{I(X;Y)}{\text{біт}} = \underbrace{\sum_x \sum_y p(x,y) \log_2 p(x|y)}_{-H(X|Y) \text{ біт}} - \underbrace{\sum_x \log_2 p(x) \sum_y p(x,y)}_{H(X) \text{ біт}}, \quad (6.7)$$

і, тому

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \quad (6.8)$$

У цьому місці знову виникає питання про сутність аксіоматичного визначення ентропії. У якості «спробного каменю» доведемо справедливості наступного твердження.

Теорема 6.3.1. Передана інформація $I(X;Y)$ завжди додатна, причому, вона дорівнює нулю тільки для незалежних джерел X і Y

$$I(X;Y) \geq 0. \quad (6.9)$$

Доказ.

Під час доказу будемо виходити з визначення $I(X;Y)$ і використаємо три прийоми. По-перше, скористаємося оцінкою функції натурального логарифма (1.19). По-друге, без обмеження спільності будемо розглядати тільки такі пари символів, імовірність яких відмінна від нуля. У третій, в аргументі логарифмічної функції з (1.19) поміняємо місцями чисельник і знаменник, що еквівалентно множенню логарифмічної функції на мінус 1, тому, нам досить довести справедливості нерівності

$$\frac{-I(X;Y)}{\text{біт}} = \sum_x \sum_y p(x,y) \ln \frac{p(x)}{p(x|y)} \leq 0.$$

Так як, у силу зробленого нами обмеження, суми беруться тільки по парах (x, y) , для яких $p(x, y) \neq 0$, аргумент логарифмічної функції $\frac{p(x)}{p(x|y)}$ завжди має відмінне від нуля скінчене додатне значення, тому використовуємо оцінку (1.19)

$$\frac{-I(X;Y)}{\text{біт}} = \sum_X \sum_Y p(x, y) \left[\frac{p(x)}{p(x|y)} - 1 \right] = \sum_X \sum_Y \underbrace{\frac{p(x, y)p(x)}{p(x|y)}}_{p(x)p(y)} - \underbrace{\sum_X \sum_Y p(x, y)}_{=1} = \sum_X \sum_Y p(x)p(y) - 1 \leq 0.$$

Якби нерівність (6.9) не виконувалася, передавання інформації не знижувало б ентропію (тобто визначеність джерела не підвищувалася б).

Те, що передана інформація завжди додатна і завжди справедливі рівності (6.8) і (6.4), зайвий раз підтверджує справедливості наступних тверджень:

Будь-яке обмеження не може підвищувати невизначеність джерела

$$H(X) \geq H(X|Y) \quad (6.10)$$

Спільна ентропія досягає свого максимуму, коли джерела незалежні

$$H(X, Y) \leq H(X) + H(Y) \quad (6.11)$$

Знайдені залежності можна наочно пояснити за допомогою діаграми потоків інформації (рис. 6.5). Діаграма допомагає усвідомити зміст умовних ентропій $H(X|Y)$ і $H(Y|X)$.

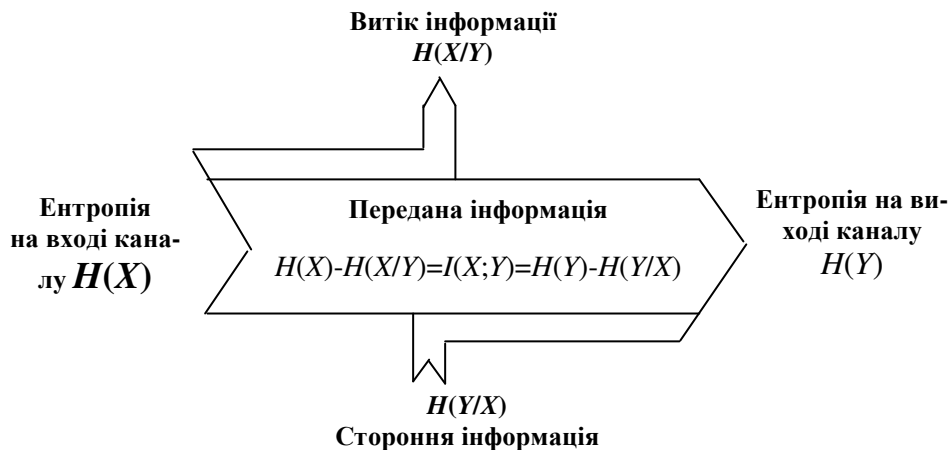


Рисунок 6.5 – Діаграма інформаційних потоків

$H(X|Y)$ – визначає середню міру невизначеності посланого символу джерела X у тому випадку, коли символи приймача джерела Y відомі, тобто невизначеність приймача, яка залишилася.

Величину $H(X|Y)$ часто називають також «витоком» інформації, тому що ентропію $H(X)$ можна інтерпретувати як власну інформацію джерела X , $H(X) = I(X; X)$.

У безшумному каналі інформація передається від джерела X до джерела Y і назад без втрат (без «витоку»). Якщо канал цілком зашумлений, то $H(X|Y) = H(X)$ і ніякої передачі інформації не відбувається («витікає» вся інформація).

$H(Y|X)$ – визначає середню невизначеність прийнятого символу при відомих посланих символах, тому її називають «сторонньою» шумовою інформацією.

Передачу інформації зашумленим каналом можна розглядати як серію випадкових експериментів, які сприяють зниженню невизначеності. З точки зору теорії інформації, канал є джерелом шумів.

Приклад: Передача інформації двійковим симетричним каналом.

Пояснимо фізичний зміст величини $I(x; y)$ на прикладі ДСК (рис. 6.2). Для двійкового симетричного каналу маємо

$$\frac{I(X;Y)}{\text{біт}} = \sum_{i=1}^2 \sum_{j=1}^2 p(x_i, y_j) \log_2 \frac{p(y_j | x_i)}{p(y_j)}$$

Як бачимо, $I(X;Y)$ залежить тільки від двох параметрів – імовірності помилки в каналі й імовірності появи символу x_1 на виході каналу $p(x_1)$. При цьому виконуються наступні співвідношення

$$\begin{aligned} p(x_1) &= p, \quad p(x_2) = 1 - p; \\ P_{Y|X} &= \begin{pmatrix} p(y_1 | x_1) & p(y_2 | x_1) \\ p(y_1 | x_2) & p(y_2 | x_2) \end{pmatrix} = \begin{pmatrix} 1 - \varepsilon & \varepsilon \\ \varepsilon & 1 - \varepsilon \end{pmatrix}; \\ p(y_1) &= (1 - \varepsilon) \cdot p + \varepsilon \cdot (1 - p), \quad p(y_2) = 1 - p(y_1); \\ p(x_i, y_j) &= p(y_j | x_i) p(x_i). \end{aligned}$$

Результати обчислень для $I(X;Y)$ при різних p, ε надані на рис. 6.6 як сімейство кривих $I(X;Y) = f(p)$ при $\varepsilon = \{0,05; 0,1; 0,2; 0,4; 0,5\}$. У каналі без шуму $\varepsilon = 0$ передача інформації відбувається без спотворень і інформація $I(X;Y)$, у цьому випадку, дорівнює ентропії $H(X)$ на вході каналу. Зі збільшенням рівня шуму, імовірність помилки ε підвищується, а кількість переданої інформації знижується, причому, відносно малий рівень шуму $\varepsilon = 0,05$ приводить до помітного зниження $I(X;Y)$. У цілком зашумленому каналі $\varepsilon = 0,5$ передача інформації неможлива.

Цікаво відзначити, що при фіксованих значеннях ε інформація $I(X;Y)$ істотно залежить від імовірності p на вході каналу. При $p = 1/2$ через канал передається максимальна кількість інформації. У розділі 6.5, в якому буде введено нове поняття – пропускна здатність каналу, цю властивість $I(X;Y)$ буде розглянуто докладно.

Приклад: Зв'язані джерела

Для додаткового пояснення змісту ентропії на числовому прикладі пропонується така конструкція зв'язаних джерел, в якій всі величини, які нас цікавлять, можуть бути досить просто підраховані.

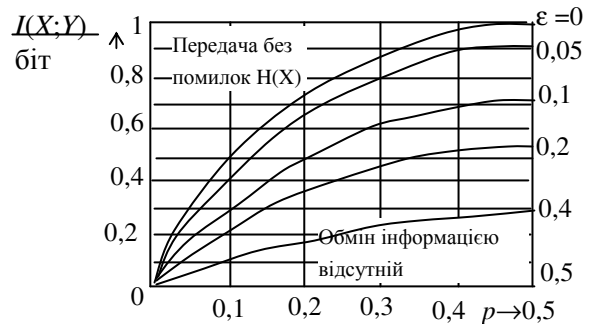


Рисунок 6.6 – Передавання інформації двійковим симетричним каналом з імовірністю помилки ε для різних значень імовірності p символу на вході каналу

У табл. 6.1 задане дискретне джерело без пам'яті Z із символами з алфавіту $\{0,1,2,3\}$ і відповідними ймовірностями символів. Кожен символ z_i кодується двійковим кодом з першим бітом x_i і другим – y_i . Ми будемо інтерпретувати ці біти як символи двох зв'язаних джерел X і Y .

Таблиця 6.1. Джерело Z і його двійкове кодування

i	z_i	$p_Z(z_i)$	x	y
1	0	1/2	0	0
2	1	1/4	0	1
3	2	1/8	1	0
4	3	1/8	1	1

Завдання для самостійної роботи

1. Описати джерела X і Y .
2. Встановити зв'язок між джерелами X і Y у формі моделі каналу, в якій джерело X є входом каналу, а джерело Y його виходом.
3. Привести для завдання 2 діаграму інформаційних потоків і знайти для цієї діаграми числові значення ентропій.
4. Знайти ентропію джерела Z .
5. Виконати завдання 2 і 3, якщо вважати джерело Y входом каналу.

Рішення.

1. Почнемо з опису джерел X і Y . Обидва джерела є дискретними джерелами без пам'яті. Використовуючи табл. 6.1, знайдемо розподіл ймовірностей символів 0 і 1 для кожного з них

$$\begin{aligned} p_X(0) &= p_Z(0) + p_Z(1) = 3/4; \\ p_X(1) &= p_Z(2) + p_Z(3) = 1/4; \\ p_Y(0) &= p_Z(0) + p_Z(2) = 5/8; \\ p_Y(1) &= p_Z(1) + p_Z(3) = 3/8. \end{aligned} \tag{6.12}$$

Згідно з формулою (1.28), ентропії джерел дорівнюють

$$\begin{aligned} \frac{H(X)}{\text{біт}} &= -\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \approx 0,8113, \\ \frac{H(Y)}{\text{біт}} &= -\frac{5}{8} \log_2 \left(\frac{5}{8} \right) - \frac{3}{8} \log_2 \left(\frac{3}{8} \right) \approx 0,9544. \end{aligned}$$

2. Модель каналу являє собою двійковий канал із символами x_1 і x_2 на вході і символами y_1 і y_2 на виході. Канал може бути заданий матрицею перехідних ймовірностей (6.1), що містить імовірності $p(y_j | x_i)$ як елементи матриці. З (6.12) і табл. 6.1 випливає, що

$$\begin{aligned}
 P_{Y|X}(0|0) &= \frac{P_{X,Y}(0,0)}{P_X(0)} = \frac{P_Z(0)}{P_X(0)} = \frac{1/2}{3/4} = \frac{2}{3}; \\
 P_{Y|X}(0|1) &= \frac{P_{X,Y}(1,0)}{P_X(1)} = \frac{P_Z(2)}{P_X(1)} = \frac{1/8}{1/4} = \frac{1}{2}; \\
 P_{Y|X}(1|0) &= \frac{P_{X,Y}(0,1)}{P_X(0)} = \frac{P_Z(1)}{P_X(0)} = \frac{1/4}{3/4} = \frac{1}{3}; \\
 P_{Y|X}(1|1) &= \frac{P_{X,Y}(1,1)}{P_X(1)} = \frac{P_Z(3)}{P_X(1)} = \frac{1/8}{1/4} = \frac{1}{2}.
 \end{aligned}$$

У результаті одержимо матрицю перехідних ймовірностей каналу

$$P_{Y|X} = \begin{pmatrix} 2/3 & 1/3 \\ 1/2 & 1/2 \end{pmatrix}.$$

Зауваження. Як і слід було очікувати, матриця є стохастичною, тому що сума ймовірностей у кожному її рядку дорівнює одиниці.

