

Верица Бакева Смиљкова

Александра Поповска-Митровиќ

Теорија на информации со дигитални комуникации

Универзитет „Св. Кирил и Методиј“ во Скопје

УНИВЕРЗИТЕТ „СВ. КИРИЛ И МЕТОДИЈ“ ВО СКОПЈЕ

ВЕРИЦА БАКЕВА СМИЉКОВА
АЛЕКСАНДРА ПОПОВСКА-МИТРОВИЌ

ТЕОРИЈА НА ИНФОРМАЦИИ СО ДИГИТАЛНИ
КОМУНИКАЦИИ

СКОПЈЕ 2025

Издавач:

Универзитет „Св. Кирил и Методиј“ во Скопје
Бул. „Гоце Делчев“ бр. 9, 1000 Скопје
www.ukim.edu.mk

Уредник за издавачка дејност на УКИМ:

проф. д-р Биљана Ангелова, ректор

Уредник на публикацијата:

проф. д-р Верица Бакева Смиљкова

проф. д-р Александра Поповска-Митровиќ

Факултет за информатички науки и компјутерско инженерство – Скопје

Рецензенти:

1. проф. д-р Смиле Марковски

2. проф. д-р Наташа Илиевска

Техничка обработка:

проф. д-р Верица Бакева Смиљкова, проф. д-р Александра Поповска-Митровиќ

Лектура на македонски јазик: м-р Соња Попоска

Илустратор:

проф. д-р Верица Бакева Смиљкова, проф. д-р Александра Поповска-Митровиќ

CIP - Каталогизација во публикација
Национална и универзитетска библиотека "Св. Климент Охридски", Скопје

519.72:621.391.037.372(075.8)

БАКЕВА Смиљкова, Верица
Теорија на информации со дигитални комуникации [Електронски извор] / Верица Бакева
Смиљкова, Александра Поповска-Митровиќ ; [илустратор Верица Бакева Смиљкова, Александра
Поповска-Митровиќ]. - Скопје :
Универзитет "Св. Кирил и Методиј", 2025

Начин на пристапување (URL):

<https://www.ukim.edu.mk/e-izdanija/FINKI/Teorija-na-informacii-so-digitalni-komunikacii.pdf>.

- Текст во PDF формат, содржи 201, [9] стр., илустр. - Наслов преземен од екранот. - Опис на изворот
на ден 13.02.2025. - Библиографија: стр.

201

ISBN 978-9989-43-525-6

1. Поповска-Митровиќ, Александра [автор] [илустратор]

а) Теорија на информации – Математичка кибернетика – Комуникациски системи – Дигитални
кодови – Високошколски учебници

COBISS.MK-ID 65263621

Посветено на проф. д-р Магдалена Георгиева

Предговор

Овој учебник е наменет, пред сè, за предметите Теорија на информации со дигитални комуникации и Основи на теорија на информации на прв циклус студии и предметот Применета теорија на информации на втор циклус студии на Факултетот за информатички науки и компјутерско инженерство. Но, учебникот може да се користи и на другите технички факултети каде што се користат избрани поглавја од теоријата на информации.

Учебникот се состои од осум глави, и тоа: Вовед, Ентропија и информација, Својство на асимптотска рамноправност, Вериги на Марков и рата на ентропија на случаен процес, Компресија на податоци, Комуникациски канал, Диференцијална ентропија и Линеарни кодови. На крајот од секоја глава има решени задачи од материјалот во таа глава, како и задачи за самостојна работа на студентите.

Во воведот на овој учебник се дадени некои историски податоци за тоа како е воведен основниот поим во теорија на информации, а тоа е ентропијата. Исто така, претставени се поединечно сите елементи на еден комуникациски систем.

Поимите за ентропија на една случајна променлива, ентропија на случаен вектор, релативна ентропија помеѓу две распределби над исто множество вредности, како и поимот за заемна информација помеѓу две случајни променливи дефинирани се во втората глава. Сите овие поими се однесуваат на дискретни случајни променливи. Разгледани се и некои основни својства на дефинираните ентропии и информација.

Во третата глава е разгледано својството за асимптотска рамномерност кое, во теоријата на информации, е аналог на законот на големите броеви. Ова својство овозможува множеството од сите пораки да се подели на две подмножества: типично множество и неговиот комплемент. Типичното множество ги содржи најверојатните пораки. При кодирање на пораките, на оние кои припаѓаат на типичното множество им се доделуваат пократки кодни зборови, а на останатите подолги кодни зборови. Со тоа, се добива код кој

овозможува компресирање на податоците.

Во четвртата глава, најпрво, е дефиниран случаен процес, како и основните елементи поврзани со него. Потоа, се разгледува специјален случаен процес наречен верига на Марков. Во продолжение, дефинирана е рата (брзина на промена) на ентропијата кога должината на пораките неограничено расте.

Кодовите на изворот, т.е. кодовите кои овозможуваат компресирање на податоците, се разгледуваат во петтата глава. Дадени се неколку алгоритми кои овозможуваат компресија на податоци, како што се: Хуфмановиот код, кодот на Шенон-Фано-Елиас и аритметичките кодови.

Во шестата глава, дефинирани се поимите за дискретен канал и канал без меморија и разгледани се неколку видови канали: бинарен канал без шум, бинарен канал со непреклопувачки излези, бинарен симетричен канал, канал со бришење и симетричен канал.

Поимот за диференцијална ентропија, како ентропија на случајна променлива од апсолутно-непрекинат тип е дефиниран во седмата глава. Во продолжение е даден посебен осврт на Гаусов канал, како еден од најкористените канали за комуникација.

Во последната глава се разгледува концептот на кодирање и декодирање. Ова се кодови кои се однесуваат на каналот и кои овозможуваат откривање или поправање на грешките кои настануваат при пренос низ каналот. Посебно се разгледани таканаречените линеарни кодови.

За полесно разбирање на темите обработени во учебникот, читателот треба да е запознат со основните поими од теорија на веројатност [11].

Би сакале да ја искажеме нашата посебна благодарност до рецензентите на овој учебник за деталното читање на трудот и корисните забелешки кои се составен дел од овој учебник.

Од авторите

Содржина

1. Вовед	1
1.1. Историски податоци	1
1.2. Комуникациски систем	3
2. Ентропија и информација	7
2.1. Ентропија на случајна променлива	7
2.2. Заедничка и условна ентропија	11
2.3. Релативна ентропија и информација	14
2.4. Својства на ентропија, релативна ентропија и заемна информација	17
2.5. Решени задачи	23
2.6. Задачи	32
3. Својство за асимптотска рамномерност	35
3.1. Закон на големите броеви	35
3.2. Својство за асимптотска рамномерност	36
3.3. Компресија на податоци	40
3.4. Решени задачи	45
3.5. Задачи	47
4. Вериги на Марков и рата на ентропија на случаен процес	49
4.1. Дефиниција на случаен процес. Стационарност	49
4.2. Вериги на Марков	53
4.3. Рата на ентропија	58
4.3.1. Рата на ентропија на верига на Марков	62
4.4. Решени задачи	63
4.5. Задачи	75
5. Компресија на податоци	79

5.1.	Дефиниција на код на изворот. Видови кодови	79
5.2.	Крафтово неравенство	82
5.3.	Моментални оптимални кодови	87
5.4.	Хафманов код	93
5.5.	Шенон–Фано–Елиас код	104
5.6.	Аритметички кодови	110
5.7.	Решени задачи	115
5.8.	Задачи	122
6.	Комуникациски канал	125
6.1.	Дискретен канал без меморија	126
6.2.	Видови дискретни комуникациски канали без меморија	127
6.2.1.	Бинарен канал без шум	127
6.2.2.	Бинарен канал со непреклопувачки излези	128
6.2.3.	Бинарен симетричен канал	129
6.2.4.	Канал со бришење	131
6.2.5.	Симетричен канал	132
6.3.	Својства на капацитет на канал	135
6.4.	Кодер и декодер на каналот	135
6.5.	Решени задачи	137
6.6.	Задачи	147
7.	Диференцијална ентропија	149
7.1.	Диференцијална ентропија	149
7.2.	Гаусов канал	159
7.2.1.	Капацитет на Гаусов канал	160
7.2.2.	SNR	162
7.2.3.	Веројатност на бит грешка при BPSK модулација	162
7.3.	Решени задачи	166
7.4.	Задачи	167
8.	Линеарни кодови	169
8.1.	Концепт на кодирање и декодирање	169
8.2.	Линеарни блок кодови	177
8.3.	Векторски простор и потпростор	180
8.4.	Конструкција на контролна матрица за даден линеарен блок код	182
8.5.	Алгоритам за корекција на грешки кај линеарен блок код	187
8.6.	Решени задачи	194

<i>СОДРЖИНА</i>	iii
8.7. Задачи	199
Литература	201

Глава 1

Вовед

Теоријата на информации одговара на две фундаментални прашања од теоријата на комуникации:

- Колку е максималната можна компресија на податоци? Одговорот е: ентропијата H .
- Колкава е максималната брзина на пренос низ комуникациски канал? Одговорот е: капацитетот на каналот C .

Од овие причини теоријата на информации се смета како подмножество од теоријата на комуникации, но тоа не е комплетно точно. Теоријата на информации има фундаментален придонес и во многу други науки, како во физика (термодинамика), во компјутерски науки (алгоритамска комплексност, комплексност на Колмогоров), во електроинженерство (теорија на комуникации) и др.

1.1. Историски податоци

Еден од главните проблеми на теоријата на информации е како квантитативно да се изрази информацијата која ја „носи“ одредена порака, односно како математички да се дефинира мерка за информација, која ќе овозможи егзактна анализа на информационите системи. Н. Никуист (H. Nyquist) во 1924 год. и Хартли (R. Hartley) во 1928 год. заклучиле дека мерката за количина информација мора да има логаритамски карактер. Хартли предложил на сигналот, кој се избира од множество од n можни сигнали, да му се придружи информација:

$$I(n) = \log n.$$

Основната причина за ваквата дефиниција е следнава: информацијата која ја носи порака составена од два сигнала, при што едниот се избира од множество со m , а другиот од множество со n можни сигнали, треба да биде еднаква на збирот на информациите кои ги носат сигналите поединечно. Такви парови сигнали има mn , и секој таков пар сигнали поединечно носи информација $I(mn)$. Притоа, треба да важи:

$$I(mn) = I(m) + I(n).$$

Така се наметнува идејата да се земе логаритамската функција, бидејќи

$$\log(mn) = \log m + \log n.$$

Ваквата дефиниција е поткрепена и со следниве констатации:

1. $\log n > 0$, за секој природен број $n > 0$;
2. Ако $m < n$, тогаш $\log m < \log n$.

Значи, информацијата секогаш е позитивна величина. Понатаму, со примање на еден од можните сигнали се отстранува одредена неизвесност која постоела пред примањето на сигналот. Според тоа, $\log n$ може да се интерпретира и како мерка на неизвесност во врска со изборот на една од n можни алтернативи. Интуитивно е прифатливо својството на поголем број алтернативи да одговара поголема неизвесност. Главниот недостаток на Хартлиевата мерка за информација е во тоа што сите сигнали (пораки), од множеството можни сигнали, се третираат рамноправно, т.е. еднакво веројатно, што не е во согласност со реалноста. Клод Шенон (Claude Shannon) воочил дека сигналите и пораките, како и пречките во комуникациските системи, имаат стохастички карактер, така што е разумно да се претпостави дека тие се појавуваат во согласност со одредена распределба на веројатност. Тој предложил сигналот, кој се појавува со веројатност $p_i > 0$, да носи информација

$$I(p_i) = \log \frac{1}{p_i} = -\log p_i > 0.$$

Ако има n можни сигнали, така што $\sum_{i=1}^n p_i = 1$, тогаш просечната (средна) количина информација, која ја носи поединечен сигнал, е:

$$H(p_1, p_2, \dots, p_n) = \sum_{i=1}^n p_i I(p_i) = -\sum_{i=1}^n p_i \log p_i.$$

Да воочиме дека Хартлиевата дефиниција за мерка на информација е специјален случај на Шеноновата, за $p_i = 1/n$. Имено, ако $p_i = 1/n$, тогаш

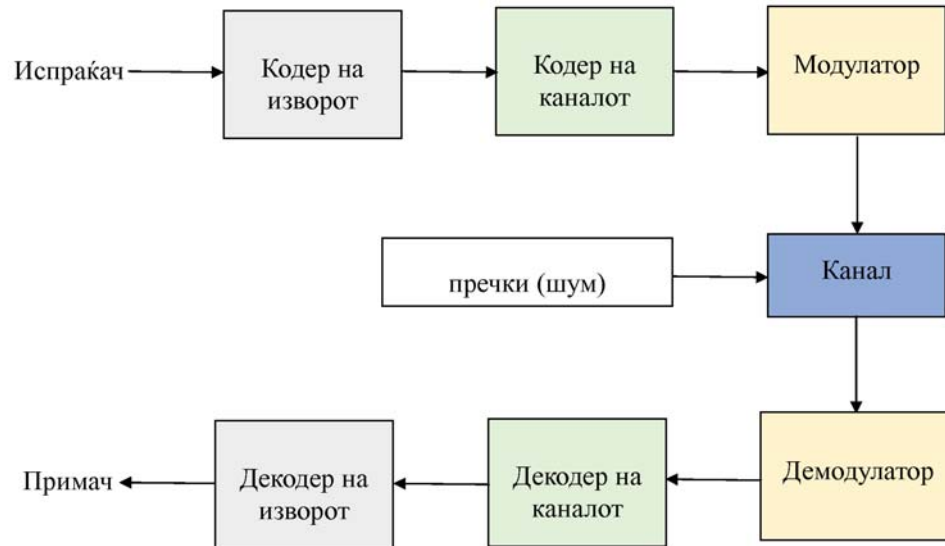
$$\begin{aligned} H(p_1, p_2, \dots, p_n) &= - \sum_{i=1}^n p_i \log p_i \\ &= - \sum_{i=1}^n \frac{1}{n} \log \frac{1}{n} \\ &= -n \frac{1}{n} \log \frac{1}{n} \\ &= \log n. \end{aligned}$$

Во Шеноновата дефиниција за мерка на информација се води сметка за фактот дека сигналите се јавуваат во согласност со некоја распределба на веројатности (p_1, p_2, \dots, p_n) , така што средната информација по сигнал зависи единствено од таа распределба, а не од некои други величини со кои се карактеризираат сигналите. Затоа Шенон, величината $H(p_1, p_2, \dots, p_n)$ ја нарекол ентропија на конечна распределба на веројатности, затоа што таа може да се интерпретира и како неизвесност во поглед на случајниот избор на една од n -те можни вредности.

1.2. Комуникациски систем

Главна цел на теоријата на информации е анализа и проучување на комуникациски системи и наоѓање на математички модел за опишување на истите. На слика 1.1 е претставен блок дијаграм на еден комуникациски систем. Во продолжение ќе ги објасниме сите елементи во овој блок дијаграм.

- *Испраќач* (или *извор*) на информации е човек или машина што генерира информација која треба да се пренесе до примачот.
- *Кодерот на изворот* ги компресира податоците, со што ја намалува големината на податоците кои треба да се пренесат низ каналот. Притоа, се користат алгоритми кои потоа, во декодерот на изворот, ќе овозможат комплетно враќање на компресираните пораки. Целта на компресијата е да се зголеми ефикасноста и брзината на пренос, како и да се намалат неговите трошоци.
- При пренос на податоците низ каналот, најчесто се појавуваат грешки кои се должат на присуството на шум во каналот. *Кодерот на каналот*



Слика 1.1: Комуникациски систем

додава на компресираната порака дополнителни, таканаречени *редундантни* симболи, кои овозможуваат да се откријат или поправат грешките при пренос на пораките низ каналот.

- *Модулаторот* го подготвува сигналот за пренос на одредено растојание. Тој го претвора дигитално кодираниот сигнал во аналоген сигнал доколку сигналот треба да се пренесе преку радиобран. Ако сигналот треба да се пренесе преку светлина, тогаш модулаторот го претвора електричниот сигнал во светлина користејќи дополнително коло на конверторот. Модулаторот, исто така, го засилува сигналот.
- *Каналот* е средство или медиум за пренесување информации од испраќачот до примачот. Одредувањето на најсоодветниот канал или медиум е од клучно значење за ефективноста на комуникацијата. Каналот може да биде електрично спроводлива линија или кабел со оптички влакна или простор. Електричните проводници или линии генерално го пренесуваат сигналот преку електрични или електронски сигнали, оптичките кабли пренесуваат светлина, а низ просторот се пренесуваат радиобранови. Во каналот дејствува шум кој доведува до промена на дел од по-

датоците, така што излезот од каналот не секогаш се совпаѓа со влезот во истиот.

- *Демодулаторот* го демодулира сигналот добиен на излезот од каналот и го враќа во првобитната (дигитална) форма.
- *Декодерот на каналот* ги користи редувантите симболи за да открие (или коригира) грешки настанати при пренос низ каналот. Ако се користат кодови кои откриваат грешки, доколку декодерот открие грешка тогаш бара од испраќачот да ја испрати пораката уште еднаш. Доколку се користат кодови кои поправаат грешки, тогаш декодерот се обидува да ги поправи сите или дел од грешките настанати при пренос низ каналот.
- *Декодерот на изворот* прави инверзен процес од оној на кодерот на изворот. Тој ја декомпресира добиената порака и ја враќа во изворната форма (онаа во која пораката е пратена од испраќачот).

Во овој учебник ќе бидат разгледани сите елементи на еден комуникациски систем, со посебен осврт на кодовите на изворот и на каналот.

Глава 2

Ентропија и информација

Во оваа глава ќе бидат дефинирани поимите за ентропија на случајна променлива, ентропија на случаен вектор, релативна ентропија помеѓу две распределби и заемна информација помеѓу две случајни променливи.

2.1. Ентропија на случајна променлива

Нека X е дискретна случајна променлива со множество вредности R_X и закон на распределба

$$p(x) = P\{X = x\}.$$

Дефиницијата на ентропија на случајната променлива X , дадена во продолжение, комплетно соодветствува на дефиницијата за ентропија на Шенон која е дадена во воведот на овој учебник.

Дефиниција 2.1. *Ентропија* $H(X)$ на дискретна случајна променлива X се дефинира со:

$$H(X) = - \sum_{x \in R_X} p(x) \log p(x).$$

Како што е претходно кажано, ентропијата на една случајна променлива може да се толкува како неизвесност во поглед на генерирање на вредност на случајната променлива или како просечна количина информација што ја носи еден симбол. Логаритамот во дефиницијата за ентропија ќе сметаме дека е со основа 2 и во тој случај единица мерка за ентропијата е бит. Се договараме $0 \cdot \log 0 = 0$, бидејќи

$$\lim_{x \rightarrow 0} x \cdot \log x = 0.$$

Ако основата на логаритамот во дефиницијата за ентропија е b , тогаш ентропијата ќе ја означуваме со $H_b(X)$. Ако логаритмот е природен (со основа e), тогаш ентропијата се мери во нити.

Пример 2.1. Монета се фрла еднаш. Нека настанот A означува дека се појавила глава. Да се определи ентропијата на случајната променлива I_A – индикатор на настанот A .

Решение: Распределбата на I_A е

$$I_A : \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}.$$

За ентропијата на I_A , се добива:

$$H(I_A) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = -\log \frac{1}{2} = 1 \text{ бит.}$$

Овој резултат е очекуван, затоа што вредностите на случајната променлива I_A се 0 и 1, т.е. секоја вредност е еден бит. Оттука, неизвесноста која постои пред да се генерира една вредност на оваа случајна променлива е 1 бит. \square

Од дефиницијата на ентропија произлегуваат следните две основни својства.

Својство 1. $H(X) \geq 0$.

Доказ:
$$H(X) = - \sum_{x \in R_X} p(x) \log p(x) = \sum_{x \in R_X} p(x) \log(1/p(x)).$$

Бидејќи $p(x) \geq 0$ и $\log(1/p(x)) \geq 0$, следува дека $H(X) \geq 0$. \square

Својство 2. $H_b(X) = (\log_b a) H_a(X)$.

Доказ:

$$\begin{aligned} H_b(X) &= - \sum_{x \in R_X} p(x) \log_b p(x) \\ &= - \sum_{x \in R_X} p(x) \frac{\log_a p(x)}{\log_a b} \\ &= \frac{1}{\log_a b} \left(- \sum_{x \in R_X} p(x) \log_a p(x) \right) \\ &= (\log_b a) H_a(X). \end{aligned}$$

\square

Пример 2.2. Нека

$$X : \begin{pmatrix} 0 & 1 \\ p & 1-p \end{pmatrix}.$$

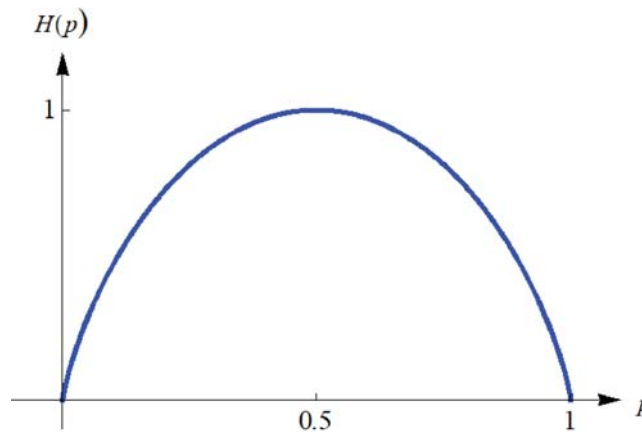
За ентропијата на X , се добива:

$$H(X) = -p \log p - (1-p) \log(1-p).$$

Нека $H(X)$ се разгледува како функција од p , за $0 \leq p \leq 1$, т.е. нека

$$H(p) = -p \log p - (1-p) \log(1-p), \quad 0 \leq p \leq 1.$$

Графикот на оваа функција е даден на слика 2.1.



Слика 2.1

Од графикот се гледа дека ентропијата е испакната функција и таа прима вредност 0, ако $p = 0$ или $p = 1$. Тоа е очекувано, бидејќи во овие два случаја нема случајност, па нема ниту неизвесност. Неизвесноста е најголема, ако $p = 1/2$, што соодветствува на максималната вредност на ентропијата. \square

Пример 2.3. Нека случајната променлива X е зададена со нејзиниот закон на распределба

$$X : \begin{pmatrix} a & b & c & d \\ 1/2 & 1/4 & 1/8 & 1/8 \end{pmatrix}.$$

Ентропијата на X е

$$H(X) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{8} \log \frac{1}{8} - \frac{1}{8} \log \frac{1}{8} = \frac{7}{4} \text{ бита} = 1.75 \text{ бита}.$$

Ги поставуваме правилата на следната игра. Компјутерот генерира една вредност на случајната променлива X . Модераторот ја знае вредноста на X , а играчот треба да ја погоди со поставување на „да–не“ прашања. Во оној момент кога играчот ќе ја погоди генерираната вредност на X , играта завршува. Прашањето е како да се постави стратегијата на поставување на прашања за очекуваниот број на поставени прашања (при поголем број на повторувања на играта) да биде минимален? Една стратегија е следната:

Првото прашање е: „Дали $X = a$?“

- Веројатноста дека играта ќе заврши по првото прашање е еднаква на веројатноста дека одговорот на ова прашање е „да“. Таа изнесува $1/2$.

Ако одговорот е „не“, тогаш се поставува второто прашање: „Дали $X = b$?“

- Веројатноста дека играта ќе заврши после второто прашање е еднаква на веројатноста дека одговорот на ова прашање ќе биде „да“, а таа веројатност е $1/4$.

Ако одговорот и на ова прашање е „не“, тогаш третото и последно прашање е: „Дали $X = c$?“

- Веројатноста дека играта ќе заврши после третото прашање е еднаква на $1/4$. Имено, ако одговорот на третото прашање е „не“, јасно е дека генерираната вредност на X ќе биде d .

Можеме да дефинираме нова случајна променлива Y – број на поставени прашања до погодување на вредноста на X . Тогаш множеството вредности на случајната променлива Y е $R_Y = \{1, 2, 3\}$, а за соодветните веројатности се добива:

$$\begin{aligned} P\{Y = 1\} &= P\{X = a\} = 1/2 \\ P\{Y = 2\} &= P\{X = b\} = 1/4 \\ P\{Y = 3\} &= P\{X \in \{c, d\}\} = 1/4, \end{aligned}$$

т.е.

$$Y : \begin{pmatrix} 1 & 2 & 3 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}.$$

Очекуваниот број на поставени прашања е

$$EY = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} = 1.75 = H(X).$$

Подоцна, во глава 5 ќе покажеме дека очекуваниот број на поставени бинарни прашања за да се погоди вредноста на X е помеѓу $H(X)$ и $H(X) + 1$. \square

2.2. Заедничка и условна ентропија

Во претходното поглавје дефиниравме ентропија на една случајна променлива. Сега, ќе ја прошириме дефиницијата за подреден пар од случајни променливи (X, Y) . Нека X и Y се дискретни случајни променливи со множества вредности R_X и R_Y , соодветно, и нека нивниот заеднички закон на распределба е даден со:

$$p(x, y) = P\{X = x, Y = y\} \geq 0, \quad \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) = 1.$$

Дефиниција 2.2. *Заедничка ентропија* $H(X, Y)$ на случајните променливи X и Y со заеднички закон на распределба $p(x, y)$ се дефинира со:

$$H(X, Y) = - \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log p(x, y).$$

Заедничката ентропија на случајните променливи X и Y може да се интерпретира како мерка за неизвесност за парот вредности (x_i, y_j) на случајниот вектор (X, Y) .

Во продолжение, ќе дефинираме условна ентропија на една случајна променлива при услов дека друга случајна променлива прима фиксна вредност.

Дефиниција 2.3.

i. Нека $p_Y(y|x) = P\{Y = y|X = x\}$ е условната распределба на Y за дадено $X = x$. *Условна ентропија на Y за дадено $X = x$* се дефинира со:

$$H(Y|X = x) = - \sum_{y \in R_Y} p_Y(y|x) \log p_Y(y|x).$$

ii. Нека $p_X(x|y) = P\{X = x|Y = y\}$ е условната распределба на X за дадено $Y = y$. *Условна ентропија на X за дадено $Y = y$* се дефинира со:

$$H(X|Y = y) = - \sum_{x \in R_X} p_X(x|y) \log p_X(x|y).$$

Условната ентропија $H(X|Y = y)$ ја толкуваме како мерка за неизвесност при генерирање на вредноста на случајната променлива X , кога е познато дека случајната променлива Y прима вредност y , т.е. $Y = y$. Аналогно, условната ентропија $H(Y|X = x)$ ја толкуваме како неизвесност при генерирање на вредноста на случајната променлива Y , кога е познато дека $X = x$.

Исто така, ќе дефинираме условна ентропија на една случајна променлива кога е дадена друга случајна променлива. Имено, ќе дефинираме $H(Y|X)$ како очекувана вредност на $H(Y|X = x)$, за $x \in R_X$.

Дефиниција 2.4. Нека (X, Y) е случаен вектор со заеднички закон на распределба $p(x, y)$. Условна ентропија $H(Y|X)$ се дефинира со:

$$\begin{aligned} H(Y|X) &= \sum_{x \in R_X} p_X(x) H(Y|X = x) \\ &= - \sum_{x \in R_X} p_X(x) \sum_{y \in R_Y} p_Y(y|x) \log p_Y(y|x) \\ &= - \sum_{x \in R_X} \sum_{y \in R_Y} p_X(x) p_Y(y|x) \log p_Y(y|x) \\ &= - \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log p_Y(y|x). \end{aligned}$$

Теорема 2.1. (Верижно правило)

$$\begin{aligned} H(X, Y) &= H(X) + H(Y|X) \\ &= H(Y) + H(X|Y). \end{aligned} \tag{2.1}$$

Доказ: При докажување на првото равенство, се користи дека

$$p(x, y) = P\{X = x, Y = y\} = P\{X = x\}P\{Y = y|X = x\} = p_X(x)p_Y(y|x).$$

Со примена на последното равенство и својствата на логаритмите, се добива:

$$\begin{aligned} H(X, Y) &= - \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log p(x, y) \\ &= - \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log p_X(x) p_Y(y|x) \\ &= - \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) [\log p_X(x) + \log p_Y(y|x)] \\ &= - \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log p_X(x) - \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log p_Y(y|x) \\ &= - \sum_{x \in R_X} \log p_X(x) \sum_{y \in R_Y} p(x, y) - \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log p_Y(y|x) \\ &= - \sum_{x \in R_X} p_X(x) \log p_X(x) - \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log p_Y(y|x) \\ &= H(X) + H(Y|X). \end{aligned}$$

Постапката за докажување на второто равенство е аналогна, со користење дека

$$p(x, y) = P\{X = x, Y = y\} = P\{Y = y\}P\{X = x|Y = y\} = p_Y(y)p_X(x|y).$$

□

Пример 2.4. Нека векторот (X, Y) е зададен со закон на заедничка распределба, даден во следната табела:

$Y \backslash X$	1	2	3	4	Σ
1	1/16	1/16	0	0	1/8
2	1/16	1/16	0	1/8	1/4
3	0	0	1/8	0	1/8
4	1/8	1/8	1/8	1/8	1/2
Σ	1/4	1/4	1/4	1/4	1

За маргиналните закони на распределба на X и на Y се добива:

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \quad Y : \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1/8 & 1/4 & 1/8 & 1/2 \end{pmatrix}$$

За ентропиите на X и на Y , имаме:

$$H(X) = -\frac{1}{4} \log \frac{1}{4} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{4} \log \frac{1}{4} = \log 4 = 2$$

$$H(Y) = -\frac{1}{8} \log \frac{1}{8} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{8} \log \frac{1}{8} - \frac{1}{2} \log \frac{1}{2} = 7/4$$

За определување на условната ентропија $H(X|Y = 1)$, потребно е да се определи условната на распределба на X при услов $\{Y = 1\}$. Познато е дека овој закон на распределба може да се добие ако сите веројатности (во заедничкиот закон на распределба) кои се наоѓаат во редицата соодветна на $\{Y = 1\}$ се поделат со сумата на таа редица. Оттука,

$$X_{\{Y=1\}} : \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix},$$

па

$$H(X|Y = 1) = H\left(\frac{1}{2}, \frac{1}{2}, 0, 0\right) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = 1.$$

На ист начин, се определуваат и следните условни ентропии:

$$H(X|Y = 2) = H\left(\frac{1}{4}, \frac{1}{4}, 0, \frac{1}{2}\right) = -\frac{1}{4} \log \frac{1}{4} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{2} \log \frac{1}{2} = 3/2$$

$$H(X|Y = 3) = H(0, 0, 1, 0) = 0$$

$$H(X|Y = 4) = H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) = 2$$

Сега, за условната ентропија на X при услов Y се добива:

$$\begin{aligned} H(X|Y) &= \sum_{i=1}^4 P(Y=i)H(X|Y=i) \\ &= \frac{1}{8} \cdot 1 + \frac{1}{4} \cdot \frac{3}{2} + \frac{1}{8} \cdot 0 + \frac{1}{2} \cdot 2 = \frac{3}{2} \end{aligned}$$

Аналогно, за условната ентропија на Y при услов X се добива:

$$\begin{aligned} H(Y|X) &= \sum_{i=1}^4 P(X=i)H(Y|X=i) \\ &= \frac{1}{4}H\left(\frac{1}{4}, \frac{1}{4}, 0, \frac{1}{2}\right) + \frac{1}{4}H\left(\frac{1}{4}, \frac{1}{4}, 0, \frac{1}{2}\right) + \frac{1}{4}H\left(0, 0, \frac{1}{2}, \frac{1}{2}\right) \\ &\quad + \frac{1}{4}H\left(0, \frac{1}{2}, 0, \frac{1}{2}\right) \\ &= \frac{1}{4} \cdot \frac{3}{2} + \frac{1}{4} \cdot \frac{3}{2} + \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot 1 \\ &= \frac{5}{4} \end{aligned}$$

Заедничката ентропија $H(X, Y)$ може да се пресмета со користење на верижното правило, па

$$H(X, Y) = H(X) + H(Y|X) = 2 + \frac{5}{4} = \frac{13}{4}.$$

Да воочиме дека $H(Y|X) \neq H(X|Y)$, но

$$H(X) - H(X|Y) = H(Y) - H(Y|X) = 1/2.$$

□

2.3. Релативна ентропија и информација

Видовме дека ентропијата на една случајна променлива е мерка за неизвесноста во поглед на појавување на вредноста на една случајна променлива. Тоа е мерка за просечната количина на информација потребна да се опише случајната променлива. Од друга страна, релативна ентропија, на некој начин, е мерка за растојанието помеѓу две распределби. Имено, релативната

ентропија $D(p||q)$ е мерка за недоволноста на претпоставката дека распределбата е q кога точната распределба е p . На пример, ако ја знаеме точната распределба p на случајната променлива, тогаш може да ја погодиме нејзината вредност со просечно $H(p)$ прашања. Но, ако погодуваме за распределбата q , тогаш ќе бидат потребни $H(p) + D(p||q)$ прашања.

Дефиниција 2.5. *Релативна ентропија или растојание на Кулбак–Лајблер (Kullback–Leibler) помеѓу два закона на распределба $p(x)$ и $q(x)$ над исто множество вредности R_X се дефинира со:*

$$D(p||q) = \sum_{x \in R_X} p(x) \log \frac{p(x)}{q(x)}.$$

Подоцна ќе покажеме дека релативната ентропија е секогаш ненегативна, и е еднаква на 0 ако и само ако $p = q$. Ова не е вистинско растојание бидејќи не е симетрична и не го задоволува неравенството на триаголник, но е корисно за релативната ентропија да се мисли како на „растојание“ помеѓу две распределби.

Пример 2.5. Нека $R_X = \{2, 3\}$ и нека p и q се две распределби над R_X дефинирани на следниот начин:

$$p : \begin{pmatrix} 2 & 3 \\ 1-t & t \end{pmatrix} \quad q : \begin{pmatrix} 2 & 3 \\ 1-s & s \end{pmatrix}$$

Согласно претходната дефиниција, релативната ентропија помеѓу p и q ќе биде

$$D(p||q) = \sum_{x=2}^3 p(x) \log \frac{p(x)}{q(x)} = (1-t) \log \frac{1-t}{1-s} + t \log \frac{t}{s},$$

а релативната ентропија помеѓу q и p е

$$D(q||p) = \sum_{x=2}^3 q(x) \log \frac{q(x)}{p(x)} = (1-s) \log \frac{1-s}{1-t} + s \log \frac{s}{t}.$$

Ако $t = 1/4$, а $s = 1/8$, за релативните ентропии се добива:

$$D(p||q) = \frac{3}{4} \log \frac{3/4}{7/8} + \frac{1}{4} \log \frac{1/4}{1/8} = 0.08 \text{ бита}$$

$$D(q||p) = \frac{7}{8} \log \frac{7/8}{3/4} + \frac{1}{8} \log \frac{1/8}{1/4} = 0.07 \text{ бита}$$

Оттука е јасно дека $D(p||q) \neq D(q||p)$. Од друга страна, ако $t = s$, т.е. ако p и q се еднакви распределби, тогаш $D(p||q) = D(q||p) = 0$. \square

Во продолжение, ќе го воведеме поимот заемна информација, што е мерка за количината на информација што една случајна променлива ја содржи за друга случајна променлива.

Дефиниција 2.6. Нека (X, Y) е случаен вектор со заеднички закон на распределба $p(x, y)$ и маргинални закони на распределба $p_X(x)$ и $p_Y(y)$, соодветно. Заемна информација $I(X; Y)$ е релативната ентропија помеѓу заедничката распределба $p(x, y)$ и производот на распределбите $p_X(x)p_Y(y)$ (што е, исто така, распределба), т.е.

$$\begin{aligned} I(X; Y) &= D(p(x, y)||p_X(x)p_Y(y)) \\ &= \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log \frac{p(x, y)}{p_X(x)p_Y(y)}. \end{aligned}$$

Од дефиницијата на заемна информација $I(X; Y)$ добиваме:

$$\begin{aligned} I(X; Y) &= \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log \frac{p(x, y)}{p_X(x)p_Y(y)} \\ &= \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log \frac{p_X(x|y)}{p_X(x)} \\ &= - \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log p_X(x) + \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log p_X(x|y) \\ &= - \sum_{x \in R_X} \log p_X(x) \sum_{y \in R_Y} p(x, y) - \left(- \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log p_X(x|y) \right) \\ &= - \sum_{x \in R_X} p_X(x) \log p_X(x) - \left(- \sum_{x \in R_X} \sum_{y \in R_Y} p(x, y) \log p_X(x|y) \right) \\ &= H(X) - H(X|Y). \end{aligned}$$

Покажавме дека

$$I(X; Y) = H(X) - H(X|Y). \quad (2.2)$$

Од последното равенство може да се заклучи дека заемната информација $I(X; Y)$ е редуција на неизвесноста на X како резултат на познавањето на Y .

Од причини на симетрија, следува дека

$$I(X; Y) = H(Y) - H(Y|X). \quad (2.3)$$

2.4. Својства на ентропија, релативна ентропија и заемна информација 17

Ако се искористи верижното правило $H(X, Y) = H(X) + H(Y|X)$, се изрази $H(Y|X)$ и се замени во равенството (2.3), тогаш се добива дека

$$I(X; Y) = H(X) + H(Y) - H(X, Y). \quad (2.4)$$

На крај, да воочиме дека

$$I(X; X) = H(X) - H(X|X) = H(X),$$

бидејќи $H(X|X) = 0$ (не постои неизвесност за X кога е дадено X). Така, заемната информација на една случајна променлива со самата себе е ентропија на таа случајна променлива. Затоа, ентропијата се нарекува и *самоинформација*.

2.4. Својства на ентропија, релативна ентропија и заемна информација

Во ова поглавје ќе бидат дадени некои поважни својства на ентропијата, релативната ентропија и заемната информација. На почеток, верижното правило, кое беше дадено во теорема 2.1 за две случајни променливи, ќе биде обопштено за произволен број случајни променливи.

