

Практична робота № 3

Дискретні джерела інформації

Мета роботи: Вивчити характеристики дискретних джерел інформації, опанувати програмні засоби їх розрахунку в MATLAB.

1.1. Теоретичні відомості

Математична модель дискретного джерела визначається статистичними характеристиками інформаційних послідовностей, що ним виробляються.

Нехай джерело з алфавітом $A = \{a_1, a_2, \dots, a_m\}$ генерує послідовність символів, причому в k -й момент часу (на k -му кроці) на виході джерела спостерігається деякий символ a_i , так що «фрагмент» довжиною $L + 1$ символів з усієї послідовності має вигляд

$$\dots, a_{i_1}, a_{i_2}, \dots, a_{i_L}, a_i, \dots \quad (3.1)$$

та описується спільним розподілом ймовірностей

$$p(a_{i_1}, a_{i_2}, \dots, a_{i_L}, a_i) = p(a_{i_1})p(a_{i_2}|a_{i_1}) \dots p(a_i|a_{i_1}, a_{i_2}, \dots, a_{i_L}), \quad (3.2)$$

де $p(a_{i_1})$ – ймовірність появи символу a_{i_1} ; $p(a_{i_2}|a_{i_1})$ – умовна ймовірність появи символу a_{i_2} після символу a_{i_1} ; $p(a_i|a_{i_1}, a_{i_2}, \dots, a_{i_L})$ – умовна ймовірність появи символу a_i за умови, що попередніми символами були символи $a_{i_1}, a_{i_2}, \dots, a_{i_L}$ у зазначеному порядку.

Дискретне джерело є стаціонарним, якщо спільний розподіл ймовірностей у виразі (3.2) за будь-яких значень L не залежить від вибору початкової точки (кроку k) відліку часу.

Стаціонарне дискретне джерело є ергодичним, якщо його ймовірнісні характеристики, отримані при дослідженні однієї, досить довгої послідовності (теоретично нескінченної довжини) з ймовірністю, близькою до одиниці (теоретично рівної одиниці), збігаються з його статистичними характеристиками (одержуваними за ансамблем всіх реалізацій, що виробляються джерелом).

Дискретне джерело з пам'яттю L -го порядку — це дискретне джерело, у якого статистичний зв'язок між символами послідовності поширюється на $L + 1$ символів і для будь-якої послідовності (3.1) описується спільним розподілом ймовірностей (3.2).

Ентропія стаціонарного ергодичного дискретного джерела з пам'яттю L -го порядку знаходиться за формулою

$$H_{L+1}(A) = - \sum_{i_1=1}^m \sum_{i_2=1}^m \dots \sum_{i_L=1}^m \sum_{i=1}^m p(a_{i_1}, a_{i_2}, \dots, a_{i_L}, a_i) \times \log p(a_i|a_{i_1}, a_{i_2}, \dots, a_{i_L}) \quad (3.3)$$

Залежно від величини порядку пам'яті джерела L виходять різні окремі випадки.

1. Дискретне джерело без пам'яті (з пам'яттю 0-го порядку) – це дискретне джерело, у якого всі символи для будь-якої послідовності (3.1) статистично незалежні і спільний розподіл ймовірностей (3.2) дорівнює добутку ймовірностей появи символів

$$p(a_{i_1}, a_{i_2}, \dots, a_{i_L}, a_i) = p(a_{i_1})p(a_{i_2}) \dots p(a_{i_L})p(a_i). \quad (3.4)$$

Таким чином, це джерело повністю задане, якщо заданий вектор (рядок) $\mathbf{p}$ розподілу ймовірностей символів a_i

$$\mathbf{p} = (p(a_1) \quad p(a_2) \quad \dots \quad p(a_m)). \quad (3.5)$$

При $L = 0$ підстановка виразу (3.4) (3.3) з урахуванням того, що умовна ймовірність $p(a_i|a_{i_1}, a_{i_2}, \dots, a_{i_L}) = p(a_i)$ і суми

$$\sum_{i_1}^m p(a_{i_1}) = \sum_{i_2}^m p(a_{i_2}) = \dots = \sum_{i_L}^m p(a_{i_L}) = 1,$$

дає формулу для ентропії джерела із незалежними символами:

$$H_1(A) = - \sum_{i=1}^m p(a_i) \log p(a_i). \quad (3.6)$$

Іноді це джерело називають дискретним джерелом Бернуллі.