Діаграма каналу з ймовірностями переходів наведена на рис. 6.7. Можна помітити її подібність з діаграмою (рис. 6.2). Однак, у нашому прикладі, вже не можна говорити про помилки в каналі.

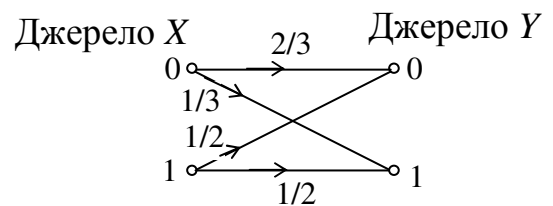


Рисунок 6.7 – Двійковий канал

3. Для побудови діаграми інформаційних потоків необхідне знання величин $H(Y|X)$, $H(X|Y)$, $I(X;Y)$. За відомими перехідними ймовірностями можна обчислити $H(Y|X)$

$$\begin{aligned}
 \frac{H(Y|X)}{\text{біт}} &= -p_Z(0) \log_2 p_{Y|X}(0|0) - p_Z(1) \log_2 p_{Y|X}(1|0) - p_Z(2) \log_2 p_{Y|X}(0|1) - \\
 &\quad - p_Z(3) \log_2 p_{Y|X}(1|1).
 \end{aligned}$$

Підставляючи числові значення, знаходимо

$$\frac{H(Y|X)}{\text{біт}} = -\frac{1}{2} \log_2 \left(\frac{2}{3} \right) - \frac{1}{4} \log_2 \left(\frac{1}{3} \right) - \frac{1}{8} \log_2 \left(\frac{1}{2} \right) - \frac{1}{8} \log_2 \left(\frac{1}{2} \right) \approx 0,9387.$$

Величини $I(X;Y)$, $H(X|Y)$ можна знайти з (6.8).

Діаграма інформаційних потоків надана на рис. 6.8.



Рисунок 6.8 – Діаграма інформаційних потоків зв'язаних джерел

4. Ентропія джерела Z дорівнює

$$\frac{H(Z)}{\text{біт}} = -\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) - \frac{2}{8} \log_2 \left(\frac{1}{8} \right) = 1,75.$$

Отриманим результатам можна дати наступну інтерпретацію. Ентропія джерела Z дорівнює спільній ентропії двійкових джерел X і Y . Ентропія джерела $H(X)$ дорівнює 0,8113, інші 0,9387 біт вносить умовна ентропія $H(Y|X)$ згідно з виразом (6.4).

5. За аналогією із завданням 2, знаходимо

$$\begin{aligned} P_{X|Y}(0|0) &= \frac{P_{X,Y}(0,0)}{P_Y(0)} = \frac{P_Z(0)}{P_Y(0)} = \frac{1/2}{5/8} = \frac{4}{5}; \\ P_{X|Y}(0|1) &= \frac{P_{X,Y}(0,1)}{P_Y(1)} = \frac{P_Z(1)}{P_Y(1)} = \frac{1/4}{3/8} = \frac{2}{3}; \\ P_{X|Y}(1|0) &= \frac{P_{X,Y}(1,0)}{P_Y(0)} = \frac{P_Z(2)}{P_Y(0)} = \frac{1/8}{5/8} = \frac{1}{5}; \\ P_{X|Y}(1|1) &= \frac{P_{X,Y}(1,1)}{P_Y(1)} = \frac{P_Z(3)}{P_Y(1)} = \frac{1/8}{3/8} = \frac{1}{3}. \end{aligned}$$

і одержуємо матрицю каналу

$$P_{Y|X} = \begin{pmatrix} 4/5 & 1/5 \\ 2/3 & 1/3 \end{pmatrix}.$$

Діаграма каналу і діаграма інформаційних потоків надані на рис. 6.9 і рис. 6.10.

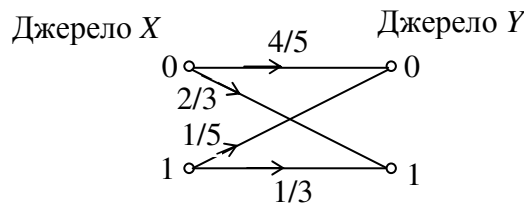


Рисунок 6.9 - Двійковий канал

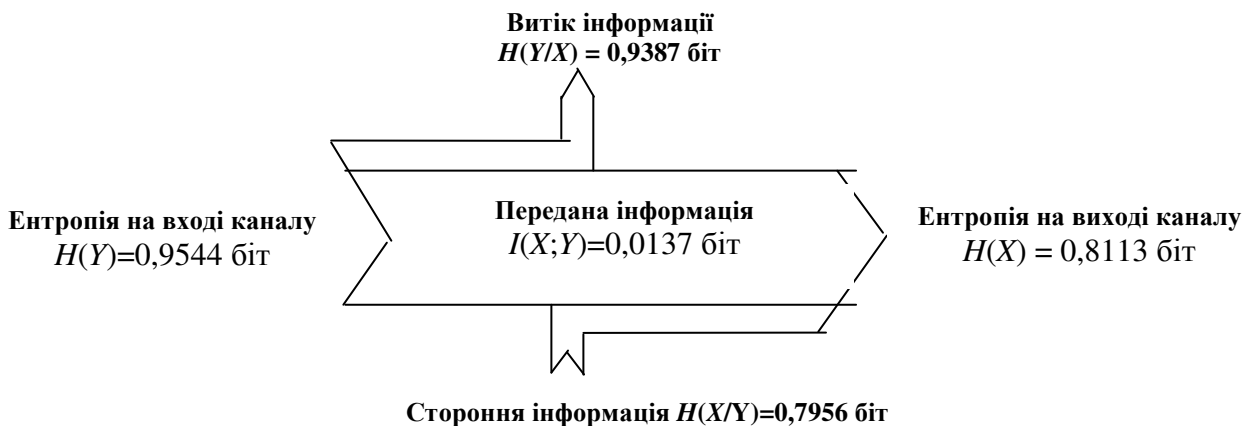


Рисунок 6.10 – Діаграма інформаційних потоків

6.4. Висновки

Всі визначення і величини, розглянуті в попередніх розділах, зібрані в табл. 6.2 і 6.3. Варто звернути особливу увагу на переходи від теорії ймовірностей до теорії інформації.

При цих переходах вихідні символи дискретних джерел без пам'яті розглядаються як вихідні події випадкових експериментів. Відповідно до ймовірностей цих подій, кожному символу приписується деяка інформаційна міра, яка дорівнює логарифму імовірності його появи. Помітимо, що ймовірності символів можна розглядати як стохастичні змінні.

При переході від ймовірностей символів до їхнього інформаційного змісту, вводиться нова величина, що не має аналога в теорії ймовірностей – взаємна інформація, яка виникає при аналізі пар спільних подій (x, y) . Взаємна інформація визначається як логарифм відношення апостеріорної імовірності символу y – $p(y|x)$ до його апріорної імовірності $p(y)$ і служить інформаційною мірою зв'язності двох подій.

Для опису джерел використовуються середні значення інформації символів, які генеруються джерелом. Таким чином, вводиться поняття ентропії як математичного сподівання кількості інформації окремих подій (символів) чи пари подій. При цьому, особливу роль грає взаємна інформація. Її математичне сподівання $I(X;Y)$ є мірою переданої інформації і характеризує зв'язок між двома джерелами X і Y , тобто описує середню кількість інформації, якою обмінюються між собою джерела каналом зв'язку. Основне значення величини $I(X;Y)$ буде докладно розкрито в наступних розділах.

Таблиця 6.2 – Дискретні джерела без пам'яті X і Y із символами

$x \in X = \{x_1, x_2, \dots, x_m\}$ і $y \in Y = \{y_1, y_2, \dots, y_N\}$

Теорія ймовірностей	Теорія інформації
Імовірність окремого символу (апріорна імовірність) $p(x)$	Інформація окремого символу $I(x) = -\log_2 p(x)$, біт
Спільна імовірність двох символів $p(x,y)$	Інформація пари символів $I(x,y) = -\log_2 p(x,y)$, біт
Умовна імовірність (апостеріорна імовірність) $p(x y) = \frac{p(x,y)}{p(y)}$; $p(y x) = \frac{p(x,y)}{p(x)}$	Умовна інформація $I(x y) = -\log_2 p(x y)$, біт $I(y x) = -\log_2 p(y x)$, біт
Немає аналогів	Взаємна інформація $I(x;y) = \log_2 \frac{\text{апостеріор. ймовірн.}}{\text{апріорная ймовірн.}} =$ $= \log_2 \frac{p(x y)}{p(x)} = \log_2 \frac{p(y x)}{p(y)}$, біт Завжди додатна.

Продовження табл. 6.2

Теорія ймовірностей	Теорія інформації
Позначення: $P(x) = P_x(x_i)$ x_i для $x_i \in X$ $P(x y) = P_{x y}(x_i y_j)$ для $x_i \in X, y_j \in Y$ $\sum_x p(x) = \sum_{i=1}^M P_x(x_i)$ $\sum_y p(y) = \sum_{j=1}^N P_y(y_j)$	Деякі важливі співвідношення, що використовуються для розрахунків $\sum_x p(x) = 1; \sum_y p(y) = 1;$ $\sum_x \sum_y p(x, y) = 1;$ $\sum_y p(x, y) = p(x); \sum_x p(x, y) = p(y);$ $\sum_y p(x y) = 1; \sum_x p(x y) = 1;$

Таблиця 6.3 – Опис “у середньому” дискретних джерел без пам'яті із символами $x \in X = \{x_1, x_2, \dots, x_m\}$ і $y \in Y = \{y_1, y_2, \dots, y_n\}$

Ентропія	$H(X) = - \sum_x p(x) \log_2 p(x), \text{ біт}$ $H(Y) = - \sum_y p(y) \log_2 p(y), \text{ біт}$
Спільна ентропія	$H(X, Y) = - \sum_x \sum_y p(x, y) \log_2 p(x, y), \text{ біт}$
Умовна ентропія	$H(X Y) = - \sum_x \sum_y p(x, y) \log_2 p(x y), \text{ біт}$ $H(Y X) = - \sum_x \sum_y p(x, y) \log_2 p(y x), \text{ біт}$
Передавання інформації	$I(X; Y) = \sum_x \sum_y p(x, y) \log_2 \frac{\text{апостеріор. ймовірн.}}{\text{апріорна ймовірн.}} =$ $= \sum_x \sum_y p(x, y) \log_2 \frac{p(y x)}{p(y)}, \text{ біт} =$ $= \sum_x \sum_y p(x, y) \log_2 \frac{p(x y)}{p(x)}, \text{ біт} \geq 0.$
Найважливіші залежності (знак рівності справедливий для незалежних джерел)	$H(X) \geq H(X Y) \text{ і } H(Y) \geq H(Y X)$ $H(X, Y) = H(X) + H(Y X) = H(Y) + H(X Y)$ $H(X, Y) = H(X) + H(Y) - I(X; Y);$ $I(X; Y) \geq 0.$ $I(X; Y) = H(X) + H(Y) - H(X, Y).$

6.5. Пропускна здатність каналу

Величина $I(X; Y)$ відіграє особливу роль у теорії інформації й описує передачу інформації каналом зв'язку. З визначення (6.8) випливає, що $I(X; Y)$ залежить як від перехідних ймовірностей каналу, так і від розподілу ймовірностей символів на вході каналу. Для подальших міркувань розглянемо дискретний канал без пам'яті з фіксованими перехідними ймовірностями і задамося питанням: Яку максимальну кількість інформації можна передати даним каналом?

Пропускна здатність каналу із заданими перехідними ймовірностями дорівнює максимуму переданої інформації з усіх вхідних розподілів символів джерела X

$$C = \max_x I(X; Y). \quad (6.13)$$

Зауваження. *Розмірність пропускної здатності – біт/символ. Якщо, наприклад, по каналі передається один символ за секунду, то можна також говорити про розмірність біт/с.*

Так як максимум шукається за всіма допустимими вхідними джерелами, то пропускна здатність залежить тільки від перехідних ймовірностей каналу.