Теорема 2.2. (Верижно правило за ентропија) Нека векторот (X_1, X_2, \dots, X_n) има закон на распределба $p(x_1, x_2, \dots, x_n)$. Тогаш

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1).$$

Доказ: Доказот на тврдењето ќе го изведеме со користење на математичка индукција. За $n = 2$, тврдењето

$$H(X_1, X_2) = H(X_1) + H(X_2|X_1)$$

следува од Теорема 2.1. Претпоставуваме дека тврдењето важи за $n = k - 1$, т.е.

$$\begin{aligned} H(X_1, X_2, \dots, X_{k-1}) &= H(X_1) + H(X_2|X_1) + \dots + H(X_{k-1}|X_{k-2}, \dots, X_1) \\ &= \sum_{i=1}^{k-1} H(X_i | X_{i-1}, \dots, X_1). \end{aligned} \quad (2.5)$$

Ќе покажеме дека тврдењето важи и за $n = k$. За да го покажеме тоа, најпрво X_1, X_2, \dots, X_{k-1} ги групираме во еден случаен вектор и тој го третираме како една (повеќедимензионална) случајна променлива, па го применуваме првото равенство од (2.1) за променливите $(X_1, X_2, \dots, X_{k-1})$ и X_k . Добиваме:

$$\begin{aligned} H(X_1, X_2, \dots, X_{k-1}, X_k) &= H((X_1, X_2, \dots, X_{k-1}), X_k) \\ &= H(X_1, X_2, \dots, X_{k-1}) + H(X_k | X_1, X_2, \dots, X_{k-1}) \end{aligned}$$

Ако се примени индуктивната претпоставка (2.5) за $H(X_1, X_2, \dots, X_{k-1})$, имаме:

$$\begin{aligned} H(X_1, X_2, \dots, X_{k-1}, X_k) &= \sum_{i=1}^{k-1} H(X_i | X_{i-1}, \dots, X_1) + H(X_k | X_{k-1}, \dots, X_1) \\ &= \sum_{i=1}^k H(X_i | X_{i-1}, \dots, X_1) \end{aligned}$$

Со ова, тврдењето е докажано. \square

Во продолжение ќе дефинираме условна заемна информација како редукција на неизвесноста на X како резултат на познавање на Y кога Z е дадена случајна променлива.

Дефиниција 2.7. Условна заемна информација на случајните променливи X и Y при дадено Z се дефинира со:

$$I(X; Y | Z) = H(X | Z) - H(X | Y, Z)$$

.

Верижно правило важи и за заемната информација. Тоа е дадено со следната теорема.

Теорема 2.3. (Верижно правило за информација)

$$I(X_1, X_2, \dots, X_n; Y) = \sum_{i=1}^n I(X_i; Y | X_{i-1}, \dots, X_1).$$

Доказ:

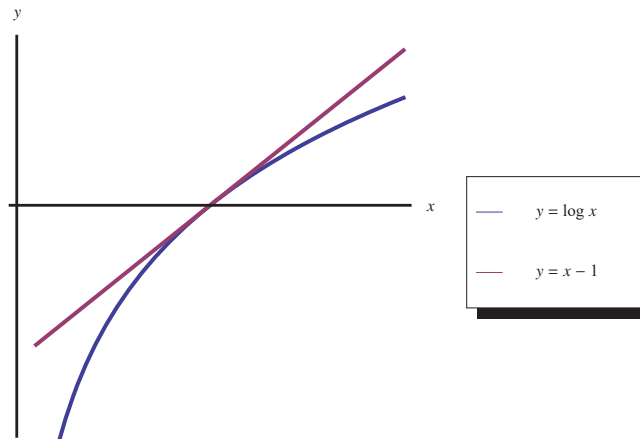
$$\begin{aligned} I(X_1, X_2, \dots, X_n; Y) &= H(X_1, X_2, \dots, X_n) - H(X_1, X_2, \dots, X_n | Y) \\ &= \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1) - \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1, Y) \\ &= \sum_{i=1}^n [H(X_i | X_{i-1}, \dots, X_1) - H(X_i | X_{i-1}, \dots, X_1, Y)] \\ &= \sum_{i=1}^n I(X_i; Y | X_{i-1}, \dots, X_1). \end{aligned}$$

□

Во продолжение, ќе покажеме дека релативната ентропија помеѓу две распределби p и q е ненегативна. За да го докажеме тоа тврдење, ќе го користиме следното неравенство:

$$\log x \leq x - 1. \quad (2.6)$$

Притоа, равенство важи, т.е. $\log x = x - 1$ ако $x = 1$. Тоа најдобро може да се илустрира на слика 2.2.



Слика 2.2

Имено, графикот на функцијата $y = \log x$ лежи комплетно под неговата тангента $y = x - 1$ во точката $x = 1$. Притоа, двете функции се допираат само во $x = 1$, па затоа во таа точка $\log x = x - 1$. Ако неравенството (2.6) се помножи со -1 , може да се запише во следната еквивалентна форма:

$$-\log x \geq 1 - x. \quad (2.7)$$

Теоремата која ја покажува ненегативноста на релативната ентропија е дадена во продолжение.

Теорема 2.4. (Ненегативност на релативната ентропија) Нека $p(x)$ и $q(x)$ се два закони на распределба над исто множество вредности R_X . Тогаш

$$D(p||q) \geq 0.$$

Притоа, $D(p||q) = 0$ ако $p(x) = q(x)$, за секое $x \in R_X$.

Доказ:

$$\begin{aligned} D(p||q) &= \sum_{x \in R_X} p(x) \log \frac{p(x)}{q(x)} \\ &= \sum_{x \in R_X} p(x) \left(-\log \frac{q(x)}{p(x)} \right) \end{aligned}$$

Сега, се применува (2.7) од кое следува дека

$$-\log \frac{q(x)}{p(x)} \geq 1 - \frac{q(x)}{p(x)}.$$

Со заменување на последното неравенство во изразот за $D(p||q)$, се добива:

$$\begin{aligned} D(p||q) &= \sum_{x \in R_X} p(x) \left(-\log \frac{q(x)}{p(x)} \right) \\ &\geq \sum_{x \in R_X} p(x) \left(1 - \frac{q(x)}{p(x)} \right) \\ &= \sum_{x \in R_X} p(x) - \sum_{x \in R_X} q(x) \\ &= 1 - 1 \\ &= 0 \end{aligned}$$

Равенство ќе важи ако $p(x)/q(x) = 1$, т.е. $p(x) = q(x)$, за секој $x \in R_X$. \square

Од ненегативноста на релативната ентропија директно следува ненегативност на заемната информација.

Последица 2.1. (Ненегативност на заемна информација) За произволни две случајни променливи X и Y важи:

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

Притоа, $I(X; Y) = 0$ ако X и Y се независни случајни променливи.

Доказ: Во Теорема 2.4 покажавме дека релативната ентропија е ненегативна, а заемната информација е дефинирана преку релативна ентропија. Затоа имаме дека $I(X; Y) = D(p(x, y)||p_X(x)p_Y(y)) \geq 0$. Притоа, равенство важи ако $p(x, y) = p_X(x)p_Y(y)$, за секој $(x, y) \in R_{(X, Y)}$, т.е. ако X и Y се независни случајни променливи. \square

2.4. Својства на ентропија, релативна ентропија и заемна информација 21

Следната теорема покажува дека од сите распределби дефинирани над исто конечно множество вредности R_X , најголема ентропија има рамномерната распределба.

Теорема 2.5. Нека $n = |R_X|$ е кардиналниот број на множеството вредности на случајна променлива X . Тогаш $H(X) \leq \log n$. Притоа, $H(X) = \log n$ ако X има рамномерна распределба над R_X .

Доказ: Нека $u(x) = 1/n$ е законот на рамномерна распределба над R_X и нека $p(x)$ е законот на распределба на случајната променлива X . Тогаш

$$\begin{aligned} D(p||u) &= \sum_{x \in R_X} p(x) \log \frac{p(x)}{u(x)} \\ &= \sum_{x \in R_X} p(x) \log p(x) - \sum_{x \in R_X} p(x) \log u(x) \\ &= -H(X) - \sum_{x \in R_X} p(x) \log \frac{1}{n} \\ &= -H(X) + \log n \sum_{x \in R_X} p(x) \\ &= -H(X) + \log n \end{aligned}$$

Бидејќи, $D(p||q) \geq 0$, од последното равенство следува дека

$$-H(X) + \log n \geq 0.$$

Значи, $H(X) \leq \log n = \log |R_X|$. □

Теорема 2.6. За произволни случајни променливи X и Y важи неравенството

$$H(X|Y) \leq H(X).$$

Притоа, $H(X|Y) = H(X)$ ако X и Y се независни случајни променливи.

Доказ: Од $I(X; Y) = H(X) - H(X|Y)$ и $I(X; Y) \geq 0$ следува дека

$$H(X) - H(X|Y) \geq 0,$$

т.е. $H(X|Y) \leq H(X)$. Равенство важи ако $I(X; Y) = 0$, т.е. ако X и Y се независни случајни променливи. □

Интуитивно, теоремата покажува дека познавањето на некоја друга променлива може да ја намали неизвесноста во поглед на појавување на вредност на случајната променлива X . Но, тоа важи само за средната неизвесност. Имено, $H(X|Y = y)$ може да биде помало или поголемо од $H(X)$, но за средната вредност се добива $H(X|Y) \leq H(X)$.

Пример 2.6. Нека законот на распределба на случајниот вектор (X, Y) е даден во следната табела:

$Y \backslash X$	0	1	Σ
0	3/4	1/8	7/8
1	1/8	0	1/8
Σ	7/8	1/8	1

Маргиналната распределба на X е

$$X : \begin{pmatrix} 0 & 1 \\ 7/8 & 1/8 \end{pmatrix},$$

па $H(X) = H\left(\frac{7}{8}, \frac{1}{8}\right) = 0.544$ бит. Од друга страна,

$$\begin{aligned} H(X|Y=0) &= H\left(\frac{6}{7}, \frac{1}{7}\right) = 0.592 > H(X) \\ H(X|Y=1) &= H(1, 0) = 0 < H(X) \end{aligned}$$

Но, за средната ентропија важи

$$H(X|Y) = \frac{7}{8}H(X|Y=0) + \frac{1}{8}H(X|Y=1) = 0.518 < H(X).$$

Значи, неизвесноста на X опаѓа кога $Y = 1$ и расте кога $Y = 0$, но просечната неизвесност $H(X|Y)$ сигурно опаѓа во споредба со $H(X)$. \square

Теорема 2.7. За произволни случајни променливи X_1, X_2, \dots, X_n важи:

$$H(X_1, X_2, \dots, X_n) \leq \sum_{i=1}^n H(X_i).$$

Притоа, равенство важи ако $X_i, i = 1, 2, \dots, n$ се независни случајни променливи.

Доказ: Од Теорема 2.6 следува дека

$$H(X_i|X_{i-1}, \dots, X_1) \leq H(X_i),$$

за произволен $i = 1, 2, \dots, n$. Со користење на ова неравенство и верижното правило се добива следното:

$$\begin{aligned} H(X_1, X_2, \dots, X_n) &= \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1) \\ &\leq \sum_{i=1}^n H(X_i). \end{aligned}$$

Притоа, равенство важи ако X_i е независна од X_{i-1}, \dots, X_1 , за секој $i = 1, 2, \dots, n$, т.е. ако X_1, X_2, \dots, X_n се независни случајни променливи. \square

2.5. Решени задачи

Задача 2.5.1. Нека (X, Y) е зададен со следниот закон на распределба:

	Y	
X \	0	1
0	1/3	1/3
1	0	1/3

Да се определи:

- а) $H(X), H(Y)$
- б) $H(Y|X), H(X|Y)$
- в) $H(X, Y)$
- г) $I(X; Y)$.

Решение:

$$X : \begin{pmatrix} 0 & 1 \\ 2/3 & 1/3 \end{pmatrix} \quad Y : \begin{pmatrix} 0 & 1 \\ 1/3 & 2/3 \end{pmatrix}$$

$$\text{а) } H(X) = H(Y) = -\frac{2}{3} \log \frac{2}{3} - \frac{1}{3} \log \frac{1}{3} = 0.918 \text{ бита}$$

б)

$$\begin{aligned} H(X|Y) &= \frac{1}{3}H(X|Y=0) + \frac{2}{3}H(X|Y=1) = \frac{1}{3}H(1,0) + \frac{2}{3}H(1/2,1/2) \\ &= \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot \left(-\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} \right) = 0.667 \text{ бита} \end{aligned}$$

$$\begin{aligned} H(Y|X) &= \frac{2}{3}H(Y|X=0) + \frac{1}{3}H(Y|X=1) = \frac{2}{3}H(1/2,1/2) + \frac{1}{3}H(1,0) \\ &= 0.667 \text{ бита} \end{aligned}$$

$$\text{в) } H(X, Y) = \left(-\frac{1}{3} \log \frac{1}{3} \right) \cdot 3 = \log 3 = 1.585 \text{ бита}$$

$$\text{г) } I(X; Y) = H(Y) - H(Y|X) = 0.251 \text{ бита} \quad \square$$

Задача 2.5.2. Монета се фрла сè додека не се појави грб. Нека X е број на изведени фрлања.

а) Да се пресмета $H(X)$.

б) Да се определи „ефикасна“ низа од *да – не* прашања од облик „Дали X припаѓа во множество S ?“. Да се спореди $H(X)$ со очекуваниот број на прашања потребни да се определи вредноста на X .

Решение:

а) Случајната променлива X има $Geo(1/2)$ распределба. Ќе ја определеме формулата за $H(X)$ во поопшт случај кога $X \sim Geo(p)$.

$$p(x) = P\{X = x\} = q^{x-1}p, \quad x = 1, 2, \dots$$

За ентропијата на X се добива:

$$\begin{aligned}
 H(X) &= -\sum_{x=1}^{+\infty} p(x) \log p(x) = -\sum_{x=1}^{+\infty} q^{x-1} p \log(q^{x-1} p) \\
 &= -\sum_{x=1}^{+\infty} q^{x-1} p(x-1) \log q - \sum_{x=1}^{+\infty} q^{x-1} p \log p \\
 &= -p \log q \sum_{x=1}^{+\infty} q^{x-1} (x-1) - p \log p \sum_{x=1}^{+\infty} q^{x-1} \\
 &= -p \log q \sum_{x=1}^{+\infty} q^{x-1} (x-1) - p \log p \frac{1}{1-q}. \tag{2.8}
 \end{aligned}$$

Сега,

$$\begin{aligned}
 A &= \sum_{x=1}^{+\infty} q^{x-1} (x-1) \\
 &= q + 2q^2 + 3q^3 + 4q^4 \dots + nq^n + \dots \\
 &= [q + q^2 + q^3 + q^4 \dots] + [q^2 + q^3 + q^4 \dots] + [q^3 + q^4 + \dots] + \dots \\
 &= \sum_{n=1}^{+\infty} \sum_{i=n}^{+\infty} q^i = \sum_{n=1}^{+\infty} \frac{q^n}{1-q} = \sum_{n=1}^{+\infty} \frac{q^n}{p} \\
 &= \frac{1}{p} \sum_{n=1}^{+\infty} q^n = \frac{1}{p} \frac{q}{1-q} = \frac{q}{p^2}.
 \end{aligned}$$

Со замена на последниот израз во (2.8), за ентропијата на случајна променлива X со $Geo(p)$ распределба се добива

$$\begin{aligned}
 H(X) &= -p(\log q) \frac{q}{p^2} - p(\log p) \frac{1}{p} = -\frac{q}{p} \log q - \log p \\
 &= \frac{-q \log q - p \log p}{p} = \frac{H(p)}{p} \text{ бита}
 \end{aligned}$$

За $p = \frac{1}{2}$ се добива $H(X) = \frac{-\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2}}{\frac{1}{2}} = 2 \log 2 = 2$ бита.

б) Интуитивно, јасно е дека најдобри прашања се оние што имаат еднакви шанси за „да“ или „не“ одговор. Оттука, една можност за погодување на

вредноста на X е серијата од облик:

- Дали $X = 1$?

Ако НЕ, дали $X = 2$?

Ако НЕ, дали $X = 3$?

.....

Нека Y е случајната променлива - број на поставени прашања до погодување на вредноста на X . Тогаш,

$$\begin{aligned} R_Y &= \{1, 2, \dots\} \\ P\{Y = 1\} &= P\{X = 1\} = p \\ P\{Y = 2\} &= P\{X = 2\} = qp \\ &\vdots \\ P\{Y = n\} &= P\{X = n\} = q^{n-1}p \\ EY &= \frac{1}{p} \end{aligned}$$

За $p = \frac{1}{2}$ се добива $EY = 2$.

Значи, $EY = 2 = H(X)$, т.е., очекуваниот број на прашања потребни да се определи вредноста на X е еднаков на ентропијата на X .

□

Задача 2.5.3. Која е минималната вредност на $H(p_1, p_2, \dots, p_n) = H(\mathbf{p})$ каде што \mathbf{p} се менува над множеството од сите n димензионални распределби? Да се определат сите \mathbf{p} за кои се постигнува тој минимум.

Решение:

$$H(\mathbf{p}) = - \sum_{i=1}^n p_i \log p_i.$$

Притоа, $-p_i \log p_i \geq 0$, при што равенство важи ако $p_i = 0$ или $p_i = 1$

Оттука, минималната вредност на $H(\mathbf{p})$ е 0 и сите можни распределби кои ја минимизираат $H(\mathbf{p})$ се оние во кои $p_i = 1$ за точно едно i , и $p_j = 0$ за $j \neq i$. Постојат n такви распределби, за кои $H(\mathbf{p})$ е 0.

□

Задача 2.5.4. Да се покаже дека процесирањето на податоците ја намалува ентропијата, т.е., ако $Y = f(X)$, тогаш $H(Y) \leq H(X)$.

Решение:

$$H(X, f(X)) = H(X) + H(f(X) | X) = H(X), \quad (2.9)$$

бидејќи $H(f(X) | X) = \sum_x p(x) H(f(X) | X = x) = \sum_x 0 = 0$, се добива

$$H(X, f(X)) = H(f(X)) + H(X | f(X)) \quad (2.10)$$

Со замена на (2.9) во (2.10), се добива

$$H(X) = H(f(X)) + H(X | f(X)) \geq H(f(X)) = H(Y).$$

□

Задача 2.5.5. Да се покаже дека ако $H(Y|X) = 0$, тогаш Y е функција од X , т.е., за сите x за кои $p(x) > 0$, постои само една вредност на y за која $p(x, y) > 0$.

Решение:

$0 = H(Y|X) = \sum_x p(x) H(Y|X = x)$, значи за сите $p(x) > 0$, мора $H(Y|X = x) = 0$, т.е.,

$$0 = H(Y|X = x) = - \sum_y p(y|x) \log p(y|x)$$

Равенството ќе биде исполнето само ако за секое y , $p(y|x) = 0$ или $p(y|x) = 1$. За $p(y|x)$ да биде распределба на веројатност мора $p(y|x) = 1$ за точно една вредност $y = y_0$, а за сите останати вредности на y , $p(y|x)$ да е еднакво на 0. За $y = y_0$ добиваме:

$$p(y_0|x) = 1$$

$$\frac{p(x, y_0)}{p(x)} = 1$$

$$p(x, y_0) = p(x) > 0, \text{ само за } y = y_0.$$

□

Задача 2.5.6. Нека X_1 и X_2 се дискретни случајни променливи со множества вредности $R_{X_1} = \{1, 2, \dots, m\}$ и $R_{X_2} = \{m + 1, m + 2, \dots, n\}$, а θ е резултатот од фрлање на неправилна монета, така што $P\{\theta = 1\} = \alpha$, $P\{\theta = 0\} = 1 - \alpha$. Нека X_1 , X_2 и θ се меѓусебно независни, и притоа

$$X = \begin{cases} X_1, & \text{ако } \theta = 1 \\ X_2, & \text{ако } \theta = 0 \end{cases} .$$

Да се изрази $H(X)$ преку $H(X_1)$, $H(X_2)$ и α .

Решение:

Бидејќи X_1 и X_2 имаат дисјунктни множества вредности, θ може да се изрази како функција од X :

$$\theta = f(X) = \begin{cases} 1, & \text{ако } X \in R_{X_1} \\ 0, & \text{ако } X \in R_{X_2} \end{cases} .$$

Сега,

$$H(X, \theta) = H(X, f(X)) = H(X) + H(f(X)|X) = H(X),$$

бидејќи согласно со претходната задача $H(f(X)|X) = 0$. Оттука, за $H(X)$ добиваме:

$$\begin{aligned} H(X) &= H(X, \theta) = H(\theta) + H(X|\theta) = \\ &= H(\theta) + P\{\theta = 1\}H(X|\theta = 1) + P\{\theta = 0\}H(X|\theta = 0) \\ &= H(\alpha, 1 - \alpha) + \alpha H(X_1) + (1 - \alpha)H(X_2). \end{aligned}$$

□

Задача 2.5.7. Две екипи А и В играат серија партии шах, сè додека една од екипите не победи 4 пати или додека не се одиграат 7 партии. Секоја партија се игра до победа. Нека случајната променлива X го означува исходот од одиграната серија помеѓу А и В. Нека Y е број на одиграни партии. При претпоставка дека А и В имаат еднакви шанси за победа во секоја од партиите, да се пресмета: $H(X)$, $H(Y)$, $H(Y|X)$ и $H(X|Y)$.

Решение:

Постојат следните можности:

- одиграни 4 партии: Постојат две такви серии (победа на А во сите 4 партии или победа на В во сите четири партии). Секоја од сериите е со веројатност $(1/2)^4 = 1/16$.
- одиграни 5 партии: Еден начин за ова ќе се случи е А да победи во 4 партии, а В во една (но во последната партија мора да победи А, бидејќи во спротивно, играта би завршила со помалку од 5 партии). Значи, А треба да победи 3 партии од првите 4 и да ја победи последната партија. Вториот начин е В да победи во 4 партии, а А во една, но таа да не биде последната партија. Оттука, вкупниот број на можности е $2 \binom{4}{3} = 2 \cdot 4 = 8$ серии, секоја со веројатност $(1/2)^5 = 1/32$.
- одиграни 6 партии: Еден начин за ова да се случи е А да победи во 4 партии, а В во две. Исто како и претходно, во последната партија мора да победи А, бидејќи во спротивно, играта би завршила со помалку од 6 партии. Значи, А треба да победи 3 партии од првите 5 и да ја победи последната шеста партија. Вториот начин е В да победи во 4 партии, а А во две, така што двете победи на А мора да се во првите 5 партии. Вкупниот број на можности е $2 \binom{5}{3} = 2 \cdot 10 = 20$ серии, секоја со веројатност $(1/2)^6 = 1/64$.
- одиграни 7 партии: Аналогно, како и претходно, еден начин е А да победи во 4 партии, а В во три, така што во последната партија мора да победи А. Значи, А треба да победи 3 партии од првите 6 и да ја победи последната шеста партија. Вториот начин е В да победи во 4 партии, а А во три, така што трите победи на А мора да се во првите 6 партии. Вкупниот број на можности е $2 \binom{6}{3} = 2 \cdot 20 = 40$ серии, секоја со веројатност $(1/2)^7 = 1/128$.

Оттука ги добиваме распределбите на случајните променливи X и Y и нивните ентропии:

$$X : \begin{pmatrix} AAAA & BBBB & BAAAA & \dots \\ \frac{1}{16} & \frac{1}{16} & \frac{1}{32} & \dots \end{pmatrix}$$

$$Y : \left(2 \cdot \frac{4}{16} = \frac{1}{8} \quad 8 \cdot \frac{5}{32} = \frac{1}{4} \quad 20 \cdot \frac{6}{64} = \frac{5}{16} \quad 40 \cdot \frac{7}{128} = \frac{5}{16} \right)$$

$$H(X) = -2 \cdot \frac{1}{16} \log \frac{1}{16} - 8 \cdot \frac{1}{32} \log \frac{1}{32} - 20 \cdot \frac{1}{64} \log \frac{1}{64} - 40 \cdot \frac{1}{128} \log \frac{1}{128}$$

$$= 5.8125 \text{ бита}$$

$$H(Y) = -\frac{1}{8} \log \frac{1}{8} - \frac{1}{4} \log \frac{1}{4} - \frac{5}{16} \log \frac{5}{16} - \frac{5}{16} \log \frac{5}{16} = 1.924 \text{ бита}$$

Y е детерминистичка функција од X , т.е., ако X е познато, во Y нема случајност. Затоа, $H(Y|X) = 0$.

Од друга страна,

$$H(X) + H(Y|X) = H(X, Y) = H(Y) + H(X|Y) \Rightarrow$$

$$H(X|Y) = H(X) + H(Y|X) - H(Y) = H(X) - H(Y) = 3.8885 \text{ бита.}$$

□

Задача 2.5.8. Во кутија се наоѓаат три коцки од кои една е правилна, а две се неправилни со распределби дадени во табелите:

Коцка 2:	i	1	2	3	4	5	6
	$p(i)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{18}$	$\frac{1}{2}$

Коцка 3:	i	1	2	3	4	5	6
	$p(i)$	$\frac{1}{2}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$

Од кутијата случајно се извлекува една коцка и се фрла. Да се определи неопределеноста на изборот на коцката, ако се знае дека паднала единица.

Решение:

Нека X е случајна променлива – редниот број на избраната коцка. Тогаш, распределбата на X е:

$$X : \begin{pmatrix} 1 & 2 & 3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

Нека A е случајниот настан – падна единица при фрлање на коцката. Треба да се определи $H(X|A)$.

$$P(A) = \sum_{i=1}^3 P(A|X=i)P(X=i) = \frac{1}{3} \left(\frac{1}{12} + \frac{1}{2} + \frac{1}{6} \right) = \frac{1}{4}$$

$$P(X=i|A) = \frac{P(A|X=i)P(X=i)}{P(A)}, \quad i = 1, 2, 3$$

$$P(X=1|A) = \frac{\frac{1}{6} \cdot \frac{1}{3}}{\frac{1}{4}} = \frac{2}{9}, \quad P(X=2|A) = \frac{\frac{1}{12} \cdot \frac{1}{3}}{\frac{1}{4}} = \frac{1}{9},$$

$$P(X=3|A) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{4}} = \frac{2}{3}$$

За бараната ентропија се добива

$$\begin{aligned} H(X|A) &= - \sum_{i=1}^3 P(X=i|A) \log P(X=i|A) \\ &= -\frac{2}{9} \log \frac{2}{9} - \frac{1}{9} \log \frac{1}{9} - \frac{2}{3} \log \frac{2}{3} = 1.224 \text{ бита.} \end{aligned}$$

□

Задача 2.5.9. Веројатноста студент да положи математичка логика во јунска сесија е 0.3. Ако студентот претходно положил математичка логика, веројатноста да положи алгебра е 0.45, а ако студентот не положил математичка логика, веројатноста да не положи алгебра е 0.9. Да се определи средната заемна информација која за тоа дали студентот положил математичка логика или не, ја дава податокот дали положил алгебра или не.

Решение:

Нека X е случајна променлива која прима вредност x_1 ако студентот положил алгебра и вредност x_2 ако не положил алгебра. Нека Y е случајна променлива, која прима вредност y_1 ако студентот положил математичка логика и вредност y_2 , ако не положил математичка логика.

Според текстот на задачата $p(y_1) = 0.3$, $p(x_1|y_1) = 0.45$, $p(x_2|y_2) = 0.9$.
Оттука

$$\begin{aligned} p(y_2) &= 1 - p(y_1) = 0.7, \\ p(x_2|y_1) &= 1 - p(x_1|y_1) = 0.55, \\ p(x_1|y_2) &= 1 - p(x_2|y_2) = 0.1. \end{aligned}$$

За заедничките веројатности се добива:

$$\begin{aligned} p(x_1, y_1) &= p(y_1)p(x_1|y_1) = 0.3 \cdot 0.45 = 0.135 \\ p(x_1, y_2) &= p(y_2)p(x_1|y_2) = 0.7 \cdot 0.1 = 0.07 \\ p(x_2, y_1) &= p(y_1)p(x_2|y_1) = 0.3 \cdot 0.55 = 0.165 \\ p(x_2, y_2) &= p(y_2)p(x_2|y_2) = 0.7 \cdot 0.9 = 0.63. \end{aligned}$$

Од заедничките веројатности ги пресметуваме веројатностите од распределбата за X :

$$p(x_1) = 0.205, \quad p(x_2) = 0.795.$$

За заемната информација се добива:

$$\begin{aligned} I(X; Y) &= \sum_{i=1}^2 \sum_{j=1}^2 p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} = 0.135 \log \frac{0.135}{0.205 \cdot 0.3} + \\ &+ 0.07 \log \frac{0.07}{0.205 \cdot 0.7} + 0.165 \log \frac{0.165}{0.795 \cdot 0.3} + 0.63 \log \frac{0.63}{0.795 \cdot 0.7} \\ &= 0.105687 \text{ бита.} \end{aligned}$$

□

2.6. Задачи

Задача 2.6.1. Експеримент се состои во фрлање тетраедар чии страни се обележани со 1, 2, 3, 4. Ако се добие број што е поголем од 2, се фрла неправилна паричка, со веројатност за паѓање на грб $1/3$. Во останатите случаи се фрла друга неправилна паричка, при што грб се појавува во 60% од случаите. Нека X е исход при фрлање на тетраедарот, а Y – исход при фрлање на

паричката. Да се определи:

- а) $H(X), H(Y)$;
- б) $H(X|Y), H(Y|X)$;
- в) $H(X, Y)$;
- г) $I(X; Y)$.

Задача 2.6.2. Веројатноста дека Ана е дома е 0.4. Ако е дома, веројатноста дека спие е 0.2. Ако не е дома, веројатноста дека Ана заборавила да го угаси светлото е 0.1. Да се определи средната заемна информација на X и Y , каде што X е еднакво на x_0 ако Ана е дома и еднакво на x_1 кога не е дома, а Y е еднакво на y_0 кога светлото е запалено и еднакво на y_1 кога светлото не е запалено. Претпоставка е дека кога Ана спие светлото е угасено.

Задача 2.6.3. Дадена е коцка чии страни се различно обоени (зелено, жолто, црвено, сино, виолетово и црно), но коцката не е правилна, па зелената страна паѓа пет пати почесто од другите бои. Дадена е и кутија со 3 зелени, 4 жолти, 2 црвени, 3 сини, 5 виолетови и 3 црни топчиња. Истовремено се фрла коцката и се извлекува едно топче. Да се определи неодреденоста на бојата ако се знае дека таа е иста и на коцката и на извлеченото топче.

Задача 2.6.4. Во два сада има бели и црни топчиња. Во првиот има 2 бели и 1 црно, а во вториот 3 бели и 2 црни. Нека A е експеримент во кој случајно се извлекува едно топче од првиот сад и се префрлува во вториот сад, а B е експеримент во кој од вториот сад случајно се извлекуваат две топчиња (прво се изведува A , па B). Да се определат случајните променливи X и Y кои одговараат соодветно на експериментите A и B и да се пресмета

- а) $H(X), H(Y)$;
- б) $H(Y|X)$;
- в) $H(X, Y)$;
- г) $I(X; Y)$.

Задача 2.6.5. Дадени се три кутии A , B и C . Во кутијата A има 2 црни, 3 бели и 5 жолти топчиња, во кутијата B има 1 црно, 7 бели и 2 жолти топчиња, а во кутијата C има 6 црни, 1 бело и 3 жолти. На случаен начин се избира една кутија и се извлекува едно топче од неа. Да се определи неодреденоста на изборот на кутија ако е извлечено бело топче.

Задача 2.6.6. Нека X и Z се независни случајни променливи со бинарни вредности, чија распределба е дадена со $P\{X = 1\} = p$ и $P\{Z = 1\} = 1/2$. Нека $Y = X \oplus Z$ (X може да се смета за порака, Z за таен клуч, а Y криптирана порака). Да се определи

- а) $H(Y)$;
- б) $H(X|Y), H(X|Z)$;
- в) $I(X; Y); I(X; Z)$.

Каква информација дава вредноста на криптираната порака Y за вредноста на оригиналната порака X ? Со интерпретација на која од горенаведените величини се добива одговор на ова прашање?

Задача 2.6.7. Дадени се следните две распределби на азбуката $\{1, 2, \dots, L + M\}$ (L и M се позитивни цели броеви):

$$p(x) = \begin{cases} p_x; & \text{ако } x \in \{1, 2, \dots, L\} \\ 0; & \text{ако } x \in \{L + 1, L + 2, \dots, L + M\} \end{cases}$$

$$q(x) = \begin{cases} \alpha p_x; & \text{ако } x \in \{1, 2, \dots, L\} \\ \frac{1 - \alpha}{M}; & \text{ако } x \in \{L + 1, L + 2, \dots, L + M\} \end{cases}$$

каде што $0 < \alpha < 1$. Да се изрази $D(p||q)$ како функција од α .

Задача 2.6.8. Во џебот имам една паричка. Паричката е или правилна или е неправилна, кај која веројатноста да се појави „глава“ е трипати поголема од веројатноста за „петка“ . Паричката е правилна или неправилна со еднаква веројатност. Нека X е индикатор на настанот „Паричката е правилна“ . Паричката се фрла десет пати. Да се определи условната ентропија на X , ако знаеме дека „глава“ се појавило 3 пати.

Глава 3

Својство за асимптотска рамномерност

3.1. Закон на големите броеви

Во овој дел, на кратко, ќе бидат дадени некои дефиниции и неравенства кои ќе бидат потребни подоцна при воведување на својството за асимптотска рамномерност.

- **(Марково неравенство)** За произволна ненегативна случајна променлива X и за произволно $\delta > 0$, важи неравенството:

$$P\{X \geq \delta\} \leq \frac{EX}{\delta}.$$

- **(Неравенство на Чебишев)** Нека Y е случајна променлива со математичко очекување μ и дисперзија σ^2 . Тогаш за произволно $\varepsilon > 0$, важи неравенството:

$$P\{|Y - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}.$$

За низата случајни променливи X_1, X_2, \dots велиме дека *конвергира по веројатност* кон случајна променлива X , ако за секој $\varepsilon > 0$, важи:

$$\lim_{n \rightarrow +\infty} P\{|X_n - X| < \varepsilon\} = 1,$$

т.е.

$$\lim_{n \rightarrow +\infty} P\{|X_n - X| \geq \varepsilon\} = 0.$$

Пишуваме,

$$X_n \xrightarrow{\text{в.}} X.$$

Нека X_1, X_2, \dots се независни и еднакво распределени случајни променливи. Се разгледува низата аритметички средини:

$$X_1, \frac{X_1 + X_2}{2}, \frac{X_1 + X_2 + X_3}{3}, \dots, \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i, \dots$$

Ако оваа низа по веројатност конвергира кон EX , тогаш ќе велиме дека за низата X_1, X_2, \dots важи *слабиот закон на големите броеви*. Прашањето е кои услови треба да бидат задоволени за тоа да се случи? Одговорот на ова прашање е даден во следната теорема.

Теорема 3.1. (Чебишев, слаб закон на големите броеви) Нека X_1, X_2, \dots е низа од независни и еднакво распределени случајни променливи со математичко очекување μ и дисперзија σ^2 и нека

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Тогаш

$$\lim_{n \rightarrow +\infty} P\{|\bar{X}_n - \mu| < \varepsilon\} = 1.$$

□

Последниот лимес покажува дека низата аритметички средини конвергира по веројатност кон $\mu = EX$ (каде што X е случајна променлива со иста распределба како X_i), т.е.

$$\bar{X}_n \xrightarrow{в.} \mu = EX.$$

3.2. Својство за асимптотска рамномерност

Во теоријата на информации, својството за асимптотска рамномерност (Asymptotic Equipartition Property, или кратко АЕР) е аналог на законот на големите броеви. Ова својство е директна последица на слабиот закон на големите броеви. Ова својство тврди дека $-(1/n) \log p(X_1, X_2, \dots, X_n)$ се доближува до ентропијата H , кога X_1, X_2, \dots, X_n се независни и еднакво распределени случајни променливи, а $p(x_1, x_2, \dots, x_n)$ е нивниот заеднички закон на распределба. Својството е формализирано во следната теорема.

Теорема 3.2. (Својство за асимптотска рамномерност) Ако X_1, X_2, \dots, X_n се независни и еднакво распределени случајни променливи со закон на распределба $p(x)$, тогаш

$$-\frac{1}{n} \log p(X_1, X_2, \dots, X_n) \xrightarrow{\text{в.}} H(X),$$

каде што X е случајна променлива со иста распределба како $X_i, i = 1, 2, \dots, n$.

Доказ: Во доказот ќе искористиме дека функции од независни случајни променливи, се исто така, независни случајни променливи. Така, бидејќи X_1, X_2, \dots, X_n се независни случајни променливи, и $Y_1 = \log p(X_1), Y_2 = \log p(X_2), \dots, Y_n = \log p(X_n)$ се, исто така, независни случајни променливи. Оттука, од слабиот закон на големите броеви (Теорема 3.1), следува дека:

$$\begin{aligned} -\frac{1}{n} \log p(X_1, X_2, \dots, X_n) &= -\frac{1}{n} \log \left(\prod_{i=1}^n p(X_i) \right) \\ &= -\frac{1}{n} \sum_i \log p(X_i) \\ &= -\frac{1}{n} \sum_i Y_i \\ \text{(од Теорема 3.1)} &\xrightarrow[\text{в.}]{n \rightarrow \infty} -EY \\ &= -E(\log p(X)) \\ &= -\sum_{x \in R_X} p(x) \log p(x) \\ &= H(X) \end{aligned}$$

□

Ова својство покажува дека кога X_1, X_2, \dots, X_n се независни и еднакво распределени случајни променливи, веројатноста $p(X_1, X_2, \dots, X_n)$ ќе се доближува до 2^{-nH} . Тоа овозможува множеството од сите низи со должина n , да се подели на две подмножества: типично множество и нетипично множество.

Дефиниција 3.1. Типично множество $A_\varepsilon^{(n)}$ во однос на законот на распределба $p(x)$ е множеството од низи $(x_1, x_2, \dots, x_n) \in R_X^n$, со својството:

$$2^{-n(H(X)+\varepsilon)} \leq p(x_1, x_2, \dots, x_n) \leq 2^{-n(H(X)-\varepsilon)}.$$

Како последица на својството за асимптотска рамномерност, следната теорема ги дава својствата на типичното множество.

Теорема 3.3. Типичното множество ги има следните својства:

1. Ако $(x_1, x_2, \dots, x_n) \in A_\varepsilon^{(n)}$, тогаш

$$H(X) - \varepsilon \leq -\frac{1}{n} \log p(x_1, x_2, \dots, x_n) \leq H(X) + \varepsilon.$$

2. $P(A_\varepsilon^{(n)}) > 1 - \varepsilon$, за доволно големо n .

3. $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$, каде $|A|$ означува број на елементи во множеството A .

4. $|A_\varepsilon^{(n)}| \geq (1 - \varepsilon)2^{n(H(X)-\varepsilon)}$, за доволно големо n .

Доказ: 1. Ова следува директно од дефиницијата на типично множество. Имено, ако $(x_1, x_2, \dots, x_n) \in A_\varepsilon^{(n)}$, тогаш важи

$$2^{-n(H(X)+\varepsilon)} \leq p(x_1, x_2, \dots, x_n) \leq 2^{-n(H(X)-\varepsilon)}.$$

Ако последниот израз се логаритмира со основа 2, се добива:

$$-n(H(X) + \varepsilon) \leq \log p(x_1, x_2, \dots, x_n) \leq -n(H(X) - \varepsilon).$$

Ако последните неравенства се поделат со $-n$ се добива тврдењето на овој дел од теоремата.