Доведено, що будь-яке стаціонарне джерело без пам'яті також є ергодичним.

2. Стаціонарне (і ергодичне) джерело без пам'яті та з рівноймовірними символами. Таке джерело повністю задане, якщо заданий об'єм m його алфавіту. Ентропія джерела визначається виразом

$$H_0(A) = \log m. \quad (3.7)$$

Іноді це джерело називають джерелом Хартлі.

Інші окремі випадки отримуються при відповідних значеннях $L > 0$.

3. Стаціонарне ергодичне дискретне джерело з пам'яттю 1-го порядку отримується із загального випадку при $L = 1$. Дане джерело повністю заданий, якщо відомі або вектор ймовірностей символів a_i

$$\mathbf{p} = (p(a_1) \quad p(a_2) \quad \dots \quad p(a_m)). \quad (3.8)$$

та матриця умовних ймовірностей

$$\mathbf{\Pi} = \begin{pmatrix} p(a_1|a_1) & p(a_2|a_1) & \dots & p(a_m|a_1) \\ p(a_1|a_2) & p(a_2|a_2) & \dots & p(a_m|a_2) \\ \dots & \dots & \dots & \dots \\ p(a_1|a_m) & p(a_2|a_m) & \dots & p(a_m|a_m) \end{pmatrix}. \quad (3.9)$$

або матриця спільних ймовірностей

$$\mathbf{P} = \begin{pmatrix} p(a_1, a_1) & p(a_1, a_2) & \cdots & p(a_1, a_m) \\ p(a_2, a_1) & p(a_2, a_2) & \cdots & p(a_2, a_m) \\ \cdots & \cdots & \cdots & \cdots \\ p(a_m, a_1) & p(a_m, a_2) & \cdots & p(a_m, a_m) \end{pmatrix}, \quad (3.10)$$

де $p(a_i)$ – ймовірність появи символу a_i ; $p(a_j|a_i)$ – умовна ймовірність появи символу a_j після символу a_i ; $p(a_i, a_j)$ – спільна ймовірність появи двійки символів a_i та a_j . Ентропія цього джерела згідно (2.9) визначається виразом

$$\begin{aligned} H_2(A) &= -\sum_{i=1}^m \sum_{j=1}^n p(a_i, a_j) \log p(a_j|a_i) = \\ &= -\sum_{i=1}^m p(a_i) \sum_{j=1}^m p(a_j|a_i) \log p(a_j|a_i), \end{aligned} \quad (3.11)$$

Ентропії, що визначаються співвідношеннями (3.3), (3.7), (3.6), (3.11) задовольняють нерівності

$$H_0(A) > H_1(A) > H_2(A) > \dots > H_{L+1}(A). \quad (3.12)$$

з яких випливає, що зі зростанням «протяжності» пам'яті між символами ентропія джерела зменшується.

Інформаційні властивості дискретних джерел оцінюються такими основними характеристиками:

1. Коефіцієнт надмірності джерела з ентропією $H(A)$ показує, яка частка максимально можливої для даного обсягу алфавіту m інформації не виробляється джерелом і визначається виразом

$$\chi = 1 - \frac{H(A)}{H_0(A)} = 1 - \frac{H(A)}{\log m}. \quad (3.13)$$

Значення коефіцієнта χ лежить у межах $0 \leq \chi \leq 1$ і є безрозмірною величиною.

2. Продуктивність джерела з ентропією $H(A)$ визначає середню кількість інформації, що створюється джерелом в одиницю часу, і знаходиться за формулою

$$I'(A) = \frac{1}{\bar{\tau}} H(A) = \bar{V} H(A), \quad (3.14)$$

де $\bar{\tau}$ – середня тривалість одного символу;

$$\bar{V} = \frac{1}{\bar{\tau}}, \quad (3.15)$$

$\bar{V}$ – середня швидкість передачі символів джерела.