З математичної точки зору, пошук пропускної здатності дискретного каналу без пам'яті зводиться до пошуку розподілу ймовірностей вхідних символів джерела, яке забезпечує максимум інформації $I(X; Y)$. При цьому, на ймовірності вхідних символів $x \in X$ накладаються обмеження

$$0 < p(x) \leq 1; \quad \sum_x p(x) = 1. \quad (6.14)$$

У принципі, визначення максимуму $I(x, y)$ при обмеженнях (6.14) можливе при використанні мультиплікативного *методу Лагранжа*. Однак, таке рішення вимагає надмірно великих затрат. В окремому випадку (симетричні канали) знайти пропускну здатність допомагає наступна теорема [10].

Теорема 6.5.1. У симетричних дискретних каналах без пам'яті пропускна здатність досягається при рівномірному розподілі ймовірностей вхідних символів джерела X .

Зауваження. У [10] приводиться також метод, який дозволяє визначити, є канал симетричним чи ні.

6.5.1. Пропускна здатність

Двійковий дискретний симетричний канал без пам'яті визначається за допомогою матриці перехідних ймовірностей каналу (6.2). Єдиним параметром, який характеризує ДСК, є ймовірність помилки ϵ . З рівномірного розподілу вхідних символів і симетрії переходів каналу впливає рівномірний розподіл вихідних символів, тобто

$$p(x_1) = p(x_2) = p(y_1) = p(y_2) = 1/2. \quad (6.15)$$

Використовуючи (6.6), одержуємо

$$\frac{C}{\text{біт}} = \sum_i \sum_j p(x_i, y_j) \log_2 \frac{p(y_j | x_i)}{p(y_j)} = \sum_i \sum_j p(x_i) p(y_j | x_i) \log_2 \frac{p(y_j | x_i)}{p(y_j)}.$$

Підставляючи числові значення, маємо

$$\frac{C}{\text{біт}} = (1 - \epsilon) \log_2 (2(1 - \epsilon)) + \epsilon \log_2 (2\epsilon) = 1 + \underbrace{(1 - \epsilon) \log_2 (1 - \epsilon) + 2 \log_2 \epsilon}_{-H_b(\epsilon)}.$$

Ентропія ДСК визначається через (1.28)

$$\frac{H_b(\epsilon)}{\text{біт}} = -(1 - \epsilon) \log_2 (1 - \epsilon) - \epsilon \log_2 (\epsilon).$$

Остаточно одержуємо пропускну здатність ДСК у компактній формі

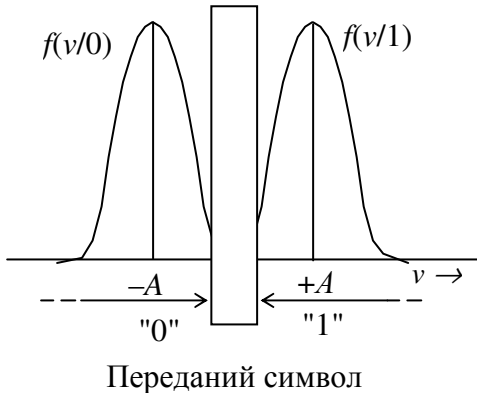
$$C^{\text{ДСК}} = 1_{\text{біт}} - H_b(\epsilon). \quad (6.16)$$

Цікавими є два граничних випадки:

1. $H_b(\epsilon = 0) = 0$; $C^{ДСК} = 1$ біт. Передача інформації безшумним каналом.
2. $H_b(\epsilon = 1/2) = 1$ біт; $C^{ДСК} = 0$ біт. Канал цілком зашумлений.

6.5.2. Пропускна здатність двійкового симетричного каналу зі стираннями

Важливим окремим випадком ДСК є двійковий симетричний канал зі стираннями (ДСКС) чи двійковий канал зі стираннями (Binary Erasure Channel, BEC – англ.).



Як і ДСК, двійковий канал зі стираннями може бути спрощеною моделлю передавання інформації каналом з адитивним білим гауссовим шумом. Правило прийняття рішення в ДСКС приведено на рис. 6.11. З рисунка видно, що поряд з рішеннями про переданий символ «0» чи «1», тут іноді приймається рішення про стирання прийнятого символу «e» (Erasure – англ.). Стирання відбувається у випадку, якщо протектований аналоговий сигнал V попадає в зону, для якої значення умовних функцій густини розподілу ймовірностей

Рисунок 6.11 – Умовні функції густини розподілу ймовірностей протектованого сигналу й області прийняття рішення

$f(V|0)$ і $f(V|1)$ виявляються близькими до нуля.

Зауваження. У двійковому каналі зі стираннями, замість однозначно «жорсткого» рішення про прийнятий символ «0» чи «1» приймається, так зване, «м'яке» рішення. У цьому випадку, ми додатково маємо деяку інформацію про надійність прийнятого двійкового символу. У зв'язку з цим, у техніці передачі даних, говорять про прийом з «жорстким» і «м'яким» рішенням. «М'яке» рішення в сполученні з відповідним кодуванням інформації дозволяє в деяких випадках здійснити більш надійну передачу даних. Один із прикладів використання «м'якого» рішення можна знайти в другій частині цього посібника.

Позначимо ймовірність стирання через q , а ймовірність помилки нестертого символу через p .

Діаграма переходів для каналу з двома вхідними і трьома вихідними символами наведена на рис. 6.12. Відповідна матриця каналу, яка містить перехідні ймовірності, має вигляд

$$P_{YX}^{ДСКС} = \begin{pmatrix} 1-p-q & q & p \\ p & q & 1-p-q \end{pmatrix} \quad (6.17)$$

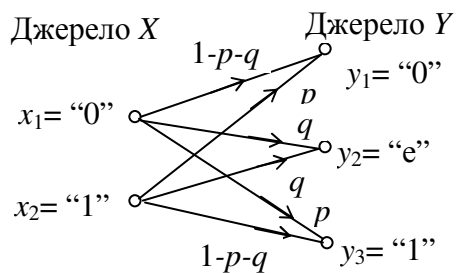


Рисунок 6.12 – Двійковий симетричний канал зі стираннями

Знайдемо пропускну здатність каналу зі стираннями. Так як канал симетричний, пропускна здатність досягається при рівномірному розподілі вхідних символів

$$p(x_1) = p(x_2) = 1/2.$$

Звідси випливає, що ймовірності вихідних символів дорівнюють

$$p(y_1) = \frac{1-q}{2}, \quad p(y_2) = q, \quad p(y_3) = \frac{1-q}{2}.$$

Тепер всі необхідні імовірності відомі. Скориставшись (6.6), маємо

$$\frac{C_{\text{ДСКС}}}{\text{біт}} = \sum_i^2 \sum_j^3 p(x_i, y_j) \log_2 \frac{p(y_j | x_i)}{p(y_j)} = \sum_i^2 \sum_j^3 p(x_i) p(y_j | x_i) \log_2 \frac{p(y_j | x_i)}{p(y_j)}. \quad (6.18)$$

Використовуючи властивість симетрії каналу, одержуємо

$$\begin{aligned} \frac{C_{\text{ДСКС}}}{\text{біт}} &= (1-p-q) \log_2 \left(2 \cdot \frac{1-p-q}{1-q} \right) + q \log_2 \frac{q}{q} + p \log_2 \left(2 \cdot \frac{p}{1-q} \right) = \\ &= 1-q + (1-p-q) \log_2 \left(\frac{1-p-q}{1-q} \right) + p \log_2 \left(\frac{p}{1-q} \right). \end{aligned} \quad (6.19)$$

Як ми бачимо, пропускна здатність каналу зі стираннями залежить тільки від ймовірностей p і q . Графік $C = f(p, q)$ являє собою просторову тривимірну поверхню, розташовану над площиною (p, q) . Тут ми обмежимося тільки розглядом двох важливих окремих випадків.

1. При $q = 0$, ми маємо двійковий симетричний канал, уже розглянутий раніше. Підставляючи $q = 0$ у вираз (6.19) ми, як і очікувалося, одержуємо формулу (6.16).

2. У каналі присутні тільки стирання, тобто при $p = 0$ – помилки не присутні, або ми ними нехтуємо. У цьому випадку

$$C_{\text{ДСКС}} = (1-q) \text{ біт}. \quad (6.20)$$

На рис. 6.13 показані пропускні здатності ДСК (6.19) і двійкового каналу зі стираннями ($p = 0$). Потрібно відзначити, що при малих ймовірностях помилки, вибором оптимальних областей стирань у ДСКС можна досягти істотно більшої пропускної здатності, чим у звичайних двійкових каналах.

Зауваження. Тут виникає питання про можливість збільшення пропускної здатності під час приймання зі стираннями на практиці. У цьому виявляється слабкість теорії інформації. Теорія інформації найчастіше не може запропонувати конструкцію, яка реалізує теоретично досяжні границі. Проте, невеликий приклад, докладно розглянутий у другій частині цього посібника показує, що введення стирань може іноді знижувати імовірність помилки. Розглянемо цей приклад на інтуїтивному рівні. Розіб'ємо потік переданої інформації на блоки, які містять 7 двійкових символів (7 біт). До кожного блоку додамо один біт («0» чи «1») перевірки на парність. Закодовані в такий спосіб блоки з восьми двійкових символів завжди будуть містити парне число одиниць. Нехай імовірність помилки в ДСК досить мала. Введемо зону стирання (рис. 6.11) таким чином, щоб помилки, в основному, переходили в стирання. При цьому, імовірність «нестертої» помилки буде зневажливо мала, а імовірність стирання буде залишатися досить малою. Ми одержимо канал зі стираннями, в якому блоки з восьми двійкових символів у переважній більшості випадків будуть прийняті правильно, або будуть містити тільки один стертий двійковий символ. Якість прийому істотно покращиться, тому що одне стирання в блоці з парним числом одиниць завжди може бути виправленим.

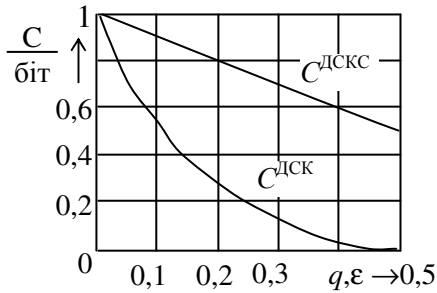


Рисунок 6.13 – Пропускна здатність двійкового симетричного каналу $C^{\text{ДСКС}}$ з імовірністю помилки ε і двійкового каналу зі стираннями $C^{\text{ДСК}}$ з імовірністю стирання q і ймовірністю помилки $p = 0$

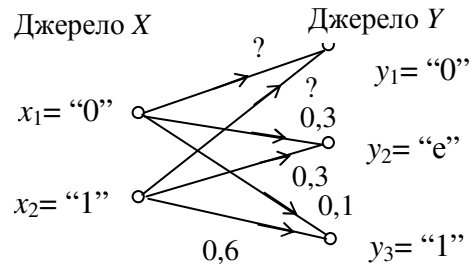


Рисунок 6.14 – Двійковий канал зі стираннями

Приклад: Двійковий симетричний канал зі стираннями.

Визначити:

1. Матрицю каналу;
2. Розподіл ймовірностей символів джерела Y , якщо відомо, що символи джерела X рівномірно розподілені, тобто $p_0 = p_1 = 1/2$;
3. Пропускна здатність каналу;
4. Діаграму інформаційних потоків з усіма ентропіями;
5. Модель каналу з матрицею $P_{Y|X}$.

Рішення.

1. Із врахуванням того, що сума ймовірностей у кожному рядку матриці дорівнює 1, одержимо

$$P_{Y|X}^{\text{ДСКС}} = \begin{pmatrix} 0,6 & 0,3 & 0,1 \\ 0,1 & 0,3 & 0,6 \end{pmatrix}.$$

2. Виходячи з рівномірного розподілу ймовірностей символів на вході, згідно з виразом (6.17), маємо

$$p(y_1) = 0,35; \quad p(y_2) = 0,3; \quad p(y_3) = 0,35.$$

3. Так як розглянутий канал симетричний, пропускна здатність досягається при рівномірному розподілі вхідних символів. З виразу (6.19) і з урахуванням (див. п. 1) маємо

$$\begin{aligned} \frac{C^{\text{ДСКС}}}{\text{біт}} &= 1 - 0,3 + (1 - 0,3 - 0,1) \log_2 \frac{1 - 0,3 - 0,1}{1 - 0,35} + 0,1 \cdot \log_2 \frac{0,1}{1 - 0,3} = 0,7 + 0,6 \cdot \log_2 \frac{0,6}{0,7} + \\ &+ 0,1 \cdot \log_2 \frac{0,1}{0,7} \approx 0,2858. \end{aligned}$$

4. Ентропія дискретного двійкового джерела без пам'яті X з рівномірним розподілом ймовірностей символів дорівнює $H(X) = 1$ біт.