2. Ова следува директно од Теорема 3.2. Имено, ако $(X_1, X_2, \dots, X_n) \in A_\varepsilon^{(n)}$, тогаш

$$-\frac{1}{n} \log p(X_1, X_2, \dots, X_n) \xrightarrow{\text{в.}} H(X).$$

Според дефиниција на конвергенција по веројатност,

$$\lim_{n \rightarrow +\infty} P \left\{ \left| -\frac{1}{n} \log p(X_1, X_2, \dots, X_n) - H(X) \right| < \varepsilon \right\} = 1.$$

Јасно е дека веројатноста тежи од лево кон 1, па за секој $\delta > 0$, постои $n_0 \in \mathbb{N}$, така што за секој $n \geq n_0$, важи:

$$P \left\{ \left| -\frac{1}{n} \log p(X_1, X_2, \dots, X_n) - H(X) \right| < \varepsilon \right\} > 1 - \delta.$$

Последното неравенство важи за секој $\delta > 0$, па и за $\delta = \varepsilon$. Со тоа се добива неравенството во вториот дел од теоремата.

3.

$$\begin{aligned}
 1 &= \sum_{(x_1, x_2, \dots, x_n) \in R_X^n} p(x_1, x_2, \dots, x_n) \\
 &= \sum_{(x_1, x_2, \dots, x_n) \in A_\varepsilon^{(n)}} p(x_1, x_2, \dots, x_n) + \sum_{(x_1, x_2, \dots, x_n) \in (A_\varepsilon^{(n)})^C} p(x_1, x_2, \dots, x_n) \\
 &\geq \sum_{(x_1, x_2, \dots, x_n) \in A_\varepsilon^{(n)}} p(x_1, x_2, \dots, x_n) \\
 &\geq \sum_{(x_1, x_2, \dots, x_n) \in A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} \\
 &= 2^{-n(H(X)+\varepsilon)} |A_\varepsilon^{(n)}|.
 \end{aligned}$$

Оттука, $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$.

4. Докажавме дека за доволно големо n , $P(A_\varepsilon^{(n)}) > 1 - \varepsilon$, така што добиваме дека

$$\begin{aligned}
 1 - \varepsilon &< P(A_\varepsilon^{(n)}) \\
 &= \sum_{(x_1, x_2, \dots, x_n) \in A_\varepsilon^{(n)}} p(x_1, x_2, \dots, x_n) \\
 &\leq \sum_{(x_1, x_2, \dots, x_n) \in A_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)} \\
 &= 2^{-n(H(X)-\varepsilon)} |A_\varepsilon^{(n)}|.
 \end{aligned}$$

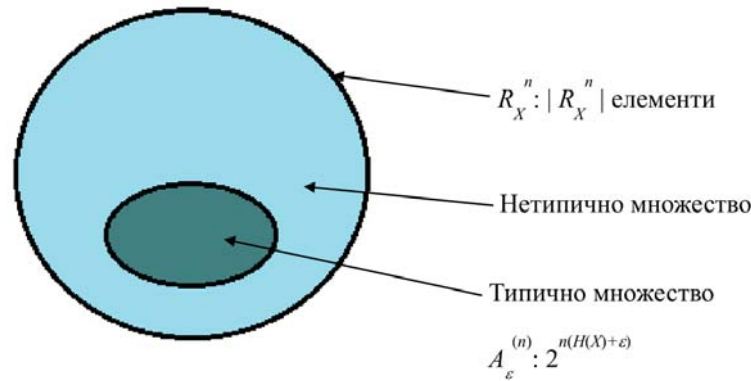
□

Според Теорема 3.3, типичното множество има веројатност блиску до 1, сите елементи од тоа множество се скоро со еднаква веројатност, и нивниот број е околу 2^{nH} .

3.3. Компресија на податоци

Својството за асимптотска рамномерност може да се искористи за компресија на податоци. Нека X_1, X_2, \dots, X_n се независни и еднакво распределени случајни променливи со закон на распределба $p(x)$. Целта е да се најде краток опис на низите од овие случајни променливи. Затоа, множеството R_X^n се дели на две подмножества (слика 3.1):

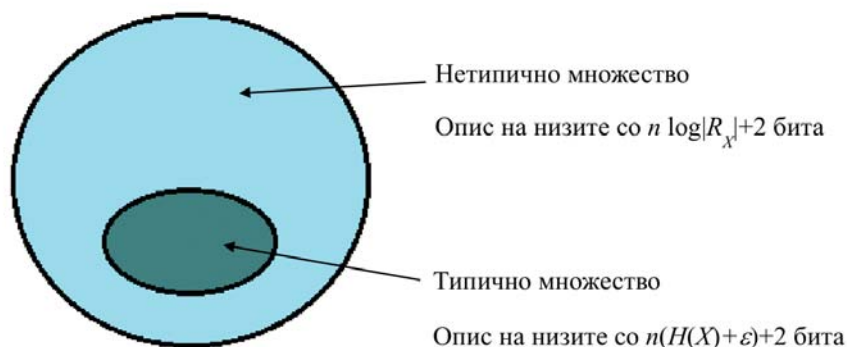
- типичното множество $A_\varepsilon^{(n)}$, и
- неговиот комплемент $(A_\varepsilon^{(n)})^c$.



Слика 3.1

Идејата за компресија на податоците е во тоа што на низите кои припаѓаат на типичното множество (кое се појавува со веројатност 1) да им се придружат пократки кодни записи. Со тоа ќе се дефинира код кој ќе овозможи очекуваната должина на кодните записи да биде најмала можна (или блиска до таа должина), со што се овозможува компресирање на податоците. Постапката за компресија на податоци е следната. Најпрво, сите елементи од типичното множество се редат во некој редослед (на пр. лексикографски). Потоа, на секоја низа од $A_\varepsilon^{(n)}$ и се доделува индекс на таа низа во множеството. Бидејќи бројот на низи во типичното множество е најмногу $2^{n(H(X)+\varepsilon)}$, за тоа индексирање ќе бидат потребни не повеќе од $n(H(X) + \varepsilon) + 1$ бит (тој дополнителен бит е потребен само во случај кога $n(H(X) + \varepsilon)$ не е цел број).

Потоа, на секој од индексите (изразени бинарно) се додава префикс 0. Со тоа се добива вкупната должина од $n(H(X) + \varepsilon) + 2$ бита за претставување на секоја низа од $A_\varepsilon^{(n)}$ (слика 3.2).



Слика 3.2

Слично, секоја низа од $(A_\varepsilon^{(n)})^c$ се индексира со не повеќе од $n \log |R_X| + 1$ бит. Се става префикс 1 на сите добиени индекси. На тој начин добиваме код за сите низи од R_X^n . Да воочиме дека имаме нумерација со груба сила (brute-force) на нетипичното множество $(A_\varepsilon^{(n)})^c$, не земајќи предвид дека бројот на елементи во $(A_\varepsilon^{(n)})^c$ е помал од бројот на елементи во R_X^n . Изненадувачки е тоа што ова е доволно добро да даде ефикасен опис.

Со ова е дефинирано едно пресликување (код) кое е инјективно, па лесно може да се декодира. Почетниот бит е на некој начин знаменце (flag) бит кој покажува колкава е должината на кодниот збор што следи, т.е. дали е коден збор за елемент од типичното или од нетипичното множество. Низите од типичното множество имаат краток опис со должина $\approx nH$.

Во продолжение, ќе ја определеме очекувана должина на кодните зборови со вака дефинираниот код. Нека $l(x_1, x_2, \dots, x_n)$ ја означува должината на кодниот збор соодветен на (x_1, x_2, \dots, x_n) . Ако n е доволно големо така што

$P(A_\varepsilon^{(n)}) > 1 - \varepsilon$, за очекуваната должина на кодните зборови се добива:

$$\begin{aligned}
E(l(X_1, X_2, \dots, X_n)) &= \sum_{(x_1, x_2, \dots, x_n)} p(x_1, x_2, \dots, x_n) l(x_1, x_2, \dots, x_n) \\
&= \sum_{(x_1, x_2, \dots, x_n) \in A_\varepsilon^{(n)}} p(x_1, x_2, \dots, x_n) l(x_1, x_2, \dots, x_n) \\
&\quad + \sum_{(x_1, x_2, \dots, x_n) \in (A_\varepsilon^{(n)})^c} p(x_1, x_2, \dots, x_n) l(x_1, x_2, \dots, x_n) \\
&\leq \sum_{(x_1, x_2, \dots, x_n) \in A_\varepsilon^{(n)}} p(x_1, x_2, \dots, x_n) [n(H + \varepsilon) + 2] \\
&\quad + \sum_{(x_1, x_2, \dots, x_n) \in (A_\varepsilon^{(n)})^c} p(x_1, x_2, \dots, x_n) [n \log |R_X| + 2] \\
&= P(A_\varepsilon^{(n)}) [n(H + \varepsilon) + 2] + P((A_\varepsilon^{(n)})^c) [n \log |R_X| + 2] \\
&\leq n(H + \varepsilon) + 2 + \varepsilon [n \log |R_X| + 2] \\
&= nH + n \left[\varepsilon + \varepsilon \log |R_X| + \frac{2 + 2\varepsilon}{n} \right] \\
&= n(H + \varepsilon'),
\end{aligned}$$

каде што $\varepsilon' = \varepsilon + \varepsilon \log |R_X| + (2 + 2\varepsilon)/n$ може да биде произволно мало со соодветен избор на ε .

Со претходното е покажана следната теорема.

Теорема 3.4. Нека X_1, X_2, \dots, X_n се независни и еднакво распределени случајни променливи со закон на распределба $p(x)$ и нека $\varepsilon > 0$. Тогаш постои код кој ги пресликува низите (x_1, x_2, \dots, x_n) со должина n , во бинарни низи таков што пресликувањето е инјективно (т.е. е инверзибилно) и

$$E \left[\frac{1}{n} l(X_1, X_2, \dots, X_n) \right] \leq H(X) + \varepsilon,$$

за доволно големо n . □

На тој начин, низите (X_1, X_2, \dots, X_n) со должина n се претставуваат со просечно $nH(X)$ бита.

Пример 3.1. Нека X_1, X_2, \dots, X_n се независни и еднакво распределени случајни променливи со распределба:

$$X : \begin{pmatrix} 0 & 1 \\ 0.4 & 0.6 \end{pmatrix}.$$

За ентропијата на X се добива:

$$H(X) = -0.4 \log 0.4 - 0.6 \log 0.6 = 0.97095 \text{ бита.}$$

Во продолжение ќе го определиме типичното множество за $n = 24$ и $\varepsilon = 0.1$. По дефиниција $A_\varepsilon^{(n)}$ се состои од сите низи (x_1, x_2, \dots, x_n) за кои $-(1/n) \log p(x_1, x_2, \dots, x_n)$ лежи во интервалот $(H(X) - \varepsilon, H(X) + \varepsilon)$. За $\varepsilon = 0.1$, тоа е интервалот $(0.87095, 1.07095)$. Во табелата во продолжение, за секој k што претставува број на единици во низа со должина 24, претставени се следните елементи:

- во колона 2 е даден број на низи со должина 24 кои содржат k единици;
- во колона 3 е дадена веројатноста за случајно да се избере низа со k единици;
- во колона 4, дадена е веројатноста $p^k q^{24-k}$, каде што $p = 0.6$, а $q = 0.4$.
- а во последната колона е дадена вредноста на $-\frac{1}{24} \log p(x_1, x_2, \dots, x_{24}) = -\frac{1}{24} \log [p^k q^{24-k}]$.

k	$\binom{24}{k}$	$\binom{24}{k} p^k q^{24-k}$	$p^k q^{24-k}$	$-\frac{1}{24} \log p(x_1, x_2, \dots, x_{24})$ $= -\frac{1}{24} \log[p^k q^{24-k}]$
0	1	2.81×10^{-10}	2.81×10^{-10}	1.321928
1	24	1.01×10^{-8}	4.22×10^{-10}	1.297555
2	276	1.75×10^{-7}	6.33×10^{-10}	1.273181
3	2024	1.92×10^{-6}	9.50×10^{-10}	1.248808
4	10626	1.51×10^{-5}	1.42×10^{-9}	1.224434
5	42504	9.09×10^{-5}	2.14×10^{-9}	1.200061
6	134596	4.32×10^{-4}	3.21×10^{-9}	1.175687
7	346104	1.67×10^{-3}	4.81×10^{-9}	1.151314
8	735471	5.31×10^{-3}	7.21×10^{-9}	1.126941
9	1307504	1.41×10^{-2}	1.08×10^{-8}	1.102567
10	1961256	3.18×10^{-2}	1.62×10^{-8}	1.078194
11	2496144	6.08×10^{-2}	2.43×10^{-8}	1.053820
12	2704156	9.88×10^{-2}	3.65×10^{-8}	1.029447
13	2496144	1.37×10^{-1}	5.48×10^{-8}	1.005073
14	1961256	1.61×10^{-1}	8.22×10^{-8}	0.980700
15	1307504	1.61×10^{-1}	1.23×10^{-7}	0.956327
16	735471	1.36×10^{-1}	1.85×10^{-7}	0.931953
17	346104	9.60×10^{-2}	2.77×10^{-7}	0.907580
18	134596	5.60×10^{-2}	4.16×10^{-7}	0.883206
19	42504	2.65×10^{-2}	6.24×10^{-7}	0.858833
20	10626	9.95×10^{-3}	9.36×10^{-7}	0.834459
21	2024	2.84×10^{-3}	1.40×10^{-6}	0.810086
22	276	5.81×10^{-4}	2.11×10^{-6}	0.785712
23	24	7.58×10^{-5}	3.16×10^{-6}	0.761339
24	1	4.74×10^{-6}	4.74×10^{-6}	0.736966

Од последната колона во табелата, може да се воочи дека $-(1/24) \log p(x_1, x_2, \dots, x_{24}) \in (0.87095, 1.07095)$, за сите низи кај кои бројот на единици е помеѓу 11 и 18. Тоа се елементите од последната колона на обоените редици.

Веројатноста на типичното множество е сума на веројатностите од третата колона за k од 11 до 18, и тоа изнесува 0.9076. Бројот на елементи во $A_{0.1}^{(24)}$ е 12 181 375.

Ако се пресметаат долната и горната граница на $A_{0.1}^{(24)}$ согласно теорема

3.3, се добива:

$$|A_{0.1}^{(24)}| \leq 2^{n(H(X)+\varepsilon)} = 2^{24 \cdot (0.97095+0.1)} = 2^{25.7} = 54\,615\,231$$

$$|A_{0.1}^{(24)}| \geq (1-\varepsilon)2^{n(H(X)-\varepsilon)} = (1-0.1)2^{24 \cdot (0.97095-0.1)} = 0.9 \cdot 2^{20.9} = 1\,764\,462.$$

Може да се воочи дека двете граници се доста лоши. \square

3.4. Решени задачи

Задача 3.4.1. Нека X_1, X_2, \dots се независни и еднакво распределени случајни променливи со закон на распределба $p(x)$. Да се определи

$$\lim_{n \rightarrow +\infty} [p(X_1, X_2, \dots, X_n)]^{1/n}.$$

Решение:

Според својството АЕР

$$-\frac{1}{n} \log p(X_1, X_2, \dots, X_n) \xrightarrow[n \rightarrow \infty]{\text{в.}} H(X)$$

т.е.,

$$\log p(X_1, X_2, \dots, X_n) \xrightarrow[n \rightarrow \infty]{\text{в.}} -nH(X)$$

$$p(X_1, X_2, \dots, X_n) \xrightarrow[n \rightarrow \infty]{\text{в.}} 2^{-nH(X)}$$

Оттука, се добива бараниот лимес:

$$(p(X_1, X_2, \dots, X_n))^{1/n} \xrightarrow[n \rightarrow \infty]{\text{в.}} 2^{-H(X)}$$

\square

Задача 3.4.2. Нека X_1, X_2, \dots се независни и еднакво распределени случајни променливи со закон на распределба $p(x)$, $x \in \{1, 2, \dots, m\}$. Значи,

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i) \text{ и } -\frac{1}{n} \log p(X_1, X_2, \dots, X_n) \xrightarrow[n \rightarrow \infty]{\text{в.}} H(X). \text{ Нека}$$

$$q(x_1, x_2, \dots, x_n) = \prod_{i=1}^n q(x_i) \text{ е друг закон на распределба над множеството } \{1, 2, \dots, m\}.$$

а) Да се определи кон што конвергира по веројатност

$$-\frac{1}{n} \log q(X_1, X_2, \dots, X_n).$$

б) Да се оцени кон што конвергира по веројатност

$$-\frac{1}{n} \log \frac{q(X_1, X_2, \dots, X_n)}{p(X_1, X_2, \dots, X_n)}.$$

Решение:

а) Бидејќи, $q(x_1, x_2, \dots, x_n) = \prod_{i=1}^n q(x_i)$ добиваме дека

$$\lim_{n \rightarrow +\infty} -\frac{1}{n} \log q(X_1, X_2, \dots, X_n) = -\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \log q(X_i).$$

Сега од слабиот закон на големите броеви, следува дека

$$\begin{aligned} & -\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \log q(X_i) \xrightarrow[n \rightarrow \infty]{\text{в.}} -E(\log q(X)) = -\sum_{x=1}^m p(x) \log q(x) \\ & = \sum_{x=1}^m p(x) \log \frac{p(x)}{q(x)} - \sum_{x=1}^m p(x) \log p(x) = D(p||q) + H(p). \end{aligned}$$

б) Слично како под а), од $p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i)$ и $q(x_1, x_2, \dots, x_n) =$

$\prod_{i=1}^n q(x_i)$ следува дека:

$$\lim_{n \rightarrow +\infty} -\frac{1}{n} \log \frac{q(X_1, X_2, \dots, X_n)}{p(X_1, X_2, \dots, X_n)} = \lim_{n \rightarrow +\infty} -\frac{1}{n} \sum_{i=1}^n \log \frac{q(X_i)}{p(X_i)}.$$

Сега,

$$\begin{aligned} & \lim_{n \rightarrow +\infty} -\frac{1}{n} \sum_{i=1}^n \log \frac{q(X_i)}{p(X_i)} \xrightarrow[n \rightarrow \infty]{\text{в.}} -E\left(\log \frac{q(X)}{p(X)}\right) = -\sum_{x=1}^m p(x) \log \frac{q(x)}{p(x)} \\ & = \sum_{x=1}^m p(x) \log \frac{p(x)}{q(x)} = D(p||q). \end{aligned}$$

□

Задача 3.4.3. Нека

$$X : \begin{pmatrix} 1 & 2 & 3 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}$$

и нека X_1, X_2, \dots се независни и еднакво распределени случајни променливи со иста распределба како X . Да се определи граничното однесување на

$$(X_1 X_2 \cdots X_n)^{1/n}.$$

Решение:

Нека $Y_n = (X_1 X_2 \cdots X_n)^{1/n}$. Тогаш

$$\log Y_n = \frac{1}{n} \log \prod_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n \log X_i \xrightarrow[n \rightarrow \infty]{\text{Б.}} E(\log X)$$

$$E(\log X) = (\log 1) \frac{1}{2} + (\log 2) \frac{1}{4} + (\log 3) \frac{1}{4} = \frac{1}{4} (1 + \log 3)$$

$$\log Y_n \xrightarrow[n \rightarrow \infty]{\text{Б.}} \frac{1}{4} (1 + \log 3).$$

Оттука, добиваме дека

$$Y_n \xrightarrow[n \rightarrow \infty]{\text{Б.}} 2^{\frac{1}{4}(1+\log 3)} = 2^{\frac{1}{4}} \cdot 2^{\log 3^{1/4}} = 2^{1/4} \cdot 3^{1/4} = 6^{1/4} = \sqrt[4]{6}.$$

□

3.5. Задачи

Задача 3.5.1. Експериментот се состои во фрлање тетраедар, чии страни се нумерирани со броевите 1, 2, 3, 4 и се набљудува страната со која тетраедарот паѓа на површината. Распределбата на паднатиот број е дадена со:

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1/4 & 1/6 & 1/4 & 1/3 \end{pmatrix}.$$

Нека $X_1, X_2, \dots, X_n, \dots$ се исходи од фрлања на тетраедарот. Да се определи граничното однесување на производот $X_1 X_2 \cdots X_n$, за големи вредности на n .

Задача 3.5.2. Нека (X_i, Y_i) се независни и еднакво распределени случајни вектори со распределба $p(x, y)$. На што е еднаква граничната вредност на

$$\frac{1}{n} \log \frac{\prod_{i=1}^n p(X_i) \prod_{i=1}^n p(Y_i)}{\prod_{i=1}^n p(X_i, Y_i)} ?$$

Глава 4

Вериги на Марков и рата на ентропија на случаен процес

Од својството за асимптотска рамномерност може да се заклучи дека $nH(X)$ бита се доволни за да се опишат n независни и еднакво распределени случајни променливи. Но, се поставува прашање, што ако случајните променливи не се независни? За да се одговори на ова прашање, најпрво ќе го дефинираме поимот случаен процес, како и поимот стационарност на случаен процес. Потоа, ќе разгледаме специјални случајни процеси, наречени вериги на Марков, каде што постои зависност од прв ред.

4.1. Дефиниција на случаен процес. Стационарност

Дефиниција 4.1. Нека (Ω, \mathcal{F}, P) е даден простор на веројатност и нека T е непразно множество. Ако X_t е случајна променлива дефинирана на просторот (Ω, \mathcal{F}, P) за секој $t \in T$, тогаш множеството $\{X_t, t \in T\}$ од случајни променливи се нарекува *случаен процес*.

Множеството T се нарекува *параметарско множество* и најчесто t е временски параметар. Да воочиме дека X_t е случајна променлива за секој t . Тоа значи дека за секој $t \in T$, X_t е функција од Ω во \mathbb{R} , која е \mathcal{F} -измерлива, т.е.

$$\{E : X_t(E) < x\} \in \mathcal{F}, \quad \text{за секој } x \in \mathbb{R}.$$

Оттука, случаен процес е функција $T \times \Omega \rightarrow \mathbb{R}$ којашто на секој пар (t, E) , го придружува реалниот број $X_t(E)$, т.е.

$$(t, E) \mapsto X_t(E) \in \mathbb{R}, \quad t \in T, E \in \Omega.$$

Ако се разгледува видот на параметарското множество, случајните процеси се делат на две групи:

- Случајни процеси со дискретно параметарско множество (случајни низи);
- Случајни процеси со непрекинато параметарско множество.

Од друга страна, ако се разгледува распределбата на случајните променливи X_t , за $t \in T$, случајните процеси, исто така, се делат на две групи:

- Случајни процеси од дискретен тип (дискретни случајни процеси), и
- Случајни процеси од апсолутно непрекинат тип.

Пример 4.1. Пациенти доаѓаат во лекарска ординација во случајни временски моменти. Нека X_n го означува времето (во часови) на чекање на n -тиот пациент пред тој да влезе на преглед кај лекарот. Тогаш $\{X_n | n \in \mathbb{N}\}$ е случаен процес од апсолутно непрекинат тип со дискретно параметарско множество. Имено, множеството вредности на случајниот процес е $R_{X_t} = \{x | x \geq 0\}$, а параметарското множество е $\mathbb{N} = \{1, 2, \dots\}$. \square

Секој случаен процес е напoлно определен со таканаречените конечно димензионални распределби.

Дефиниција 4.2. Функција на распределба од прв ред за случајниот процес $\{X_t, t \in T\}$ се дефинира со

$$F_1(t; x) = P\{X_t < x\},$$

за секој $x \in \mathbb{R}$ и секој $t \in T$,

Да воочиме дека за фиксно $t \in T$, $F_1(t; x)$ е функција на распределба на случајната променлива X_t .

Дефиниција 4.3. Функција на распределба од n -ти ред ($n \in \mathbb{N}$) за случајниот процес $\{X_t, t \in T\}$ се дефинира со

$$F_n(t_1, \dots, t_n; x_1, \dots, x_n) = P\{X_{t_1} < x_1, \dots, X_{t_n} < x_n\},$$

за секои $t_1, \dots, t_n \in T$ и секои $x_1, \dots, x_n \in \mathbb{R}$.

Од дефиницијата може да се заклучи дека за фиксни $t_1, \dots, t_n \in T$, $F_n(t_1, \dots, t_n; x_1, \dots, x_n)$ е функција на распределба на случајниот вектор $(X_{t_1}, \dots, X_{t_n})$.

Еден случаен процес е напoлно определен, ако се познати неговите функции на распределба од ред n , за секој $n \in \mathbb{N}$. Ако случајниот процес е од дискретен тип, тогаш распределбата од n -ти ред може да се изрази и преку соодветниот закон на распределба, т.е. со веројатностите

$$P\{X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_n} = x_n\},$$

за секои $t_1, t_2, \dots, t_n \in T$ и секои $x_1, \dots, x_n \in R_X$.

Едно од најкорисните својства кои може да ги има еден случаен процес е неговата стационарност.

Дефиниција 4.4. Случајниот процес $\{X_t, t \in T\}$ е *строго стационарен*, ако за произволен $n \in \mathbb{N}$, произволни $t_1, t_2, \dots, t_n \in T$ и произволен реален број h така што $t_1 + h, \dots, t_n + h \in T$, случајните вектори $(X_{t_1}, \dots, X_{t_n})$ и $(X_{t_1+h}, \dots, X_{t_n+h})$ имаат иста распределба, т.е.

$$F_n(t_1 + h, \dots, t_n + h; x_1, \dots, x_n) = F_n(t_1, \dots, t_n; x_1, \dots, x_n). \quad (4.1)$$

Строгата стационарност на случаен процес значи дека распределбата од n -ти ред не се менува со поместување по временската оска, за секој $n = 1, 2, \dots$

Пример 4.2. Да го разгледаме случајниот процес $\{X_t | t \in \mathbb{N}\}$, каде што $X_t = A$, а A е случајна променлива со функција на распределба $F_A(x)$. Ќе покажеме дека овој случаен процес е строго стационарен. Најпрво, за произволен $n \in \mathbb{N}$ и за произволни $t_1 < t_2 < \dots < t_n$, за функцијата на распределба од n -ти ред на овој случаен процес, се добива:

$$\begin{aligned} F_n(t_1, t_2, \dots, t_n; x_1, x_2, \dots, x_n) &= P\{X_{t_1} < x_1, X_{t_2} < x_2, \dots, X_{t_n} < x_n\} \\ &= P\{A < x_1, A < x_2, \dots, A < x_n\} \\ &= P\{A < \min\{x_1, x_2, \dots, x_n\}\} \\ &= F_A(\min\{x_1, x_2, \dots, x_n\}). \end{aligned}$$

Од друга страна, за произволен $h \in \mathbb{N}$,

$$\begin{aligned} F_n(t_1 + h, t_2 + h, \dots, t_n + h; x_1, x_2, \dots, x_n) &= P\{X_{t_1+h} < x_1, X_{t_2+h} < x_2, \dots, X_{t_n+h} < x_n\} \\ &= P\{A < x_1, A < x_2, \dots, A < x_n\} \\ &= P\{A < \min\{x_1, x_2, \dots, x_n\}\} \\ &= F_A(\min\{x_1, x_2, \dots, x_n\}) \\ &= F_n(t_1, t_2, \dots, t_n; x_1, x_2, \dots, x_n). \end{aligned}$$

Значи, распределбата од произволен n -ти ред е инваријантна во однос на транслација по временската оска, т.е. разгледуваниот процес е строго стационарен. \square

Пример 4.3. Да го разгледаме, сега, случајниот процес $\{X_t | t \in \mathbb{N}\}$, каде што $X_t = tA$, а $A \sim U(3, 7)$. Ќе покажеме дека овој процес не е строго стационарен. Имено,

$$\begin{aligned} F_n(t_1, t_2, \dots, t_n; x_1, x_2, \dots, x_n) &= P\{X_{t_1} < x_1, X_{t_2} < x_2, \dots, X_{t_n} < x_n\} \\ &= P\{t_1 A < x_1, t_2 A < x_2, \dots, t_n A < x_n\} \\ &= P\{A < x_1/t_1, A < x_2/t_2, \dots, A < x_n/t_n\} \\ &= P\{A < \min\{x_1/t_1, x_2/t_2, \dots, x_n/t_n\}\} \\ &= F_A(\min\{x_1/t_1, x_2/t_2, \dots, x_n/t_n\}). \end{aligned}$$

Од друга страна, за произволен $h \in \mathbb{N}$, имаме:

$$\begin{aligned} F_n(t_1 + h, t_2 + h, \dots, t_n + h; x_1, x_2, \dots, x_n) &= P\{X_{t_1+h} < x_1, X_{t_2+h} < x_2, \dots, X_{t_n+h} < x_n\} \\ &= P\{(t_1 + h)A < x_1, (t_2 + h)A < x_2, \dots, (t_n + h)A < x_n\} \\ &= P\{A < \min\{x_1/(t_1 + h), x_2/(t_2 + h), \dots, x_n/(t_n + h)\}\} \\ &= F_A(\min\{x_1/(t_1 + h), x_2/(t_2 + h), \dots, x_n/(t_n + h)\}). \end{aligned}$$

Оттука е јасно дека во општ случај,

$$F_n(t_1 + h, t_2 + h, \dots, t_n + h; x_1, x_2, \dots, x_n) \neq F_n(t_1, t_2, \dots, t_n; x_1, x_2, \dots, x_n).$$

Имено, ако се избере $x_1 = x_2 = \dots = x_n = 8$, $t_1 = t_2 = \dots = t_n = 2$ и $h = 2$, тогаш $\min\{x_1/t_1, x_2/t_2, \dots, x_n/t_n\} = 4$, а $F_A(4) = \frac{4-3}{7-3} = \frac{1}{4}$. Од друга страна, $\min\{x_1/(t_1+h), x_2/(t_2+h), \dots, x_n/(t_n+h)\} = 2$, а $F_A(2) = 0$, па имаме еден пример кога равенството (4.1) не важи, што е доволно да се заклучи дека тоа не важи во општ случај. Ова покажува дека распределбата од произволен n -ти ред не е инваријантна во однос на транслација по временската оска, т.е. разгледуваниот процес не е строго стационарен. \square

Ако разгледуваниот случаен процес е дискретен со дискретно параметарско множество, равенството (4.1) може да се запише во облик

$$P\{X_{1+h} = x_1, X_{2+h} = x_2, \dots, X_{n+h} = x_n\} = P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\},$$

за секој $h \in \mathbb{N}$.

4.2. Вериги на Марков

Наједноставен пример на случаен процес во кој постои зависност е дискретен случаен процес X_1, X_2, \dots во кој секоја случајна променлива зависи само од променливата што е пред неа, а не зависи од останатите променливи. Таквите процеси се нарекуваат верици на Марков.

Дефиниција 4.5. Дискретен случаен процес $\{X_n | n \in N\}$ се нарекува *верига на Марков*, ако за секој $n \in N$, и за секои $x_1, x_2, \dots, x_{n+1} \in R_X$, е точно равенството:

$$\begin{aligned} P\{X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_1 = x_1\} \\ = P\{X_{n+1} = x_{n+1} | X_n = x_n\} \end{aligned} \quad (4.2)$$

Равенството (4.2) покажува дека кај верига на Марков, иднината X_{n+1} не зависи од минатото X_{n-1} во однос на сегашноста X_n . Ако со веригата на Марков се опишува еден систем, тогаш X_n се нарекува состојба на системот во момент n . Во овој случај, равенството (4.2) означува дека распределбата на состојбите во моментот $n + 1$ зависи само од тоа во која состојба се наоѓа системот во моментот n , а не зависи од состојбите во кои се наоѓал системот во претходните моменти, т.е. не зависи од тоа како системот стасал до соодветната состојба во n -тиот момент.

Означуваме $p_{ij}^{(n)} = P\{X_n = j | X_{n-1} = i\}$. Овие веројатности се нарекуваат *веројатности на премин* од состојба i во состојба j во n -тиот момент на промена (n -тиот чекор). Ако овие веројатности се стават во матрица $\mathbf{P}^{(n)} = [p_{ij}^{(n)}]$, таа матрица се нарекува *матрица на преодни веројатности во n -тиот момент на промена*. Согласно со дефиницијата на верига на Марков, ако процесот има s состојби, елементите на оваа матрица ги задоволуваат следните својства:

$$0 \leq p_{ij}^{(n)} \leq 1, \quad i, j = 1, 2, \dots, s, \quad (4.3)$$

$$\sum_{j=1}^s p_{ij}^{(n)} = 1, \quad i = 1, 2, \dots, s. \quad (4.4)$$

Покрај ова, кај веригите на Марков, од интерес е и распределбата на веројатностите на состојбите во n -тиот момент. Означуваме,

$$p_j^{(n)} = P\{X_n = j\}, \quad j = 1, \dots, s,$$

каде

$$\sum_{j=1}^s p_j^{(n)} = 1, \quad n = 1, 2, \dots$$

Со ова е определен вектор $\mathbf{p}^{(n)} = (p_1^{(n)}, \dots, p_s^{(n)})$. За $p_j^{(n)}, j = 1, \dots, s$, имаме:

$$\begin{aligned} p_j^{(n)} &= P\{X_n = j\} = \sum_{i=1}^s P\{X_{n-1} = i, X_n = j\} \\ &= \sum_{i=1}^s P\{X_{n-1} = i\} P\{X_n = j \mid X_{n-1} = i\} = \sum_{i=1}^s p_i^{(n-1)} p_{ij}^{(n)} \end{aligned}$$

за $j = 1, \dots, s$. Овие равенства може да се претстават во матрична форма на следниот начин:

$$\begin{bmatrix} p_1^{(n)} & p_2^{(n)} & \dots & p_s^{(n)} \end{bmatrix} = \begin{bmatrix} p_1^{(n-1)} & p_2^{(n-1)} & \dots & p_s^{(n-1)} \end{bmatrix} \begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \dots & p_{1s}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} & \dots & p_{2s}^{(n)} \\ \dots & \dots & \ddots & \dots \\ p_{s1}^{(n)} & p_{s2}^{(n)} & \dots & p_{ss}^{(n)} \end{bmatrix}$$

или

$$\mathbf{p}^{(n)} = \mathbf{p}^{(n-1)} \cdot \mathbf{P}^{(n)}, \quad n = 1, 2, \dots \quad (4.5)$$

Нека $\mathbf{p}^{(0)} = (p_1^{(0)}, \dots, p_s^{(0)})$ е вектор на веројатности на почетната состојба на системот. Со последователна примена на формулата (4.5) се добива:

$$\mathbf{p}^{(n)} = \mathbf{p}^{(0)} \mathbf{P}^{(1)} \dots \mathbf{P}^{(n-1)} \mathbf{P}^{(n)}. \quad (4.6)$$

Дефиниција 4.6. Веригата на Марков се нарекува *хомогена* (или *временски инваријантна*), ако условните веројатности $p_{ij}^{(n)}$ не зависат од n , т.е. за секој $n = 1, 2, \dots$ важи

$$P\{X_n = j \mid X_{n-1} = i\} = P\{X_2 = j \mid X_1 = i\} = p_{ij}, \quad (4.7)$$

за сите $i, j \in R_X$.

Веројатностите $p_{ij}, i, j = 1, \dots, s$, се нарекуваат *веројатности на премин* од состојба i во состојба j за еден чекор. Матрицата формирана од овие веројатности се означува со $\mathbf{P} = [p_{ij}]$. Во овој случај, формулата (4.6) добива облик:

$$\mathbf{p}^{(n)} = \mathbf{p}^{(0)} \mathbf{P}^n \quad (4.8)$$

Во хомогена верига на Марков, означуваме:

$$p_{ij}(n) = P\{X_n = j | X_0 = i\}. \quad (4.9)$$

Тоа е веројатноста дека системот ќе помине од состојба i во состојба j за n чекори. Матрицата $\mathbf{P}(n) = [p_{ij}(n)]$ се нарекува *матрица на преодни веројатности за n чекори*.

За веројатностите на премин за 0 чекори, дефинираме

$$p_{ij}(0) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}.$$

Ваквото дефинирање на веројатностите за 0 чекори е логично, затоа што за 0 чекори системот не може да ја смени состојбата со позитивна веројатност. Тој мора да остане во состојбата во која се наоѓа со веројатност 1. Оттука, матрицата на премин за 0 чекори е единечна матрица, т.е. $\mathbf{P}(0) = \mathbf{E}$.

Веројатноста на премин од состојба i во состојба j за n чекори, може да се изрази на следниот начин:

$$\begin{aligned} p_{ij}(n) &= P\{X_n = j | X_0 = i\} \\ &= \sum_{k=1}^s P\{X_n = j, X_{n-1} = k | X_0 = i\} \\ &= \sum_{k=1}^s P\{X_{n-1} = k | X_0 = i\} P\{X_n = j | X_{n-1} = k, X_0 = i\}. \end{aligned}$$

Ако во последниот израз се искористи дека разгледуваниот случаен процес е верига на Марков, се добива дека

$$P\{X_n = j | X_{n-1} = k, X_0 = i\} = P\{X_n = j | X_{n-1} = k\},$$

па

$$\begin{aligned} p_{ij}(n) &= \sum_{k=1}^s P\{X_{n-1} = k | X_0 = i\} P\{X_n = j | X_{n-1} = k, X_0 = i\} \\ &= \sum_{k=1}^s P\{X_{n-1} = k | X_0 = i\} P\{X_n = j | X_{n-1} = k\} \\ &= \sum_{k=1}^s p_{ik}(n-1) p_{kj}. \end{aligned}$$

Овие равенства се познати како *равенства на Чепман–Колмогоров* и може да се протолкуваат на следниот начин: Системот ќе помине од состојба i во

состојба j за n чекори, ако помине од состојба i во состојба k за $n - 1$ чекори и во последниот (n -тиот) чекор помине од состојба k во состојба j . Притоа, k може да биде која било состојба од множеството $\{1, 2, \dots, s\}$, па затоа во последниот израз се сумира за сите $k = 1, 2, \dots, s$.

Овој систем равенства може да се запише во матрична форма:

$$\mathbf{P}(n) = \mathbf{P}(n - 1) \cdot \mathbf{P}, \quad \text{за секој } n \in \mathbb{N}. \quad (4.10)$$

Со рекурзивна примена на последното равенство се добива дека $\mathbf{P}(n) = \mathbf{P}^n$. Според ова, секоја конечна хомогена верига на Марков е определена, ако е познат векторот $\mathbf{p}(0)$ на веројатности на почетната состојба и матрицата $\mathbf{P} = [p_{ij}]$ на преодни веројатности за еден чекор.