Величина $\bar{\tau}$ для джерела незалежних символів з різними тривалостями τ_i знаходиться за формулою математичного очікування

$$\bar{\tau} = \sum_{i=1}^m \tau_i p(a_i). \quad (3.16)$$

Якщо символи рівноймовірні і в (3.16) $p(a_i) = \frac{1}{m}$, то $\bar{\tau} = \frac{1}{m} \sum_{i=1}^m \tau_i$.

Якщо тривалості всіх символів однакові та рівні τ , то відповідно до (3.16) $\bar{\tau} = \tau$.

Величина $\bar{V}$ визначає, скільки елементарних символів (сигналів) передається в одиницю часу і називається також технічною швидкістю передачі або швидкістю модуляції (маніпуляції). Швидкість модуляції $\bar{V}$ вимірюється у бодах. 1 Бод – це швидкість, коли за одну секунду передається один символ. При цьому очевидно, що продуктивність $I'(A)$ вимірюється у величинах біт за секунду.

3.2. Розрахункові завдання

Завдання 3.1.

Імовірності появи символів джерела з алфавітом A становлять: $p(a_1) = 0,5$; $p(a_2) = 0,25$; $p(a_3) = p(a_4) = 0,125$. Обчислити: ентропію та надмірність джерела, середню тривалість символу, швидкість модуляції та продуктивність джерела, якщо тривалості символів $\tau_1 = 0,1$; $\tau_2 = 0,2$; $\tau_3 = 0,3$; $\tau_4 = 0,4$ мілісекунд.

```
%Script_3_1
%Розрахунок характеристик дискретного джерела без
пам'яті
clear all; close all; clc
%Завдання анонімної функції
%-----
fun=@(x) -x.*log2(x+(x==0));
%-----
%Завдання ймовірностей символів
p=[.5 .25 .125 .125];
disp('Ймовірності символів:')
disp(p)
%Сума ймовірностей символів
s=sum(p);
disp('Сума ймовірностей:')
disp(s)
%Визначення обсягу алфавіту
m=length(p);
disp('Обсяг алфавіту:')
disp(m)
```

```

%Розрахунок ентропії
H=sum(fun(p));
disp('Ентропія джерела, біт/символ:')
disp(H)
%Розрахунок максимальної ентропії
H0=log2(m);
disp('Максимальна ентропія, біт/символ:')
disp(H0)
%Розрахунок коефіцієнта надмірності
x=1-H/H0;
disp('Коефіцієнт надмірності:')
disp(x)
%Визначення тривалостей символів
Ti=[.1 .2 .3 .4]*10^(-3);
disp('Тривалості символів, с:')
disp(Ti)
%Розрахунок середньої тривалості символів,
швидкості модуляції
%і продуктивності
T=sum(Ti.*p);
disp('Середня тривалість символів, с:')
disp(T)
V=1/T;
disp('Швидкість модуляції, Бод:')
disp(V)
I=V*H;
disp('Продуктивність, біт/с:')
disp(I)

```

Завдання 3.2. Дискретне джерело з пам'яттю 1–го порядку задано матрицею спільних ймовірностей символів

$$\mathbf{P} = \begin{pmatrix} p(a_1, a_1) & p(a_1, a_2) & p(a_1, a_3) \\ p(a_2, a_1) & p(a_2, a_2) & p(a_2, a_3) \\ p(a_3, a_1) & p(a_3, a_2) & p(a_3, a_3) \end{pmatrix} = \begin{pmatrix} 0,04 & 0,08 & 0,08 \\ 0,1 & 0,4 & 0 \\ 0,06 & 0,02 & 0,22 \end{pmatrix}.$$

Визначити: ентропію та надмірність джерела; продуктивність джерела, якщо тривалості символів однакові і становлять $\tau = 800$ мкс.

```

%Script_3_2
%Розрахунок показників дискретного джерела з
пам'яттю 1-го порядку
%при заданій матриці спільних ймовірностей P
clear all; close all; clc
%Завдання анонімної функції