Ентропія джерела Y дорівнює

$$\frac{H(Y)}{\text{біт}} = -2 \cdot 0,35 \log_2 0,35 - 0,35 \log_2 0,3 \approx 1,5813.$$

Так як в симетричному каналі з рівномірним розподілом вхідних символів $I(X;Y)$ збігається з пропускну здатністю C (див. п. 3), спільну ентропію і дві

умовні ентропії можна підрахувати, використовуючи таблицю 6.3. Діаграма інформаційних потоків зображена на рис. 6.15.

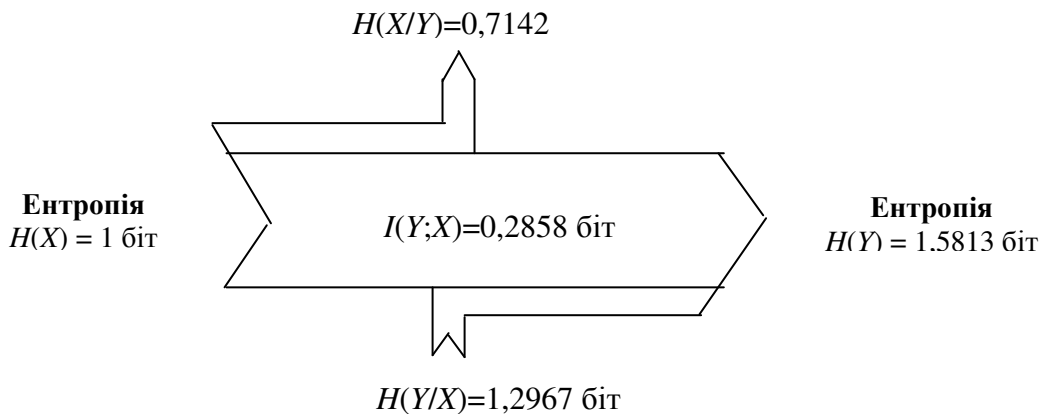


Рисунок 6.15 – Діаграма інформаційних потоків двійкового симетричного каналу зі стираннями

5. Перерахування матриці перехідних ймовірностей каналу $P_{Y|X}$ в матрицю $P_{X|Y}$ надаємо зробити читачу як самостійну вправу. Діаграма каналу з вхідним джерелом Y і вихідним X наведена на рис. 6.16 для контролю.

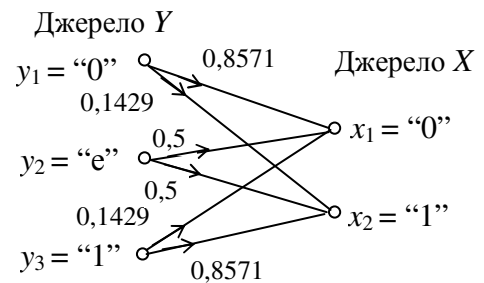


Рисунок 6.16 – Двійковий симетричний канал зі стираннями

6.6. Теорема кодування для дискретних каналів без пам'яті

Розглянемо дискретний канал без пам'яті з пропускнуою здатністю C [біт/символ], в якому кожен символ передається протягом T_s сек. Для цього каналу $C[\text{біт}/\%] = C[\text{біт}/\text{символ}] / T_s$.

Нехай ентропія деякого джерела X , вимірювана на протязі часу T_s сек. складає $H(X)$ біт. Тоді має місце наступна теорема.

Теорема 6.6.1. *Теорема кодування для каналу (теорема Шеннона).*

Для джерела X зі швидкістю $R = H(X)/T_s$ [біт/с] і $R < C$ існує деякий код, за допомогою якого інформація джерела X може бути передана каналом зв'язку з пропускнуою здатністю C [біт/с] з як завгодно малою імовірністю помилки. (Теорема кодування справедлива не тільки для дискретних каналів, вона також вірна і під час передачі дискретних повідомлень неперервними каналами).

Доказ теореми кодування для каналу (див., наприклад, [10]) досить складний і виходить за рамки цього посібника, тому обмежимося тут наступними зауваженнями.

- Доказ теореми кодування припускає використання випадкових кодів нескінченної довжини і декодера максимальної правдоподібності, який забезпечує мінімальну ймовірність помилки. Доказ не використовує ніяких конструктивних рішень. У ньому використовуються тільки статистичні

властивості і граничні переходи для блокових кодів з довжиною блоків, яка прямує до нескінченності. Доказ не дає ніяких указівок на конструкцію оптимальних кодів.

- Теорема кодування визначає також верхню границю для швидкості передачі R (Тут необхідно зробити роз'яснення. Існує зворотна теорема кодування, яка говорить про те, що при $R < C$ не існує ніякого методу кодування, який дозволяє передавати інформацію з як завгодно малою ймовірністю помилки).
- Під час доказу теореми вводиться показник експонентної оцінки R_0 , який може бути використаний для оцінки технічно досяжної швидкості передачі даних [4].

РОЗДІЛ 7

НЕПЕРЕРВНІ ДЖЕРЕЛА І КАНАЛИ

У розділі 1 дано визначення ентропії як міри невизначеності джерела. При цьому передбачалося, що ентропія вимірюється за допомогою випадкових експериментів. У даній главі ми будемо застосовувати аналогічний підхід до неперервних джерел.

Замість джерел з кінцевим алфавітом символів будемо розглядати джерела, виходом яких є неперервні сигнали.

Прикладом таких сигналів може служити напруга, яка змінюється в часі, у телефонних каналах і т.д. На рисунку 7.1 представлено неперервне джерело X , виходом якого є

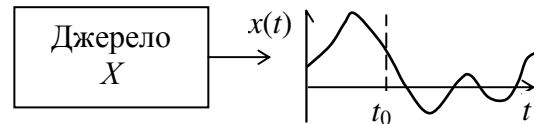


Рисунок 7.1 – Сигнал неперервного джерела

аналоговий сигнал $x(t)$, який є деякою випадковою функцією від часу t . Будемо розглядати значення $x(t)$ у деякі фіксовані моменти часу як випадкові експерименти, які несуть деяку інформацію про джерело X .

7.1. Диференціальна ентропія

На рисунку 7.2 показані два неперервних джерела X і Y , зв'язані каналом (аналогічно рис. 6.4). Тут, замість ймовірностей, стоять функції густини розподілу ймовірностей стохастичних змінних.

Використання стохастичних змінних і їхньої функції густини розподілу ймовірностей дозволяє вводити поняття інформації, ентропії, умовної і взаємної ентропії для двох неперервних джерел за аналогією з дискретними джерелами.

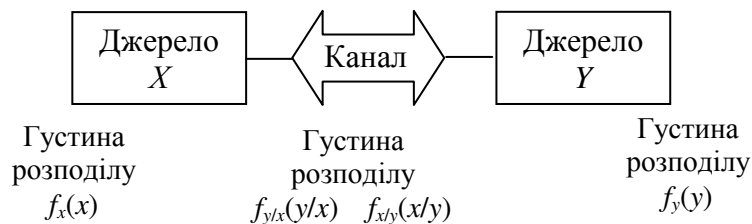
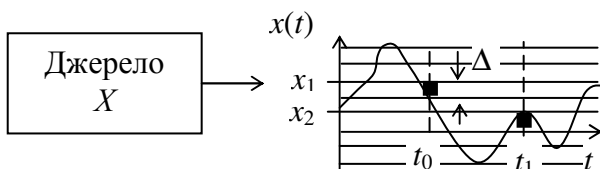


Рисунок 7.2 – Два неперервних джерела без пам'яті, зв'язаних каналом

Рисунок 7.3 – Оцифрування неперервного джерела з інтервалом квантування Δ в моменти спостереження t_0, t_1 і т.д.

Перетворимо неперервне джерело X у дискретне. Для цього проквантуємо значення аналогового виходу джерела з кроком Δ (рис. 7.3).

Крім цього, як це звичайно робиться в теорії інформації, зробимо дискретизацію джерела за часом. В результаті, одержимо послідовність стохастичних змінних X_0, X_1, X_2, \dots . Виходячи з таблиці 6.2,

визначимо взаємну інформацію символів x_i і x_j , де x_i – значення вихідного символу в момент часу t_m , а x_j – у момент часу t_n

$$\begin{aligned} \frac{I_{X_m X_n}(x_i; x_j)}{\text{біт}} &= \log_2 \frac{P([x_i - \Delta \leq X_m < x_i] \cap [x_j - \Delta \leq X_n < x_j])}{P(x_i - \Delta \leq X_m < x_i)P(x_j - \Delta \leq X_n < x_j)} = \\ &= \log_2 \frac{\int_{x_i - \Delta}^{x_i} \int_{x_j - \Delta}^{x_j} f_{X_m X_n}(x_1, x_2) dx_1 dx_2}{\int_{x_i - \Delta}^{x_i} f_{X_m}(x_1) dx_1 \int_{x_j - \Delta}^{x_j} f_{X_n}(x_2) dx_2}. \end{aligned} \quad (7.1)$$

Взаємну інформацію можна трактувати як «зняту» (утрачену) невизначеність попадання змінної X_n в інтервалі $[x_j - \Delta, x_j)$, коли відомо, що змінна X_m належить інтервалові $[x_i - \Delta, x_i)$ або навпаки. Будемо вважати функцію густини розподілу імовірності неперервною функцією. Тоді, спрямовуючи ширину інтервалу квантування до нуля, одержимо

$$\begin{aligned} \lim_{\Delta \rightarrow \infty} \log_2 \frac{\int_{x_i - \Delta}^{x_i} \int_{x_j - \Delta}^{x_j} f_{X_m X_n}(x_1, x_2) dx_1 dx_2}{\int_{x_i - \Delta}^{x_i} f_{X_m}(x_1) dx_1 \int_{x_j - \Delta}^{x_j} f_{X_n}(x_2) dx_2} &= \lim_{\Delta \rightarrow \infty} \log_2 \frac{\Delta \Delta f_{X_m X_n}(x_i, x_j)}{\Delta f_{X_m}(x_i) \Delta f_{X_n}(x_j)} = \\ &= \log_2 \frac{f_{X_m X_n}(x_i, x_j)}{f_{X_m}(x_i) f_{X_n}(x_j)}, \end{aligned}$$

тобто результат, аналогічний виразу взаємної інформації для дискретних джерел. *Передачу інформацію* можна визначити як математичне сподівання

$$\frac{I(X; Y)}{\text{біт}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) \log_2 \frac{f_{XY}(x, y)}{f_X(x) f_Y(y)} dx dy \quad (7.2)$$

Зауваження. Тут, для приведення у відповідність позначень цього розділу з результатами таблиці 6.2, замість X_m використовується X , а замість Y_n – Y .

Інформація джерела визначається виходячи з аналогічних міркувань

$$\begin{aligned} \frac{I(x_i)}{\text{біт}} &= -\log_2 P([x_i - \Delta \leq X < x_i]) = -\log_2 \int_{x_i - \Delta}^{x_i} f_X(x) dx \stackrel{\Delta \ll 1}{\approx} -\log_2 (\Delta f_X(x_i)) = \\ &= -\log_2 \Delta - \log_2 f_X(x_i). \end{aligned}$$

На відміну від виразу (7.3) для взаємної інформації, у (7.4) з'являється доданок, який залежить від інтервалу квантування Δ

При $\Delta \rightarrow \infty$ величина $\log_2(\Delta)$ також прямує до нескінченності. В результаті, вираз для $I(x_i)$ також прямує до ∞ . Це не дивно, тому що зі зменшенням кроку квантування, число окремих подій (символів алфавіту джерела) зростає і, отже, невизначеність джерела також росте.

Величина $\log_2(\Delta)$ не залежить від джерела і зовсім не придатна для його опису, тому, здається цілком природно використовувати тільки функцію густини розподілу імовірності неперервного джерела. Таким чином, ми переходимо до наступного визначення.