Нека $\{X_n | n = 1, 2, \dots\}$ е хомогена верига на Марков. Тогаш за распределбата на случајниот вектор (X_0, X_1, \dots, X_n) се добива:

$$\begin{aligned} p(x_0, x_1, \dots, x_n) &= P\{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\} \\ &= P\{X_0 = x_0\} \cdot P\{X_1 = x_1 | X_0 = x_0\} \cdot P\{X_2 = x_2 | X_1 = x_1, X_0 = x_0\} \cdot \dots \\ &\quad \dots \cdot P\{X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0\} \\ &= P\{X_0 = x_0\} \cdot P\{X_1 = x_1 | X_0 = x_0\} \cdot P\{X_2 = x_2 | X_1 = x_1\} \cdot \dots \\ &\quad \dots \cdot P\{X_n = x_n | X_{n-1} = x_{n-1}\} \\ &= p_{x_0} \cdot p_{x_0 x_1} \cdot p_{x_1 x_2} \cdot \dots \cdot p_{x_{n-1} x_n}. \end{aligned}$$

Ако е можно со позитивна веројатност од произволна состојба на веригата да се стигне во која било друга состојба со конечен број чекори, тогаш велиме дека веригата е *иредуцибилна*.

Потребен и доволен услов за стационарност на една хомогена верига на Марков е даден со следната теорема.

Теорема 4.1. Конечна хомогена верига на Марков е строго стационарен процес, ако

$$\mathbf{p}^{(n)} = \mathbf{p}^{(0)}, \quad (4.11)$$

за секој $n = 1, 2, \dots$ □

Значи, хомогена верига на Марков е стационарен процес, ако векторот на веројатности на состојби во кој било момент се совпаѓа со векторот на веројатности на состојби во почетниот момент. Ако означиме $\mathbf{p}^{(n)} = \mathbf{p}^*$, за секој $n = 1, 2, \dots$, равенството (4.5), добива облик

$$\mathbf{p}^* = \mathbf{p}^* \mathbf{P}. \quad (4.12)$$

Според Теорема 4.1, конечна хомогена верига на Марков е стационарна, ако векторот $\mathbf{p}^{(0)} = (p_1^{(0)}, \dots, p_s^{(0)})$ на веројатности на почетната состојба е решение на матричната равенка (4.12). Во развиена форма, таа матрична равенка определува, всушност, хомоген систем од s линеарни равенки:

$$p_j^* = \sum_{k=1}^s p_k^* p_{kj}, \quad j = 1, \dots, s.$$

За да решението на овој систем равенки биде распределба на веројатност, мора да биде задоволен условот:

$$\sum_{j=1}^s p_j^* = 1. \quad (4.13)$$

Дефиниција 4.7. Секоја распределба $\mathbf{p}^* = (p_1^*, p_2^*, \dots, p_s^*)$ која го задоволува системот равенки

$$p_j^* = \sum_{i=1}^s p_i^* p_{ij} \quad (4.14)$$

се нарекува *стационарна распределба* на веригата, а веројатностите p_j^* , $j = 1, \dots, s$ се нарекуваат *стационарни* или *финални веројатности*.

Да воочиме дека системот линеарни равенки (4.14) може да има и повеќе решенија, т.е. да постојат повеќе стационарни распределби. Потребен услов за постоење на единствена стационарна распределба е матрицата \mathbf{P} на преодни веројатности за еден чекор да биде *регуларна*, а тоа значи дека постои природен број k , така што сите елементи во матрицата \mathbf{P}^k се строго позитивни. Тоа значи дека за k чекори, системот од која било состојба i може да помине во произволна состојба j , за $i, j = 1, 2, \dots, s$.

Пример 4.4. Нека веригата на Марков има две состојби и е зададена со матрица на преодни веројатности

$$\mathbf{P} = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix}.$$

Нека стационарната распределба е определена со вектор $\mathbf{p}^* = (p_1^*, p_2^*)$, каде p_i^* е стационарната веројатност на i -тата состојба, $i = 1, 2$. За определување на овие веројатности се добива систем во матрична форма:

$$\begin{bmatrix} p_1^* & p_2^* \end{bmatrix} = \begin{bmatrix} p_1^* & p_2^* \end{bmatrix} \cdot \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix},$$

или во развиена форма, заедно со нормирачкото равенство (4.13):

$$\begin{cases} p_1^* = (1 - \alpha)p_1^* + \beta p_2^* \\ p_2^* = \alpha p_1^* + (1 - \beta)p_2^* \\ p_1^* + p_2^* = 1 \end{cases}$$

Со едноставни пресметки, овој систем линеарни равенки добива облик:

$$\begin{cases} \alpha p_1^* = \beta p_2^* \\ p_1^* + p_2^* = 1 \end{cases}.$$

и неговите решенија се:

$$p_1^* = \frac{\beta}{\alpha + \beta}, \quad p_2^* = \frac{\alpha}{\alpha + \beta}.$$

Ако веригата на Марков има почетна состојба распределена согласно со стационарната распределба, тогаш разгледуваната верига е стационарен случаен процес.

Ентропијата на состојбата X_n во момент n е:

$$H(X_n) = H\left(\frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta}\right).$$

Но, ова не е рата (брзина) со која ентропијата $H(X_n)$ расте со зголемување на n . Зависноста помеѓу X_i има свое влијание во определувањето на таа рата. \square

4.3. Рата на ентропија

Ако е дадена низа од n случајни променливи, тогаш се поставува прашањето: Како се менува ентропијата со растење на n ? Затоа се дефинира рата на ентропија како рата (брзина) на промена. Дефинираме две величини:

$$H_n = \frac{1}{n} H(X_1, X_2, \dots, X_n) \quad (4.15)$$

$$h_n = H(X_n | X_{n-1}, X_{n-2}, \dots, X_1) \quad (4.16)$$

Величината H_n може да се толкува како просечна ентропија по буква во n -члена порака, а h_n како условна ентропија на n -тиот симбол кога се познати претходните $(n - 1)$ симболи.

Дефиниција 4.8. Рата на ентропија на случајниот процес $\{X_n | n = 1, 2, \dots\}$ се дефинира со:

$$H = \lim_{n \rightarrow +\infty} H_n = \lim_{n \rightarrow +\infty} \frac{1}{n} H(X_1, X_2, \dots, X_n), \quad (4.17)$$

Во продолжение ќе разгледаме неколку примери на случајни процеси и ќе ги определиме нивните рати на ентропија.

Пример 4.5. *Генератор на случајни броеви* (ГСБ). Нека ГСБ има m можни еднакверојатни излези, т.е. може да генерира која било буква од множеството $\{1, 2, \dots, m\}$ со иста веројатност. Тогаш ГСБ може да генерира m^n можни пораки со должина n , и сите ќе бидат генерирани со еднаква веројатност. Оттука, ако X_i е i -тата генерирана буква, тогаш $H(X_1, X_2, \dots, X_n) = \log m^n = n \log m$. Просечната ентропија по буква ќе биде:

$$H_n = \frac{1}{n} H(X_1, X_2, \dots, X_n) = \log m,$$

а бидејќи таа е константна за секој n , за ратата на ентропија се добива:

$$H = \lim_{n \rightarrow +\infty} H_n = \log m.$$

Значи, во овој случај, за било кој конечен број m , ратата на ентропија постои и е еднаква на $\log m$. \square

Пример 4.6. *Низа од независни и еднакво распределени случајни променливи.* Нека X_1, X_2, \dots се независни и еднакво распределени случајни променливи. Тогаш од верижното правило за ентропијата, независноста и еднаквата распределба на случајните променливи, добиваме:

$$\begin{aligned} H(X_1, X_2, \dots, X_n) &= \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1) \\ &\stackrel{\text{нез.}}{=} \sum_{i=1}^n H(X_i) \\ &\stackrel{\text{едн.}}{=} nH(X_1). \end{aligned}$$

Оттука,

$$H_n = \frac{1}{n} H(X_1, X_2, \dots, X_n) = H(X_1),$$

па и ратата на ентропија е $H = H(X_1)$. Може да се заклучи дека и во овој случај, ратата на ентропија постои и е еднаква на ентропијата на секоја случајна променлива. \square

Пример 4.7. Низа од независни, но не еднакви случајни променливи. Нека X_1, X_2, \dots се независни, но не и еднакво распределени случајни променливи. Тогаш

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1) \stackrel{\text{нез.}}{=} \sum_{i=1}^n H(X_i).$$

Во овој случај, бидејќи случајните променливи не се еднакво распределени, $H(X_i)$ не се еднакви. Ќе покажеме дека може да се избере низа случајни променливи X_1, X_2, \dots со одредени распределби, такви што $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(X_i)$ не постои. Нека случајните променливи X_i имаат распределба:

$$X_i : \begin{pmatrix} 0 & 1 \\ 1 - p_i & p_i \end{pmatrix},$$

каде што p_i не е константна вредност за секој i , туку е функција од i . Избираме:

$$p_i = \begin{cases} 0.5, & \text{ако } 2^{2k} < i \leq 2^{2k+1} \\ 0, & \text{ако } 2^{2k+1} < i \leq 2^{2k+2} \end{cases}, \quad k = 0, 1, 2, \dots$$

Ако $p_i = 0.5$, тогаш $H(X_i) = H(1/2, 1/2) = 1$, а ако $p_i = 0$, тогаш $H(X_i) = H(0, 1) = 0$. Во овој случај, може да се генерира експоненцијално долга низа каде што $H(X_i) = 1$, а по неа да следува експоненцијално долга низа каде што $H(X_i) = 0$. Значи, просечната вредност $\frac{1}{n} \sum_{i=1}^n H(X_i)$ ќе осцилира

помеѓу 0 и 1, па лимесот $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(X_i)$ нема да постои. Може да заклучиме дека, ратата H за овој процес не е дефинирана. \square

Во продолжение, ќе покажеме дека за стационарен случаен процес $\{X_n | n = 1, 2, \dots\}$, постои и друг начин на дефинирање на ратата на ентропија. Нека

$$h = \lim_{n \rightarrow +\infty} h_n = \lim_{n \rightarrow +\infty} H(X_n | X_{n-1}, X_{n-2}, \dots, X_1), \quad (4.18)$$

ако лимесот постои.

Ќе покажеме дека при одредени услови, $H = h$, но најпрво, во следната теорема, ќе покажеме дека лимесот од $H(X_n | X_{n-1}, \dots, X_1)$ постои, ако случајниот процес $\{X_n | n = 1, 2, \dots\}$ е стационарен.

Теорема 4.2. За стационарен случаен процес, низата условни ентропии $\{h_n\}_{n=1}^{+\infty}$ е опаѓачка низа и има лимес h .

Доказ: Со примена на Теорема 2.6 за условно редуцирање на ентропијата, се добива:

$$h_{n+1} = H(X_{n+1}|X_n, \dots, X_2, X_1) \leq H(X_{n+1}|X_n, \dots, X_2),$$

а ако се искористи стационарноста на случајниот процес, добиваме дека

$$\begin{aligned} h_{n+1} &= H(X_{n+1}|X_n, \dots, X_2, X_1) \leq H(X_{n+1}|X_n, \dots, X_2) \\ &= H(X_n|X_{n-1}, \dots, X_1) = h_n, \end{aligned}$$

за секој $n = 1, 2, \dots$. Оттука, низата $\{h_n\}_{n=1}^{+\infty}$ е опаѓачка низа. Исто така, таа е ограничена од долу со 0, бидејќи ентропијата е ненегативна величина. Значи, низата $\{h_n\}_{n=1}^{+\infty}$ е опаѓачка и ограничена од долу, па таа е конвергентна, т.е. нејзин лимес h постои. \square

Следната теорема е помошна теорема во овој дел. Ќе ја искористиме во доказот дека за стационарен случаен процес, $H = h$, па затоа овде ќе ја дадеме без доказ.

Теорема 4.3. (*Cesaro*) Нека $\{a_n\}_{n=1}^{+\infty}$ и $\{b_n\}_{n=1}^{+\infty}$ се две низи, така што

$$b_n = \frac{1}{n} \sum_{i=1}^n a_i,$$

за $n = 1, 2, \dots$. Ако $\lim_{n \rightarrow +\infty} a_n = a$, тогаш и $\lim_{n \rightarrow +\infty} b_n = a$. \square

Со ова имаме комплетна подготовка за да ја покажеме главната теорема во овој дел.

Теорема 4.4. За стационарен случаен процес, лимесите за H и h постојат и се еднакви, т.е. $H = h$.

Доказ: Со користење на верижното правило, се добива:

$$H_n = \frac{1}{n} H(X_1, X_2, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n H(X_i|X_{i-1}, \dots, X_1) = \frac{1}{n} \sum_{i=1}^n h_i,$$

т.е. просечната ентропија по буква е средина од условните ентропии. Но, условните ентропии h_n тежат кон h , кога $n \rightarrow +\infty$, па според Теорема 4.3 (Cesaro), добиваме:

$$H = \lim_{n \rightarrow \infty} H_n = \lim_{n \rightarrow \infty} h_n = h.$$

□

Значи, за стационарен случаен процес, ратата на ентропија може да се пресметува со една од формулите (4.17) или (4.17).

4.3.1. Рата на ентропија на верига на Марков

За стационарна верига на Марков, ратата на ентропија е дадена со:

$$H = h = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}, \dots, X_1) = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}) = H(X_2 | X_1),$$

каде што условната ентропија е пресметана со користење на дадената стационарна распределба. Овој резултат е изразен во следната теорема.

Теорема 4.5. Нека $\{X_n | n = 1, 2, \dots\}$ е стационарна верига на Марков со стационарна распределба $\mathbf{p}^* = (p_1^*, p_2^*, \dots, p_s^*)$. Тогаш ратата на ентропија е:

$$H = - \sum_i p_i^* \sum_j p_{ij} \log p_{ij}. \quad (4.19)$$

Доказ: Претходно добивме дека за стационарна верига на Марков, важи $H = H(X_2 | X_1)$. Со примена на дефиниција за условна ентропија во однос на распределбата \mathbf{p}^* , се добива:

$$H = H(X_2 | X_1) = \sum_i p_i^* H(X_2 | X_1 = i) = - \sum_i p_i^* \sum_j p_{ij} \log p_{ij}.$$

□

Пример 4.8. Во Пример 1, за верига на Марков со две состојби зададена со матрица на преодни веројатности:

$$\mathbf{P} = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix},$$

утврдиме дека стационарната распределба е

$$p_1^* = \frac{\beta}{\alpha + \beta}, \quad p_2^* = \frac{\alpha}{\alpha + \beta}.$$

Согласно со равенството (4.19), ратата на ентропија на оваа верига е

$$H = \frac{\beta}{\alpha + \beta} H(1 - \alpha, \alpha) + \frac{\alpha}{\alpha + \beta} H(\beta, 1 - \beta).$$

□

Ако веригата на Марков е редуцибилна и непериодична, тогаш се покажува дека таа има единствена стационарна распределба, т.е. системот линеарни равенки (4.14), заедно со нормирачката равенка, има единствено решение. Во случај кога таа стационарна распределба се совпаѓа со распределбата на состојби во почетниот момент, тогаш веригата е строго стационарна. Во спротивно, велиме дека постои *гранична стационарност*. Тогаш, која и да е почетната распределба, распределбата на веригата ќе тежи кон стационарната, кога $n \rightarrow +\infty$. Во тој случај, ратата на ентропија се пресметува со истите формули како и претходно.

4.4. Решени задачи

Задача 4.4.1. Нека X_0, X_1, X_2, \dots е низа независни дискретни случајни променливи и нека

$$Y_0 = X_0, \quad Y_n = X_0 + \dots + X_n, \quad n = 1, 2, \dots$$

Да се провери дали низата Y_0, Y_1, Y_2, \dots е верига на Марков.

Решение:

Од начинот како е дефинирана низата $Y_n, n = 0, 1, \dots$, веднаш може да се забележи дека $Y_k = Y_{k-1} + X_k$, т.е. $X_k = Y_k - Y_{k-1}$, за $k = 0, 1, \dots$. Нека $b_i \in R_{Y_i}$, за $i = 0, \dots, n$ и $n \geq 0$. Тогаш:

$$\begin{aligned}
& P\{Y_{n+1} = b_{n+1} \mid Y_n = b_n, \dots, Y_0 = b_0\} \\
&= \frac{P\{Y_0 = b_0, \dots, Y_n = b_n, Y_{n+1} = b_{n+1}\}}{P\{Y_0 = b_0, \dots, Y_n = b_n\}} \\
&= \frac{P\{Y_0 = b_0, Y_1 - Y_0 = b_1 - b_0, \dots, Y_{n+1} - Y_n = b_{n+1} - b_n\}}{P\{Y_0 = b_0, Y_1 - Y_0 = b_1 - b_0, \dots, Y_n - Y_{n-1} = b_n - b_{n-1}\}} \\
&= \frac{P\{X_0 = b_0, X_1 = b_1 - b_0, \dots, X_n = b_n - b_{n-1}, X_{n+1} = b_{n+1} - b_n\}}{P\{X_0 = b_0, X_1 = b_1 - b_0, \dots, X_n = b_n - b_{n-1}\}}
\end{aligned}$$

Ако во последното равенство се искористи независноста на случајните променливи X_0, X_1, X_2, \dots се добива:

$$\begin{aligned}
& P\{Y_{n+1} = b_{n+1} \mid Y_n = b_n, \dots, Y_0 = b_0\} \\
&= \frac{P\{X_0 = b_0\}P\{X_1 = b_1 - b_0\} \dots P\{X_n = b_n - b_{n-1}\}P\{X_{n+1} = b_{n+1} - b_n\}}{P\{X_0 = b_0\}P\{X_1 = b_1 - b_0\} \dots P\{X_n = b_n - b_{n-1}\}},
\end{aligned}$$

а по кратењето, имаме:

$$\begin{aligned}
& P\{Y_{n+1} = b_{n+1} \mid Y_n = b_n, \dots, Y_0 = b_0\} \\
&= P\{X_{n+1} = b_{n+1} - b_n\} \\
&\stackrel{\text{нез.}}{=} P\{X_{n+1} = b_{n+1} - b_n \mid Y_n = b_n\} \\
&= P\{Y_{n+1} - Y_n = b_{n+1} - b_n \mid Y_n = b_n\} \\
&= P\{Y_{n+1} = b_{n+1} \mid Y_n = b_n\}.
\end{aligned}$$

Согласно со дефиницијата на верига на Марков следува дека низата Y_0, Y_1, Y_2, \dots е верига на Марков. \square

Задача 4.4.2. Матрицата на преодни веројатности на хомогена верига на Марков е:

$$\mathbf{P} = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 \end{bmatrix}.$$

Во колку чекори може да се помине:

а) од втора во трета состојба,

б) од втора до четврта.

Решение:

Бројот на чекори за кој може да се помине од состојба i во состојба j е најмалиот број k , таков што веројатноста $p_{ij}(k) > 0$. Се користат равенствата на Чепмен–Колмогоров:

$$p_{ij}(n) = \sum_k p_{ik}(n-1)p_{kj}.$$

а)

$$p_{23}(1) = p_{23} = 0$$

$$p_{23}(2) = \sum_{k=1}^4 p_{2k}p_{k3} = \frac{1}{6} > 0,$$

што значи дека од втора во трета состојба може да се помине во два чекора.

б)

$$p_{24}(1) = p_{24} = 0,$$

$$p_{24}(2) = \sum_{r=1}^4 p_{2r}p_{r4} = 0,$$

$$p_{24}(3) = \sum_{k=1}^4 p_{2k}(2)p_{k4} = \sum_{k=1}^4 \left(\sum_{l=1}^4 p_{2l}p_{lk} \right) p_{k4} = \frac{1}{12} > 0,$$

т.е. од состојба 2 во состојба 4 може да се помине во 3 чекора.

□

Задача 4.4.3. Да претпоставиме дека секој човек според образованието може да се класифицира во три категории: високообразован, занаетчија и неквалификуван работник. Познато е дека од децата на високообразован човек, 80 %

се високообразовани, 10 % се занаетчии и 10 % се неквалификувани работници. Од децата на занаетчиите, 60 % се занаетчии, 20 % се високообразовани и 20 % се неквалификувани работници. На крај, од децата на неквалификуваните работници, 50 % се неквалификувани работници, а по 25 % се во другите две категории. Да претпоставиме дека секој човек има барем едно дете.

- а) Да се формира верига на Марков следејќи ја професијата на случајно избрано дете од дадено семејство низ неколку генерации. Да се определи матрицата на преодни веројатности за еден чекор.
- б) Да се определи веројатноста дека случајно избран внук на неквалификуван работник е високообразован човек.

Решение:

а) Да воочиме дека од условите на задачата, ако е позната категоријата на образованието на родителот, на категоријата на образование на детето не влијае категоријата на образование на роднините од повисоко колено (баба–дедо, итн.). Оттука, овој процес може да се разгледува како верига на Марков. Соодветната матрица на преодни веројатности е

$$\mathbf{P} = \begin{array}{c} \begin{array}{c} VO \\ Z \\ NR \end{array} \begin{array}{c} VO \\ Z \\ NR \end{array} \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.25 & 0.25 & 0.5 \end{bmatrix} \end{array} .$$

б) Веројатноста дека случајно избран внук на неквалификуван работник е високообразован човек е, всушност, веројатност на премин од состојба 3 во состојба 1 за два чекори и тоа е елементот $p_{31}(2)$ во матрицата \mathbf{P}^2 на преодни веројатности за два чекори. Бидејќи,

$$\mathbf{P}^2 = \begin{bmatrix} 0.685 & 0.165 & 0.150 \\ 0.330 & 0.430 & 0.240 \\ 0.375 & 0.300 & 0.325 \end{bmatrix} ,$$

наоѓаме дека $p_{31}(2) = 0.375$.

Да воочиме дека оваа веројатност може да се определи и директно со користење на равенствата на Чепман–Колмогоров (како во претходната задача),

без да се пресметува целата матрица \mathbf{P}^2 . Имено,

$$\begin{aligned} p_{31}(2) &= \sum_{k=1}^3 p_{3k}p_{k1} \\ &= 0.25 \cdot 0.8 + 0.25 \cdot 0.2 + 0.5 \cdot 0.25 \\ &= 0.375. \end{aligned}$$

□

Задача 4.4.4. (Еренфестов модел на дифузија.) Во затворен сад, поделен со мембрана на два дела (А и В), се наоѓаат вкупно $2a$ молекули. Секоја секунда, случајно, една молекула поминува од едниот во другиот дел од садот. Нека X_n го означува бројот на молекули во делот А од садот во моментот $n \in \mathbb{N}$. Да се утврди дека $\{X_n, n \in \mathbb{N}\}$ е хомогена верига на Марков, да се определи множеството вредности и матрицата на преодни веројатности за еден чекор.

Решение:

Да воочиме дека во која било секунда, бројот на молекули во делот А од садот зависи само од тоа колку молекули имало во тој дел во претходната секунда, а не и од начинот на кој молекулите дошле во тој дел од садот до тој момент. Оттука, јасно е дека состојбата на системот во секоја секунда зависи само од тоа во која состојба се наоѓал тој во претходната, т.е. разгледуваниот процес е верига на Марков.

Множеството вредности на веригата, т.е., бројот на молекули кои може да се најдат во делот А е $R_{X_n} = \{0, 1, \dots, 2a\}$.

Ќе ги дефинираме следните настани:

$C_i = \{\text{една молекула од делот А (во кој има } i \text{ молекули) поминува во В}\},$
 $D_i = \{\text{една молекула од делот В (во кој има } 2a - i \text{ молекули) поминува во А}\}.$

За преодните веројатности за еден чекор, имаме:

$$\begin{aligned} p_{i,i-1} &= P(C_i) = \frac{i}{2a}, & i &= 1, 2, \dots, 2a \\ p_{i,i+1} &= P(D_i) = \frac{2a-i}{2a}, & i &= 0, 1, \dots, 2a-1 \\ p_{ii} &= 0 \\ p_{ij} &= 0, & |i-j| &> 1 \end{aligned}$$

Матрицата на преодни веројатности е:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ \frac{1}{2a} & 0 & \frac{2a-1}{2a} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{2}{2a} & 0 & \frac{2a-2}{2a} & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \frac{2a-1}{2a} & 0 & \frac{1}{2a} \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix}.$$

Може да се воочи дека преодните веројатности (матрицата \mathbf{P}) не зависат од n , од каде што следува дека веригата на Марков е хомогена. Исто така, може да се воочи дека состојбата 0 и состојбата $2a$ се рефлектирачки состојби на системот. Секогаш кога системот ќе дојде во една од нив, во следниот момент, со веројатност 1, поминува во соседната состојба.

□

Задача 4.4.5. (Генетски модел) Се разгледува популација од $2m$ живи индивидуи и нејзиниот генетски развој. Нека $k \in \{0, 1, \dots, 2m\}$ го означува бројот на индивидуи кои имаат одредена особина А, додека $2m-k$ ја немаат таа особина. Идната генерација се формира по моделот „извлекување со враќање“ $2m$ пати од претходната генерација. Нека X_n е број на индивидуи со особина А во n -тото поколение. Да се утврди дека X_n е верига на Марков, да се определи множеството вредности и матрицата на преодни веројатности.

Решение:

Бидејќи секоја следна генерација се формира од претходната (и само од неа), јасно е дека разгледуваната низа од случајни променливи $\{X_n \mid n = 1, 2, \dots\}$ е верига на Марков. Исто така, бројот на индивидуи со особина А во секоја генерација не зависи од редниот број на генерацијата, туку само од бројот на такви индивидуи во претходната генерација. Значи, станува збор за хомогена верига на Марков. Множеството вредности на X_n , за $n = 1, 2, \dots$ е $R_{X_n} = \{0, 1, \dots, 2m\}$. За преодните веројатности за еден чекор добиваме:

$$p_{ij} = P\{X_n = j \mid X_{n-1} = i\} = \binom{2m}{j} p_i^j q_i^{2m-j}, \quad i, j = 0, 1, \dots, 2m,$$

каде што p_i е веројатноста на настанот дека ќе се извлече индивидуа со особина А, кога во популацијата има i такви индивидуи. Оттука,

$$p_i = \frac{i}{2m}, \quad i = 0, 1, \dots, 2m.$$

□

Задача 4.4.6. Во метеоролошки поглед денот може да се класифицира како ден со врнежи (состојба 1) и ден без врнежи (состојба 2). Врз основа на метеоролошките набљудувања, добиено е дека за временските промени во одредено место, важи следната законитост. Веројатноста дека по дождлив ден, ќе следува повторно дождлив ден е $1/4$. Веројатноста дека по сув ден ќе следува дождлив ден е $1/2$. Се претпоставува дека временската состојба на секој ден влијае само на временската состојба на следниот ден. Нека X_n е временската состојба n дена по денешниот (почетен) ден.

- а) Да се утврди дека X_n е хомогена верига на Марков и да се определи матрицата на преодни веројатности за еден чекор;
- б) Да се определат веројатностите дека третиот ден по денешниот ќе биде дождлив или сув, ако денешниот ден е дождлив;
- в) Да се определи веројатноста дека по денешниот дождлив ден и следните два ќе бидат дождливи;
- г) Да се определи векторот на стационарни веројатности, ако постои.

Решение: а) Со оглед на тоа што утрешната состојба зависи само од денешната, јасно е дека станува збор за верига на Марков со множество вредности $\{1, 2\}$. Од условите на задачата, $p_{11} = 1/4$ и $p_{21} = 1/2$. Притоа,

$$\begin{aligned} p_{12} &= 1 - p_{11} = 3/4 \\ p_{22} &= 1 - p_{21} = 1/2 \end{aligned}$$

Оттука, матрицата на преодни веројатности е:

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

и бидејќи не зависи од n , јасно е дека веригата е хомогена.

б) Ако денешниот (почетниот) ден е дождлив, тогаш векторот на почетни

веројатности е $\mathbf{p}^{(0)} = (1, 0)$, па векторот на веројатности $\mathbf{p}^{(3)}$, кој ги определува веројатностите третиот ден по денешниот да биде дождлив или сув, се определува со:

$$\mathbf{p}^{(3)} = \mathbf{p}^{(0)}\mathbf{P}^3 = [1 \ 0] \begin{bmatrix} 25/64 & 39/64 \\ 13/32 & 19/32 \end{bmatrix} = [25/64 \ 39/64].$$

Значи, веројатноста дека третиот ден по денешниот дождлив ден, ќе биде повторно дождлив е $25/64$, а $39/64$ е веројатноста дека третиот ден после денешниот дождлив ќе биде сув.

в) Треба да се определи $P\{X_1 = 1, X_2 = 1 \mid X_0 = 1\}$.

$$\begin{aligned} P\{X_1 = 1, X_2 = 1 \mid X_0 = 1\} &= \frac{P\{X_0 = 1, X_1 = 1, X_2 = 1\}}{P\{X_0 = 1\}} \\ &= \frac{P\{X_0 = 1\}P\{X_1 = 1 \mid X_0 = 1\}P\{X_2 = 1 \mid X_1 = 1\}}{P\{X_0 = 1\}} \\ &= p_{11} \cdot p_{11} = \frac{1}{16}. \end{aligned}$$

г) Во матрицата на преодни веројатности \mathbf{P} сите елементи се позитивни, па таа е регуларна матрица. Оттука, стационарните веројатности постојат и векторот на стационарни веројатности $\mathbf{p}^* = [p_1^* \ p_2^*]$ може да се добие како решение на системот линеарни равенки $\mathbf{p}^* = \mathbf{p}^*\mathbf{P}$ или во развиена форма:

$$\begin{cases} \frac{1}{4}p_1^* + \frac{1}{2}p_2^* = p_1^* \\ \frac{3}{4}p_1^* + \frac{1}{2}p_2^* = p_2^* \\ p_1^* + p_2^* = 1 \end{cases} .$$

Со решавање на овој систем се добива дека $p_1^* = \frac{2}{5}, p_2^* = \frac{3}{5}$.

□

Задача 4.4.7. Да се определи ратата на ентропија на верига на Марков, чија матрица на преодни веројатности е:

$$\mathbf{P} = \begin{bmatrix} 1 - \alpha & \alpha \\ 1 & 0 \end{bmatrix}.$$

Решение:

Прво се определува векторот на стационарни веројатности со решавање на следниот систем равенки, кој во матрична форма има облик:

$$[p_1^* \ p_2^*] = [p_1^* \ p_2^*] \begin{bmatrix} 1 - \alpha & \alpha \\ 1 & 0 \end{bmatrix},$$

или во развиена форма:

$$\begin{cases} p_1^* = (1 - \alpha)p_1^* + p_2^* \\ p_2^* = \alpha p_1^* \\ p_1^* + p_2^* = 1 \end{cases} \iff \begin{cases} \alpha p_1^* = p_2^* \\ p_1^* + p_2^* = 1 \end{cases}.$$

Со замена на првата равенка во втората се добива:

$$p_1^* + \alpha p_1^* = 1,$$

па оттука решенијата на системот равенки се $p_1^* = \frac{1}{1 + \alpha}$ и $p_2^* = 1 - p_1^* = \frac{\alpha}{1 + \alpha}$. Ратата на ентропија на веригата на Марков е:

$$H = \frac{1}{1 + \alpha} H(1 - \alpha, \alpha) + \frac{\alpha}{1 + \alpha} H(1, 0) = \frac{1}{1 + \alpha} [-(1 - \alpha) \log(1 - \alpha) - \alpha \log \alpha].$$

□

Задача 4.4.8. Дадена е хомогена верига на Марков со множество вредности $S = \{0, 1, 2\}$ и матрица на преодни веројатности:

$$\mathbf{P} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \end{bmatrix}.$$

За векторот на почетни веројатности $p^{(0)} = \left(\frac{1}{2} \ \frac{1}{3} \ \frac{1}{6}\right)$, да се пресмета:

- $P\{X_0 = 0, X_1 = 0, X_2 = 1, X_3 = 2\}$;
- $P\{X_2 = 1, X_3 = 2 \mid X_0 = 0, X_1 = 0\}$;
- $P\{X_3 = 2 \mid X_0 = 0\}$;

г) $\mathbf{p}^{(3)}$.

д) Да се определи рата на ентропија на дадената верига.

Решение:

а) Бидејќи низата $\{X_n, n \in \mathbb{N}\}$ е верига на Марков се добива:

$$\begin{aligned} & P\{X_0 = 0, X_1 = 0, X_2 = 1, X_3 = 2\} \\ &= P\{X_0 = 0\}P\{X_1 = 0 \mid X_0 = 0\}P\{X_2 = 1 \mid X_1 = 0\}P\{X_3 = 2 \mid X_2 = 1\} \\ &= p_0^{(0)} p_{00} p_{01} p_{12} \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{24}. \end{aligned}$$

б) Со примена на формулата за условна веројатност и резултатот под а), се добива:

$$\begin{aligned} P\{X_2 = 1, X_3 = 2 \mid X_0 = 0, X_1 = 0\} &= \frac{P\{X_0 = 0, X_1 = 0, X_2 = 1, X_3 = 2\}}{P\{X_0 = 0, X_1 = 0\}} \\ &= \frac{1/24}{P\{X_0 = 0\}P\{X_1 = 0 \mid X_0 = 0\}} \\ &= \frac{1/24}{p_0^{(0)} p_{00}} = \frac{1/24}{1/2 \cdot 1/2} = \frac{1}{6}. \end{aligned}$$

в) Бидејќи веригата на Марков е хомогена $P\{X_3 = 2 \mid X_0 = 0\} = p_{02}^{(3)}$ е всушност, елементот $p_{02}^{(3)}$ од матрицата \mathbf{P}^3 . Притоа,

$$\mathbf{P}^3 = \begin{bmatrix} 25/72 & 31/72 & 16/72 \\ 31/108 & 46/108 & 31/108 \\ 16/72 & 31/72 & 25/72 \end{bmatrix}.$$

$$\text{Значи, } p_{02}^{(3)} = \frac{16}{72} = \frac{2}{9}.$$

г)

$$\mathbf{p}^{(3)} = \mathbf{p}^{(0)}\mathbf{P}^3 = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} 25/72 & 31/72 & 16/72 \\ 31/108 & 46/108 & 31/108 \\ 16/72 & 31/72 & 25/72 \end{bmatrix},$$

$$\text{т.е., } \mathbf{p}^{(3)} = \begin{bmatrix} 397 & 463 & 436 \\ 1296 & 1296 & 1296 \end{bmatrix}.$$

д) Прво, се определува векторот на стационарните веројатности $\mathbf{p}^* = (p_0^*, p_1^*, p_2^*)$, кој се добива како решение на следниот систем равенки:

$$[p_0^* \ p_1^* \ p_2^*] = [p_0^* \ p_1^* \ p_2^*] \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \end{bmatrix},$$

или во развиена форма

$$\begin{cases} p_0^* = \frac{1}{2}p_0^* + \frac{1}{3}p_1^* \\ p_1^* = \frac{1}{2}p_0^* + \frac{1}{3}p_1^* + \frac{1}{2}p_2^* \\ p_2^* = \frac{1}{3}p_1^* + \frac{1}{2}p_2^* \\ p_0^* + p_1^* + p_2^* = 1 \end{cases}.$$

Решенијата на овој систем се: $p_0^* = \frac{2}{7}, p_1^* = \frac{3}{7}, p_2^* = \frac{2}{7}$. За ратата на ентропија се добива:

$$H = \frac{2}{7}H\left(\frac{1}{2}, \frac{1}{2}, 0\right) + \frac{3}{7}H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) + \frac{2}{7}H\left(0, \frac{1}{2}, \frac{1}{2}\right) = \frac{2}{7} + \frac{3}{7}\log 3 + \frac{2}{7} = 1.2507.$$

□

Задача 4.4.9. N црни и N бели топчиња се распоредени на случаен начин во две кутии, така што во секоја од нив има по N топчиња. Бројот на бели топчиња во првата кутија ја дефинира состојбата на системот. Секоја секунда се зема на случаен начин по едно топче истовремено од двете кутии и тие си ги менуваат местата.

- а) Да се определи множеството состојби на системот и матрицата на преходни веројатности;

- б) Да се определи системот равенки со кој се пресметуваат стационарните веројатности.

Решение:

а) Бројот на бели топчиња во првата кутија може да биде $0, 1, \dots, N$, па оттука множеството состојби на системот е $\{0, 1, \dots, N\}$. За преодните веројатности за еден чекор, се добива:

$$p_{ii} = 2 \frac{i}{N} \frac{N-i}{N}, \quad i = 0, 1, \dots, N,$$

$$p_{i,i-1} = \left(\frac{i}{N} \right)^2, \quad i = 1, \dots, N,$$

$$p_{i,i+1} = \left(\frac{N-i}{N} \right)^2, \quad i = 0, 1, \dots, N-1,$$

$$p_{i,j} = 0, \quad |i-j| > 1,$$

т.е., матрицата на преодни веројатности е:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ \frac{1}{N^2} & 2 \frac{1}{N} \frac{N-1}{N} & \frac{(N-1)^2}{N^2} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{2^2}{N^2} & 2 \frac{2}{N} \frac{N-2}{N} & \frac{(N-2)^2}{N^2} & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \frac{(N-1)^2}{N^2} & 2 \frac{N-1}{N} \frac{1}{N} & \frac{1}{N^2} \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix}.$$

- б) Векторот на стационарни веројатности $\mathbf{p}^* = [p_0^* \dots p_N^*]$ се добива како решение на матричната равенка $\mathbf{p}^* = \mathbf{p}^* \mathbf{P}$. За матрицата \mathbf{P} , системот добива облик:

$$\left\{ \begin{array}{l} p_0^* = \frac{1}{N^2} p_1^* \\ p_k^* = \frac{(N-k+1)^2}{N^2} p_{k-1}^* + 2 \frac{k}{N} \frac{N-k}{N} p_k^* + \frac{(k+1)^2}{N^2} p_{k+1}^*, \quad k = 1, 2, \dots, N-1 \\ p_N^* = \frac{1}{N^2} p_{N-1}^* \\ \sum_{k=0}^N p_k^* = 1 \end{array} \right.$$

□

4.5. Задачи

Задача 4.5.1. За хомогена верига на Марков со множество состојби $S = \{0, 1, 2\}$, матрицата на преодни веројатности е:

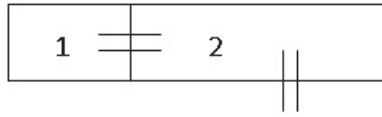
$$\mathbf{P} = \begin{bmatrix} 2/5 & 1/5 & 2/5 \\ 1/5 & 4/5 & 0 \\ 2/5 & 0 & 3/5 \end{bmatrix}.$$

Ако векторот на почетните веројатности е $\mathbf{p}^{(0)} = [1/3 \ 1/2 \ 1/6]$, да се определи:

- $P\{X_0 = 1, X_1 = 1, X_2 = 0, X_3 = 2\}$;
- $P\{X_2 = 0, X_3 = 2 | X_0 = 1, X_1 = 1\}$;
- $P\{X_2 = 2 | X_0 = 0\}$;
- $\mathbf{p}^{(2)}$;
- ратата на ентропија.