```

```

%-----
fun=@(x) -x.*log2(x+(x==0));
%-----
%Завдання матриці спільних ймовірностей
P=[.04 .08 0.08;.1 .4 0;0.06 0.02 .22];
disp('Матриця спільних ймовірностей P:')
disp(P)
%Розрахунок суми елементів матриці спільних
ймовірностей
s=sum(sum(P));
disp('Сума ймовірностей:')
disp(s)
%Визначення обсягу алфавіту
[m, n]=size(P);
disp('Обсяг алфавіту:')
disp(m)
%Перевірка рівності m = n
if m~=n
disp('Матриця має бути квадратною!')
end
%Розрахунок ймовірностей символів
p=sum(P,2);
disp('Ймовірності символів алфавіту:')
disp(p)
%Розрахунок елементів матриці умовних ймовірностей
П
PI=bsxfun(@rdivide,P,p(:));
disp('Матриця умовних ймовірностей П:')
disp(PI)
%Розрахунок ентропії джерела
H=sum(p.*sum(fun(PI),2));
disp('Ентропія джерела, біт/символ:')
disp(H)
%Розрахунок коефіцієнта надмірності
x=1-H/log2(m);
disp('Коефіцієнт надмірності:')
disp(x)
%Визначення тривалості символів, с
T=800*10^(-6);
%Розрахунок продуктивності джерела
I=H/T;
disp('Продуктивність, біт/с:')
disp(I)

```

Завдання 3.3. Дискретне джерело з пам'яттю 1-го порядку задано матрицею спільних ймовірностей символів

$$\mathbf{P} = \begin{pmatrix} 0,08 & 0,1 & 0,12 \\ 0,02 & 0,05 & 0,18 \\ 0,2 & 0,1 & 0,15 \end{pmatrix}.$$

Визначити: ентропію та надмірність джерела; продуктивність джерела, якщо тривалості символів $\tau_1 = 1/8$ мс; $\tau_2 = 1/4$ мс; $\tau_3 = 1/3$ мс.

Вказівка: використовувати Script_3_1, Script_3_2.

Завдання 3.4. Дискретне джерело з пам'яттю 1-го порядку задано матрицею спільних ймовірностей символів

$$\mathbf{P} = \begin{pmatrix} 0,1 & 0,1 & 0,05 \\ 0,05 & 0,2 & 0,15 \\ 0,1 & 0,1 & 0,15 \end{pmatrix}.$$

Визначити: ентропію та надмірність джерела; продуктивність джерела, якщо тривалості символів $\tau_1 = 1/2$ мс; $\tau_2 = 1/4$ мс; $\tau_3 = 1/4$ мс.

Вказівка: використовувати Script_3_1, Script_3_2.

Завдання 3.5. Дискретне джерело з пам'яттю 1-го порядку задано матрицею умовних ймовірностей символів

$$\mathbf{\Pi} = \begin{pmatrix} p(a_1|a_1) & p(a_2|a_1) & p(a_3|a_1) \\ p(a_1|a_2) & p(a_2|a_2) & p(a_3|a_2) \\ p(a_1|a_3) & p(a_2|a_3) & p(a_3|a_3) \end{pmatrix} = \begin{pmatrix} 0 & 0,4 & 0,6 \\ 0,4 & 0 & 0,6 \\ 0,075 & 0,075 & 0,85 \end{pmatrix}..$$