Середня інформація неперервного джерела, так звана *диференціальна ентропія*, визначається як

$$\frac{H(X)}{\text{біт}} = - \int_{-\infty}^{\infty} f(x) \log_2 f(x) dx. \quad (7.3)$$

Насамперед відзначимо, що таке довільне визначення диференціальної ентропії підтверджує свою придатність тим, що ентропійні відносини для дискретних джерел виявляються справедливими і для випадку неперервних джерел і каналів. Зокрема, для неперервних джерел мають місце співвідношення (6.39) – (6.42).

Таким чином, диференціальна ентропія неперервного джерела залежить тільки від функції густини розподілу імовірності, яка у загальному випадку є нескінченною величиною, тому, поставимо питання про те, наскільки великим може бути значення диференціальної ентропії. Насамперед відзначимо, що характеристиками стохастичного процесу є дві величини: середнє значення, яке приймає стохастична змінна (з властивістю лінійності) μ і стандартне відхилення стохастичної змінної σ .

Середнє значення або математичне сподівання μ не має ніякого впливу на диференціальну ентропію. Зі зростом σ невизначеність джерела зростає, що призводить також до зростання диференціальної ентропії. У зв'язку з цим, порівняння різних функцій густини розподілу ймовірностей щодо відповідних їм ентропій має сенс робити при однакових σ .

Зауваження. В інформаційній техніці за вихідний параметр приймають σ^2 – дисперсію, яка визначає середню потужність стохастичного процесу, [10]. Ясно, що зі збільшенням потужності передавача, кількість переданої інформації збільшується і, навпаки, зі збільшенням потужності шумів, зростає невизначеність, тобто в одиницю часу передається менше інформації.

З теорії інформації випливає, що диференціальна ентропія досягає свого максимуму при гауссовому розподілі імовірності.

Теорема 7.1.1. При заданій дисперсії σ^2 , максимальну диференціальну ентропію має джерело з гауссовим розподілом імовірності, причому,

$$H_{\text{Gaus}}(X) = \frac{1}{2} \cdot \frac{\ln(2\pi\sigma^2 e)}{\ln 2} \text{ біт}. \quad (7.4)$$

Приклад: Диференціальна ентропія гауссового джерела.

З (7.3) випливає, що диференціальна ентропія гауссового джерела дорівнює

$$\begin{aligned} \frac{H(X)}{\text{нат}} &= - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \log_2 \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \right] dx = \\ &= - \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) \log_2 \left[\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{x^2}{2\sigma^2} \right] dx. \end{aligned}$$

Вираз у квадратних дужках може бути розкладений на два інтеграли. Таким чином, остаточно маємо

$$\frac{H(X)}{\text{нат}} = \ln \sqrt{2\pi\sigma^2} + \frac{1}{2\sigma^2} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \frac{1}{2} \ln(2\pi\sigma^2) + \frac{1}{2} = \frac{1}{2} \ln(2\pi\sigma^2 e). \quad (7.5)$$

Кількісні приклади для трьох, найбільш поширених розподілів, приведені в таблиці 7.1.

Приклад: Телефонія.

Практична користь приведених вище результатів може бути наочно показана за допомогою оцінки швидкості передачі інформації (у бітах) в цифрових телефонних лініях. Сучасні стандартні методи цифрової передачі мови (логарифмічні PCM) вимагають витрати 8 біт на кодування одного відліку, при частоті відліків 8 кГц. Таким чином, швидкість передачі мови складає 64 Кбіт/сек.

Таблиця 7.1. Приклад диференціальної ентропії

Розподіл	Функція густини розподілу імовірності	Диференціальна ентропія
Рівномірний	$f(x) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{для } x \leq \sqrt{3}; \\ 0, & \text{const.} \end{cases}$	$\frac{\ln(2\sqrt{3}\sigma)}{\ln 2} = 1,79$
Лапласа	$f(x) = \frac{1}{\sqrt{2}\sigma^2} \exp\left(-\sqrt{2} \frac{x}{\sigma}\right)$	$\frac{\ln(\sqrt{2}\sigma e)}{\ln 2} = 1,94$
Гауссовий	$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right)$	$\frac{\ln(2\pi\sigma^2 e)}{2\ln 2} = 2,04$

Виходячи з рівномірного розподілу ймовірностей в інтервалі $[-1, 1]$, дослідним шляхом одержимо $\sigma^2 = 1/3$. Таким чином, диференціальна ентропія на один відлік складає

$$\frac{H(X)}{\text{біт}} = \frac{\ln(2\sqrt{3}\sigma)}{\ln 2} \Big|_{\sigma^2 = \frac{1}{3}} = 1.$$

Так як відліки беруться з частотою 8 кГц, отримуємо, що необхідна швидкість передачі мови складає 8 Кбіт/с. При оцінці ентропії ми не брали до уваги зв'язок між сусідніми відліками (пам'ять джерела) і тому, реальна диференціальна ентропія джерела мови буде ще менше. Насправді, ми знаємо, що сучасні алгоритми кодування мови дозволяють здійснювати передачу мовного сигналу зі швидкістю близько 8 Кбіт/с при якості, порівнянній зі стандартним PCM.

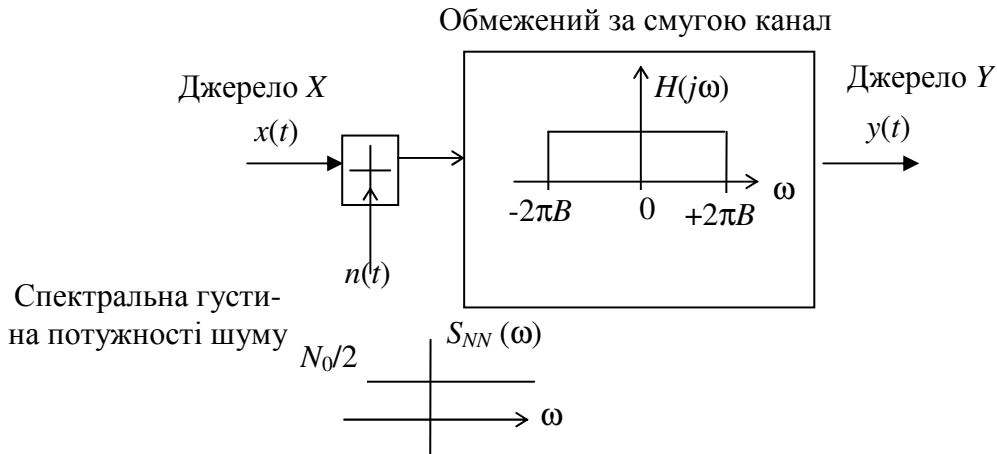
7.2. Пропускна здатність каналу і границя Шеннона

Аналогічно дискретним каналам, можна визначити пропускну здатність для неперервних каналів. Будемо шукати, як і раніше, найбільше значення інформації, яка переноситься, по всіх можливих функціях густини розподілу ймовірностей.

$$C = \sup_{f(x)} I(X;Y). \quad (7.6)$$

Пошук точної верхньої грані, у загальному випадку, являє собою досить складну задачу. Розглянемо найважливіший частковий випадок - передачу інформації по каналу з адитивним білим гауссовим шумом з обмеженою смугою. Модель такого каналу зображена на рис. 7.4. Для гауссового розподілу імовірності

сті амплітуди сигналу на вході каналу отримуємо формулу пропускної здатності каналу, добре відому в теорії інформації.



Теорема 7.2.1. *Пропускна здатність каналу (Хартлі – Шеннон).*

Пропускна здатність ідеального каналу із шириною смуги пропускання B і адитивним білим гауссовим шумом потужності $N = N_0B$ дорівнює

$$\frac{C}{\text{біт/сек}} = B \cdot \log_2 \left(1 + \frac{S}{N} \right) \quad (7.7)$$

Тут, як і раніше, S – потужність сигналу в смузі пропускання каналу. Розмірність C – біт/сек.

Для інтерпретації пропускної здатності неперервного каналу C , розглянемо вираз (7.7) при ширині смуги пропускання, яка дорівнює 1 Гц (рис. 7.5). Можна відзначити асимптотично лінійний характер поведінки функції $C/B = \log_2 f(S/N)$ (рис. 7.5 виконано у напівлогарифмічному масштабі).

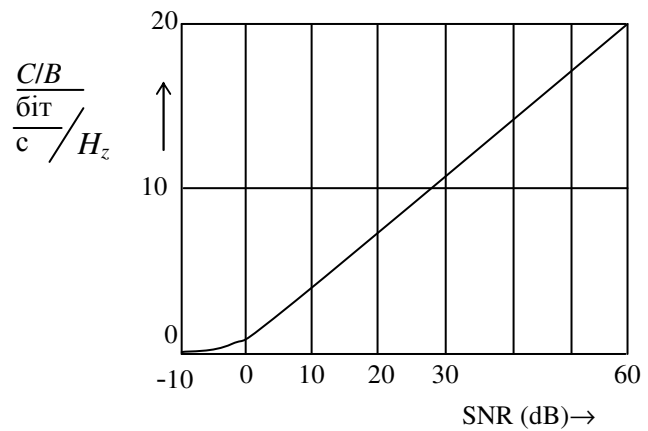


Рисунок 7.5. Пропускна здатність на 1 Гц смуги пропускання, як функція відношення сигнал/шум

Нехай задане відношення сигнал/шум (Signal to Noise Ratio – SNR) більше 0 дБ. З формули (7.7) випливає, що подвоєння пропускної здатності вимагає квадратичного збільшення відношення сигнал/шум SNR, тобто квадрата потужності передавача при постійному шумі.

Нехай задані відношення сигнал/шум, смуга пропускання B и час передачі t . Тоді можна визначити пропускну здатність каналу і максимальний обсяг інформації, що передається за заданий проміжок часу. Помітимо так само, що при фіксованому обсязі інформації v біт, можна довільно варіювати два з трьох параметрів (SNR, B , t), а третій параметр буде визначатися із співвідношення (7.7) і умови $v = C \cdot t$. Усе вищесказане ілюструє рис. 7.6.

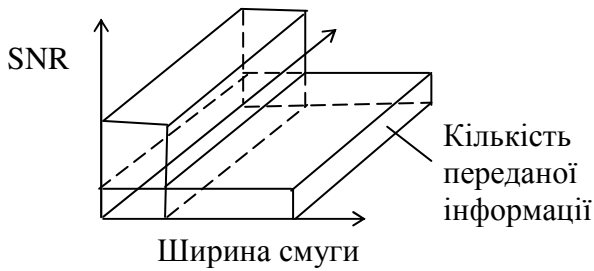


Рисунок 7.6 – Співвідношення між шириною смуги пропускання, SNR, часом передавання і максимальною кількістю переданої інформації

рення смуги викликає збільшення потужності шуму N . Потужність шуму при постійній спектральній густині N_0 пропорційна ширині смуги пропускання. Чим ширше смуга, тим більше шум на виході приймача (рис. 7.7), так як $N = N_0 \cdot B$.

Дослідимо граничний перехід

$$\frac{C_\infty}{\text{біт/сек}} = \lim_{B \rightarrow \infty} \frac{C}{\text{біт/сек}} = \lim_{B \rightarrow \infty} B \ln\left(1 + \frac{S}{N}\right) \cdot \log_2 e. \quad (7.8)$$

Зауваження. Ми використовували натуральний логарифм для спрощення обчислення межі. Перехід від розмірності пропускної здатності нат/сек до розмірності біт/сек досягається множенням правої частини виразу (7.8) на $\log_2(e)$.

Підставляючи $N = N_0 \cdot B$ в граничний перехід, отримуємо невизначеність

$$\lim_{B \rightarrow \infty} B \ln\left(1 + \frac{S}{N_0 B}\right) = \lim_{B \rightarrow \infty} \frac{\ln\left(1 + \frac{S}{N_0 B}\right)}{1/B}.$$

Розкриваючи невизначеність за правилом Лопіталія, отримуємо кінцеве значення межі

$$\lim_{B \rightarrow \infty} B \ln\left(1 + \frac{S}{N_0 B}\right) = \lim_{B \rightarrow \infty} \frac{\left(\frac{S}{N_0}\right)\left(-\frac{1}{B^2}\right)}{\left(1 + \frac{S}{N_0 B}\right)\left(-\frac{1}{B^2}\right)} = \frac{S}{N_0}. \quad (7.9)$$

Вираз (7.9) разом з (7.8) визначає **границю Шеннона**.