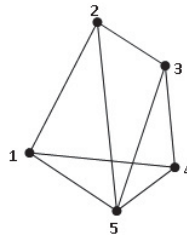
Задача 4.5.2. На случаен начин се движиме низ две простории со следниот облик:

При движењето просторијата секогаш се менува со веројатности кои се пропорционални на бројот на излези на просторијата во која се наоѓаме. Кога ќе се излезе надвор, нема повеќе враќање внатре.



- а) Нека состојбата на системот е просторијата (состојба 1 или 2) во која се наоѓаме, или надвор (состојба 0). Дали станува збор за верига на Марков?
- б) Да се најде матрицата на преодни веројатности за еден чекор.

Задача 4.5.3. На случаен начин се движиме низ графот даден на сликата. Притоа, во секој момент се случува движење од еден јазол во друг со веројатности кои се пропорционални на бројот на рабови кои излегуваат од јазолот во кој се наоѓаме.



- а) Нека состојбата на системот е јазолот во кој се наоѓаме. Дали станува збор за верига на Марков? (одговорот да се образложи)
- б) Да се најде матрицата на преодни веројатности за еден чекор.
- в) Да се определи ратата на ентропија.

Задача 4.5.4. Се разгледува случајно движење на крал на следната 3×3 табла за шах.

1	2	3
4	5	6
7	8	9

Кралот не може да остане во иста состојба и во секој момент се движи од едно поле во некое од соседните полиња со веројатности кои се пропорционални на бројот на соседни полиња.

- а) Нека состојбата на системот е бројот на полето во кое се наоѓа кралот. Дали станува збор за верига на Марков? (одговорот да се образложи)
- б) Да се определи матрицата на преодни веројатности за еден чекор.
- в) Да се определи ратата на ентропија.

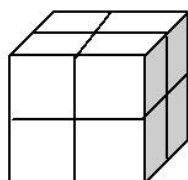
Задача 4.5.5. Дефинирана е следната постапка: во првиот чекор, коцката е поставена на страната со бројот 6; ако по n чекори коцката е на страна со бројот j , таа во $(n + 1)$ -от чекор се превртува на некоја од останатите пет страни со еднаква веројатност. Нека X_n е број на страната на која е коцката по n чекори.

- а) Да се определи множеството вредности на процесот $\{X_n, n \in \{1, 2, \dots\}\}$ и матрицата на веројатности за еден чекор.
- б) Дали $\{X_n, n \in \{1, 2, \dots\}\}$ е хомогена верига на Марков? (одговорот да се образложи)
- в) Да се определи $P\{X_2 = 3, X_3 = 4, X_5 = 3\}$.

Задача 4.5.6. Џубоксот во еден ресторан има диско, џез и рок музика. Ако е пуштена диско музика тогаш следниот избор ќе биде џез во 20 % од времето и рок во 30 %. Ако моменталниот избор е џез, следниот ќе биде џез во 40 % од времето и рок во 20 %. Ако моменталниот избор е рок, тогаш следниот ќе биде џез во 10 % од времето и рок во 60 %.

- а) Нека состојбата на системот е видот (жанрот) на песната која е пуштена на џубоксот. Дали станува збор за верига на Марков? (одговорот да се образложи)
- б) Да се определи матрицата на преодни веројатности за еден чекор.
- в) Ако во моментот е пуштена рок-музика, колкава е веројатноста дека по две песни (третата по моменталната) на џубоксот ќе биде пуштена диско-музика.

Задача 4.5.7. Една птица е изгубена во $2 \times 2 \times 2$ коцка (со 8 соби). Птицата лета од една во друга соседна соба поминувајќи со еднаква веројатност низ секој од ѕидовите. Во секој момент преминува од собата каде што е во една од соседните и притоа не може да излезе надвор од коцката. Коцката и начинот на нумерирање на собите се претставени на слика 4.1.



Прв кат

4	3
1	2

Втор кат

8	7
5	6

Слика 4.1

- а) Нека состојбата на системот е бројот на собата во која се наоѓа птицата. Дали станува збор за верига на Марков? (одговорот да се образложи)
- б) Да се оредели матрицата на преодни веројатности за еден чекор.

Задача 4.5.8. Точка може да се најде во некоја од положбите $1, 2, \dots, n$. Во моментите $0 < t_1 < t_2 < \dots < t_k < \dots$, точката може да има скок во левата соседна положба со веројатност $2/3$ или во десната соседна положба со веројатност $1/3$, ако се наоѓа во некоја од положбите $2, \dots, n - 1$. Ако точката е во положба 1, со веројатност $2/3$ останува во неа и со веројатност $1/3$ преминува во положба 2. Ако точката е во положба n , со веројатност $1/3$ останува во n и со веројатност $2/3$ преминува во $n - 1$.

- а) Да се определи матрицата на преодни веројатности.
- б) Да се определат стационарните веројатности за $n = 4$.

Задача 4.5.9. Се изведува низа од независни експерименти во кои настанот A се појавува со веројатност $p = 0.6$. Сметаме дека системот се наоѓа во состојба E_1 , ако во $(n - 1)$ -иот и n -тиот експеримент се реализира настанот $\bar{A}\bar{A}$, во состојба E_2 , ако се реализира $\bar{A}A$, во состојба E_3 , ако се реализира $A\bar{A}$ и во состојба E_4 , ако се реализира AA . Да се определат:

- а) преодните веројатности за еден чекор;
- б) преодните веројатности за n чекори;
- в) стационарните веројатности.

Глава 5

Компресија на податоци

Компресија на податоци е процес на кодирање на информациите со цел намалување на бројот на симболи потребни за претставување на податоците. Ваквото кодирање се нарекува *кодирање на изворот* и тоа се прави пред податоците да се пренесуваат или складираат. Компресирањето податоци може да заштеди простор за складирање, да го забрза преносот на датотеки и да ги намали трошоците за хардверот за складирање и за пропусниот опсег на мрежата.

Во оваа глава, ентропијата ќе биде искористена за поставување на граница до каде што може да оди компресирањето на податоци. Логично е дека компресирање податоци може да се постигне ако на пофреквентните излези од изворот на податоци се придружи пократок опис, а на помалку фреквентните излези – подолг опис. На пример, во Морзевата азбука, на најфреквентниот симбол придружена е една точка.

5.1. Дефиниција на код на изворот. Видови кодови

Дефиниција 5.1. Код C на изворот за случајна променлива X е пресликување од множеството вредности R_X на X во множеството $D^* = D \cup D^2 \cup D^3 \cup \dots$ (множеството од сите конечни стрингови над множеството D).

Притоа, D е конечно множество и се нарекува *азбука на кодот* или *кодна азбука*. Со $C(x)$ ќе го означуваме кодниот збор придружен на симболот $x \in R_X$, а со $l(x)$ ќе ја означуваме должината на $C(x)$.

На пример, ако $R_X = \{a, b, c\}$, тогаш со $C(a) = 00$, $C(b) = 11$ и $C(c) = 101$ е дефиниран код на изворот со кодна азбука $D = \{0, 1\}$.

Дефиниција 5.2. Очекувана должина $L(C)$ на кодот C на изворот за случајната променлива X со закон на распределба $p(x)$ се определува со:

$$L(C) = \sum_{x \in R_X} l(x)P\{X = x\} = \sum_{x \in R_X} p(x)l(x),$$

каде $l(x)$ е должината на кодниот збор придружен на x .

Без губење на општоста, ќе претпоставиме дека кодната азбука е $D = \{0, 1, \dots, d - 1\}$.

Пример 5.1. Во табелата подолу се дадени распределбата на случајната променлива X и кодните зборови придружени на елементите од R_X .

Распределба на X	Коден збор
$P\{X = 1\} = 1/2$	$C(1) = 0$
$P\{X = 2\} = 1/4$	$C(2) = 10$
$P\{X = 3\} = 1/8$	$C(3) = 110$
$P\{X = 4\} = 1/8$	$C(4) = 111$

За ентропијата на X , се добива:

$$\begin{aligned} H(X) &= -(1/2) \log(1/2) - (1/4) \log(1/4) - (1/8) \log(1/8) - (1/8) \log(1/8) \\ &= 1/2 + (1/4) \cdot 2 + (1/8) \cdot 3 + (1/8) \cdot 3 = 7/4 = 1.75 \text{ бита.} \end{aligned}$$

За очекуваната должина на кодните зборови, се добива:

$$L(C) = (1/2) \cdot 1 + (1/4) \cdot 2 + (1/8) \cdot 3 + (1/8) \cdot 3 = 1.75 \text{ бита.}$$

Значи, во овој случај $L(C) = H(X)$. Да напоменеме дека секоја низа од нули и единици може да биде еднозначно декодирана со симболи од X . На пример, низата 011110101100 се декодира со 142231. \square

Пример 5.2. Исто како и во претходниот пример, во табелата се дадени распределбата на случајната променлива X и кодните зборови придружени на елементите од R_X .

Распределба на X	Коден збор
$P\{X = 1\} = 1/3$	$C(1) = 0$
$P\{X = 2\} = 1/3$	$C(2) = 10$
$P\{X = 3\} = 1/3$	$C(3) = 11$

За ентропијата на X , се добива:

$$H(X) = \log 3 = 1.58 \text{ бита},$$

а очекуваната должина на кодните зборови е

$$L(C) = (1/3) \cdot 1 + (1/3) \cdot 2 + (1/3) \cdot 2 = 1.66 \text{ бита}.$$

Во овој случај $L(C) > H(X)$. Исто како и во претходниот пример, секоја низа од нули и единици може да биде еднозначно декодирана со симболи од X . \square

Во продолжение, ќе дефинираме видови на кодови кои задоволуваат сè построги услови.

Дефиниција 5.3. За еден код велиме дека е *несингуларен*, ако секој елемент од R_X се пресликува во различен елемент од D^* , т.е.

$$x_i \neq x_j \quad \Rightarrow \quad C(x_i) \neq C(x_j).$$

Несингуларноста значи дека на различни елементи од X се придружуваат различни кодни зборови. Но, обично се испраќаат цели низи од елементи на X . Во тој случај, проблем при декодирање на испратената низа е да се одвојат кодните зборови. Еден начин да се реши овој проблем е да се додаде некој специјален знак (на пример, запирка) со кој ќе се одделуваат кодните зборови еден од друг. Но, со тоа, очекуваната должина на кодот станува поголема и кодот е помалку ефикасен (за складирање на кодираната порака е потребен поголем простор или преносот на пораката трае подолго). Затоа се развива идејата за саморазделувачки или моментални кодови.

Дефиниција 5.4. *Проширување* C^* на кодот C е пресликување од множество стрингови над R_X со конечна должина над множество стрингови од D^* со конечна должина, дефинирано со:

$$C^*(x_1x_2 \dots x_n) = C(x_1)||C(x_2)||\dots||C(x_n),$$

каде $C(x_1)||C(x_2)||\dots||C(x_n)$ е конкатенација на соодветните кодни зборови.

На пример, ако $C(a) = 0$, а $C(b) = 11$, тогаш $C^*(ab) = 011$.

Дефиниција 5.5. За еден код се вели дека овозможува *еднозначно декодирање*, ако неговото проширување е несингуларно, т.е. за секоја низа \mathbf{y} од D^* , постои единствена низа \mathbf{x} со елементи од R_X која со C^* се пресликува во \mathbf{y} , т.е. $\mathbf{y} = C^*(\mathbf{x})$.

Дефиниција 5.6. Еден код има *својство на префикс* или се нарекува *моментален код*, ако не постои коден збор кој е префикс на друг коден збор.

Кај моментален код, симболот x_i може да се декодира веднаш кога ќе се прими кодниот збор кој соодветствува на него. Не е потребно да се види симболот (симболите) кој доаѓа потоа. Моменталниот код е саморазделувачки.

Пример 5.3. Во табелата подолу се дадени кодови кои задоволуваат некои од условите дефинирани претходно.

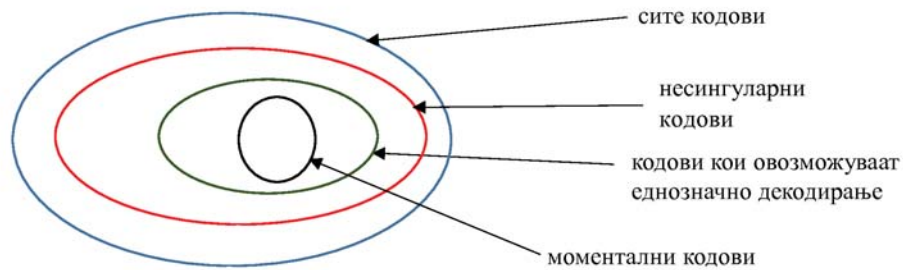
Вредности на X	Несингуларен код кој не овозможува еднозначно декодирање	Код кој овозможува еднозначно декодирање, но, не е моментален	Моментален код
1	0	10	0
2	1	00	10
3	00	11	110
4	11	110	111

Да воочиме дека првиот код не овозможува еднозначно декодирање. Имено, $C^*(1234) = 010011$, $C^*(121122) = 010011$, $C^*(12114) = 010011$ итн. Значи, постојат повеќе влезни низи кои се кодираат во 010011, па, ако на излез се добие 010011, нема да се знае што е пратено на влез. Вториот код овозможува еднозначно декодирање, т.е. секоја низа со елементи од $D = \{0, 1\}$ може еднозначно да се декодира, но тој не е моментален. Имено, $C(3) = 11$ е префикс на $C(4)$. Последниот код овозможува еднозначно декодирање и има својство на префикс, т.е. е моментален код.

Од дефинициите на кодовите и претходниот пример, може да се воочи дека класите од кодови се вгнездуваат една во друга. Тоа вгнездување е дадено на слика 5.1. □

5.2. Крафтово неравенство

Нашата цел во оваа глава е да најдеме начин за конструкција на моментален код кој ќе има најмала просечна должина на кодните зборови. Јасно е дека не може на сите симболи од азбуката на изворот да им придружиме најкраток



Слика 5.1

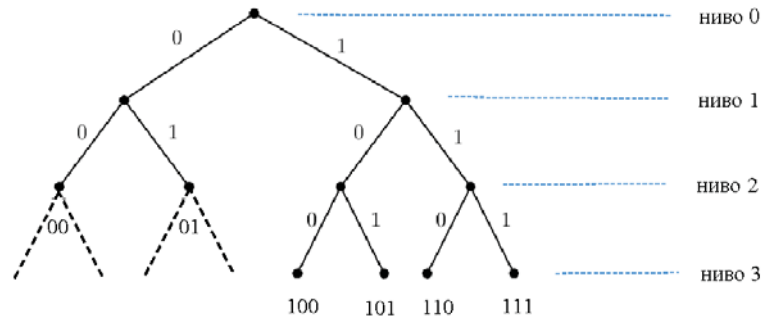
коден збор, и да го задржиме својството на префикс. Следната теорема дава потребен и доволен услов кој треба да го задоволуваат должините на кодните зборови за да се обезбеди моментален код.

Теорема 5.1. (Крафтово неравенство) За моментален код (код со својство на префикс) над азбука со d симболи, должините на кодните зборови l_1, l_2, \dots, l_m треба да го задоволуваат неравенството:

$$\sum_{i=1}^m d^{-l_i} \leq 1.$$

Обратно, за дадени должини l_1, l_2, \dots, l_m кои го задоволуваат горното неравенство, постои моментален код со тие должини на кодните зборови.

Доказ: Нека е даден моментален код над азбука со d симболи со должини на кодните зборови l_1, l_2, \dots, l_m . Се разгледува d -арно дрво во кое секој јазол има d деца. Нека гранките на дрвото ги претставуваат симболите во кодниот збор. На пример, d гранки кои потекнуваат од коренот претставуваат d можности за прв симбол во кодниот збор. Тогаш кодниот збор е претставен со лист од дрвото. Патот од коренот до листот ги трасира симболите во еден коден збор. На слика 5.2 е претставено бинарно дрво ($d = 2$). Бидејќи се разгледува код со својство на префикс, не треба да постои ни еден коден збор што е потомок на некој друг коден збор (во спротивно, вториот би бил префикс на првиот). Затоа, секој коден збор ги елиминира неговите наследници во дрвото како можни кодни зборови. Нека l_{max} е најголемата должина на коден збор во множеството од сите кодни зборови. Се разгледуваат сите



Слика 5.2

јазли на дрвото на ниво l_{max} . Некои од нив се кодни зборови, а некои се наследници на кодни зборови (па самите не се кодни зборови). Секој коден збор на ниво l_i има $d^{l_{max}-l_i}$ наследници на ниво l_{max} . Сите тие множества наследници (на сите кодни зборови) мора да се дисјунктни. Исто така, вкупниот број на јазли во тие множества мора да биде помал или еднаков на вкупниот број на листови $d^{l_{max}}$. Оттука, сумирајќи по сите кодни зборови, се добива:

$$\sum_i d^{l_{max}-l_i} \leq d^{l_{max}}.$$

Ако последното неравенство се подели со $d^{l_{max}}$, се добива:

$$\sum_i d^{-l_i} \leq 1,$$

што е Крафтовото неравенство.

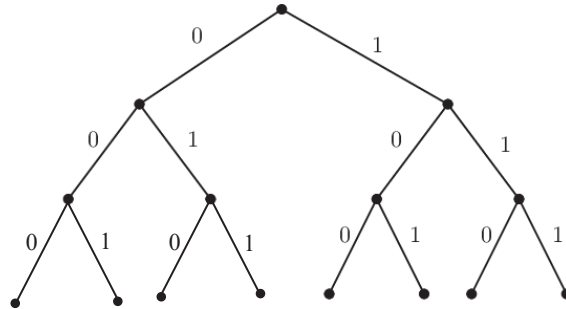
Обратно, нека е дадено множество должини l_1, l_2, \dots, l_m на кодни зборови кои го задоволуваат Крафтовото неравенство. Тогаш може да се конструира дрво, исто како претходното. Првиот јазол (лексикографски) на длабочина l_1 се зема како прв коден збор. Неговите наследници се отстрануваат од дрвото. Од преостанатите јазли, се зема првиот на длабочина l_2 за втор коден збор, итн. Со продолжување на постапката се добива код со својство на префикс со специфицирана должина на кодните зборови l_1, l_2, \dots, l_m . Крафтовото неравенство гарантира дека постапката ќе може да се спроведе до крај. \square

Пример 5.4. Ќе илустрираме како може да се конструира бинарен код со својство на префикс, ако должините на кодните зборови се 2, 2, 3. Да воочиме

дека овие должини го задоволуваат Крафтовото неравенство:

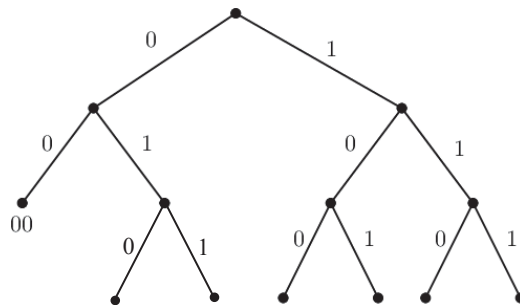
$$2^{-2} + 2^{-2} + 2^{-3} = \frac{1}{4} + \frac{1}{4} + \frac{1}{8} = \frac{5}{8} < 1.$$

Затоа што најголемата должина на кодниот збор е 3, доволно е да се разгледа бинарно дрво со длабочина 3, т.е. до ниво 3 (слика 5.3).



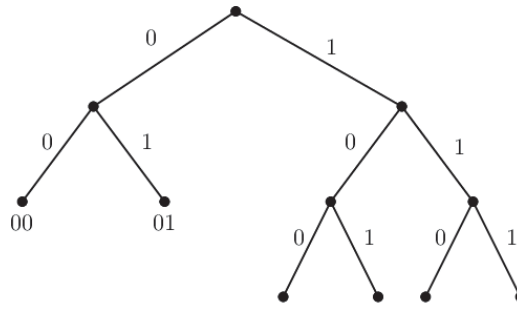
Слика 5.3

Првата должина е $l_1 = 2$, па се зема првиот јазол на длабочина 2 за коднен збор $C(x_1) = 00$. Неговите наследници се отстрануваат од дрвото (слика 5.4).



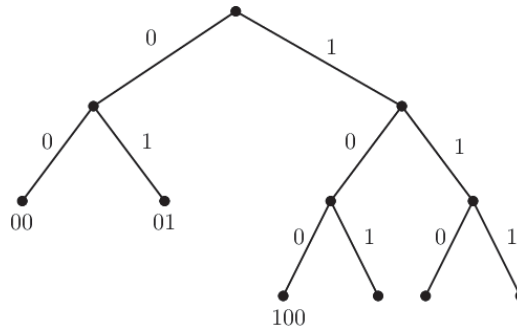
Слика 5.4

Втората должина е $l_2 = 2$. Од преостанатите јазли, се зема првиот на длабочина 2 за втор коднен збор $C(x_2) = 01$ и неговите наследници се отстрануваат од дрвото (слика 5.5).



Слика 5.5

Последната должина е $l_3 = 3$. Од преостанатите јазли, се зема првиот на длабочина 3 за трет коден збор $C(x_3) = 100$ (слика 5.6).



Слика 5.6

□

Следната теорема покажува дека бројот на кодни зборови може да биде бесконечен, а Крафтовото неравенство за нивните должини повторно да важи.

Теорема 5.2. (Проширено Крафтово неравенство) За секое преброиво множество кодни зборови за код со својство на префикс над азбука со d симболи, должините на кодните зборови l_1, l_2, \dots мора да го задоволуваат неравенството:

$$\sum_i d^{-l_i} \leq 1.$$

Обратно, за дадени должини l_1, l_2, \dots кои го задоволуваат горното неравенство, постои моментален код со тие должини на кодните зборови.

Доказ: Нека кодната азбука е $D = \{0, 1, \dots, d-1\}$. Го разгледуваме i -тиот коден збор $y_1y_2 \dots y_{l_i}$. Нека $0.y_1y_2 \dots y_{l_i}$ е реален број чиј развој во d -арен броен систем е:

$$0.y_1y_2 \dots y_{l_i} = \sum_{j=1}^{l_i} y_j d^{-j}.$$

Тој коден збор соодветствува на интервал

$$\left(0.y_1y_2 \dots y_{l_i}, 0.y_1y_2 \dots y_{l_i} + \frac{1}{d^{l_i}}\right).$$

Овој интервал ги содржи сите реални броеви чиј развој во d -арен броен систем започнува со $0.y_1y_2 \dots y_{l_i}$ и има должина d^{-l_i} . Овој интервал е подинтервал од $[0, 1]$. Поради својството на префикс овие интервали, соодветни на различни кодни зборови, се дисјунктни. Оттука, сумата на нивните должини е најмногу еднаква на 1. \square

5.3. Моментални оптимални кодови

Во претходното поглавје покажавме дека секое множество кодни зборови кое има својство на префикс, го задоволува и Крафтовото неравенство, т.е. Крафтовото неравенство е доволен услов за постоење код со специфицирана должина на кодните замени. Сега, се поставува следната цел, а тоа е определување на код со својство на префикс кој има минимална просечна должина на кодните зборови. Тоа значи дека се бараат должини l_1, l_2, \dots, l_m кои го задоволуваат Крафтовото неравенство и за кои очекуваната должина на кодните замени

$$L = \sum_{i=1}^m p_i l_i$$

е помала од очекуваната должина на кој било друг код со својство на префикс. Постапката за минимизирање на L е следната. Треба да се минимизира

$$L = \sum_{i=1}^m p_i l_i$$

по сите l_1, l_2, \dots, l_m кои го задоволуваат условот

$$\sum_{i=1}^m d^{-l_i} \leq 1.$$

Без губење на општоста во натамошните пресметки ќе претпоставиме дека

$$\sum_{i=1}^m d^{-l_i} = 1. \quad (5.1)$$

Во процесот на минимизација ќе го користиме методот со Лагранжови множителите за определување на врзан екстрем. Затоа, се минимизира функцијата:

$$J = \sum_{i=1}^m p_i l_i + \lambda \left(\sum_{i=1}^m d^{-l_i} \right).$$

Со диференцирање по l_i се добива:

$$\frac{\partial J}{\partial l_i} = p_i - \lambda d^{-l_i} \ln d.$$

Ако изводите се изедначат со 0, добиваме:

$$\frac{\partial J}{\partial l_i} = p_i - \lambda d^{-l_i} \ln d = 0,$$

т.е.

$$d^{-l_i} = \frac{p_i}{\lambda \ln d}, \quad i = 1, 2, \dots, m. \quad (5.2)$$

Со сумирање на претходните равенства за $i = 1, 2, \dots, m$ и со примена на ограничувањето 5.1, се добива

$$\sum_{i=1}^m d^{-l_i} = \sum_{i=1}^m \frac{p_i}{\lambda \ln d}$$

$$1 = \frac{1}{\lambda \ln d} \sum_{i=1}^m p_i$$

$$1 = \frac{1}{\lambda \ln d}$$

$$\lambda = \frac{1}{\ln d}.$$

Ако се замени последниот израз за λ во (5.2), се добива $p_i = d^{-l_i}$, па за оптималната должина на кодните зборови се добива:

$$l_i^* = -\log_d p_i. \quad (5.3)$$

Ако кодните замени се избераат со овие должини (доколку се цели броеви), просечната должина на кодните зборови ќе биде:

$$L^* = \sum_{i=1}^m p_i l_i^* = - \sum_{i=1}^m p_i \log_d p_i = H_d(X).$$

Но, за да важи последното равенство l_i мора да се цели броеви. Но, не сме секогаш во можност да избереме должини како во (5.3). Во тој случај, се избираат должини кои се „блиски“ до оптималните. Во продолжение, наместо со користење втори изводи да покажеме дека функцијата L има глобален минимум во l_i^* , $i = 1, 2, \dots, m$, тоа ќе го покажеме со следната теорема.

Теорема 5.3. Очекуваната должина L на моментален d -арен код за случајна променлива X , е поголема или еднаква од ентропијата $H_d(X)$, т.е.

$$L \geq H_d(X).$$

Притоа, равенство важи акко $d^{-l_i} = p_i$.

Доказ: Ќе покажеме дека разликата $L - H_d(X) \geq 0$, од каде што следува тврдењето.

$$\begin{aligned} L - H_d(X) &= \sum_{i=1}^m p_i l_i + \sum_{i=1}^m p_i \log_d p_i \\ &= - \sum_{i=1}^m p_i \log_d d^{-l_i} + \sum_{i=1}^m p_i \log_d p_i = \sum_{i=1}^m p_i (\log_d p_i - \log_d d^{-l_i}) \\ &= \sum_{i=1}^m p_i \log_d \frac{p_i}{d^{-l_i}} = \sum_{i=1}^m p_i \log_d \left(\frac{p_i}{d^{-l_i}} \cdot \frac{\sum_{j=1}^m d^{-l_j}}{\sum_{j=1}^m d^{-l_j}} \right) \\ &= \sum_{i=1}^m p_i \log_d \frac{p_i}{\sum_{j=1}^m d^{-l_j}} - \sum_{i=1}^m p_i \log_d \left(\sum_{j=1}^m d^{-l_j} \right) \\ &= \sum_{i=1}^m p_i \log_d \frac{p_i}{r_i} - \sum_{i=1}^m p_i \log_d c, \end{aligned}$$

каде $r_i = \frac{d^{-l_i}}{\sum_{j=1}^m d^{-l_j}}$, $c = \sum_{j=1}^m d^{-l_j} \leq 1$. Така,

$$\begin{aligned} L - H_d(X) &= D(p||r) - \log_d c \\ &= D(p||r) + \log_d \frac{1}{c} \geq 0. \end{aligned}$$

Да воочиме дека двата собирока во последниот ред се ненегативни, од каде што следува дека $L - H_d(X) \geq 0$. Притоа, равенство ќе важи ако $d^{-l_i} = p_i$, што е јасно од претходното изведување. \square

Дефиниција 5.7. Една распределба се нарекува d -арна, ако сите веројатности се од облик d^{-n} , за некој природен број n .

Значи, во Теорема 5.3 ќе важи равенството $L = H_d(X)$ ако распределбата на X е d -арна.

Во продолжение, ќе покажеме како може да се изберат должините на кодните зборови за да се добие код чија очекувана должина L на кодните зборови е помалку од 1 бит од долната граница, т.е.

$$H(X) \leq L < H(X) + 1.$$

Да повториме дека да се минимизира L , значи да се најдат должини l_1, l_2, \dots, l_m кои го задоволуваат Крафтовото неравенство и за кои очекуваната должина на кодните замени

$$L = \sum_{i=1}^m p_i l_i$$

е помала од очекуваната должина L на кој било друг код со својство на префикс. Утврдивме дека $L \geq H_d(X)$ и равенство важи ако $l_i = -\log_d p_i$, $i = 1, 2, \dots, m$. Но, бидејќи $-\log_d p_i$ не мора да е цел број, можеме да придружиме должина која е блиску до овој број, т.е.

$$l_i = \lceil -\log_d p_i \rceil, \quad (5.4)$$

каде што $\lceil x \rceil$ го означува најмалиот цел број што е поголем или еднаков на x .

Ќе покажеме дека овие должини го задоволуваат Крафтовото неравенство. Имено, од

$$-\log_d p_i \leq \lceil -\log_d p_i \rceil$$

следува дека

$$-(-\log_d p_i) \geq -\lceil -\log_d p_i \rceil,$$

т.е.

$$d^{-\lceil -\log_d p_i \rceil} \leq d^{-(-\log_d p_i)}.$$

Со примена на последното неравенство, добиваме:

$$\sum_{i=1}^m d^{-\lceil -\log_d p_i \rceil} \leq \sum_{i=1}^m d^{-(-\log_d p_i)} = \sum_{i=1}^m d^{\log_d p_i} = \sum_{i=1}^m p_i = 1,$$

што значи дека вака избраните должини го задоволуваат Крафтовото неравенство.

Од друга страна, од (5.4) следува дека

$$-\log_d p_i \leq l_i < -\log_d p_i + 1,$$

за секој $i = 1, 2, \dots, m$. Ако се помножат овие неравенствата со соодветното p_i и се сумира по i , се добива:

$$-\sum_{i=1}^m p_i \log_d p_i \leq \sum_{i=1}^m p_i l_i < -\sum_{i=1}^m p_i \log_d p_i + \sum_{i=1}^m p_i,$$

т.е.

$$H_d(X) \leq L < H_d(X) + 1.$$

Но, оптималниот код може само да биде подобар од кодот чии должини се определени со (5.4). Затоа важи следната теорема.

Теорема 5.4. Нека $l_1^*, l_2^*, \dots, l_m^*$ се оптимални должини на кодните зборови за распределба \mathbf{p} на изворот и нека L^* е очекуваната должина на оптималниот d -арен код. Тогаш

$$H_d(X) \leq L^* < H_d(X) + 1.$$

Доказ: Нека $l_i = \lceil -\log p_i \rceil$, $i = 1, 2, \dots, m$. Тогаш l_i го задоволуваат Крафтовото неравенство и притоа важи дека

$$H_d(X) \leq L < H_d(X) + 1.$$

Но, ако L^* е очекуваната должина на оптималниот код, тогаш $L^* \leq L$ и според Теорема 5.3, $L^* \geq H_d$. Оттука, добиваме дека

$$H_d(X) \leq L^* \leq L < H_d(X) + 1.$$

□

Од претходната теорема можеме да заклучиме дека очекуваната должина на кодот ја надминува долната граница за најмногу 1 бит. Тоа надминување на долната граница може да се редуцира, ако не се кодира симбол по симбол, туку се кодира цела низа од n симболи. Таквата низа од n симболи ја разгледуваме како еден *суперсимбол* од азбуката R_X^n . За почеток, претпоставуваме дека сите симболи во таа низа се независни и еднакво распределени. Нека L_n е очекуван број на кодни симболи по еден симбол од влезната порака, т.е. ако $l(x_1, x_2, \dots, x_n)$ е должината на кодниот збор придружен на (x_1, x_2, \dots, x_n) , тогаш

$$L_n = \frac{1}{n} \sum p(x_1, x_2, \dots, x_n) l(x_1, x_2, \dots, x_n) = \frac{1}{n} E(l(X_1, X_2, \dots, X_n)).$$

Сега, може да се применат границите изведени претходно, па

$$H(X_1, X_2, \dots, X_n) \leq E(l(X_1, X_2, \dots, X_n)) < H(X_1, X_2, \dots, X_n) + 1.$$

Ако X_1, X_2, \dots, X_n се независни и еднакво распределени случајни променливи, добиваме дека

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i) = nH(X).$$

Оттука,

$$nH(X) \leq E(l(X_1, X_2, \dots, X_n)) < nH(X) + 1,$$

па

$$H(X) \leq \frac{1}{n} E(l(X_1, X_2, \dots, X_n)) < H(X) + \frac{1}{n},$$

т.е.

$$H(X) \leq L_n < H(X) + \frac{1}{n}.$$

Добивме дека со кодирање на блок со поголема должина, очекуваниот број на кодни симболи по еден симбол од влезната порака се доближува до ентропијата. Колку n е поголемо, тоа приближување е подобро.

Истото може да се направи и за низа од симболи кои не мора да се независни и еднакво распределени. Имено, поаѓаме од неравенствата

$$H(X_1, X_2, \dots, X_n) \leq E(l(X_1, X_2, \dots, X_n)) < H(X_1, X_2, \dots, X_n) + 1.$$

Ако се подели со n се добива:

$$\frac{H(X_1, X_2, \dots, X_n)}{n} \leq L_n < \frac{H(X_1, X_2, \dots, X_n)}{n} + \frac{1}{n}.$$

Ако случајниот процес е стационарен, тогаш

$$\frac{1}{n} H(X_1, X_2, \dots, X_n) \rightarrow H$$

и очекуваната должина на кодните зборови по симбол, тежи кон ратата на ентропијата кога $n \rightarrow +\infty$. Со тоа е докажана следната теорема.

Теорема 5.5. Минималната очекувана должина по симбол во една n -члена порака задоволува:

$$\frac{H(X_1, X_2, \dots, X_n)}{n} \leq L_n < \frac{H(X_1, X_2, \dots, X_n)}{n} + \frac{1}{n}.$$

Уште повеќе, ако X_1, X_2, \dots, X_n е стационарен случаен процес, тогаш

$$L_n^* \rightarrow H,$$

каде што H е ратата на ентропија на случајниот процес.

Оваа теорема дава друго толкување на дефиницијата на ратата на ентропија. Имено, тоа е очекуван број на битови по симбол потребни да се опише процесот.

5.4. Хафманов код

Во продолжение ќе разгледаме еден едноставен алгоритам за конструкција на оптимален моментален код, т.е. оптимален код со својство на префикс. Овој алгоритам е познат како алгоритам на Хафман [5]. Подоцна ќе покажеме дека кој било друг код над иста азбука не може да има помала должина на кодните замени од кодот конструиран со овој алгоритам. Најпрво, Хафмановиот код ќе го воведеме на неколку примери.

Пример 5.5. Се разгледува случајна променлива X со распределба:

$$X : \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 \\ 0.25 & 0.25 & 0.2 & 0.15 & 0.15 \end{pmatrix}.$$

Се очекува оптималниот бинарен код за X да има најдолги кодни зборови за симболите α_4 и α_5 . Исто така, се очекува дека должината на овие два кодни збора мора да биде еднаква, бидејќи во спротивно ќе може да се скрати 1 бит од подолгиот коден збор и ќе се добие код со помала очекувана должина од претходниот.

Хафмановиот алгоритам е претставен во табелата подолу.

A	p_i	A_1	p_i	A_2	p_i	A_3	p_i
α_1	0.25	$\alpha_{4,5}$	0.3	$\alpha_{2,3}$	0.45	$\alpha_{1,4,5}$ 0	0.55
α_2	0.25	α_1	0.25	$\alpha_{4,5}$ 0	0.3	$\alpha_{2,3}$ 1	0.45
α_3	0.2	α_2 0	0.25	α_1 1	0.25		
α_4 0	0.15	α_3 1	0.2				
α_5 1	0.15						

Во првата колона на табелата се дадени буквите (симболите) од азбуката $A = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\}$, т.е. вредностите на случајната променлива X , подредени во нерастечки редослед на нивните веројатности, а во втората соодветните веројатности. На последните две букви α_4 и α_5 (оние со најмала веројатност) им се придружува по еден симбол од кодната азбука $D = \{0, 1\}$. Во следниот чекор тие две букви се спојуваат во една која се означува со $\alpha_{4,5}$ и на која се придружува веројатност 0.3, која е збир од веројатностите на споените симболи. Со тоа се формира нова азбука $A_1 = \{\alpha_1, \alpha_2, \alpha_3, \alpha_{4,5}\}$. Потоа, постапката се повторува со азбуката A_1 . Во следните две колони од табелата се дадени симболите од A_1 (подредени во нерастечки редослед на нивните веројатности) и соодветните веројатности. На последните две букви α_2 и α_3 им се придружува по еден симбол 0 и 1, соодветно. Во следниот чекор тие две букви се спојуваат во една која се означува со $\alpha_{2,3}$ и на која се придружува веројатност 0.45, која е збир од веројатностите на споените симболи. Со тоа се формира нова азбука $A_2 = \{\alpha_1, \alpha_{2,3}, \alpha_{4,5}\}$, итн. Постапката завршува кога ќе се добие азбука која содржи толку букви колку што има и кодната азбука D . Во овој случај, постапката завршува кога ќе се добие азбука со две букви. На тие букви им се придружува по еден симбол 0 и 1, соодветно. Кодниот

збор за симболот α_i се формира така што се земаат по ред, почнувајќи од десно кон лево во табелата, сите симболи од азбуката D кои се придружени кон индексот i . Така добиваме:

$$C(\alpha_1) = 01, \quad C(\alpha_2) = 10, \quad C(\alpha_3) = 11, \quad C(\alpha_4) = 000, \quad C(\alpha_5) = 001.$$

Очекуваната должина на овој код е

$$L = 2 \cdot 0.25 + 2 \cdot 0.25 + 2 \cdot 0.2 + 3 \cdot 0.15 + 3 \cdot 0.15 = 2.3 \text{ бита.}$$

□

Пример 5.6. Да се конструира тернарен код за истата случајна променлива од Пример 5.5. Се повторува истата постапка како во претходниот пример, само што сега, во секој чекор на последните три симбола од секоја азбука (оние со најмала веројатност) им се придружува по една буква од кодната азбука $D = \{0, 1, 2\}$ и тие се спојуваат во еден симбол. Кодирањето е дадено во следната табела.