Визначити: ентропію, надмірність джерела та продуктивність джерела, якщо імовірності появи символів становлять: $p(a_1) = p(a_2) = 0,1$; $p(a_3) = 0,8$, а їх тривалості становлять $\tau_1 = \tau_2 = 0,4$ мс; $\tau_3 = 0,1$ мс.

```
%Script_3_5
%Розрахунок характеристик дискретного джерела з
пам'яттю 1-го порядку
%при заданій матриці умовних ймовірностей П
%i векторі ймовірностей символів\
clear all; close all; clc
%Завдання анонімної функції
%-----
fun=@(x) -x.*log2(x+(x==0));
%-----
%Завдання матриці умовних ймовірностей
PI=[0.4 .6;.4 0 .6;.075 .075 .85];
```

```

disp('Матриця умовних ймовірностей П:')
disp(PI)
%Знадання вектора ймовірностей символів
p=[.1; .1; .8];
disp('Ймовірності символів:')
disp(p)
%Визначення обсягу алфавіту
[m, n]=size(PI);
disp('Обсяг алфавіту:')
disp(m)
%Перевірка рівності m = n
if m~=n
disp('Матриця має бути квадратною!')
end
%Розрахунок ентропії джерела
H=sum(p.*sum(fun(PU),2));
disp('Ентропія джерела, біт/символ:')
disp(H)
%Розрахунок коефіцієнта надмірності
x=1-H/log2(m);
disp('Коефіцієнт надмірності:')
disp(x)
%Визначення тривалостей символів
Ti=[.4; .4; .1]*10^(-3);
disp('Тривалості символів, с:')
disp(Ti)
%Розрахунок середньої тривалості символів,
швидкості модуляції
%та продуктивності
T=sum(Ti.*p);
disp('Середня тривалість символів, с:')
disp(T)
V= 1/T;
disp('Швидкість модуляції, Бод:')
disp(V)
I=V*H;
disp('Продуктивність, біт/с:')
disp(I)

```

Завдання 3.6. Дискретне джерело з пам'яттю 1–го порядку задано матрицею умовних ймовірностей символів

$$\mathbf{P} = \begin{pmatrix} 0,5 & 0,5 & 0 \\ 0 & 0,5 & 0,5 \\ 1 & 0 & 0 \end{pmatrix}..$$

Розрахувати ентропію, надмірність та продуктивність джерела, якщо імовірності появи символів становлять: $p(a_1) = p(a_2) = 0,4$; $p(a_3) = 0,2$, а їх тривалості становлять $\tau_1 = \tau_2 = 0,45$ мс; $\tau_3 = 0,2$ мс.

Вказівка: використовувати Script_3_5.

Завдання 3.7. Розрахувати ентропію, надмірність та продуктивність дискретного джерела з пам'яттю 1-го порядку, якщо матриця умовних ймовірностей символів має вигляд

$$\mathbf{P} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}..$$

імовірності появи символів становлять: $p(a_1) = 3/11$; $p(a_2) = p(a_3) = 4/11$, а їх тривалості становлять $\tau_1 = 0,2$ мс; $\tau_2 = \tau_3 = 0,1$ мс.

Вказівка: використовувати Script_3_5.

3.3. Звіт та контрольні питання

Звіт з лабораторної роботи оформляється у вигляді документа MS Word і повинен містити:

1. Назва лабораторної роботи (шрифт Times New Roman).
2. Результати виконання всіх розрахункових завдань, включаючи лістинги m-файлів, що копіюються з вікна **Editor**, результати їх виконання, що копіюються з вікна **Command Window** (шрифт Courier New), а також отримані графіки, що копіюються з вікна **Figure** за командою **Edit | Copy Figure**.

Захист лабораторної роботи проводиться на основі поданого звіту та відповідей на контрольні питання з наступного списку:

1. Запишіть вираз для спільної ймовірності послідовності з $L + 1$ символів, що виробляється дискретним джерелом з алфавітом $A = \{a_1, a_2, \dots, a_m\}$ на k -му кроці.
2. Що таке дискретне стаціонарне джерело?
3. Дайте визначення дискретного стаціонарного ергодичного джерела.
4. Що таке дискретне джерело із пам'яттю L -го порядку?
5. Запишіть вираз для ентропії стаціонарного ергодичного дискретного джерела з пам'яттю L -го порядку.

6. Запишіть вираз для ентропії стаціонарного ергодичного дискретного джерела без пам'яті.

7. Запишіть вираз для ентропії стаціонарного ергодичного дискретного джерела з пам'яттю 1-го порядку.

8. Що таке коефіцієнт надмірності дискретного джерела? У яких межах він змінюється?

9. Дайте визначення продуктивності дискретного джерела інформації.

10. Що таке середня швидкість передачі символів? У яких одиницях вона вимірюється?