Пропускна здатність неперервного каналу з АБГШ і необмеженою смугою пропускання дорівнює

$$\frac{C_\infty}{\text{біт/сек}} \approx \frac{S}{N_0} \log_2 e \approx 1,44 \frac{S}{N_0}.$$

В техніці зв'язку при передаванні цифрової інформації часто використовується відносна величина – *енергія сигналу* E_b , яка приходить на біт переданої інформації. Так як максимальна швидкість передачі інформаційних біт/сек визначається як:

$$R_{\max} = C_\infty. \quad (7.10)$$

Мінімальна тривалість передачі одного біта дорівнює

З рис. 7.6 видно, що обсяг переданої інформації за визначений час можна представити як об'єм паралелепіпеда у координатах SNR, B і t .

При уважному розгляді співвідношення (7.7), виникає важливе питання: **як буде змінюватися пропускна здатність каналу при прямуванні ширини смуги пропускання до нескінченності?**

Відповідь не очевидна, тому що розши-

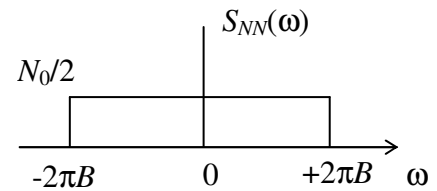


Рисунок 7.7 – Спектр білого гауссовського шуму в обмеженій смузі

$$T_b = \frac{1}{C_\infty}. \quad (7.11)$$

Енергія, затрачувана на передачу біта інформації, визначається добутком $E_b = S \cdot T_b$. Використовуючи попередні викладення, переходимо до наступного твердження.

Для передачі одного біта цифрової інформації необхідно, щоб відношення енергії на біт E_b до *спектральної густини потужності* білого гауссовського шуму N_0 дорівнювало, як мінімум

$$\left. \frac{E_b}{N_0} \right|_{\min} = \frac{1}{\log_2 e} \approx 0,69 \cong -1,59 \text{ дБ}. \quad (7.12)$$

Зауваження. Помітимо, що в літературі *спектральна густина шуму* іноді визначається не як $\frac{N_0}{2}$, а як N_0 . Це приводить до появи додаткового доданка в правій частині (7.12), що дорівнює 1,42 дБ.

Наочно зв'язок між SNR, B й бітовою швидкістю R можна представити у вигляді діаграми. Для цього, насамперед, формально підставимо R замість C і $N_0 B$ замість N у (7.7), розв'яжемо рівність відносно $\frac{S}{N_0 B}$ й одержимо

$$\frac{S}{N_0 B} = 2^{\frac{R}{B}} - 1.$$

Усунемо залежність лівої частини від смуги пропускання B шляхом множення обох частин на $\frac{B}{R}$ одержимо

$$\frac{S/R}{N_0} = \frac{B}{R} \cdot (2^{\frac{R}{B}} - 1). \quad (7.13)$$

Енергія E_b , затрачувана на передачу одного біта інформації, визначається відношенням $E_b = \frac{S}{R}$. Отримана залежність між $\frac{E_b}{N_0}$ і $\frac{B}{R}$ приведена на рис. 7.8.

Рисунок 7.8 можна розглядати як стандарт, що дозволяє оцінити ефективність обраного методу кодування в реальних системах зв'язку. Нехай передача інформації здійснюється при деяких фіксованих значеннях R , B и SNR. Цим значенням відповідає деяка точка на діаграмі рис. 7.8. Відповідно до теореми кодування для каналу, швидкість R не повинна перевищувати пропускну здатність каналу C , тому наша точка завжди лежить вище кривої, яка задається (7.13). **Відстань від граничної лінії, яка відповідає значенню $R = 3$, дозволяє оцінити потенційну можливість поліпшення обраного нами методу кодування.**

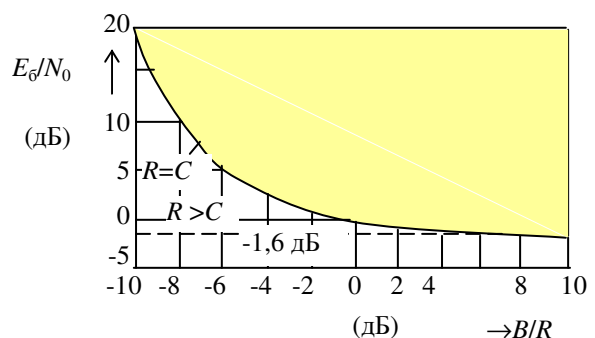


Рисунок 7.8 – Залежність між шириною смуги пропускання і SNR при передачі інформації

Зауваження. Сучасні цифрові системи зв'язку використовують прогресивні методи кодування, такі, наприклад, як турбо-коди. Застосування таких конструкцій дозволяє наблизитися до граничної кривої ціною деякої затримки декодування. При оцінці співвідношення ціна - ефективність, враховується багато додаткових факторів, тому оптимальність системи, у смислу наближення швидкості, до пропускну здатності каналу, іноді відходить на другий план.

7.3. Приклади

Приклад: Відеотелефонія

Розглянемо можливість передачі зображень між абонентами телефонної мережі стандарту ISDN. У цій системі для передачі двовимірних зображень використовується канал зі швидкістю 64 Кбіт/с, тому, необхідно, щоб потік відеоінформації був стиснутий до цієї величини. При цьому, алгоритм стиску має усувати як надмірність з погляду теорії інформації, так і деталі зображення, не істотні для абонента або багаторазово повторювані. Однією з таких деталей може бути, наприклад, не змінний тривалий час другий план зображення і т.д. Надалі, ми розглянемо приклад ідеального стиску відеосигналу.

При обчисленні необхідної швидкості передачі даних, ми будемо виходити зі стандарту QCIF (Quarter Common Intermediate Format).

У цьому стандарті для передачі кольорових зображень використовується один сигнал яскравості (Y) і два сигнали кольоровості (U, V). Параметри цих сигналів приведені в таблиці 7.2 (Відеосигнал у стандартах QCIF PAL і SECAM містить – 288 рядків і 352 точки в рядку для сигналу яскравості (Y) і два сигнали кольоровості (U, V) по 176 рядків і 144 точки на кожний, для відеосигналу в стандарті QCIF NTSC – 240 рядків по 352 точки в рядку (Y) і, відповідно, 120 рядків і 144 точки (U, V)).

Таблиця 7.2. Приклад сигналів для передавання зображень.

	Кількість точок у рядку	Кількість рядків у кадрі	Кількість біт на точку	Частота зміни кадрів
Сигнал яскравості (Y)	176	144(120)	8	5...15 Гц
Сигнал кольоровості (U, V)	88	72 (66)	8	5...15 Гц

З табл. 7.2 випливає, що відображення кольорового зображення на екрані досягається за допомогою чотириточкових матриць, кожна з яких око сприймає як одну точку. У середині матриці дві точки належать сигналові однієї кольоровості, а дві – іншому. Яскравість кожної точки визначається сигналом Y , який квантується на $2^8 = 256$ рівнів. Будемо вважати, що зміна кадрів відбувається з частотою 10 Гц. При швидкому переміщенні камери і порівняно повільній зміні кадрів, на екрані виникають, так звані, розмивання зображення.

1. Знайдіть інформаційний потік у відеоканалі за умови, що частота зміни кадрів дорівнює 10 Гц. Будемо виходити з наступної спрощеної моделі: проква-

нтовані відеосигнали мають рівномірний розподіл імовірності у часі і в просторі і взаємно незалежні.

2. Знайдіть мінімальну ширину смуги пропускання ідеального телевізійного приймача, якщо необхідне для задовільної якості зображення відношення сигнал/шум складає 30 дБ.

Рішення.

1. З табл. 7.2 маємо, що число символів, необхідних для передачі кадру, дорівнює

$$\frac{N_s}{\text{символ}} = 176 \cdot 144 + 2 \cdot 88 \cdot 72 = 38016.$$

Так як частота зміни кадрів дорівнює 10 Гц, швидкість передачі символів складає

$$r_s = N_s \cdot 10 \text{ Гц} = 380160 \text{ символ/с.}$$

З огляду на те, що імовірності $2^8 = 256$ можливих квантованих значень амплітуд сигналів Y, U, V розподілені рівномірно, інформація кожного символу дорівнює

$$I_s = -\log_2 \frac{1 \text{ біт}}{2^8 \text{ символ}} = 8 \text{ біт/символ.}$$

Таким чином інформаційний потік, який шукали, складає

$$I = r_s \cdot I_s = 380160 \frac{\text{символ}}{\text{с}} \cdot 8 \frac{\text{біт}}{\text{символ}} = 3,04128 \text{ Мбіт/с.}$$

Зауваження. Помітимо, що інформація символу дорівнює довжині двійкового запису символу тільки у випадку рівномірного розподілу ймовірностей символів. У реальних каналах передачі цифрових відеосигналів інформація одного символу істотно менше 8 біт. Порівняння останнього результату з ISDN-B каналом, по якому інформація передається зі швидкістю 64 Кбіт/с показує, що фактор стиснення в цьому каналі повинен дорівнювати, приблизно, 50.

2. Для передачі інформаційного потоку 3,04128 Мбіт/с повинен використовуватися канал з більшою пропускною здатністю. Із шеннонівської формули (7.7) отримуємо

$$B > \frac{I}{\log_2 \left(1 + \frac{S}{N} \right)} = \frac{3,04128 \cdot 10^6}{\log_2 (1 + 10^3)} \cdot \frac{1}{\text{с}} \approx 305 \text{ кГц.}$$

Приклад: Імпульсно-кодова модуляція.

Аналоговий сигнал з верхньою спектральною частотою 4 кГц підлягає дискретизації в часі з частотою відліків, у 1,25 разів перевищуючою мінімально необхідну для відновлення сигналу. Кожен відлік квантується на 256 рівнів символами, що містять 8 біт. Передбачається, що символи незалежні і мають рівномірний розподіл ймовірностей.

1. Знайдіть інформаційний потік оцифрованого джерела.

2. Чи може цей інформаційний потік бути переданий по каналі з АБГШ зі смугою $B = 10 \text{ КГц}$ і $\text{SNR} = 20 \text{ дБ}$ із нехтовно малою імовірністю помилки?

3. Яке повинно бути мінімальне SNR , необхідне для передачі інформаційного потоку в смузі $B = 10 \text{ кГц}$ із нехтовно малою імовірністю помилки?

4. Яке мінімальне значення смуги пропускання B , яке забезпечує передачу інформації по каналу з АБГШ при $\text{SNR} = 20$ дБ із нехтовно малою імовірністю помилки?

Рішення.

1. Частота відліків дорівнює

$$f_s = 1,25 \cdot 4 \cdot 2 = 10 \text{ кГц.}$$

Швидкість передачі символів дорівнює $r_s = 10^4$ сим/с.

Інформація одного символу – $I_s = 8$ біт.

Інформаційний потік – $I = I_s \cdot r_s = 80$ Кбіт/с.

2. Пропускна здатність каналу дорівнює

$$C = 10 \text{ кГц} \cdot \log_2(1 + 10^2) \approx 66 \text{ Кбіт/с.}$$

таким чином, $C < I$. **Безпомилкова передача неможлива.**

3. Мінімумально необхідне відношення сигнал/шум визначається з (7.7)

$$10 \text{ кГц} \cdot \log_2\left(1 + \frac{S}{N}\right) = 80 \text{ Кбіт/сек} \quad \text{і дорівнює} \quad \frac{S}{N}\Big|_{\min} = 255 \approx 24,1 \text{ дБ.}$$

4. Мінімумально необхідне значення смуги пропускання обчислюється, виходячи з (7.7)

$$B \log_2(1 + 10^3) = 80 \text{ Кбіт/с.}$$

Таким чином, одержуємо $B\Big|_{\min} = 12 \text{ кГц.}$

Приклад: Телефонний канал.

Стандартний аналоговий телефонний канал має смугу пропускання від 300 Гц до 3,4 кГц. Будемо припускати, що телефонний вузол зв'язку забезпечує на виході задане відношення сигнал/шум.

Кожен абонент зв'язаний з комутатором парою проводів. Загасання сигналу в смузі до 4 кГц складає, приблизно, 2 дБ/км.