A	p_i	A_1	p_i
α_1	0.25	$\alpha_{3,4,5}$ $\boxed{0}$	0.5
α_2	0.25	α_1 $\boxed{1}$	0.25
α_3 $\boxed{0}$	0.2	α_2 $\boxed{2}$	0.25
α_4 $\boxed{1}$	0.15		
α_5 $\boxed{2}$	0.15		

Кодните зборови придружени на буквите од азбуката A се:

$$C(\alpha_1) = 1, \quad C(\alpha_2) = 2, \quad C(\alpha_3) = 00, \quad C(\alpha_4) = 01, \quad C(\alpha_5) = 02,$$

а очекуваната должина на кодот е:

$$L = 1 \cdot 0.25 + 1 \cdot 0.25 + 2 \cdot 0.2 + 2 \cdot 0.15 + 2 \cdot 0.15 = 1.5$$

тернарни симболи. □

Од претходните примери може да се воочи дека во секој следен чекор, бројот на симболи во азбуката на изворот се намалува за $d-1$, а во последниот чекор, остануваат d симболи во азбуката на изворот. Ако $d \geq 3$, тогаш може да се случи во последниот чекор да нема d симболи, туку помалку. Во тој случај, нема да се добие оптимален код. За да се избегне тоа, во азбуката на изворот се додаваат фиктивни симболи. Ним им се придружува веројатност 0. Бројот на фиктивни симболи се определува на следниот начин:

- Ако m е бројот на симболи во азбуката на изворот, тогаш треба $m - 1$ да е деливо со $d - 1$.
- Ако овој услов не е исполнет, тогаш во азбуката на изворот се додаваат најмал број на симболи така што новодобиениот број на симболи го задоволува тој услов. Имено, бројот на фиктивни букви m' се определува така што се наоѓа остатокот c од делењето на $m - 1$ со $d - 1$, и тој остаток се одзема од $d - 1$, т.е.

$$m' = d - 1 - c. \quad (5.5)$$

Во продолжение е даден Хафмановиот алгоритам по чекори.

Хафманов алгоритам 1

Чекор 1. Ако m и d се такви што не постои природен број r , така што ќе важи $m - 1 = r(d - 1)$, т.е. $m = r(d - 1) + 1$, тогаш се додаваат толку фиктивни букви колку што е потребно равенството да важи. Бројот на фиктивни симболи m' се определува со равенството (5.5).

Чекор 2. Азбуката на изворот се подредува така што ќе важи $p_1 \geq p_2 \geq \dots \geq p_m$.

Чекор 3. На последните d симболи од подредената азбука на биективен начин им се придружува по еден симбол од кодната азбука D .

Чекор 4. Се формира нова азбука A_1 на изворот, така што се здружуваат последните d симболи од азбуката во еден симбол (се означува со индексите од сите здружени симболи) и му се придружува веројатност добиена со собирање на веројатностите на здружените симболи. Бројот на симболи во редуцираната азбука A_1 ќе биде

$$m_1 = m + m' - (d - 1) = m - (d + 1 - m') = m - c.$$

Чекор 5. Ако бројот на букви во A_1 е поголем од d , се оди на Чекор 2. Во спротивно, се оди на Чекор 6.

Чекор 6. Кодниот збор за симболот α_i се формира така што се земаат по ред, почнувајќи од десно кон лево во табелата, сите симболи од азбуката D кои се придружени кон индексот i .

После r чекори (каде r е количникот при делење на $m + m' - 1$ со $d - 1$) се доаѓа до редуцирана азбука A_r во која бројот на симболи е:

$$\begin{aligned} m_r &= m_{r-1} - (d - 1) \\ &= m_{r-2} - 2(d - 1) \\ &= \dots \\ &= m_1 - (r - 1)(d - 1) \\ &= m - c - (r - 1)(d - 1). \end{aligned}$$

Од $m - 1 = r(d - 1) + c$, се добива дека $m - c = r(d - 1) + 1$. Со замена во горното равенство, наоѓаме дека

$$m_r = r(d - 1) + 1 - (r - 1)(d - 1) = d - 1 + 1 = d.$$

Значи, во последниот чекор остануваат точно d симболи. Да напоменеме дека ако азбуката на кодот е бинарна, т.е. $d = 2$, тогаш нема потреба од додавање на фиктивни симболи, затоа што за кој било m , $m - 1$ е деливо со $d - 1 = 1$. Во тој случај, секогаш ќе постои цел број r , така што $m - 1 = r(d - 1)$, т.е. $r = m - 1$.

Пример 5.7. Ќе конструираме тернарен код за распределбата

$$X : \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 & \alpha_7 & \alpha_8 \\ 0.3 & 0.2 & 0.15 & 0.1 & 0.1 & 0.08 & 0.05 & 0.02 \end{pmatrix}.$$

Да воочиме дека во овој случај, $m = 8$, а $d = 3$ и $m - 1 = 7$ не е деливо со $d - 1 = 2$. Остатокот при делење на овие два броја е $c = 1$. За да добиеме оптимален код со користење на Хафмановиот алгоритам, треба да додадеме фиктивни симболи. Нивниот број е $m' = d - 1 - c = 3 - 1 - 1 = 1$, т.е. треба да се додаде еден фиктивен симбол на кој ќе се придружи веројатност 0.

A	p_i	A_1	p_i	A_2	p_i	A_3	p_i
α_1	0.30	α_1	0.3	α_1	0.30	$\alpha_{2,3,4}$ 0	0.45
α_2	0.20	α_2	0.20	$\alpha_{5,6,7,8}$	0.25	α_1 1	0.30
α_3	0.15	α_3	0.15	α_2 0	0.20	$\alpha_{5,6,7,8}$ 2	0.25
α_4	0.10	α_4	0.10	α_3 1	0.15		
α_5	0.10	α_5 0	0.10	α_4 2	0.10		
α_6	0.08	α_6 1	0.08				
α_7 0	0.05	$\alpha_{7,8}$ 2	0.07				
α_8 1	0.02						
— 2	0						

Сега, кодните зборови се:

$$C(\alpha_1) = 1, \quad C(\alpha_2) = 00, \quad C(\alpha_3) = 01, \quad C(\alpha_4) = 02, \\ C(\alpha_5) = 20, \quad C(\alpha_6) = 21, \quad C(\alpha_7) = 220, \quad C(\alpha_8) = 221.$$

Очекуваната должина на кодот е

$$L = 1 \cdot 0.30 + 2 \cdot 0.20 + 2 \cdot 0.15 + 2 \cdot 0.10 + 2 \cdot 0.10 + 2 \cdot 0.08 + 3 \cdot 0.05 + 3 \cdot 0.02 = 1.75$$

тернарни симболи. □

Постои и друга варијанта на Хафмановиот алгоритам во кој пред да се формираат кодните зборови се конструира дрво во кое буквите од азбуката на изворот се листови. Оваа варијанта на алгоритмот е дадена во продолжение.

Хафманов алгоритам 2

Чекор 1. Се земаат последните d најмалку веројатни симболи $\alpha_{m-d+1}, \alpha_{m-d+2}, \dots, \alpha_m$ од азбуката на изворот и се креира поддрво за кое овие d симболи се листови. Коренот на поддрвото се означува со сите индекси на споените симболи и му се доделува веројатност која е збир од веројатностите на споените јазли. На ребрата на поддрвото се додаваат симболите $0, 1, \dots, d-1$.

Чекор 2. Од азбуката се отстрануваат симболите $\alpha_{m-d+1}, \alpha_{m-d+2}, \dots, \alpha_m$, а се додава новиот симбол, со што се формира нова азбука A_1 .

Чекор 3. Ако во азбуката A_1 има повеќе од еден симбол, тогаш се оди на Чекор 1. Во спротивно, се оди на Чекор 4.

Чекор 4. Коднот збор за α_i се формира со спојување на симболите на сите ребра кои се на патот од коренот на дрвото во листот α_i .

Пример 5.8. Со конструкција на бинарно дрво ќе најдеме бинарен Хафманов код за изворот определен со случајната променлива

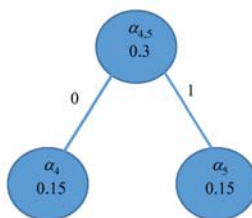
$$X : \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 \\ 0.25 & 0.25 & 0.2 & 0.15 & 0.15 \end{pmatrix}.$$

Буквите од азбуката ги претставуваме со јазли кои всушност ќе бидат листови на дрвото кое ќе го конструираме (слика 5.7).



Слика 5.7

Се земаат последните 2 најмалку веројатни симбола α_4 и α_5 од азбуката на изворот и се креира подрво, за кое овие 2 симбола се листови. Коренот на подрвото се означува $\alpha_{4,5}$ и му се доделува веројатност 0.3 која е збир од веројатностите на споените јазли. На ребрата на подрвото се додаваат симболите 0 и 1 (слика 5.8).



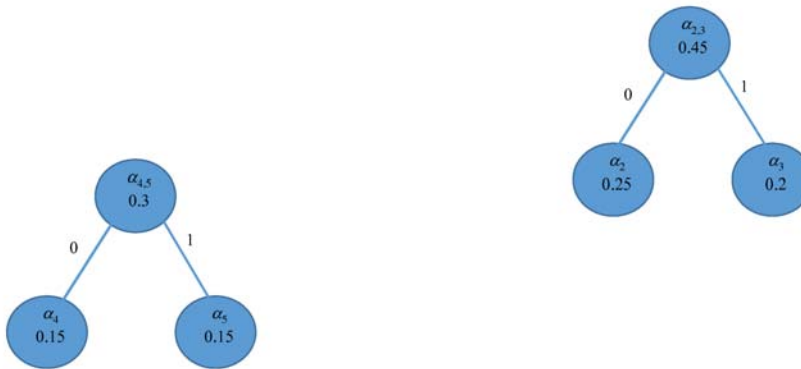
Слика 5.8

Се формира новата азбука $A_1 = \{\alpha_{4,5}, \alpha_1, \alpha_2, \alpha_3\}$ (слика 5.9). Се земаат последните 2 најмалку веројатни симбола α_2 и α_3 од азбуката A_1 и се креира подрво, за кое овие 2 симбола се листови. Коренот на подрвото се



Слика 5.9

означува $\alpha_{2,3}$ и му се доделува веројатност 0.45 која е збир од веројатностите на споените јазли. На ребрата на поддрвото се додаваат симболите 0 и 1 (слика 5.10). Со продолжување на постапката, сè додека не се добие азбука



Слика 5.10

со еден симбол, се добива дрвото на слика 5.11.

Кодниот збор за симболот α_i се добива со трасирање на симболите од коренот до листот соодветен на разгледуваниот симбол. Така,

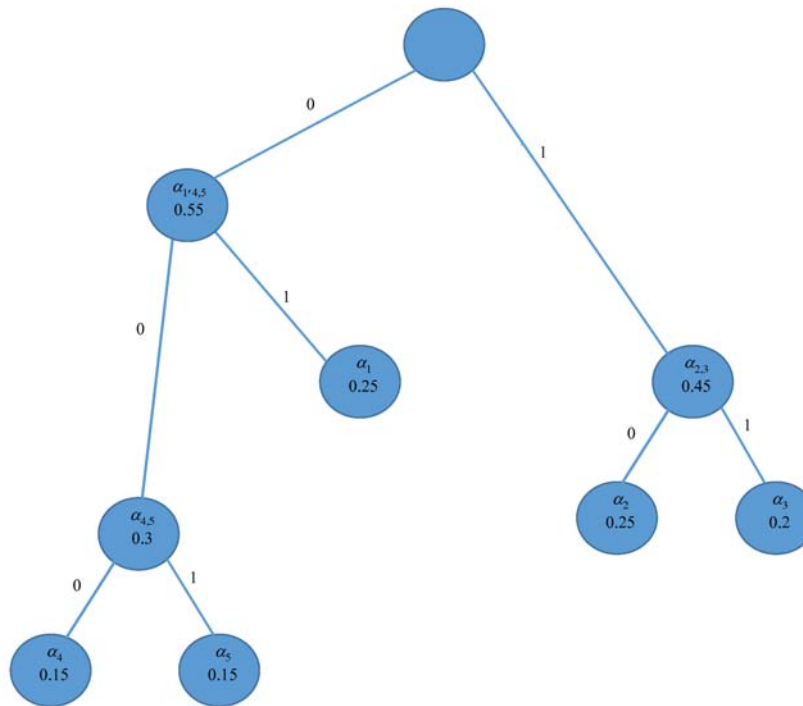
$$C(\alpha_1) = 01, \quad C(\alpha_2) = 10, \quad C(\alpha_3) = 11, \quad C(\alpha_4) = 000, \quad C(\alpha_5) = 001.$$

□

Од конструкцијата на дрвото, јасно е дека Хафмановиот код е моментален код. Сите кодни зборови се добиваат со трасирање на симболите од коренот до листовите, па нема коден збор кој е префикс на друг коден збор.

Со индукција ќе покажеме дека Хафмановиот код е оптимален. Значајно е да се нагласи дека постојат многу оптимални кодови:

- Ако се променат битовите (во бинарен код), 0 со 1, и обратно, тогаш повторно се добива оптимален код;
- Ако два кодни збора со иста должина си ги заменат местата, повторно се добива оптимален код.



Слика 5.11

Теорема 5.6. За произволна распределба, постои оптимален моментален бинарен код (со минимална очекувана должина) кој ги задоволува следните својства:

- i)* ако $p_j > p_k$, тогаш $l_j \geq l_k$;
- ii)* двата најдолги кодни збора имаат иста должина;
- iii)* двата најдолги кодни збора се разликуваат само во последниот бит и соодветствуваат на двата најмалку веројатни симбола.

Доказ: *i)* Нека $p_j > p_k$, за дадени индекси j и k . Да разгледаме нов код C'_m кој се добива кога кодните зборови j и k од C_m се заменуваат меѓусебе. Тогаш

$l'_j = l_k$, $l'_k = l_j$, а $l'_i = l_i$, за $i \neq j, k$. Сега, добиваме:

$$\begin{aligned} L(C'_m) - L(C_m) &= \sum_{i=1}^m p_i l'_i - \sum_{i=1}^m p_i l_i \\ &= p_j l'_j + p_k l'_k - p_j l_j - p_k l_k \\ &= p_j l_k + p_k l_j - p_j l_j - p_k l_k \\ &= p_j (l_k - l_j) - p_k (l_k - l_j) \\ &= (p_j - p_k)(l_k - l_j). \end{aligned}$$

Бидејќи C_m е оптимален код, мора $L(C'_m) - L(C_m) \geq 0$. Но, $p_j - p_k > 0$, па мора $l_k - l_j \geq 0$, т.е. $l_k \geq l_j$.

ii) Ако двата најдолги кодни збора немаат иста должина, тогаш може да се избрише последниот бит од најдолгиот збор (притоа запазувајќи го својството на префикс), со што ќе се добие код со помала очекувана должина на кодните зборови, што е контрадикција со оптималноста на кодот. Затоа, двата најдолги кодни збора мора да имаат иста должина.

iii) Ова својство не важи за сите оптимални кодови, но со преуредување на кодните зборови може да се најде код за кој тоа ќе важи. Кодните зборови кои се разликуваат само во последниот бит, ќе ги нарекуваме „браќа“. Ако постои коден збор со максимална должина кој нема „брат“, тогаш со бришење на последниот бит од него (притоа запазувајќи го својството на префикс), се добива код со помала очекувана должина на кодните зборови, што е повторно контрадикција со оптималноста на кодот. Значи, најдолгиот коден збор мора да има „брат“. \square

Да воочиме дека во Теорема 5.6 се тврди дека за произволна распределба може да се најде моментален оптимален бинарен код со разгледуваните својства. Оваа теорема може да се обопшти и за случај кога кодната азбука има d симболи. Во тој случај, d најдолги кодни збора имаат иста должина и сите тие се разликуваат само во последниот симбол.

Да воочиме дека Хафмановиот алгоритам обезбедува код кој ги задоволува својствата од теоремата. Имено, ако се конструира кодот со дрво, тогаш се поаѓа од симболите кои имаат најмала веројатност и тие се ставаат на најголема длабочина во дрвото. Колку веројатноста на симболот е поголема, толку тој симбол се наоѓа повисоко во дрвото, а со тоа должината на соодветниот коден збор е помала.

Исто така, јасно е дека на последно ниво се наоѓаат два јазли (ако станува збор за бинарен код) кои соодветствуваат на двата најмалку веројатни симбола

од азбуката на изворот. Па, кодните зборови на тие симболи ќе имаат иста должина и ќе се разликуваат само во последниот бит.

Во продолжение, ќе покажеме дека Хафмановиот код е оптимален код. Ќе разгледаме код во кој кодната азбука е бинарна, но заклучоците може да се обопштат за која било азбука со d симболи. Нека C_m е кодот формиран со Хафмановиот алгоритам (m е бројот на симболи во азбуката на изворот). Без губење на општоста, претпоставуваме дека $p_1 \geq p_2 \geq \dots \geq p_m$. Нека C_{m-1} е нов код за азбуката со $m - 1$ симбол, кој се добива кога од кодните зборови (кои се доделени со кодот C_m) на двата најмалку веројатни симбола се земе заедничкиот префикс и тој се придружи на симбол со веројатност $p_{m-1} + p_m$. Останатите кодни зборови од кодот C_m остануваат исти и во C_{m-1} . Веројатностите за појавување на соодветните симболи, кодните зборови и нивните должини за кодовите C_{m-1} и C_m се дадени во следната табела.

C_{m-1}			C_m		
веројатност	коден збор	должина	веројатност	коден збор	должина
p_1	w'_1	l'_1	p_1	$w_1 = w'_1$	$l_1 = l'_1$
p_2	w'_2	l'_2	p_2	$w_2 = w'_2$	$l_2 = l'_2$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
p_{m-2}	w'_{m-2}	l'_{m-2}	p_{m-2}	$w_{m-2} = w'_{m-2}$	$l_{m-2} = l'_{m-2}$
$p_{m-1} + p_m$	w'_{m-1}	l'_{m-1}	p_{m-1}	$w_{m-1} = w'_{m-1}0$	$l_{m-1} = l'_{m-1} + 1$
			p_m	$w_m = w'_{m-1}1$	$l_m = l'_{m-1} + 1$

За очекуваната должина на C_m се добива:

$$\begin{aligned}
 L(C_m) &= \sum_{i=1}^m p_i l_i \\
 &= \sum_{i=1}^{m-2} p_i l'_i + p_{m-1}(l'_{m-1} + 1) + p_m(l'_{m-1} + 1) \\
 &= \sum_{i=1}^{m-2} p_i l'_i + l'_{m-1}(p_{m-1} + p_m) + p_{m-1} + p_m \\
 &= L(C_{m-1}) + p_{m-1} + p_m.
 \end{aligned}$$

Добивме дека очекуваната должина на C_m се разликува од очекуваната должина на C_{m-1} за фиксна вредност која не зависи од C_{m-1} . Сега, минимизирањето на $L(C_m)$ се сведува на минимизирање на $L(C_{m-1})$. Со ова проблемот е редуциран на проблем со $m - 1$ симбол и закон на распределба $(p_1, p_2, \dots, p_{m-2}, p_{m-1} + p_m)$. Потоа, се продолжува на тој начин што се бара

оптимален код за $m - 2$ симбола со соодветен закон на распределба, што се добива со спојување на двата збора со најмала веројатност од претходно споената листа. Продолжувајќи на тој начин, на крај, проблемот се редуцира на проблем со 2 симбола, чие решение е очигледно: се придружува 0 на едниот и 1 на другиот симбол.

Бидејќи во секој чекор на редукција е застапена оптималноста, конструираниот код за m симболи, ќе биде оптимален. Со ова е покажана следната теорема.

Теорема 5.7. Хафмановиот код е оптимален, т.е. ако C^* е Хафманов код и C' е некој друг код, тогаш

$$L(C^*) \leq L(C').$$

□

Да напоменеме уште еднаш дека доказот е изведен за бинарна кодна азбука. Тој може да се обопшти со конструкција на оптималност за d -арна азбука на ист начин како претходно, само што сега на последните d симболи со најмала веројатност ќе им се доделуваат кодни зборови кои се разликуваат во последниот симбол.

5.5. Шенон–Фано–Елиас код

Во ова поглавје, ќе воведеме нов код кој има својство на префикс (го задоволува Крафтовото неравенство) и чија очекувана должина е блиска до оптималната. Претходно покажавме дека ако се стави должините на кодните зборови да се $l(x) = \lceil -\log p(x) \rceil$, тогаш тие го задоволуваат Крафтовото неравенство и може да се конструира код на изворот кој овозможува еднозначно декодирање.

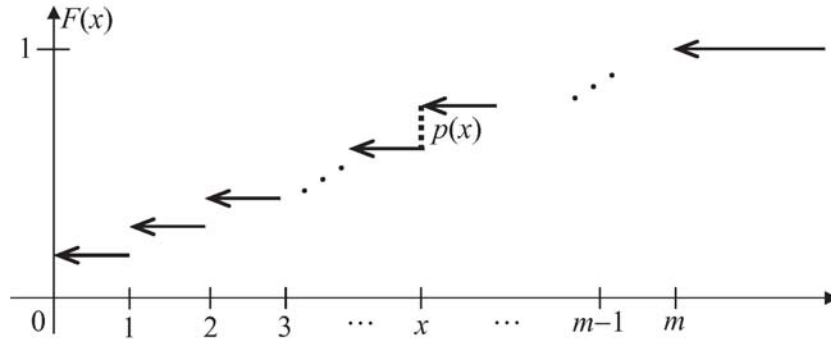
Нека симболите од азбуката на изворот се означат со $1, 2, \dots, m$, т.е. $R_X = \{1, 2, \dots, m\}$. Претпоставуваме дека $p(x) > 0$, за сите x . Функцијата на распределба на X е дефинирана со:

$$F(x) = \sum_{a < x} p(a).$$

Графикот на $F(x)$ е даден на слика 5.12.

Разгледуваме нова функција $F_1(x)$ дефинирана со:

$$F_1(x) = \sum_{a < x} p(a) + \frac{1}{2}p(x),$$



Слика 5.12

каде што $F_1(x)$ ја означува сумата на веројатностите на сите симболи помали од x плус половина од веројатноста на симболот x , т.е.

$$F_1(x) = F(x) + \frac{p(x)}{2}.$$

Да воочиме дека $p(x) = 0$, за секој $x \notin R_X$, т.е. $F_1(x) = F(x)$, за секој $x \notin R_X$. Оттука, графикот на функцијата $F_1(x)$ ќе се разликува од графикот на функцијата $F(x)$ само во точките $x \in R_X$, каде што за дадено x , ќе прима вредност $F(x) + \frac{p(x)}{2}$.

Бидејќи сите веројатности $p(x)$, за $x \in R_X$ се позитивни, $F_1(a) \neq F_1(b)$, ако $a \neq b$, за $a, b \in R_X$. Оттука, ако е позната $F_1(x)$, може еднозначно да се определи $x \in R_X$. Затоа, децималниот дел од вредноста на $F_1(x)$ (претставена во бинарен броен систем) може да се користи како коден збор за $x \in R_X$. Но, во општ случај, $F_1(x)$ е реален број со бесконечен број на битови (децимали). Затоа, не е ефикасно да се користи точната вредност на $F_1(x)$ како коден збор за x . Се поставуваат следните прашања: Ако се користи приближна вредност на $F_1(x)$, дали добиениот код ќе ги има бараните својства? И како да се избере $l(x)$ за да се добијат тие својства?

Да претпоставиме дека $F_1(x)$ е претставен во бинарен броен систем и е заокружен на $l(x)$ децимали. Тоа го означуваме со $\lfloor F_1(x) \rfloor_{l(x)}$. Нека, првите $l(x)$ бинарни децимали на $F_1(x)$ ги користиме како коден збор за x . Од начинот на кој е дефинирано заокружувањето, добиваме:

$$F_1(x) - \lfloor F_1(x) \rfloor_{l(x)} < \frac{1}{2^{l(x)}}. \quad (5.6)$$

Ако $l(x) = \lceil -\log_2 p(x) \rceil + 1$, тогаш се добива:

$$\begin{aligned} l(x) &> -\log_2 p(x) + 1 \\ 1 - l(x) &< \log_2 p(x) \\ 2^{1-l(x)} &< p(x), \end{aligned}$$

т.е.

$$\frac{1}{2^{l(x)}} < \frac{p(x)}{2} = F_1(x) - F(x) \quad (5.7)$$

Првото прашање кое се поставува за вака дефинираниот код за случајна променлива X е дали кодот овозможува еднозначно декодирање, т.е. дали $l(x) = \lceil -\log_2 p(x) \rceil + 1$ децимали се доволни за да се овозможи еднозначно декодирање? Во продолжение, ќе покажеме дека одговорот на ова прашање е потврден. Имено, од (5.6) следува дека

$$F_1(x) - \frac{1}{2^{l(x)}} < \lfloor F_1(x) \rfloor_{l(x)}, \quad (5.8)$$

а од неравенството (5.7) добиваме:

$$F(x) < F_1(x) - \frac{1}{2^{l(x)}}. \quad (5.9)$$

Со комбинирање на (5.6) и (5.7), се добива:

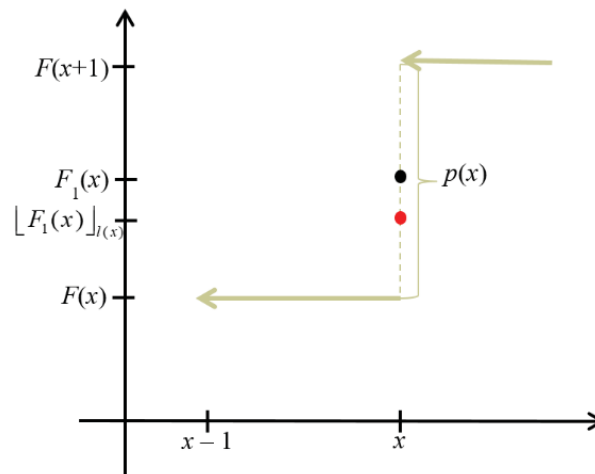
$$F(x) < F_1(x) - \frac{1}{2^{l(x)}} < \lfloor F_1(x) \rfloor_{l(x)} \leq F_1(x).$$

Така, $\lfloor F_1(x) \rfloor_{l(x)}$ лежи во чекорот кој соодветствува на x (слика 5.13), т.е. ако е познато $\lfloor F_1(x) \rfloor_{l(x)}$, тогаш еднозначно може да се определи x . Тоа значи дека $l(x)$ децимали се доволно за да се добие еднозначно декодирање на x .

Во продолжение, ќе покажеме дека вака дефинираниот код е моментален код, т.е. код со својство на префикс. За да го покажеме тоа, секој коден збор $\lfloor F_1 \rfloor_{l(x)} = z_1 z_2 \dots z_{l(x)}$ го претставуваме со интервалот

$$\left[0.z_1 z_2 \dots z_{l(x)}, 0.z_1 z_2 \dots z_{l(x)} + \frac{1}{2^{l(x)}} \right) = \left[\lfloor F_1 \rfloor_{l(x)}, \lfloor F_1 \rfloor_{l(x)} + \frac{1}{2^{l(x)}} \right).$$

Овој интервал ги содржи сите реални броеви кои што имаат исти први $l(x)$ децимали во бинарен запис. Кодот има својство на префикс ако интервалите придружени на кодните зборови се дисјунктни. Според (5.7), интервалот



Слика 5.13

придружен на x има должина $2^{-l(x)}$, што е помалку од половина од висината на чекорот придружен на x , бидејќи важи

$$\frac{1}{2^{l(x)}} < \frac{p(x)}{2}.$$

Од друга страна,

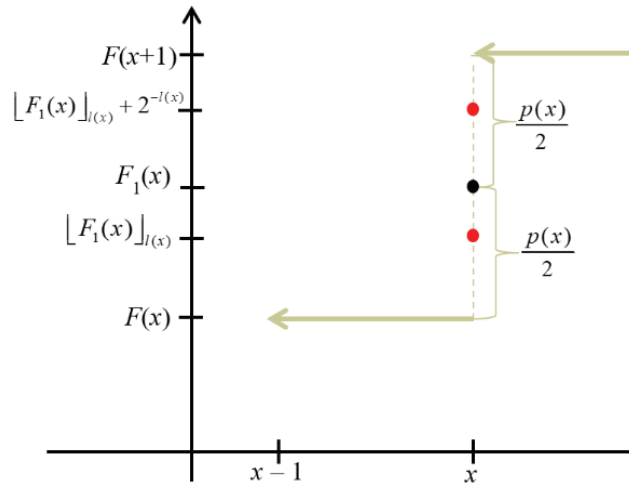
$$\frac{1}{2^{l(x)}} < \frac{p(x)}{2} = F(x+1) - F_1(x).$$

Оттука,

$$F(x+1) > F_1(x) + \frac{1}{2^{l(x)}} > [F_1(x)]_{l(x)} + \frac{1}{2^{l(x)}}.$$

Тоа значи, дека интервалот $\left[[F_1]_{l(x)}, [F_1]_{l(x)} + \frac{1}{2^{l(x)}} \right)$ е комплетно внатре во интервалот $[F(x), F(x+1)]$ што соодветствува на интервалот за симболот x (слика 5.14). Тоа покажува дека интервалите придружени на кодните зборови нема да имаат пресек, т.е. тие се дисјунктни интервали, па кодот има својство на префикс.

Во продолжение е даден по чекори алгоритмот за кодирање на Шенон–Фано–Елијас, кој овозможува еднозначно декодирање и има својство на префикс.



Слика 5.14

Шенон–Фано–Елијас алгоритам за кодирање

Чекор 1. За дадена распределба на изворот X се наоѓа функцијата на распределба $F(x)$.

Чекор 2. Се пресметуваат вредностите $F_1(x) = F(x) + \frac{p(x)}{2}$, за $x \in R_X$.

Чекор 3. Децималниот дел од $F_1(x)$ се претставува во бинарен броен систем, за секој $x \in R_X$.

Чекор 4. Коден збор на $x \in R_X$ се првите $l(x) = \lceil -\log_2 p(x) \rceil + 1$ децимали во бинарното претставување на децималниот дел од $F_1(x)$.

Да воочиме дека процедурата во овој алгоритам не бара символите да се подредени по веројатност. Бидејќи, за претставување на x се користат $l(x) = \lceil -\log_2 p(x) \rceil + 1$ бита, очекуваната должина на кодот е:

$$\begin{aligned} L &= \sum_x p(x)l(x) = \sum_x p(x) (\lceil -\log p(x) \rceil + 1) \\ &< \sum_x p(x) (-\log p(x) + 2) = H(X) + 2. \end{aligned}$$

Ова значи дека овој алгоритам за кодирање дава просечна должина на кодот, што е најмногу 2 бита над ентропијата.

Пример 5.9. Прво го разгледуваме примерот каде што распределбата е 2-арна, т.е. сите веројатности се од облик 2^{-n} , за некој природен број n .

	$p(x)$	$F(x)$	$F_1(x)$	$F_1(x)$ бинарно	$l = \lceil -\log_2 p_i \rceil + 1$	коден збор
α_1	0.25	0	0.125	0.001	3	001
α_2	0.5	0.25	0.5	0.10	2	10
α_3	0.125	0.75	0.8125	0.1101	4	1101
α_4	0.125	0.875	0.9375	0.1111	4	1111
		1				

Просечната должина на кодните замени е

$$L = 3 \cdot 0.25 + 2 \cdot 0.5 + 4 \cdot 0.125 + 4 \cdot 0.125 = 2.75 \text{ бита,}$$

а ентропијата на X е $H(X) = 1.75$ бита.

Може да се воочи дека овој код има одредени неефикасности. На пример, може да се отстрани последниот бит од двата последни кодни зборови. Но, не може да се отстрани по еден бит од секој коден збор, затоа што во тој случај кодот нема да има својство на префикс. \square

Пример 5.10. Сега, разгледуваме пример каде распределбата не е 2-арна. Во тој случај, во бинарната презентација на $F_1(x)$ може да се појават бесконечен број на битови.

	$p(x)$	$F(x)$	$F_1(x)$	$F_1(x)$ бинарно	$l = \lceil -\log_2 p_i \rceil + 1$	коден збор
α_1	0.25	0	0.125	0.001	3	001
α_2	0.25	0.25	0.375	0.011	3	011
α_3	0.2	0.5	0.6	0.1(0011)	4	1001
α_4	0.15	0.7	0.775	0.110(0011)	4	1100
α_5	0.15	0.85	0.925	0.111(0110)	4	1110
		1				

\square

5.6. Аритметички кодови

Претходно утврдивме дека ако должината на кодниот збор придружен на симболот x е $\lceil -\log p(x) \rceil$, тогаш кодот е блиску до оптималниот, т.е. очекуваната должина на кодните зборови е најмногу за 1 бит над ентропијата. Видовме дека кодовите добиени со Хафмановиот алгоритам се оптимални кодови. Ако азбуката на изворот е мала, тогаш за постигнување на ефикасно кодирање мора да се кодираат блокови од повеќе симболи. На пример, ако изворот е бинарен и ако се кодира секој симбол посебно, тогаш просечната должина на кодните зборови е до еден бит по симбол поголема од ентропијата на изворот. Ако се користат подолги блокови (суперсимболи), тогаш може да се достигне очекувана должина по симбол која е блиску до ратата на ентропија на изворот. Тоа го покажавме во поглавјето 5.3.

Затоа е пожелно да се има ефикасна процедура за кодирање која работи со подолги блокови од симболи од изворот. Хафмановиот код не е погоден за тоа, бидејќи тоа е процедура оддолу-нагоре, која бара пресметување на веројатностите на сите низи од изворот за дадена должина на блокот. Подобра шема е онаа во која може лесно да се прошири должината на блокот без да се прават сите тие пресметки повторно. Аритметичките кодови, кои се директно проширување на Шенон–Фано–Елиас кодовите, ја постигнуваат таа цел.

Основната идеја на аритметичките кодови е ефикасно пресметување на законот на распределба $p(x_1, x_2, \dots, x_n)$ и функцијата на распределба $F(x_1, x_2, \dots, x_n)$, за низата од изворот (x_1, x_2, \dots, x_n) . Со користење на идејата на Шенон–Фано–Елиас кодовите, како коден збор за (x_1, x_2, \dots, x_n) може да се користи еден број од интервалот

$$[F(x_1, x_2, \dots, x_n), F(x_1, x_2, \dots, x_n) + p(x_1, x_2, \dots, x_n)).$$

На пример, изразувајќи ја вредноста на $F(x_1, x_2, \dots, x_n)$ со точност $\lceil -\log p(x_1, x_2, \dots, x_n) \rceil + 1$, се добива код за (x_1, x_2, \dots, x_n) . Според претходната дискусија за кодовите на Шенон–Фано–Елиас, следува дека кодниот збор соодветен на секоја низа лежи во еден чекор на функцијата на распределба соодветна на таа низа. Така, кодните зборови за различни зборови со должина n се различни.

Ќе разгледаме поедноставена верзија на алгоритмот за аритметички кодови за да ги илустрираме главните идеи. Претпоставуваме дека имаме фиксна должина n на низата која се кодира која е позната и на кодерот и на деко-

дерот. Претпоставуваме дека изворот генерира независни и еднакво распределени случајни променливи, па

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i).$$

Алгоритмот за кодирање ќе биде опишан на следниот пример.

Пример 5.11. Нека изворот е определен со случајната променлива

$$X : \begin{pmatrix} a & b & c \\ 0.5 & 0.25 & 0.25 \end{pmatrix}.$$

Интервалот $[0,1)$ се дели на три подинтервали соодветни на секој од симболите на изворот согласно со веројатноста на секој од нив:

симбол	веројатност	интервал
a	0.5	$[0, 0.5)$
b	0.25	$[0.5, 0.75)$
c	0.25	$[0.75, 1)$

Ќе определиме коден збор за низата $baca$, ако азбуката на кодот е бинарна. Веројатноста за генерирање на оваа низа е:

$$p(\mathbf{x}) = p(baca) = p(b)p(a)p(c)p(a) = 0.015625.$$

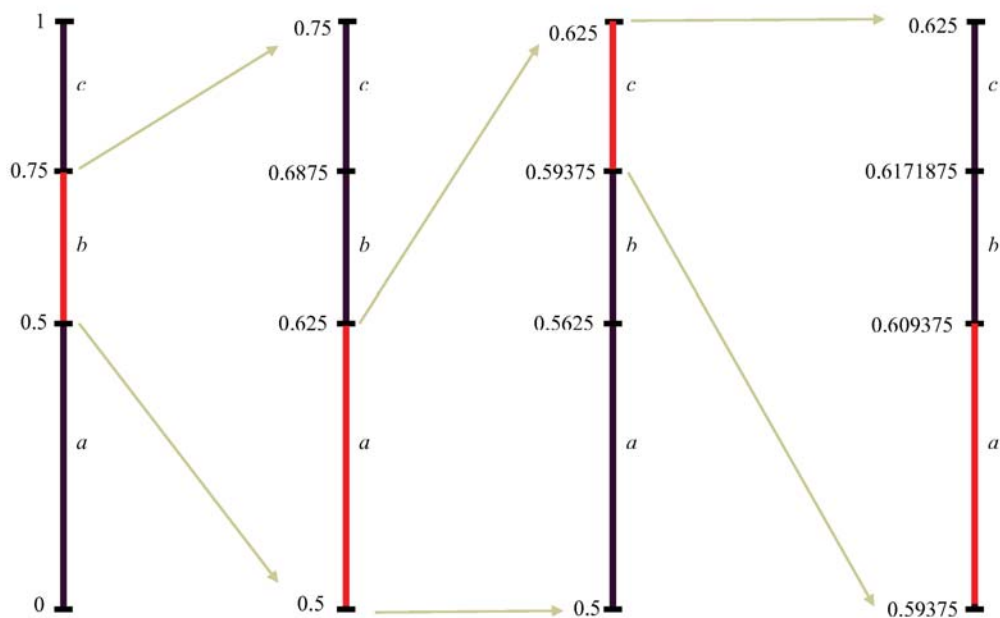
Должината на кодниот збор треба да биде:

$$\lceil -\log p(\mathbf{x}) \rceil + 1 = 7.$$

Кодирањето на $baca$ е претставено на слика 5.15.

Постапката за кодирање е следнава:

- Најпрво, прв симбол во низата која се кодира е b . Интервалот соодветен на b е $[0.5, 0.75)$.
- Во следниот чекор се поаѓа од тој интервал и тој се дели на подинтервали согласно со распределбата на X . Значи, првата половина на интервалот соодветствува на ba , следната четвртина на bb и последната четвртина на bc . Значи, на ba соодветствува подинтервал $[0.5, 0.625)$, на bb - подинтервал $[0.625, 0.875)$ и на bc - подинтервал $[0.685, 0.75)$. Затоа што првите два симбола во низата која се кодира се ba , во следниот чекор, се разгледува подинтервалот $[0.5, 0.625)$.



Слика 5.15

- Сега, овој подинтервал се дели согласно со распределбата на X , па на baa соодветствува подинтервал $[0.5, 0.5625)$, на bab - подинтервал $[0.5625, 0.59375)$, а на bac - подинтервал $[0.59375, 0.625)$. Првите три симбола во низата која ја кодираме е bac , па се разгледува интервалот $[0.59375, 0.625)$.
- Постапката со делење на подинтервали се повторува и на $[0.59375, 0.625)$ на истиот начин. Последниот симбол од низата која ја кодираме е a , па интервалот соодветен на $baca$ во овој чекор е $[0.59375, 0.609375)$.