Побудуйте графік пропускну здатності каналу, як функцію віддаленості абонента від вузла зв'язку.

Рішення.

Підставимо у вираз для пропускну здатності (7.7) значення відношення сигнал/шум як функції відстані l . Будемо вважати ширину смуги пропускання 4 кГц, тоді

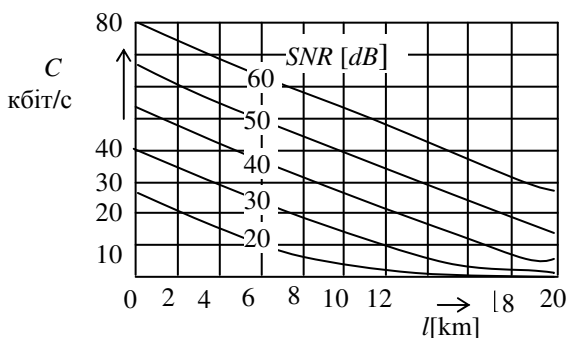


Рисунок 7.9 – Залежність пропускну здатності від віддалі підключення

$$C = 4 \log_2 \left(1 + 10^{0,1 \left[\frac{\text{SNR} - 2l}{\text{км}} \right]} \right) \text{ Кбіт/сек.}$$

Графіки залежності пропускну здатності від відстані приведені на рис. 7.9.

Значення пропускну здатності C , як функції l , що відповідає відношенню сигнал/шум 60 дБ, падає майже лінійно з ростом $0 < l < 20$ км із 80 Кбіт/с до 26 Кбіт/с. Подібним же чином поведуться криві, що відповідають меншим початковим відно-

шенням сигнал/шум, однак, зі зменшенням SNR, властивість лінійності втрачається.

Зауваження. Статистика показує, що 99,5% абонентів, підключених до центрального вузла зв'язку, були віддалені від нього не більше ніж на 8 км. Фактично, пропускна здатність каналів зв'язку істотно вище, ніж приведена в прикладі. Це пояснюється тим фактом, що ширина смуги пропускання пари телефонних проводів у багато разів вище 4 кГц.

Сучасні методи передачі інформації по провідних лініях зв'язку (xDSL) дозволяють реалізувати швидкість передачі даних до декількох Мбіт/с. Наприклад, у стандарті ADSL, передача інформації здійснюється зі швидкістю вище 64 Кбіт/с і використовується смуга частот від 26... 138 кГц. При з'єднанні абонентів у мережу зі швидкістю 6 Мбіт/с, смуга частот складає 26... 1104 кГц.

Приклад: Телеграфія.

У телеграфії для передачі повідомлень використовуються два символи – крапка і дефіс. Окремі символи розділені паузами. Будемо вважати, що тривалість дефісу складає 1 с, а тривалість крапки і паузи – 1/3 с.

Знайдіть середній інформаційний потік телеграфного джерела.

Рішення.

Відомо, що відношення частот з'явлення крапок і дефісів в коді Морзе дорівнює, приблизно, 2:1 (з урахуванням ймовірностей букв алфавіту в тексті). Виходить, імовірності появи крапки і дефісу зв'язані співвідношенням

$$P_{\text{крапка}} = 2P_{\text{дефіс}} \quad \text{і} \quad P_{\text{крапка}} + P_{\text{дефіс}} = 1.$$

Отже,

$$P_{\text{крапка}} = 2/3 \quad \text{і} \quad P_{\text{дефіс}} = 1/3.$$

Ентропія двійкового телеграфного джерела

$$\frac{H(x)}{\text{біт}} = -P_{\text{крапка}} \cdot \log_2 P_{\text{крапка}} - P_{\text{дефіс}} \log_2 P_{\text{дефіс}} = 0,92.$$

Середня тривалість одного символу, включаючи наступну паузу, складає

$$\bar{T} = 2/3 t_{\text{крапка}} + 1/3 t_{\text{дефіс}} + t_{\text{пауза}} = 2/3 \cdot 1/3 \text{ с} + 1/3 \text{ с} + 1/3 \text{ с} = 8/9 \text{ с}.$$

Таким чином, середній інформаційний потік дорівнює

$$I = \frac{H(X)}{T} = 1,04 \text{ біт/с}.$$

<p>ЕФЕКТИВНЕ КОДУВАННЯ ДЖЕРЕЛА ПОВІДОМЛЕНЬ КОДОМ ХАФФМЕНА ВИКОРИСТОВУЄТЬСЯ ДЛЯ</p> <p>1. підвищення завадостійкості передавання повідомлень 2. підвищення швидкості передавання повідомлень</p> <p>Відповідь: 2</p>
<p>МІНІМАЛЬНА ДОПУСТИМА ЧАСТОТА ДИСКРЕТИЗАЦІЇ СИГНАЛУ ЗІ СМУГОЮ 300 – 3400 Гц ДЛЯ ПЕРЕДАВАННЯ ЙОГО ВІДЛІКАМИ</p> <p>1. 6200 Гц 2. 6800 Гц 3. 8000 Гц</p> <p>Відповідь: 2</p>
<p>У РАЗІ НЕСКІНЧЕННОГО ЗБІЛЬШЕННЯ СМУГИ ЧАСТОТ ІНФОРМАЦІЙНОГО КАНАЛУ ($F_K \rightarrow \infty$) ЙОГО ПРОПУСКНА ЗДАТНІСТЬ</p> <p>1. $\rightarrow \infty$ 2. залишається незмінною 3. $\rightarrow 0$ 4. прямує до певного значення</p> <p>Відповідь: 4</p>
<p>ЗА ТЕОРЕМОЮ ШЕННОНА ШУМИ В ІНФОРМАЦІЙНОМУ КАНАЛІ</p> <p>1. обмежують якість передавання інформації 2. обмежують швидкість передавання інформації</p> <p>Відповідь: 2</p>
<p>13. ПРОПУСКНА ЗДАТНІСТЬ ГАУССОВОГО КАНАЛУ ПРИ ЙОГО СМУЗІ 8 кГц ТА ВІДНОШЕННІ СИГНАЛ/ШУМ, РІВНОМУ СЕМИ, ДОРІВНЮЄ, біт/с</p> <p>1. 8000 2. 16000 3. 24000</p> <p>Відповідь: 3</p>
<p>14. ЗА ТЕОРЕМОЮ ШЕННОНА ЯКІСНЕ ПЕРЕДАВАННЯ ІНФОРМАЦІЇ КАНАЛОМ МОЖЛИВЕ, ЯКЩО МІЖ ПРОДУКТИВНІСТЮ ДЖЕРЕЛА R_d ТА ПРОПУСКНОЮ ЗДАТНІСТЮ КАНАЛУ C_k ВИКОНУЄТЬСЯ СПІВВІДНОШЕННЯ</p> <p>1. $R_d = C_k$ 2. $R_d > C_k$ 3. $R_d < C_k$</p> <p>Відповідь: 3</p>
<p>15. ШВИДКІСТЬ КОДУ (n, k) ВИЗНАЧАЄТЬСЯ</p> <p>1. n/k 2. k/n 3. $1 - k/n$</p> <p>Відповідь: 2</p>
<p>16. ДЛЯ ШУМУ З ГУСТИНОЮ ЙМОВІРНОСТІ</p> <p>$p(x) = \frac{1}{2\sqrt{2\pi}} \exp(-\frac{(x-3)^2}{8})$ МАТЕМАТИЧНЕ СПОДІВАННЯ МАЄ ЗНАЧЕННЯ</p> <p>1. 2 2. 3 3. 4 4. 8</p> <p>Відповідь: 2</p>
<p>17. ДЛЯ ШУМУ З ГУСТИНОЮ ЙМОВІРНОСТІ</p> <p>$p(x) = \frac{1}{2\sqrt{2\pi}} \exp(-\frac{(x-3)^2}{8})$ ДИСПЕРСІЯ МАЄ ЗНАЧЕННЯ</p> <p>1. 2 2. 3 3. 4 4. 8</p> <p>Відповідь: 3</p>

18. ТВЕРДЖЕННЯ ПРО ТЕ, ЩО В КАНАЛІ З ОБМЕЖЕНОЮ СМУГОЮ ЧАСТОТ І ЗАВАДАМИ МОЖНА ДОСЯГТИ БУДЬ-ЯКОЇ ВИСОКОЇ ЯКОСТІ ПЕРЕДАЧІ ІНФОРМАЦІЇ ЗА УМОВИ, ЩО $R_d < C_k$

1. вірне

2. невірне

Відповідь: 1

19. ТВЕРДЖЕННЯ ПРО ТЕ, ЩО В КАНАЛІ З ОБМЕЖЕНОЮ СМУГОЮ ЧАСТОТ І ЗАВАДАМИ МОЖНА ДОСЯГТИ БУДЬ-ЯКОЇ ВЕЛИКОЇ ШВИДКОСТІ ПЕРЕДАЧІ ІНФОРМАЦІЇ

1. вірне

2. невірне

Відповідь: 2

20. СТИСНЕННЯ ІНФОРМАЦІЇ ПРИЗНАЧЕНЕ ДЛЯ

1. підвищення завадостійкості

2. підвищення швидкості передачі

3. захисту інформації

Відповідь: 2

21. АЛГОРИТМ СТИСНЕННЯ ІНФОРМАЦІЇ ЛЕМПЕЛА-ЗІВА Є

1. словниково-орієнтований

2. матричний

3. арифметичний

Відповідь: 1

22. КРИПТОГРАФІЯ ВИКОРИСТОВУЄТЬСЯ ДЛЯ

1. підвищення швидкості передачі

2. захисту від завад

3. захисту від несанкціонованого доступу

Відповідь: 3

23. КОД ХЕММІНГА Є

1. простий

2. коректуючий

3. ефективний

Відповідь: 2

Встановіть відповідність у вигляді комбінації цифр і букв, записавши, наприклад, 1С, 2D, ...

24. ОСНОВНІ ОДИНИЦІ ВИМІРЮВАННЯ В ІНФОРМАЦІЙНИХ СИСТЕМАХ

ВеличиниНайменування одиниці

1. Кількість інформації

А. дБ

2. Швидкість модуляції

В. біт/с

3. Швидкість передачі інформації

С. біт

4. Відносна величина (рівень)

D. Бод

Відповідь: 1С, 2D, 3В, 4А

25. ВЗАЄМОЗВ'ЯЗОК МІЖ КОДОВОЮ ВІДДАЛЮ d_{\min} ТА КОРЕКТУЮЧОЮ ЗДАТНІСТЮ КОДУ

Число помилок q Кодова віддаль d_{\min}

1. що виявляються

А. $2q + 1$

2. що виправляються

В. $q + 1$

Відповідь: 1В, 2А

26. ТЕОРЕМИ ТЕОРІЇ ІНФОРМАЦІЇ	
<u>Теорема</u>	<u>Встановлюють співвідношення</u>
1. Котельникова	А. Між продуктивністю джерела та пропускною здатністю інформаційного каналу
2. Шеннона	В. Між максимальною частотою спектра сигналу та інтервалом дискретизації
3. Хінчина–Вінера	С. Між функцією кореляції та спектральною густиною потужності випадкового сигналу
Відповідь: 1В, 2А, 3С	

27. КІЛЬКІСНА МІРА ЗАВАДОСТІЙКОСТІ ПРИЙМАННЯ	
<u>Сигнали</u>	<u>Величина</u>
1. Аналогові	А. Імовірність помилки
	В. Відношення сигнал/шум
2. Цифрові	С. Коефіцієнт помилок
	Д. Середньоквадратична похибка
Відповідь: 1ВD, 2АС	

28. ПРИ ПОРІВНЯННІ ХАРАКТЕРИСТИК ПРИСТРОЇВ ІНФОРМАЦІЙНИХ СИСТЕМ ВИКОРИСТОВУЮТЬСЯ	
<u>Характеристики</u>	<u>Пристрій системи</u>
1. Енергетичні втрати	А. Кодек коректуючого коду
2. Виграш демодулятора	В. Модем сигналів аналогових модуляцій
3. Енергетичний виграш кодування	С. Модем сигналів цифрових модуляцій
Відповідь: 1С, 2В, 3А	