Кодирањето може да се претстави и со следнава табела.

Симбол	Долна граница на интервал	Горна граница на интервал	Ранг
	0	1	1
b	0.5	0.75	0.25
a	0.5	0.625	0.125
c	0.59375	0.609375	0.015625

Во табелата претходно, ранг е разлика помеѓу горната и доланат граница на интервалот, т.е. должината на интервалот.

Како коден збор за дадената низа може да се земе децималниот дел на кој било број од интервалот $[0.59375, 0.609375)$, на пример, $0.59375 = 0.1001100_2$. Претходно утврдивме дека должината на кодниот збор треба да е 7, па затоа за коден збор на *baca* се зема 1001100. \square

Алгоритмот за кодирање може да се претстави на следниот начин.

Алгоритам за кодирање

Влез: Низа симболи од R_X која треба да се кодира

Чекор 1. Постави $DolnaGranica := 0$
 $GornaGranica := 1$

Чекор 2. Повторувај додека се добијат сите симболи од низата:
 $Rang := GornaGranica - DolnaGranica$
 $GornaGranica := DolnaGranica$
 $+ Rang \cdot GornaGranicaNaIntervalotNaSimbolot$
 $DolnaGranica := DolnaGranica$
 $+ Rang \cdot DolnaGranicaNaIntervalotNaSimbolot$

Соодветниот алгоритам за декодирање е следен:

Алгоритам за декодирање

Влез: $Broj = 0.z_1z_2 \dots z_n$,
каде што $z_1z_2 \dots z_n$ е низата која треба да се декодира

Чекор 1. Определи го *Symbol* согласно со интервалот на кој припаѓа веројатноста.

Чекор 2. Повторувај за сите симболи од низата:
 $Rang := GornaGranicaNaIntervalotNaSimbolot -$
 $DolnaGranicaNaIntervalotNaSimbolot,$
 $Broj := Broj - DolnaGranicaNaIntervalotNaSimbolot,$
 $Broj := Broj / Rang.$

Пример 5.12. Во пример 5.11 добивме коден збор 1001100. Ќе видиме како овој коден збор може да се декодира со користење на претходниот алгоритам за декодирање. Бидејќи $0.1001100_2 = 0.59375_{10}$, влез во овој алгоритам е $Broj = 0.59375$ и е познато дека влезната порака има должина 4.

Поделбата на интервалот $[0, 1)$ на подинтервали согласно со распределбата на X е следна:

симбол	веројатност	интервал
a	0.5	$[0, 0.5)$
b	0.25	$[0.5, 0.75)$
c	0.25	$[0.75, 1)$

$Broj = 0.59375$ припаѓа на интервалот соодветен на симболот b .

$Broj$	$Simbol$	$Rang$
0.59375	b	0.25
$(0.59375 - 0.5)/0.25 = 0.375$	a	0.5
$(0.375 - 0)/0.5 = 0.75$	c	0.25
$(0.75 - 0.75)/0.25 = 0$	a	

□

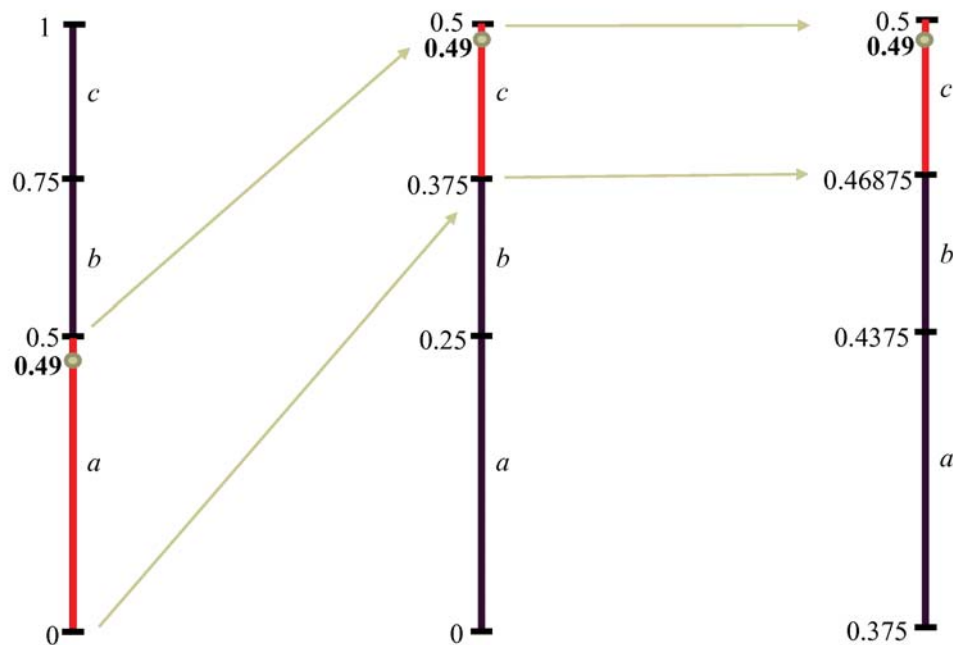
И алгоритмот за декодирање може да се претстави визуелно со претставување на поделбата на интервалите согласно со дадената распределба. Тоа ќе го покажеме во следниот пример.

Пример 5.13. Нека распределбата на изворот X е иста како во пример 5.11. Ќе видиме како може да се декодира бројот 0.49 претставен декадно, ако е познато дека пораката има должина 3. Декодирањето е претставено на слика 5.16.

Постапката по чекори е следна:

- Најпрво се утврдува дека 0.49 припаѓа на интервалот $[0, 0.5)$ соодветен на симболот a . Затоа, првиот симбол во декодираната порака е a .
- Сега, интервалот $[0, 0.5)$ соодветен на a се дели согласно распределбата на X . Подинтервалот $[0, 0.25)$ соодветствува на aa , подинтервалот $[0.25, 0.375)$ соодветствува на ab , а $[0.375, 0.5)$ соодветствува на ac . Бројот 0.49 припаѓа на подинтервалот соодветен на ac , па тоа се првите два симбола во декодираната порака.
- Сега, интервалот $[0.375, 0.5)$ соодветен на ac се дели согласно распределбата на X . Подинтервалот $[0.375, 0.4375)$ соодветствува на aca , подинтервалот $[0.4375, 0.46875)$ соодветствува на acb , а $[0.46875, 0.5)$ соодветствува на acc . Бројот 0.49 припаѓа на подинтервалот соодветен на acc , па декодираната порака (која има три симбола) е acc .

□



Слика 5.16

5.7. Решени задачи

Задача 5.7.1. Нека случајната променлива X има множество на вредности $R_X = \{1, 2, 3, 4, 5\}$. Нека $p(x)$ и $q(x)$ се две распределби и $C_1(x)$ и $C_2(x)$ се два кода за $p(x)$ и $q(x)$, соодветно.

симбол	$p(x)$	$q(x)$	$C_1(x)$	$C_2(x)$
1	$1/2$	$1/2$	0	0
2	$1/4$	$1/8$	10	100
3	$1/8$	$1/8$	110	101
4	$1/16$	$1/8$	1110	110
5	$1/16$	$1/8$	1111	111

- Да се пресмета $H(p)$, $H(q)$, $D(p||q)$ и $D(q||p)$.
- Да се провери дали C_1 е оптимален за p , а C_2 за q .

в) Нека C_2 е код кој се користи кога распределбата е p . Која е просечната должина на кодните зборови? За колку ја надминува ентропијата на p ?

г) Колку се губи, ако за распределбата q се користи кодот C_1 ?

Решение:

а)

$$H(p) = H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\right) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{8} \log \frac{1}{8} - 2 \cdot \frac{1}{16} \log \frac{1}{16} = 1.875$$

$$H(q) = H\left(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right) = -\frac{1}{2} \log \frac{1}{2} - 4 \cdot \frac{1}{8} \log \frac{1}{8} = 2$$

$$\begin{aligned} D(p||q) &= \sum_{x \in R_X} p(x) \log \frac{p(x)}{q(x)} \\ &= \frac{1}{2} \log \frac{1/2}{1/2} + \frac{1}{4} \log \frac{1/4}{1/8} + \frac{1}{8} \log \frac{1/8}{1/8} + \frac{1}{16} \log \frac{1/16}{1/8} + \frac{1}{16} \log \frac{1/16}{1/8} \\ &= 0.125 \end{aligned}$$

$$\begin{aligned} D(q||p) &= \sum_{x \in R_X} q(x) \log \frac{q(x)}{p(x)} \\ &= \frac{1}{2} \log \frac{1/2}{1/2} + \frac{1}{8} \log \frac{1/8}{1/4} + \frac{1}{8} \log \frac{1/8}{1/8} + \frac{1}{8} \log \frac{1/8}{1/16} + \frac{1}{8} \log \frac{1/8}{1/16} \\ &= 0.125 \end{aligned}$$

б)

$$l_1 = L_p(C_1) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + 4 \cdot \frac{1}{16} = \frac{15}{8} = 1.875 = H(p)$$

Значи, C_1 е оптимален код за распределбата p .

$$l_2 = L_q(C_2) = 1 \cdot \frac{1}{2} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} = 2 = H(q),$$

т.е. C_2 е оптимален код за распределбата q .

в)

$$L_p(C_2) = 1 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{16} + 3 \cdot \frac{1}{16} = \frac{32}{16} = 2$$

$$L_p(C_2) - H(p) = 0.125 = D(p||q)$$

г)

$$L_q(C_1) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} = 2.125$$

$$L_q(C_1) - H(q) = 0.125 = D(q||p)$$

Од решенијата под в) и г) може да се заклучи дека релативната ентропија е „цена“ за кодирање на погрешна распределба.

□

Задача 5.7.2. Дадени се 6 шишиња вино. Познато е дека едно вино е расипано (има лош вкус). Со испитување на шишињата, утврдено е дека веројатностите p_i – „виното во i - тото шише е расипано“ се следниве:

$$(p_1, p_2, \dots, p_6) = \left(\frac{8}{23}, \frac{6}{23}, \frac{4}{23}, \frac{2}{23}, \frac{2}{23}, \frac{1}{23} \right).$$

Се претпоставува дека вината се пробуваат по еднаш. Да се одреди редот на дегустација за да се минимизира очекуваниот број на пробувања потребни да се открие расипаното вино. Кој е очекуваниот број на пробувања?

Решение:

Ќе конструираме код, така што на i - тата позиција на кодниот збор ќе стои 1, ако виното од соодветното шише се пробува во i - тиот обид, а на претходните позиции ќе стои 0. Бидејќи веројатноста p_1 за првото вино е најголема, прво се пробува тоа вино, потоа второто, и.т.н

$$1 \rightarrow 1$$

$$2 \rightarrow 01$$

$$3 \rightarrow 001$$

$$4 \rightarrow 0001$$

$$5 \rightarrow 00001$$

$$6 \rightarrow 00000 \text{ (ако првите 5 вина се добри, 6 – тото не мора да се проба).}$$

За очекуваниот број на пробувања, потребни за да се открие расипаното вино, се добива:

$$\sum_{i=1}^6 l_i p_i = 1 \cdot \frac{8}{23} + 2 \cdot \frac{6}{23} + 3 \cdot \frac{4}{23} + 4 \cdot \frac{2}{23} + 5 \cdot \frac{2}{23} + 5 \cdot \frac{1}{23} = \frac{55}{23} = 2.39.$$

□

Задача 5.7.3. Во услови на претходната задача, се спроведува следната стратегија на дегустација: се мешаат некои од вината во чиста чаша и потоа се пробува мешавината. Постапката се повторува сè додека расипаното вино не биде откриено.

- а) Кој е очекуваниот број на пробувања потребен да се открие расипаното вино?
- б) Која мешавина ќе се проба прво?

Решение:

Бидејќи Хафмановиот код е оптимален, може да се користи токму тој код. Повторно 1 на i -тата позиција ќе означува дека виното од соодветното шише е ставено во мешавината во i -тиот обид.

A	p_i	A_1	p_i	A_2	p_i	A_3	p_i	A_4	p_i
1	$\frac{8}{23}$	1	$\frac{8}{23}$	1	$\frac{8}{23}$	3,4,5,6	$\frac{9}{23}$	1,2	$\frac{14}{23}$
2	$\frac{6}{23}$	2	$\frac{6}{23}$	2	$\frac{6}{23}$	1	$\frac{8}{23}$	3,4,5,6	$\frac{9}{23}$
3	$\frac{4}{23}$	3	$\frac{4}{23}$	4,5,6	$\frac{5}{23}$	2	$\frac{6}{23}$		
4	$\frac{2}{23}$	5,6	$\frac{3}{23}$	3	$\frac{4}{23}$				
5	$\frac{2}{23}$	4	$\frac{2}{23}$						
6	$\frac{1}{23}$								

Се добиваат следните кодни зборови за вината (1 на i -тата позиција во кодот на виното значи дека тоа вино е ставено во мешавината во i -тиот обид).

- 1 → 11
 2 → 10
 3 → 00
 4 → 010
 5 → 0111
 6 → 0110

- а) За очекуваниот број на пробувања потребен да се открие расипаното вино се добива

$$\sum_{i=1}^6 l_i p_i = 2 \cdot \frac{8}{23} + 2 \cdot \frac{6}{23} + 2 \cdot \frac{4}{23} + 3 \cdot \frac{2}{23} + 4 \cdot \frac{2}{23} + 4 \cdot \frac{1}{23} = \frac{54}{23} = 2.35$$

б) Прво се пробува мешавина од првото и второто шише, бидејќи единица се јавува на прва позиција во кодните зборови соодветни на првото и второто шише.

□

Задача 5.7.4. Случајната променлива има закон на распределба:

$$X : \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 \\ \frac{10}{28} & \frac{6}{28} & \frac{5}{28} & \frac{4}{28} & \frac{2}{28} & \frac{1}{28} \end{pmatrix}.$$

- а) Да се определи бинарен Хафманов код за X .
- б) Да се најде распределба на случајна променлива Y , каде што $R_X = R_Y$, за која со должините на кодните зборови добиени со кодот под а) очекуваната должина на кодните зборови ќе биде најмала можна.
- в) Без пресметување на вредностите, да се споредат $H(X)$ и $H(Y)$. Која од овие ентропии е поголема?

Решение:

а)

A	p_i	A_1	p_i	A_2	p_i	A_3	p_i	A_4	p_i
α_1	$\frac{10}{28}$	α_1	$\frac{10}{28}$	α_1	$\frac{10}{28}$	$\alpha_{2,3}$	$\frac{11}{28}$	$\alpha_{1,4,5,6}$	$\frac{17}{28}$
α_2	$\frac{6}{28}$	α_2	$\frac{6}{28}$	$\alpha_{4,5,6}$	$\frac{7}{28}$	α_1	$\frac{10}{28}$	$\alpha_{2,3}$	$\frac{11}{28}$
α_3	$\frac{5}{28}$	α_3	$\frac{5}{28}$	α_2	$\frac{6}{28}$	$\alpha_{4,5,6}$	$\frac{7}{28}$		
α_4	$\frac{4}{28}$	α_4	$\frac{4}{28}$	α_3	$\frac{5}{28}$				
α_5	$\frac{2}{28}$	$\alpha_{5,6}$	$\frac{3}{28}$						
α_6	$\frac{1}{28}$								

Се добиваат следните кодни зборови:

$\alpha_1 \rightarrow 00$
 $\alpha_2 \rightarrow 10$
 $\alpha_3 \rightarrow 11$
 $\alpha_4 \rightarrow 010$
 $\alpha_5 \rightarrow 0110$
 $\alpha_6 \rightarrow 0111$

б) Должините на кодните зборови во Хафмановиот код се 2, 2, 2, 3, 4, 4. Најмала очекувана должина на кодните замени (која е еднаква на ентропијата на случајната променлива) се добива ако $l_i = -\log p_i$, т.е., $p_i = 2^{-l_i}$, $i = 1, \dots, 6$. Оттука, ја добиваме следната распределба за Y

$$Y : \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \frac{1}{16} \end{pmatrix}.$$

в) Ако L_1 е очекуваната должина на кодните замени на кодот под а), а L_2 очекуваната должина на кодните замени ако наместо распределбата на X ја земеме распределбата на Y , тогаш

$$L_1 = 2 \cdot \frac{10}{28} + 2 \cdot \frac{6}{28} + 2 \cdot \frac{5}{28} + 3 \cdot \frac{4}{28} + 4 \cdot \frac{2}{28} + 4 \cdot \frac{1}{28} = \frac{66}{28} = \frac{33}{14}$$

$$L_2 = 2 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + 4 \cdot \frac{1}{16} = \frac{19}{8} = \frac{133}{56} (= H(Y)).$$

Значи, $L_1 < L_2$. За очекуваната должина на секој код, точно е следното неравенство

$$H(X) \leq L < H(X) + 1.$$

Оттука,

$$H(X) \leq L_1 < L_2 = H(Y).$$

Значи, $H(X) < H(Y)$. □

Задача 5.7.5. Извор емитура симболи a и b со веројатности 0.8 и 0.2.

- а) Да се определи оптимален код со Хафманов алгоритам за зборовите со должина три од азбуката $A = \{a, b\}$, ако кодната азбука е $D = \{0, 1, 2, 3\}$ и да се определи просечната должина на кодните замени.
- б) Зборовите со должина три од азбуката $A = \{a, b\}$ да се кодираат со Шенон–Фано–Елиас кодот.

Решение:

Постојат 8 зборови со должина 3 со симболи од азбуката $A = \{a, b\}$. Во продолжение се дадени тие зборови, заедно со соодветните веројатности:

$\alpha_1 = aaa$	$p(\alpha_1) = 0.8^3 = 0.512;$
$\alpha_2 = aba$	$p(\alpha_2) = 0.8^2 \cdot 0.2 = 0.128;$
$\alpha_3 = baa$	$p(\alpha_3) = 0.8^2 \cdot 0.2 = 0.128;$
$\alpha_4 = aab$	$p(\alpha_4) = 0.8^2 \cdot 0.2 = 0.128;$
$\alpha_5 = abb$	$p(\alpha_5) = 0.2^2 \cdot 0.8 = 0.032;$
$\alpha_6 = bab$	$p(\alpha_6) = 0.2^2 \cdot 0.8 = 0.032;$
$\alpha_7 = bba$	$p(\alpha_7) = 0.2^2 \cdot 0.8 = 0.032;$
$\alpha_8 = bbb$	$p(\alpha_8) = 0.2^3 = 0.008$

- а) Бројот на симболи (зборови) кои треба да се кодираат е $m = 8$, а број на кодни симболи $d = 4$. Оттука, $c = (m-1) \bmod (d-1) = 7 \bmod 3 = 1 \neq 0$, што значи дека треба да се додадат $m' = d-1-c = 2$ фиктивни симболи со веројатност 0.

A	p_i	A_1	p_i	A_2	p_i
α_1	0.512	α_1	0.512	α_1 0	0.512
α_2	0.128	α_2	0.128	$\alpha_{4,5,6,7,8}$ 1	0.232
α_3	0.128	α_3	0.128	α_2 2	0.128
α_4	0.128	α_4 0	0.128	α_3 3	0.128
α_5	0.032	$\alpha_{7,8}$ 1	0.04		
α_6	0.032	α_5 2	0.032		
α_7 0	0.032	α_6 3	0.032		
α_8 1	0.008				
– 2	0				
– 3	0				

Се добиваат следните кодни зборови:

$$\begin{aligned} \alpha_1 &\rightarrow 0, & \alpha_2 &\rightarrow 2, & \alpha_3 &\rightarrow 3, & \alpha_4 &\rightarrow 10, \\ \alpha_5 &\rightarrow 12, & \alpha_6 &\rightarrow 13, & \alpha_7 &\rightarrow 110, & \alpha_8 &\rightarrow 111. \end{aligned}$$

За просечната должина на кодните зборови се добива:

$$\begin{aligned} L &= 1 \cdot (0.512 + 0.128 + 0.128) + 2 \cdot (0.128 + 0.032 + 0.032) + 3 \cdot (0.032 + 0.008) \\ &= 1.272. \end{aligned}$$

б) За Шенон–Фано–Елиас кодот се добива:

	p_i	$F(x)$	$F_1(x)$	$F_1(x)$ бинарно	$l = \lceil -\log_2 p_i \rceil + 1$	коден збор
α_1	0.512	0	0.256	0.01000001	2	01
α_2	0.128	0.512	0.576	0.1001001	4	1001
α_3	0.128	0.64	0.704	0.1011010	4	1011
α_4	0.128	0.768	0.832	0.1101010	4	1101
α_5	0.032	0.896	0.912	0.1110100	6	111010
α_6	0.032	0.928	0.944	0.1111000	6	111100
α_7	0.032	0.96	0.976	0.1111100	6	111110
α_8	0.008	0.992	0.996	0.11111110	8	11111110
		1				

Кодните зборови со овој код се:

$$\begin{aligned} \alpha_1 &\rightarrow 01, & \alpha_2 &\rightarrow 1001, & \alpha_3 &\rightarrow 1011, & \alpha_4 &\rightarrow 1101, \\ \alpha_5 &\rightarrow 111010, & \alpha_6 &\rightarrow 111100, & \alpha_7 &\rightarrow 111110, & \alpha_8 &\rightarrow 11111110. \end{aligned}$$

□

5.8. Задачи

Задача 5.8.1. Да се конструира тернарен Хафманов код за распределбата:

$$X : \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 & \alpha_7 \\ 0.25 & 0.25 & 0.1 & 0.15 & 0.07 & 0.08 & 0.1 \end{pmatrix}.$$

Задача 5.8.2. Во текстот „ТЕОРИЈА НА ВЕРОЈАТНОСТ И ИНФОРМАЦИИ“ да се определи фреквенцијата на буквите и со користење на Хафмановиот алгоритам да се определат кодни зборови за сите букви од текстот (без празно место) со користење на кодната азбука:

а) $B = \{0, 1, 2\}$

б) $B = \{0, 1\}$.

Да се споредат просечните должини на кодните замени за добиените кодови.

Задача 5.8.3. Да се конструира код на Шенон–Фано–Елиас, за распределбата:

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0.1 & 0.2 & 0.3 & 0.15 & 0.25 \end{pmatrix}.$$

Задача 5.8.4. Во текстот „ТЕОРИЈА НА КОДИРАЊЕ И КРИПТОГРАФИЈА“ да се одредат фреквенциите на појавување на буквите (без празното место) и да се кодираат со Шенон–Фано–Елиас кодот.

Задача 5.8.5. Нека изворот е определен со случајната променлива:

$$X : \begin{pmatrix} a & b & c \\ 0.3 & 0.4 & 0.3 \end{pmatrix}.$$

Со користење на аритметичко кодирање, да се определи коден збор за $aabcb$.

Задача 5.8.6. Нека изворот е дефиниран со случајната променлива

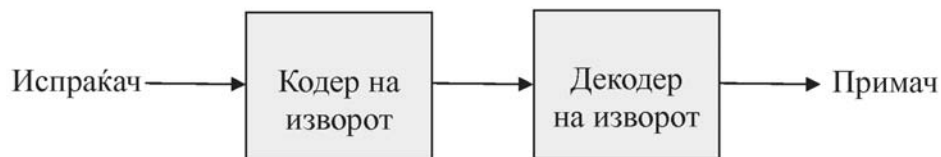
$$X : \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ 0.2 & 0.1 & 0.4 & 0.3 \end{pmatrix}.$$

- а) Да се конструира Шенон–Фано–Елиас код за симболите a_1, a_2, a_3, a_4 .
- б) Ако изворот ги емитура симболите a_1, a_2, a_3, a_4 независно, да се конструира оптимален код со Хафманов алгоритам со кодна азбука $D = \{0, 1, 2, 3\}$ за зборовите со должина 2.
- в) За истиот извор (распределба) да се конструира коден збор за $a_3a_2a_1$ со користење на аритметичко кодирање.

Глава 6

Комуникациски канал

Досега го проучувавме проблемот со компресија на податоци (кодирање на изворот), шематски претставен на дијаграмот на слика 6.1



Слика 6.1

Пред да поминеме на проблемот на кодирање на каналот (глава 8), ќе воведеме еден чекор помеѓу излезот од кодерот на изворот (компресорот) и влезот во декодерот на изворот (декомпресорот). Тоа е каналот во комуникацискиот систем. Каналот е средство или медиум за пренесување информации од испраќачот до примачот. Одредувањето на најсоодветниот канал или медиум е од клучно значење за ефективноста на комуникацијата. Во најголем број од каналите, при пренос на информации, се јавуваат одредени пречки (шум), што доведува до тоа пораката која се добива на излезот на каналот да не е иста како таа што е пуштена на влезот во каналот. Некои примери на канал со шум се: безжична врска, телефонска линија, уред за складирање во кој може да се појават грешки, IP-мрежа со потенцијал за загуба на пакети, итн.

Во оваа глава ќе разгледаме некои примери на комуникациски канали и ќе го определиме нивниот капацитет. Ќе се задржиме на дискретни канали без меморија.

6.1. Дискретен канал без меморија

Во следните две дефиниции ќе дефинираме најпрво што е дискретен канал, а потоа и што е канал без меморија.

Дефиниција 6.1. *Дискретен канал* е тројката (R_X, Π, R_Y) , каде R_X е влезна азбука, R_Y е излезна азбука, а $\Pi = [p(y|x)]$ е матрица на каналот. Притоа, $p(y|x)$ е веројатноста дека на излезот ќе се добие симбол y , ако на влезот е пратен симбол x , и

$$\sum_{y \in R_Y} p(y|x) = 1. \quad (6.1)$$

Да воочиме дека за фиксно $x \in R_X$, $p(y|x)$ се сите елементи од редицата во матрицата Π која е соодветна на влезот x . Поради равенството (6.1) следува дека сумата на секоја редица од матрицата на каналот е 1. Ваквите матрици се нарекуваат *стохастички матрици*.

Во овој случај, со користење на теоремата за тотална веројатност, за секој $y \in R_Y$, се добива:

$$P\{Y = y\} = \sum_{x \in R_X} P\{X = x\}P\{Y = y|X = x\} = \sum_{x \in R_X} P\{X = x\}p(y|x). \quad (6.2)$$

Ако во вектор редица \mathbf{p}_X се стават веројатностите од законот на распределба на X , а во вектор редица \mathbf{p}_Y се стават веројатностите од законот на распределба на Y , тогаш равенствата (6.2) може да се запишат во матрична форма

$$\mathbf{p}_Y = \mathbf{p}_X \Pi. \quad (6.3)$$

Дефиниција 6.2. Каналот се нарекува *канал без меморија*, ако распределбата на излезот зависи само од влезот во тој момент, а не и од претходните влезови и излези на каналот, т.е. ако $\mathbf{x} = (x_1, x_2, \dots, x_n) \in R_X^n$, а $\mathbf{y} = (y_1, y_2, \dots, y_n) \in R_Y^n$, тогаш

$$P(\mathbf{y}|\mathbf{x}) = P((y_1, y_2, \dots, y_n)|(x_1, x_2, \dots, x_n)) = P(y_1|x_1)P(y_2|x_2) \dots P(y_n|x_n).$$

Дефиниција 6.3. „Информациски“ капацитет на дискретен канал без меморија се дефинира со:

$$C = \max_{p(x)} I(X; Y),$$

каде максимумот се зема по сите можни влезни распределби $p(x)$.

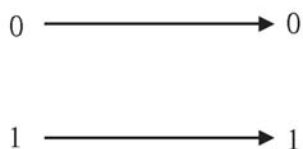
Всушност, информацискиот капацитет се дефинира како максимална информација која излезот од каналот ја дава за влезот. Тоа е, всушност, максималната количина на информација што може да се пренесе низ каналот без грешка.

6.2. Видови дискретни комуникациски канали без меморија

Во ова поглавје ќе разгледаме неколку видови дискретни комуникациски канали без меморија.

6.2.1. Бинарен канал без шум

Ова е канал, кој за кој било бинарен влез го дава точно истиот излез (слика 6.2).



Слика 6.2

Оттука, 1 бит без грешка ќе биде пренесен при секое користење на каналот. Матрицата на овој канал е единечна матрица од ред 2, т.е.

$$\Pi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Интуитивно е јасно дека информацискиот капацитет на овој канал ќе биде 1 бит. За да се пресмета информацискиот капацитет на каналот, да воочиме дека за произволна распределба на X , имаме:

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y|X) \\ &= H(Y) \quad (X \text{ еднозначно го определува } Y, \text{ па } H(Y|X) = 0) \\ &= H(X) \quad (\text{бидејќи } X \text{ и } Y \text{ имаат иста распределба}). \end{aligned}$$

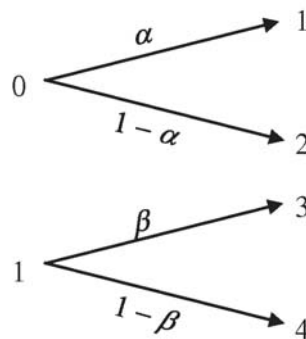
Сега, максимумот на $I(X; Y)$ се совпаѓа со максимумот на $H(X)$, а знаеме дека тоа се постигнува за рамномерна распределба на X на множеството

$R_X = \{0, 1\}$. Оттука,

$$C = \max_{p(x)} I(X; Y) = \max_{p(x)} H(X) = H(1/2, 1/2) = 1 \text{ бит.}$$

6.2.2. Бинарен канал со непреклопувачки излези

Овој канал има два (или повеќе) можни излези за секој од двата влеза, но множествата излези соодветни на секој влез се дисјунктни (слика 6.3).



Слика 6.3

Влезот во овој канал може еднозначно да биде определен од излезот, и оттука секој испратен бит може да се открие без грешка. Неговата матрица е од облик:

$$\Pi = \begin{bmatrix} \alpha & 1 - \alpha & 0 & 0 \\ 0 & 0 & \beta & 1 - \beta \end{bmatrix}.$$

За капацитетот на овој канал, добиваме:

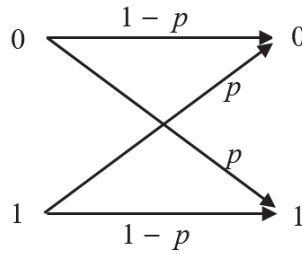
$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= H(X) \quad (Y \text{ еднозначно го определува } X, \text{ па } H(X|Y) = 0). \end{aligned}$$

Оттука, од исти причини како и претходно, добиваме дека

$$C = \max_{p(x)} I(X; Y) = \max_{p(x)} H(X) = H(1/2, 1/2) = 1 \text{ бит.}$$

6.2.3. Бинарен симетричен канал

Ова е наједноставен пример на канал со пречки, каде секој бит може да биде погрешно пренесен со веројатност p ($0 < p < 1/2$). Грешка ќе се случи кога 0 ќе се пренесе како 1, или обратно (Слика 6.4). Затоа, од примениот бит не може да се открие дали се појавила грешка или не. Оттука, сите примени битови се неверодостојни.



Слика 6.4

Матрицата на овој канал е:

$$\Pi = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}.$$

За дадена вредност на веројатноста p за погрешно пренесен бит, за капацитетот на бинарен симетричен канал, наоѓаме:

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_x p(x) H(Y|X=x) \\ &= H(Y) - \sum_x p(x) H(p, 1-p) \\ &= H(Y) - H(p, 1-p) \sum_x p(x) \\ &= H(Y) - H(p, 1-p) \\ &\leq 1 - H(p, 1-p). \end{aligned}$$

Последното неравенство доаѓа од тоа што Y е бинарната случајна променлива, па $H(Y) \leq \log 2 = 1$. Во овој случај, равенство ќе важи ако Y има рамномерна распределба. Затоа се поставува прашањето дали за некоја распределба на влезот X , излезот Y ќе има рамномерна распределба на $R_Y =$

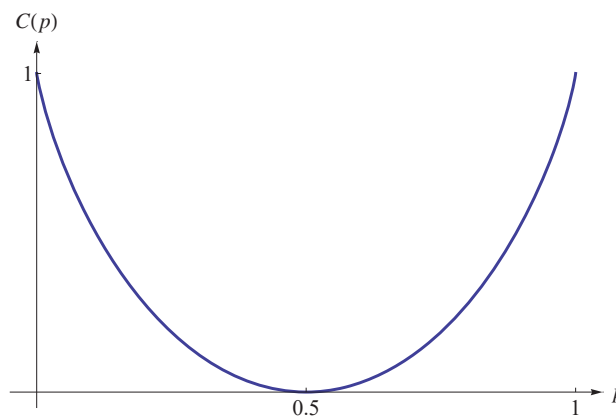
$\{0, 1\}$? Ќе покажеме дека тоа се случува за рамномерна распределба на X на $R_X = \{0, 1\}$. Имено,

$$\begin{aligned} P\{Y = 0\} &= P\{X = 0\}P\{Y = 0|X = 0\} + P\{X = 1\}P\{Y = 0|X = 1\} \\ &= \frac{1}{2} \cdot (1 - p) + \frac{1}{2} \cdot p = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} P\{Y = 1\} &= P\{X = 0\}P\{Y = 1|X = 0\} + P\{X = 1\}P\{Y = 1|X = 1\} \\ &= \frac{1}{2} \cdot p + \frac{1}{2} \cdot (1 - p) = \frac{1}{2}. \end{aligned}$$

Значи, $H(Y)$ ќе ја достигне горната граница $H(Y) = 1$, па оттука информацискиот капацитет на каналот ја достигнува горната граница, т.е. $C = 1 - H(p, 1 - p)$ бита. Во овој случај, $H(p, 1 - p)$ може да се интерпретира како изгубена информација по еден пренесен бит низ каналот.

Графикот на функцијата $C(p) = 1 - H(p, 1 - p)$ е даден на слика 6.5.



Слика 6.5

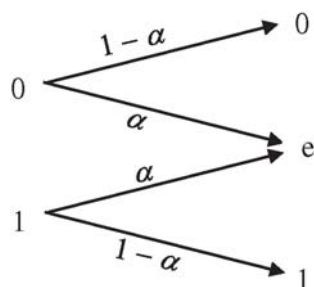
Од графикот може да се воочи следното:

- Ако $p = 0$ или $p = 1$, капацитетот на каналот е максимален и изнесува 1. Ова е очекувано затоа што ако $p = 0$, тогаш во каналот нема пречки. Ако, пак, $p = 1$, тогаш ако на влез во каналот се пушти 0, на излез се добива 1, и обратно. Во овој случај, декодерот ќе треба само да замени 0 со 1 и обратно.
- Ако $p = 1/2$, тогаш капацитетот на каналот е 0 и низ него практично не се пренесува никаква информација. И ова е очекувано, затоа што во тој

случај, секој бит се пренесува точно или погрешно со иста веројатност $1/2$.

6.2.4. Канал со бришење

Овој тип на канал е сличен на бинарен симетричен канал, само што во овој случај некои битови ($\alpha \cdot 100\%$, каде што $0 \leq \alpha \leq 1$) може да се избришат (наместо да бидат погрешно пренесени како во претходниот случај). На слика 6.6 е даден ваков тип на канал.



Слика 6.6

Матрицата на овој канал е:

$$\begin{bmatrix} 1 - \alpha & \alpha & 0 \\ 0 & \alpha & 1 - \alpha \end{bmatrix}$$

За определување на капацитетот на овој канал, се тргнува од равенството $I(X; Y) = H(X) - H(X|Y)$. За определување на $H(X|Y)$, ќе го искористиме следното:

- Ако $Y = 0$ или $Y = 1$, тогаш влезот X е еднозначно определен. Затоа, $H(X|Y = 0) = 0$ и $H(X|Y = 1) = 0$.
- Ако $Y = e$, тогаш X може да биде или 0 или 1, со иста веројатност. Затоа, $H(X|Y = e) = H(X)$.

Сега,

$$\begin{aligned} H(X|Y) &= P\{Y = 0\}H(X|Y = 0) + P\{Y = e\}H(X|Y = e) \\ &\quad + P\{Y = 1\}H(X|Y = 1) \\ &= P\{Y = 0\} \cdot 0 + P\{Y = e\} \cdot H(X) + P\{Y = 1\} \cdot 0 \\ &= P\{Y = e\} \cdot H(X). \end{aligned}$$

Од друга страна,

$$\begin{aligned} P\{Y = e\} &= P\{X = 0\}P\{Y = e|X = 0\} + P\{X = 1\}P\{Y = e|X = 1\} \\ &= P\{X = 0\} \cdot \alpha + P\{X = 1\} \cdot \alpha \\ &= \alpha. \end{aligned}$$

Оттука,

$$H(X|Y) = \alpha H(X).$$

Сега, за заемната информација помеѓу X и Y , имаме:

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= H(X) - \alpha H(X) \\ &= (1 - \alpha)H(X). \end{aligned}$$

Конечно, за капацитетот на овој канал добиваме:

$$C = \max_{p(x)} I(X; Y) = \max_{p(x)} (1 - \alpha)H(X) = 1 - \alpha,$$

бидејќи $\max_{p(x)} H(X) = H(1/2, 1/2) = 1$ бит и се добива за рамномерна распределба на X на множеството $R_X = \{0, 1\}$.

На изразот $C = 1 - \alpha$ може да му дадеме и одредено интуитивно значење. Бидејќи дел α од битовите се изгубени во каналот, ние можеме да откриеме $1 - \alpha$ од битовите. Оттука, капацитетот на каналот е најмногу $1 - \alpha$.

Во многу практични случаи, испраќачот добива повратна информација од примачот. Ако тоа е случај кај каналите со бришење, повратната информација би била дали битот е добиен или избришан. Ако е избришан, тој треба да се испрати повторно. Бидејќи еден бит се пренесува коректно со веројатност $1 - \alpha$, ефективната брзина на пренос е $1 - \alpha$. На тој начин, со користење на повратна информација е полесно да се постигне капацитет $1 - \alpha$.

6.2.5. Симетричен канал

Во ова поглавје ќе го обопштиме поимот за бинарен симетричен канал и ќе дефинираме симетричен канал каде азбуката R_X на влезот и азбуката R_Y на излезот не се бинарни.

Дефиниција 6.4. Еден канал се нарекува *симетричен*, ако редиците на неговата матрица Π на преодни веројатности се пермутации една на друга, и колоните се пермутации една на друга.

Поради ова својство, кај симетричен канал не само што сумите по редици се еднакви на 1, туку и сумите по колони се, исто така, еднакви на 1. Да воочиме дека кај симетричен канал, $|R_X| = |R_Y|$, т.е. матрицата Π е квадратна.

На пример, каналот чијашто матрица е

$$\Pi = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

е симетричен канал. Исто така, ако X и Z се независни случајни променливи и имаат некои распределби на множеството вредности $\{0, 1, \dots, c-1\}$, се покажува дека ако X е влез во каналот, Z е шум, а $Y = X + Z \pmod{c}$ е излез од каналот, тогаш каналот е симетричен (види Задача 6.5.4).