29. РОЗРАХУНКОВІ ФОРМУЛИ ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНИХ СИСТЕМ	
<u>Назва</u>	<u>Формула</u>
1. Інформаційна η	А. R_k / C
2. Енергетична β	В. R_k / F
3. Частотна γ	С. R_k / ρ_0
Відповідь: 1А, 2С, 3В	

30. ОСНОВНІ СПІВВІДНОШЕННЯ ТЕОРІЇ ІНФОРМАЦІЇ	
<u>Поняття</u>	<u>Формула</u>
1. Кількість інформації	А. $-\sum p(a_i) \log_2 p(a_i)$
2. Ентропія джерела	В. $-\log_2 p(a_i)$
3. Пропускна здатність інформаційного каналу	С. $-F_k \log_2 (1 + P_c / P_{\text{ш}})$
$p(a_i)$ – імовірність сигналу, F_k – смуга частот каналу, $P_c / P_{\text{ш}}$ – відношення сигнал/шум	
Відповідь: 1В, 2А, 3С	

Доповніть твердження, записавши пропущене слово:

31. НЕПЕРЕРВНИЙ КАНАЛ, В ЯКОМУ ДІЄ ЗАВАДА З НОРМАЛЬНИМ РОЗПОДІЛОМ ІМОВІРНОСТЕЙ, ДІСТАВ НАЗВУ – _____ КАНАЛ
Відповідь: гауссів

32. ДВІЙКОВИЙ КАНАЛ, В ЯКОМУ ЙМОВІРНОСТІ ПОМИЛОК СИМВОЛІВ 0 та 1 ОДНАКОВІ, ДІСТАВ НАЗВУ – ДВІЙКОВИЙ _____ КАНАЛ

Відповідь: симетричний

Запишіть номери правильних, на Ваш погляд, відповідей:

33. В АЦП, ЯК МІНІМУМ, ЗАСТОСОВУЮТЬСЯ БЛОКИ

1. фільтр нижніх частот 2. кодер 3. декодер
4. дискретизатор 5. квантувач

Відповідь: 2, 4, 5

34. У ЦАП, ЯК МІНІМУМ, ЗАСТОСОВУЮТЬСЯ БЛОКИ

1. фільтр нижніх частот 2. кодер 3. декодер
4. дискретизатор 5. квантувач

Відповідь: 1, 3

35. ПРИ ВИКОРИСТАННІ В КАНАЛІ ЗАВАДОСТІЙКОГО КОДУВАННЯ

1. зменшується необхідна смуга частот каналу
2. зменшується ймовірність помилки переданого цифрового сигналу
3. збільшується необхідна смуга частот каналу

Відповідь: 2, 3

36. КОДУ (6,5) З ПАРНИМ ЧИСЛОМ ОДИНИЦЬ НАЛЕЖАТЬ КОМБІНАЦІЇ

1. 110010 2. 110011
3. 101010 4. 011101

Відповідь: 2, 4

37. КОД ШЕННОНА-ФАНО Є

1. простий 2. коректуючий 3. ефективний

Відповідь: 1,3

38. КОД ХАФФМЕНА Є

1. простий 2. коректуючий 3. ефективний

Відповідь: 1,3

Д.2. Завдання для поглибленого вивчення програми модуля

ЗАВДАННЯ 1

1. Маємо Марківське джерело з матрицею перехідних ймовірностей

$$3. P = \begin{bmatrix} 1/4 & 0 & 3/4 \\ 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

who chatters to you will chatter about you

ЗАВДАННЯ 2

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/3 & 0 & 2/3 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

ехал грека через реку, видит грека в реке рак

ЗАВДАННЯ 3

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/4 & 0 & 3/4 \\ 0 & 2/3 & 1/3 \\ 0 & 1/3 & 1/3 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

сунул Грека руку в реку, рак за руку Греку цап!

ЗАВДАННЯ 4

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/4 & 3/4 & 0 \\ 0 & 1/4 & 3/4 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

шел козел с косой козой, шла коза с босым козлом

ЗАВДАННЯ 5

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 0 & 1/4 & 3/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

либо дождик, либо снег, либо будет, либо нет

ЗАВДАННЯ 6

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/3 & 1/2 & 1/6 \\ 0 & 1/4 & 3/4 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

на острую косу много и покосу! покоси-ка, коса!

ЗАВДАННЯ 7

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/3 & 1/2 & 1/6 \\ 0 & 3/4 & 1/4 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

два щенка щeka к щеке грызли щетку в уголке

ЗАВДАННЯ 8

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$4. P = \begin{bmatrix} 1/4 & 3/4 & 0 \\ 0 & 1/4 & 3/4 \\ 3/4 & 0 & 1/4 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

кукушка кукушонку купила капюшон, как в капюшоне он смешон!

ЗАВДАННЯ 9

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$5. P = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 0 & 1/2 & 1/2 \\ 3/4 & 0 & 1/4 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

корабли лавировали, лавировали, да не вылавировали!

ЗАВДАННЯ 10

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \\ 3/4 & 0 & 1/4 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

на дворе трава, на траве дрова. Не руби дрова на траве двора!

ЗАВДАННЯ 11

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/3 & 0 & 2/3 \\ 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

не узнавай друга в три дня, узнавай в три года.

ЗАВДАННЯ 12

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/4 & 3/4 & 0 \\ 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

better late than never but better never late

ЗАВДАННЯ 13

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/4 & 0 & 3/4 \\ 0 & 1/2 & 1/2 \\ 1/6 & 1/3 & 1/2 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

men make houses but women make homes

ЗАВДАННЯ 14

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/4 & 3/4 & 0 \\ 0 & 1/6 & 5/6 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

кукушка хвалит петуха за то, что хвалит он кукушку

ЗАВДАННЯ 15

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 0 & 1/4 & 3/4 \\ 1/8 & 1/4 & 5/8 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

рапортовал, да не дорапортовал, а стал дорапортовывать – зарпортовался

ЗАВДАННЯ 16

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/3 & 1/2 & 1/6 \\ 0 & 1/4 & 3/4 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

четыре чертенка чертили черными чернилами чертеж

ЗАВДАННЯ 17

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/3 & 1/2 & 1/6 \\ 0 & 1/4 & 3/4 \\ 1/2 & 1/4 & 1/4 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

can you can a can as a canner can can a can?

ЗАВДАННЯ 18

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/4 & 3/4 & 0 \\ 0 & 1/4 & 3/4 \\ 5/8 & 0 & 3/8 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздалегідь відомих частотах появи літер.

early to bed and early to rise makes a man wise

ЗАВДАННЯ 19

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 0 & 1/2 & 1/2 \\ 1/8 & 0 & 7/8 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздальгідь відомих частотах появи літер.

от умного научишься, от глупого разучишься

ЗАВДАННЯ 20

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \\ 1/4 & 0 & 3/4 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздальгідь відомих частотах появи літер.

do not trouble trouble until trouble troubles you

ЗАВДАННЯ 21

1. Маємо Марківське джерело з матрицею перехідних імовірностей

$$P = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \\ 1/16 & 0 & 15/16 \end{bmatrix}.$$

Знайти $H(X)$, $H(X | X^\infty)$, $H_2(X)$. Побудувати коди Хаффмена для ансамблів X , X^2 . Указати найкращий алгоритм кодування для даного джерела.

2. Визначити частоти появи літер у приказці, побудувати для заданих частот код Хаффмена, знайти середню довжину кодових слів, визначити витрати на передачу приказки при заздальгідь відомих частотах появи літер.

не имей сто рублей, а имей сто друзей

Завдання для самостійної роботи з розділу "Арифметичне кодування"

1. Арифметичне кодування використане для кодування послідовності довжини 5 на виході двійкового постійного джерела з імовірністю одиниці 0,4. Кодове слово на виході арифметичного кодера має вид **101010**. Знайти послідовність на вході кодера.

23. Арифметичне кодування використане для кодування послідовності довжини 5 на виході двійкового постійного джерела імовірністю одиниці $1/3$. Кодове слово на виході арифметичного кодера має вид **1011100**. Знайти послідовність на вході кодера.

ЛІТЕРАТУРА

1. C. E. Shannon: A mathematical theory of communication Bell Sys. Tech. J., vol 27, 1948, S.379-423, (Имеется перевод: Шеннон К. Математическая теория связи.-В сб.: Работы по теории информации и кибернетики. – М.: ИЛ, 1963.
2. C. Berrou, A. Glavieux: "Near Optimum Error-Correcting Coding and Decoding: Turbo Codes."IEEE Trans. Commun. Vol. 44, 1996, S.1261-1270.
3. C. Berrou, A. Glavieux, P.Thitimajshima: "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes."Proc. IEEE Int. Conf. Commun., May 1993, S. 1064-1070.
4. R. E. Blahut: Principles and practice of information theory, Reading, Mass.: Addison-Wesley Publishing Company, 1987.
5. R. E. Blahut: Principles and practice of error control coding, Addison-Wesley Publishing Company, 1984 (Имеется перевод: Блей-хут. Р. Теория и практика кодов, контролирующих ошибки. – М.: Мир, 1986.
6. D. A. Huffman: "A method for the Construction of Minimum Redundancy Codes."Proc. IRE, vol. 40,1952,S.1098-1101, (Имеется перевод: Хаффман Д. А. Методы построения кодов с минимальной избыточностью. – Кибернетический сб. вып. 3. – М.: ИЛ, 1961.
7. S. Lin, D. J. Costello: Error control coding: Fundamentals and Applications, Englewood Cliffs, NJ,: Prentice- Hall Inc., 1983.
8. F. J. MacWilliams, N. J. Sloane: The Theory of Error-Correcting Codes. Amsterdam: North-Holland, 1977, (Имеется перевод: Мак-Вильямс Ф. Дж., Слоэн Н. Дж. А. Теория кодов исправляющих ошибки. – М.: Связь, 1979.
9. N. Wiener: Cybernetics or Control and Communication in the Animal and the Mashine. Paris: Hermann, 1948.
10. R. G. Gallager: Information Theory and Reliable Communication. New York: John Willey and Sons Inc. 1968, (Имеется перевод: Галагер Р. Теория информации и надежная связь. – М.: Сов. радио, 1974.
11. W. W. Peterson, E. J. Weldon: Error- Correcting Codes. 2nd ed. -Cambridge (Mass.): MIT Press, 1971 (Имеется перевод: Питерсен У., Уэлдон Э. Коды, исправляющие ошибки. – М.: Мир, 1976.
12. G. Ungerbock: Channel Coding with Multilevel/phase Signals. IEEE Trans. Inform. Theory, IT-28,1982, S.55-67.
13. J. G.Proakis: Digital Communications. – New York: McGraw-Hill, 2000.
14. I. Chisar, J. Kerner: Information theory, Coding theorems for discrete memoryless systems. Akademiai Budapest, 1981 (Имеется перевод: Чисар И., Кернер Я. Теория информации, теоремы кодирования для дискретных систем без памяти. – М.: Мир, 1985.
15. W. Feller: An introduction to probabilistic theory and its application. John Willey k. Sons, NY, 1970 (Имеется перевод: Феллер В. Введение в теорию вероятностей и ее приложения. – М.: Мир, 1984.

16. J. M. Wozencraft, I. Jacobs: Principles of communication engineering. John Willey & Sons, NY, 1965 (Имеется перевод: Возенкрафт Дж., Джекобс И. Теоретические основы техники связи. – М.: Мир, 1969).

17. J. J. Stiffler: Theory of synchronous communications. NJ: Prentice Hall Inc., 1971, (Имеется перевод: Стиффлер Дж. Теория синхронной связи. – М.: Связь, 1978).

18. G. C. Klark, J. B. Cain: Error correction coding for digital communication, Plenum press, 1982 (Имеется перевод: Кларк Дж., Кейн Дж. Кодирование с исправлением ошибок в системах цифровой связи. – М.: Радио и связь, 1987).

19. Д. Сэломон: Практическое руководство по методам сжатия данных. – М.: Техносфера, 2003.

20. E. R. Berlekamp: Algebraic Coding Theory. NY, McGraw-Hill, 1968 (Имеется перевод: Берлекэмп Э. Алгебраическая теория кодирования. – М.: Мир, 1971).

21. G. D. Jr. Forney: Concatenated Codes, Cambridge, Mass., MIT Press, 1963 (Имеется перевод: Форни Д. Каскадные коды. – М.: Мир, 1970).