Нека \mathbf{r} е произволна редица од матрицата на произволен симетричен канал. Тогаш, заедната информација помеѓу влезот X и излезот Y на тој канал е:

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y|X) \\ &= H(Y) - H(\mathbf{r}) \\ &\leq \log |R_Y| - H(\mathbf{r}). \end{aligned}$$

Последното неравенство ќе помине во равенство, ако распределбата на излезот е рамномерна. И повторно, како кај бинарен симетричен канал, се поставува прашањето дали за некоја распределба на влезот, распределбата на излезот ќе биде рамномерна. Ќе покажеме дека ако X има рамномерна распределба на R_X , тогаш и Y ќе има рамномерна распределба на R_Y . Имено, ако распределбата на X е рамномерна, т.е. $P\{X = x\} = 1/|R_X|$, за секој $x \in R_X$, тогаш за распределбата на Y се добива:

$$\begin{aligned} P\{Y = y\} &= \sum_{x \in R_X} P\{X = x\} P\{Y = y|X = x\} \\ &= \sum_{x \in R_X} \frac{1}{|R_X|} P\{Y = y|X = x\} \\ &= \frac{1}{|R_X|} \sum_{x \in R_X} P\{Y = y|X = x\} \\ &= \frac{1}{|R_X|} = \frac{1}{|R_Y|} \end{aligned}$$

за секој $y \in R_Y$. Оттука, Y има рамномерна распределба на R_Y , па за капацитетот на симетричен канал се добива:

$$C = \max_{p(x)} I(X; Y) = \max_{p(x)} H(Y) - H(\mathbf{r}) = \log |R_Y| - H(\mathbf{r}).$$

Пример 6.1. Нека матрицата на каналот е

$$\Pi = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

Претходно видовме дека ова е матрица на симетричен канал. Согласно претходно изведеното, неговиот капацитет е

$$C = \log 3 - H(0.5, 0.3, 0.2).$$

□

Дефиниција 6.5. Каналот се нарекува *слабо симетричен*, ако редиците на неговата матрица Π на преодни веројатности се пермутации една на друга, а сумата на елементите во секоја колона е еднаква.

На пример, каналот чија матрица е

$$\Pi = \begin{bmatrix} 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{bmatrix},$$

е слабо симетричен, но не е симетричен.

Ќе покажеме дека за слабо симетричен канал, капацитетот е повторно $C = \log |R_Y| - H(\mathbf{r})$, каде \mathbf{r} е една редица од матрицата Π , и тој се постигнува за рамномерна распределба на влезот. Имено, ако распределбата на влезот е рамномерна, т.е. $p(x) = 1/|R_X|$, тогаш за распределбата на излезот Y се добива:

$$\begin{aligned} P\{Y = y\} &= \sum_{x \in R_X} P\{X = x\} P\{Y = y | X = x\} \\ &= \sum_{x \in R_X} \frac{1}{|R_X|} P\{Y = y | X = x\} \\ &= \frac{1}{|R_X|} \sum_{x \in R_X} P\{Y = y | X = x\} \\ &= \frac{1}{|R_X|} \cdot c = \frac{1}{|R_Y|} \end{aligned}$$

за секој $y \in R_Y$. Овде, c е сумата на елементите во една колона во матрицата на каналот. Значи, Y има рамномерна распределба на R_Y . Оттука, исто како претходно, се добива дека капацитетот на слабо симетричен канал е:

$$C = \max_{p(x)} I(X; Y) = \max_{p(x)} H(Y) - H(\mathbf{r}) = \log |R_Y| - H(\mathbf{r}).$$

6.3. Својства на капацитет на канал

Видовме дека информацискиот капацитет на еден канал се определува со

$$C = \max_{p(x)} I(X; Y),$$

каде што максимумот се зема по сите распределби $p(x)$ на влезот. Капацитетот на еден канал ги има следниве својства:

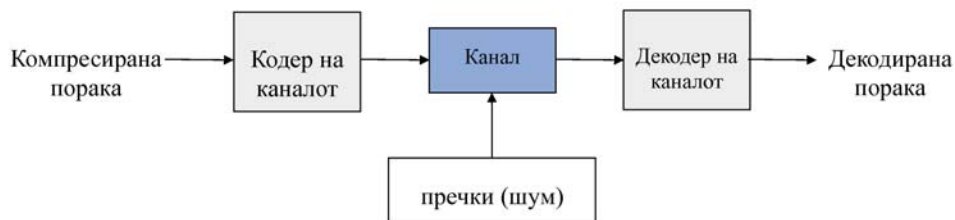
1. Капацитетот $C \geq 0$, бидејќи $I(X; Y) \geq 0$.
2. Капацитетот $C \leq \log |R_X|$, бидејќи

$$C = \max I(X; Y) = \max(H(X) - H(X|Y)) \leq \max H(X) = \log |R_X|.$$

3. Од исти причини, $C \leq \log |R_Y|$.

6.4. Кодер и декодер на каналот

Како што видовме претходно, најголем број од каналите во комуникациските системи се канали со шум. Поради дејството на шумот, излезната порака од каналот не мора да соодветствува со влезната. Затоа, пред каналот треба да постои *кодер на каналот*, а по каналот *декодер на каналот* (слика 6.7). Целта на кодерот е на компресираната порака да додаде дополнителни симболи, кои се нарекуваат *редундантни* симболи, кои ќе помогнат во декодерот да се поправат грешките при преносот и да се врати влезната порака, ако тоа е можно. Декодерот на секоја излезна порака од каналот придружува, според одредено правило, една од можните влезни пораки.



Слика 6.7

Дефиниција 6.6. (k, n) блок код на каналот (R_X, Π, R_Y) се состои од следното:

1. Индексно множество $\{1, 2, \dots, k\}$.
2. Функција за кодирање $X : \{1, 2, \dots, k\} \rightarrow R_X^n$, која произведува кодни зборови $X(1), X(2), \dots, X(k)$.
3. Функција за декодирање $g : R_Y^n \rightarrow \{1, 2, \dots, k\}$, која е детерминистичко правило кое за секој можен излез придружува еден влезен вектор.

Доколку се појави грешка при пренос, декодерот може да врати порака која се разликува од влезната. Тогаш се појавува грешка при декодирање. Веројатноста на таквата грешка е дадена со следната дефиниција.

Дефиниција 6.7. Веројатноста за грешка при декодирање на симболот i се дефинира со:

$$\lambda_i = P \{g(\mathbf{Y}) \neq i | \mathbf{X} = i\}.$$

Дефиниција 6.8. Максимална веројатност $\lambda^{(n)}$ на грешка при декодирање за кодот (k, n) се дефинира со:

$$\lambda^{(n)} = \max_{i \in \{1, 2, \dots, k\}} \lambda_i.$$

Значи, максималната веројатност на грешка при декодирање е максималната веројатност на грешка при декодирање на некој симбол i , за $i = 1, 2, \dots, k$.

Дефиниција 6.9. Коефициент на пренос или рата на кодот или брзина на даден блок код (k, n) се дефинира со:

$$R = \frac{\log k}{n} \quad \text{бита по пренос.}$$

Дефиниција 6.10. Ратата R е остварлива ако постои низа од $(2^{nR}; n)$ кодови така што максималната веројатност за грешка при преносот за која било порака со должина n конвергира кон 0, ако $n \rightarrow +\infty$.

Дефиниција 6.11. Оперативен капацитет на комуникациски канал е супремум од сите остварливи рати на тој канал.

Се покажува дека оперативниот капацитет на каналот се совпаѓа со информацискиот капацитет. Тоа следува од следнава теорема.

Теорема 6.1. (Теорема на Шенон за кодирање на канал со шум) За произволен дискретен канал без меморија, точни се следните тврдења:

- Информацискиот капацитет на каналот го задоволува следното својство. За произволен $\varepsilon > 0$, и рата $R < C$, за доволно големо N , постои код со должина N и рата $\geq R$ и алгоритам за декодирање, така што максималната веројатност на блок грешка е помала од ε .
- Ако веројатноста за бит грешка во каналот е p_b , тогаш може да се постигнат рати до

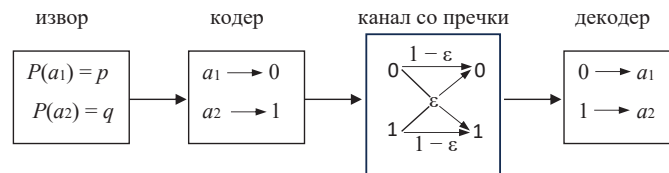
$$R(p_b) = \frac{C}{1 - H(p_b, 1 - p_b)}.$$

- За произволно p_b , рата поголема од $R(p_b)$ не е остварлива.

Оваа теорема тврди дека преку канал со шум со произволно мала веројатност за бит грешка, може да се комуницира кога ратата на пренос е под максимумот кој е константен за каналот.

6.5. Решени задачи

Задача 6.5.1. Извор на информации во фиксни временски интервали емитира по една од две можни пораки a_1 и a_2 со веројатности $P(a_1) = p$ и $P(a_2) = 1 - p = q$. Во кодерот на пораката a_1 и се придружува симбол 0, а на a_2 симбол 1, така што низ каналот се пренесуваат нули и единици. Каналот е бинарен симетричен со веројатност ε ($0 < \varepsilon < 1/2$) за погрешно пренесен бит. Шематски тоа може да се претстави на следниот начин:



Да се определи веројатноста за грешка во преносот.

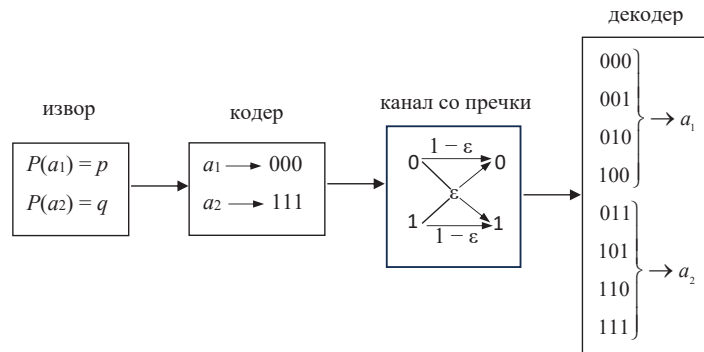
Решение:

Дефинираме случаен настан E : се појавила грешка при преносот (од декодерот е добиена порака различна од таа што ја емитирал изворот). Да воочиме дека ако се испраќа порака a_1 , таа се кодира со 0, па грешка при пренос ќе се појави ако 0 се пренесе како 1, бидејќи 1 се декодира во a_2 . Веројатноста за тоа е ε . Оттука, $P(E|a_1) = \varepsilon$. Од исти причини, и $P(E|a_2) = \varepsilon$. Оттука, за веројатноста на настанот E , т.е. за веројатноста за грешка во преносот низ овој канал се добива:

$$P_1(E) = P(E|a_1)P(a_1) + P(E|a_2)P(a_2) = p \cdot \varepsilon + q \cdot \varepsilon = (p + q)\varepsilon = \varepsilon.$$

Интерпретација на резултатот: Ако изворот емитира една порака во секунда, а низ каналот се пренесува еден бит во секунда, тогаш примачот коректно ќе прими $P_1(E) = 1 - \varepsilon$ пораки во секунда. □

Задача 6.5.2. Нека изворот е дефиниран исто како во претходната задача. Кодирањето и декодирањето шематски се претставени на следната слика:



Да се определи веројатноста за грешка во преносот и да се спореди со $P_1(E)$ од задача 1.

Решение:

Го разгледуваме истиот случаен настан E : се појавила грешка при преносот. Во овој случај, ако се испраќа порака a_1 , таа се кодира со 000, па грешка при пренос ќе се појави, ако на излез од каналот се добие една од пораките

011, 101, 110, 111, кои се декодираат во a_2 . За да се случи ова при преносот на 000, има две можности:

- две нули да се пренесат како единици, едната нула да биде коректно пренесена. Веројатноста за тоа е $\binom{3}{2}\varepsilon^2(1-\varepsilon)$. Биномниот коефициент се јавува затоа што 2 нули (од вкупно 3) може да се избераат на $\binom{3}{2}$ начини.
- сите три нули да бидат погрешно пренесени (како единици). Веројатноста за тоа е ε^3 .

Оттука,

$$P(E|a_1) = 3\varepsilon^2(1-\varepsilon) + \varepsilon^3 = \varepsilon^2(3-2\varepsilon).$$

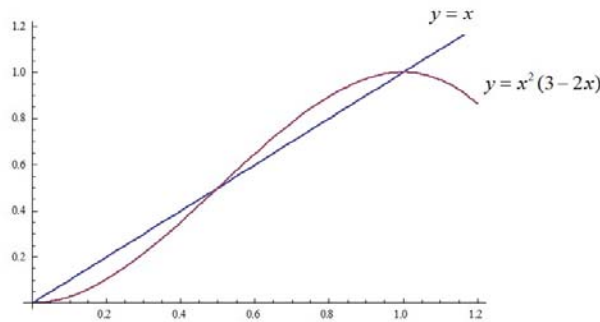
Од исти причини, и

$$P(E|a_2) = 3\varepsilon^2(1-\varepsilon) + \varepsilon^3 = \varepsilon^2(3-2\varepsilon).$$

Сега, за веројатноста на настанот E , грешка при пренос низ каналот, се добива:

$$P_3(E) = P(a_1)P(E|a_1) + P(a_2)P(E|a_2) = (p+q)\varepsilon^2(3-2\varepsilon) = \varepsilon^2(3-2\varepsilon)$$

За да ги споредиме $P_1(E)$ и $P_3(E)$ ќе ги разгледаме графиците на функциите $f(x) = x^2(3-2x)$ и $f(x) = x$.



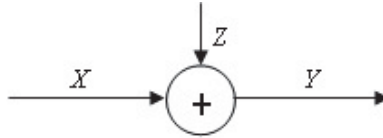
Од сликата се гледа дека за $0 \leq x \leq 1/2$, $x^2(3-2x) \leq x$ и равенство важи за $x = 0$ и $x = 1/2$. Оттука, следува дека $P_3(E) < P_1(E)$ за $0 < \varepsilon <$

1/2. Да воочиме дека со зголемување на должината на кодните замени (во кодерот) вкупната веројатност за грешка при пренос на порака е намалена, но намалена е и брзината на пренос. Имено, во првиот случај точно се пренесува $P_1(E) = 1 - \varepsilon$ пораки во секунда, а во вториот случај се пренесуваат $\frac{1 - P_3(E)}{3}$ пораки во секунда. Притоа, за $0 < \varepsilon < \frac{1}{2}$,

$$\frac{1 - P_3(E)}{3} = \frac{1}{3} (1 - \varepsilon^2(\varepsilon - 2\varepsilon)) < 1 - \varepsilon.$$

□

Задача 6.5.3. (Канал со адитивен шум) Да се определи капацитетот на следниот дискретен канал без меморија:



каде $P\{Z = 0\} = P\{Z = a\} = 1/2$. Азбуката на изворот е $R_X = \{0, 1\}$ и Z е независно од X . Да се определи како капацитетот на каналот зависи од a .

Решение:

Нека $a = 0$. Тогаш $Y = X$ и затоа $C = \max_{p(x)} I(X; Y) = \max_{p(x)} H(X) = 1$, т.е., капацитетот е 1 бит по испраќање.

Нека $a = 1$. Тогаш $R_Y = \{0, 1, 2\}$

$$\begin{aligned} p(0|0) &= P\{Y = 0|X = 0\} = P\{X + Z = 0|X = 0\} = P\{Z = 0\} = 1/2 \\ p(1|0) &= P\{Y = 1|X = 0\} = P\{X + Z = 1|X = 0\} = P\{Z = 1\} = 1/2 \\ p(2|0) &= P\{Y = 2|X = 0\} = P\{X + Z = 2|X = 0\} = 0 \\ p(0|1) &= P\{Y = 0|X = 1\} = P\{X + Z = 0|X = 1\} = 0 \\ p(1|1) &= P\{Y = 1|X = 1\} = P\{X + Z = 1|X = 1\} = P\{Z = 0\} = 1/2 \\ p(2|1) &= P\{Y = 2|X = 1\} = P\{X + Z = 2|X = 1\} = P\{Z = 1\} = 1/2. \end{aligned}$$

Значи, матрицата на каналот е:

$$\Pi = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

Оваа матрица е иста со матрицата на бинарен канал со бришење каде што $\alpha = 1/2$. Оттука, капацитетот на овој канал е $C = 1 - \alpha = 1/2$ бит по пренос.

Нека $a = -1$. Тогаш $R_Y = \{-1, 0, 1\}$. Овој случај е сличен на претходниот кога $a = 1$. Значи, капацитетот е $C = 1 - \alpha = 1/2$ бит по пренос.

Нека $a \neq 0, 1$ и -1 . Тогаш $R_Y = \{0, 1, a, a + 1\}$

$$\begin{aligned} p(0|0) &= P\{Y = 0|X = 0\} = P\{X + Z = 0|X = 0\} = P\{Z = 0\} = 1/2 \\ p(1|0) &= P\{Y = 1|X = 0\} = P\{X + Z = 1|X = 0\} = 0 \\ p(a|0) &= P\{Y = a|X = 0\} = P\{X + Z = a|X = 0\} = P\{Z = a\} = 1/2 \\ p(a + 1|0) &= P\{Y = a + 1|X = 0\} = P\{X + Z = a + 1|X = 0\} = 0 \end{aligned}$$

$$\begin{aligned} p(0|1) &= P\{Y = 0|X = 1\} = P\{X + Z = 0|X = 1\} = 0 \\ p(1|1) &= P\{Y = 1|X = 1\} = P\{X + Z = 1|X = 1\} = P\{Z = 0\} = 1/2 \\ p(a|1) &= P\{Y = a|X = 1\} = P\{X + Z = a|X = 1\} = 0 \\ p(a + 1|1) &= P\{Y = a + 1|X = 1\} = P\{X + Z = a + 1|X = 1\} \\ &= P\{Z = a\} = 1/2. \end{aligned}$$

Матрицата на каналот е:

$$\Pi = \begin{bmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{bmatrix}$$

Значи, ако е познато Y , знаеме кое X е пратено. Оттука, $H(X|Y) = 0$, па

$$C = \max_{p(x)} H(X) = 1$$

и се постигнува за рамномерна распределба на X .

□

Задача 6.5.4. Се разгледува дискретен канал без меморија каде што $Y = X + Z \pmod{11}$,

$$Z : \begin{pmatrix} 1 & 2 & 3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

и $X \in \{0, 1, \dots, 10\}$. Претпоставуваме дека Z не зависи од X .

а) Да се определи капацитетот на каналот.

б) За која распределба на влезот, се постигнува тој капацитет?

Решение:

а) Множеството вредности на случајната променлива Y е $R_Y = \{0, 1, \dots, 10\}$. За дадено $x \in R_X = \{0, 1, \dots, 10\}$ се добива:

$$\begin{aligned} p_Y(x + 1(\bmod 11)|x) &= P\{X + Z = x + 1|X = x\} = P\{Z = 1\} = 1/3 \\ p_Y(x + 2(\bmod 11)|x) &= P\{X + Z = x + 2|X = x\} = P\{Z = 2\} = 1/3 \\ p_Y(x + 3(\bmod 11)|x) &= P\{X + Z = x + 3|X = x\} = P\{Z = 3\} = 1/3. \end{aligned}$$

За $i \neq 1, 2, 3$, се добива:

$$p_Y(x + i(\bmod 11)|x) = P\{X + Z = x + i|X = x\} = P\{Z = i\} = 0.$$

Матрицата на каналот е:

$$\Pi = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Значи, станува збор за симетричен канал, бидејќи сите редици се пермутации една на друга, а и сите колони се пермутации една на друга. Оттука,

$$C \leq H(Y) - H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = H(Y) - \log 3.$$

б) Капацитетот е максимален, ако распределбата на Y е рамномерна. Ќе покажеме дека ако распределбата на влезот е рамномерна, т.е. $p(x) = 1/|R_X|$, тогаш и распределбата на излезот е рамномерна. Имено,

$$p_Y(y) = \sum_{x \in R_X} p_Y(y|x)p_X(x) = \frac{1}{|R_X|} \sum_{x \in R_X} p_Y(y|x) = \frac{1}{|R_X|} = \frac{1}{|R_Y|}.$$

Во овој случај,

$$C = \log 11 - \log 3 = \log \frac{11}{3}.$$

□

Задача 6.5.5. (Z -канал) Z -каналот има бинарен влез и излез и матрица

$$\Pi = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}.$$

Да се определи капацитетот на Z -каналот и распределбата на влезот за која се постигнува.

Решение:

Нека $p = P\{X = 0\}$. За условната ентропија $H(Y|X)$, се добива:

$$\begin{aligned} H(Y|X) &= P\{X = 0\} H(Y|X = 0) + P\{X = 1\} H(Y|X = 1) \\ &= pH(1, 0) + (1 - p)H(1/2, 1/2) = p \cdot 0 + (1 - p) \cdot 1 = 1 - p \end{aligned}$$

За распределбата на Y , имаме:

$$\begin{aligned} P\{Y = 0\} &= p \cdot P\{Y = 0|X = 0\} + (1 - p) \cdot P\{Y = 0|X = 1\} \\ &= p + (1 - p) \cdot \frac{1}{2} = \frac{1 + p}{2} \end{aligned}$$

$$\begin{aligned} P\{Y = 1\} &= p \cdot P\{Y = 1|X = 0\} + (1 - p) \cdot P\{Y = 1|X = 1\} \\ &= p \cdot 0 + (1 - p) \cdot \frac{1}{2} = \frac{1 - p}{2}. \end{aligned}$$

Сега, ентропијата на Y е:

$$H(Y) = -\frac{1-p}{2} \log \frac{1-p}{2} - \frac{1+p}{2} \log \frac{1+p}{2}$$

Оттука, заемната информација помеѓу X и Y е

$$I(X; Y) = H(Y) - H(Y|X) = -\frac{1-p}{2} \log \frac{1-p}{2} - \frac{1+p}{2} \log \frac{1+p}{2} - (1-p).$$

Ако $p = 0$ или $p = 1$, тогаш $I(X; Y) = 0$. Значи, $I(X; Y)$ има максимум за $0 < p < 1$. Следно, треба да се определи вредноста на p (стационарната точка) за која функцијата

$$f(p) = I(X; Y) = -\frac{1-p}{2} \log \frac{1-p}{2} - \frac{1+p}{2} \log \frac{1+p}{2} - (1-p)$$

има максимум. Со диференцирање на $f(p)$, се добива:

$$\begin{aligned} f'(p) &= \frac{1}{2} \log_2 \frac{1-p}{2} + \frac{1-p}{2} \cdot \frac{2}{1-p} \cdot \frac{1}{\ln 2} \cdot \frac{1}{2} \\ &\quad - \frac{1}{2} \log_2 \frac{1+p}{2} - \frac{1+p}{2} \cdot \frac{2}{1+p} \cdot \frac{1}{\ln 2} \cdot \frac{1}{2} + 1 \\ &= \frac{1}{2} \log_2 \frac{1-p}{2} - \frac{1}{2} \log_2 \frac{1+p}{2} + 1 = \frac{1}{2} \log_2 \frac{1-p}{1+p} + 1 \end{aligned}$$

Со изедначување на последниот извод со 0, се добива:

$$f'(p) = 0,$$

т.е.

$$\frac{1}{2} \log_2 \frac{1-p}{1+p} + 1 = 0.$$

Оттука,

$$\frac{1}{2} \log_2 \frac{1-p}{1+p} = -1$$

$$\log_2 \frac{1-p}{1+p} = -2.$$

Со решавање на логаритамската равенка се добива $p = \frac{3}{5}$. За оваа вредност, функцијата $f(p) = I(X; Y)$ има максимум и тој изнесува 0.322, т.е. $C = 0.322$. \square

Задача 6.5.6. Еден извор емитира симболи А, В, С со веројатности 0.3, 0.3 и 0.4 соодветно. На излезот од каналот симболите се појавуваат со веројатности 0.3, 0.4 и 0.3. Веројатноста дека симболот В ќе биде добро пренесен низ каналот е 0.6, за симболот А е 0.7, за симболот С е 0.6. Веројатноста дека при преносот симболот А ќе се промени во В е 0.3. Да се определи:

- а) матрицата на каналот;
- б) веројатноста за грешка;
- в) $H(X)$, $H(Y)$, $H(X, Y)$, $I(X; Y)$, $H(Y|X)$;

Решение: Од условите на задачата распределбите на влезот X и излезот Y на каналот, се дадени со:

$$\mathbf{p}_X = [0.3 \ 0.3 \ 0.4], \quad \mathbf{p}_Y = [0.3 \ 0.4 \ 0.3].$$

а) Матрицата на каналот е:

$$\Pi = \begin{bmatrix} 0.7 & 0.3 & p_1 \\ p_2 & 0.6 & p_3 \\ p_4 & p_5 & 0.6 \end{bmatrix}.$$

Непознатите веројатности во матрицата на каналот се определуваат така што да биде задоволено равенството (6.3), т.е. $\mathbf{p}_X \Pi = \mathbf{p}_Y$ и сумата на веројатностите во секоја редица во матрицата да биде еднаква на 1. Односно, од следниот систем равенки:

$$\begin{cases} 0.21 + 0.3p_2 + 0.4p_4 = 0.3 \\ 0.09 + 0.18 + 0.4p_5 = 0.4 \\ 0.3p_1 + 0.3p_3 + 0.24 = 0.3 \\ 0.7 + 0.3 + p_1 = 1 \\ p_2 + 0.6 + p_3 = 1 \\ p_4 + p_5 + 0.6 = 1 \end{cases}$$

Решение на овој систем равенки е:

$$\begin{cases} p_1 = 0 \\ p_2 = 0.2 \\ p_3 = 0.2 \\ p_4 = 0.075 \\ p_5 = 0.325 \end{cases}$$

Значи, матрицата на каналот е:

$$\Pi = \begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.2 & 0.6 & 0.2 \\ 0.075 & 0.325 & 0.6 \end{bmatrix}.$$

б) Треба да се определи веројатноста на случајниот настан E – погрешен пренос низ каналот (добиеен е симбол различен од пратениот симбол).

$$\begin{aligned} P(E) &= \sum_{\alpha_i \in R_X} P\{E|X = \alpha_i\}P\{X = \alpha_i\} \\ &= 0.3(0.3 + 0) + 0.3(0.2 + 0.2) + 0.4(0.075 + 0.325) = 0.37. \end{aligned}$$

в) Од \mathbf{p}_X и \mathbf{p}_Y е јасно дека ентропиите на X и Y се еднакви.

$$\begin{aligned} H(X) &= H(Y) = H(0.3, 0.3, 0.4) \\ &= -0.3 \log 0.3 - 0.3 \log 0.3 - 0.4 \log 0.4 = 1.57 \text{ бита.} \end{aligned}$$

Законот на распределба на векторот (X, Y) се добива со множење на секоја редица во матрицата со соодветната веројатност од распределбата на влезот X , затоа што

$$p(i, j) = P\{X = i, Y = j\} = P\{X = i\} \cdot P\{Y = j|X = i\}.$$

$X \backslash Y$	y_A	y_B	y_C
x_A	0.21	0.09	0
x_B	0.06	0.18	0.06
x_C	0.03	0.13	0.24

Оттука, се добива:

$$\begin{aligned} H(X, Y) &= -0.21 \log 0.21 - 0.09 \log 0.09 - 0 \log 0 \\ &\quad - 0.06 \log 0.06 - 0.18 \log 0.18 - 0.06 \log 0.06 \\ &\quad - 0.03 \log 0.03 - 0.13 \log 0.13 - 0.24 \log 0.24 \\ &= 2.7464 \text{ бита} \end{aligned}$$

Сега, заемната информација на X и Y е

$$I(X; Y) = H(X) + H(Y) - H(X, Y) = 1.57 + 1.57 - 2.7464 = 0.3936 \text{ бита}$$

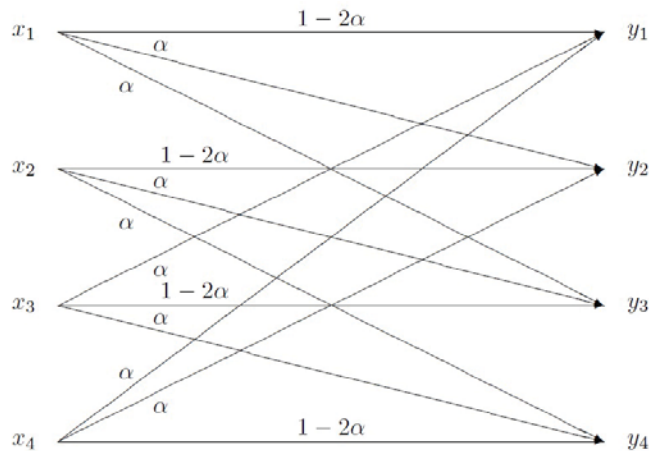
и

$$H(Y|X) = H(X, Y) - H(X) = 2.7464 - 1.57 = 1.1764 \text{ бита.}$$

□

6.6. Задачи

Задача 6.6.1. Извор емитира симболи x_1, x_2, x_3 и x_4 со еднаква веројатност. Потоа, симболите се пренесуваат низ канал без меморија шематски претставен на следната слика.



Ако $\alpha = 0.4$ да се определи:

- матрицата на каналот
- веројатноста за погрешен пренос.

Задача 6.6.2. На влезот на бинарен симетричен канал, влегуваат бинарни пораки во кои симболот 0 се појавува трипати почесто од симболот 1. На излезот од каналот нулата се појавува со веројатност 0.6. Да се определи:

- веројатноста за грешка;

б) $H(X), H(Y), H(X|Y), I(X; Y)$;

в) капацитет на каналот.

Задача 6.6.3. Еден извор емитира симболи 0, 1 и 2 со веројатности 0.2, 0.5 и 0.3 соодветно. Веројатноста дека симболот 0 ќе биде добро пренесен низ каналот е 0.3, за симболот 1 е 0.5, а за симболот 2 е 0.2. Веројатноста дека при преносот 0 ќе се промени во 2 е 0.2, дека 1 ќе се промени во 2 е 0.3, дека 2 ќе се промени во 0 е 0.3. Да се определи:

а) матрицата на каналот;

б) веројатноста за грешка при преносот;

в) $H(Y), H(Y|X)$ и $H(X, Y)$.

Задача 6.6.4. Пораките што се пренесуваат низ еден канал се составени од симболите a_1, a_2 и a_3 , кои на влезот на каналот се јавуваат во однос 3 : 4 : 5. Веројатностите за премин на симболите од еден во друг, при преносот, се дадени со матрицата на каналот:

$$\Pi = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.7 & 0.1 & 0.2 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}.$$

Да се определат ентропиите на влезот и излезот и веројатноста за грешка при преносот низ каналот.

Задача 6.6.5. На влезот на бинарен канал со бришење, влегуваат бинарни пораки во кои симболот 1 се појавува двапати пати почесто од симболот 0. На излезот од каналот нулата се појавува со веројатност 0.3. Да се определи:

а) веројатноста за грешка (бришење);

б) $H(X), H(Y), H(X|Y)$ и $I(X; Y)$;

в) капацитет на каналот.

Глава 7

Диференцијална ентропија

Во поглавјето 2.1 беше дефинирана ентропија на случајна променлива од дискретен тип. Во оваа глава ќе дефинираме ентропија на случајна променлива од апсолутно непрекинат тип. Таквата ентропија се нарекува *диференцијална ентропија*.

7.1. Диференцијална ентропија

Нека X е случајна променлива од апсолутно непрекинат тип, зададена со нејзината густина на распределба $p(x)$.

Дефиниција 7.1. *Диференцијална ентропија* $h(X)$ на случајната променлива X од апсолутно непрекинат тип се дефинира со:

$$h(X) = - \int_{-\infty}^{+\infty} p(x) \log p(x) dx,$$

ако интегралот постои.

Диференцијалната ентропијата се изразува во исти единици како и ентропијата во дискретен случај. Ако основата на логаритамот е 2, мерка за ентропијата е бит, а ако основата е бројот e , мерката е нит.

Пример 7.1. Ќе определиме диференцијална ентропија на случајна променлива X која има рамномерна $U(0, a)$ распределба. Густината на X е зададена со:

$$p(x) = \begin{cases} \frac{1}{a}, & x \in (0, a) \\ 0, & \text{инаку} \end{cases}.$$

За диференцијалната ентропија на X , се добива следното:

$$h(X) = - \int_{-\infty}^{+\infty} p(x) \log p(x) dx = - \int_0^a \frac{1}{a} \log \frac{1}{a} dx = \frac{1}{a} (\log a) \int_0^a dx = \log a.$$

Да воочиме дека за $a < 1$, $\log a < 0$, што покажува дека диференцијалната ентропија може да биде и негативна, за разлика од дискретниот случај кога ентропијата е секогаш ненегативна. \square

Забелешка 7.1. Диференцијалната ентропија не е мерка за просечната количина на информација (самоинформација) што ја содржи една случајна променлива од апсолутно непрекинат тип. Во општ случај, случајна променлива од апсолутно непрекинат тип содржи бесконечна количина на информација. Од ова може да се заклучи дека диференцијалната ентропија не е аналог на ентропијата во дискретен случај. Таа сама по себе нема некое физичко значење, но според својствата што ги има, го оправдува терминот ентропија во своето име.

Во следниот пример, ќе воочиме дека случајна променлива од апсолутно непрекинат тип содржи бесконечна количина на информација.

Пример 7.2. Нека $X \sim U[0, 1)$ распределба. Тогаш X може да се запише како $X = 0.X_1X_2X_3\dots$, каде $X_i \sim U(\{0, 1, 2, \dots, 9\})$, $i = 1, 2, \dots$ и сите X_i се независни случајни променливи. Со користење на верижното правило за ентропија, се добива:

$$\begin{aligned} H(X) &= H(X_1, X_2, X_3, \dots) \\ &= \sum_{i=1}^{+\infty} H(X_i | X_{i-1}, \dots, X_1) \\ &= \sum_{i=1}^{+\infty} H(X_i) \\ &= \sum_{i=1}^{+\infty} \log 10 \\ &= +\infty. \end{aligned}$$

\square

Следната теорема покажува дека диференцијалната ентропија на случајна променлива X од апсолутно непрекинат тип не се менува ако на X се додаде некоја фиксна константа c .

Теорема 7.1. Нека X е случајна променлива од апсолутно непрекинат тип зададена со густина $p(x)$, а c е дадена константа. Тогаш

$$h(X + c) = h(X).$$

Доказ: Нека $Y = X + c$. За функцијата на распределба на Y , се добива:

$$F_Y(y) = P\{Y < y\} = P\{X + c < y\} = P\{X < y - c\} = F_X(y - c).$$

Со диференцирање се добива густината на распределба на Y . Имено,

$$p_Y(y) = F'_Y(y) = F'_X(y - c) = p_X(y - c).$$

Сега, за диференцијалната ентропија на Y , добиваме:

$$h(Y) = \int_{-\infty}^{+\infty} p_Y(y) \log p_Y(y) dy = \int_{-\infty}^{+\infty} p_X(y - c) \log p_X(y - c) dy.$$

Во последниот интеграл се воведува смена $x = y - c$. Тогаш $dx = dy$, а границите на интеграција остануваат непроменети. Оттука,

$$h(Y) = \int_{-\infty}^{+\infty} p_X(y - c) \log p_X(y - c) dy = \int_{-\infty}^{+\infty} p_X(x) \log p_X(x) dx = h(X).$$

□

Со оваа теорема покажавме дека диференцијалната ентропија е инваријантна на додавање на константа на случајна променлива од апсолутно непрекинат тип.

Пример 7.3. Нека $X \sim N(a, \sigma^2)$ распределба. Густината на X е зададена со:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-a)^2/2\sigma^2}.$$

За диференцијалната ентропија се добива:

$$\begin{aligned}
h(X) &= - \int_{-\infty}^{+\infty} p(x) \log p(x) dx \\
&= - \int_{-\infty}^{+\infty} p(x) \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-a)^2/2\sigma^2} \right] dx \\
&= - \int_{-\infty}^{+\infty} p(x) \left[-\log(\sqrt{2\pi\sigma^2}) - \frac{(x-a)^2}{2\sigma^2} \log e \right] dx \\
&= \log(\sqrt{2\pi\sigma^2}) \int_{-\infty}^{+\infty} p(x) dx + \frac{1}{2\sigma^2} (\log e) \int_{-\infty}^{+\infty} (x-a)^2 p(x) dx \\
&= \frac{1}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} DX(\log e) \\
&= \frac{1}{2} \log(2\pi\sigma^2) + \frac{1}{2} \log e \\
&= \frac{1}{2} \log(2\pi e\sigma^2).
\end{aligned}$$

Добивме дека за $X \sim N(a, \sigma^2)$ распределба, диференцијалната ентропија е

$$h(X) = \frac{1}{2} \log(2\pi e\sigma^2).$$

Може да се воочи дека таа не зависи од параметарот a , т.е. од математичкото очекување на случајната променлива X , туку само од нејзината дисперзија. \square

Во следната дефиниција е дадено обопштувањето на дефиницијата на диференцијална ентропија на една случајна променлива во диференцијална ентропија на случаен вектор (X, Y) .

Дефиниција 7.2. Диференцијална ентропија $h(X, Y)$ на случајниот вектор (X, Y) од апсолутно непрекинат тип, зададен со густина на распределба $p(x, y)$, се дефинира со:

$$h(X, Y) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \log p(x, y) dx dy,$$

ако интегралот постои.

Дефиниција 7.3. а) Условна диференцијална ентропија $h(X|Y)$ се дефинира со:

$$h(X|Y) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \log p_X(x|y) dx dy.$$

б) Условна диференцијална ентропија $h(Y|X)$ се дефинира со:

$$h(Y|X) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \log p_Y(y|x) dx dy.$$

Бидејќи $p_X(x|y) = \frac{p(x, y)}{p_Y(y)}$, за условната диференцијална ентропија се добива:

$$\begin{aligned} h(X|Y) &= - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \log p_X(x|y) dx dy \\ &= - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \log \frac{p(x, y)}{p_Y(y)} dx dy \\ &= - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \log p(x, y) dx dy + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \log p_Y(y) dx dy \\ &= h(X, Y) - \left[- \int_{-\infty}^{+\infty} \log p_Y(y) \left(\int_{-\infty}^{+\infty} p(x, y) dx \right) dy \right] \\ &= h(X, Y) - \left[- \int_{-\infty}^{+\infty} p_Y(y) \log p_Y(y) dy \right] \\ &= h(X, Y) - h(Y). \end{aligned}$$

Значи,

$$h(X|Y) = h(X, Y) - h(Y). \quad (7.1)$$

Во продолжение се дадени дефинициите за релативна ентропија и заемна информација на случајни променливи од апсолутно непрекинат тип.

Дефиниција 7.4. Релативна ентропија (или растојание на Кулбак-Леиблер) на две распределби p и q од апсолутно непрекинат тип се дефинира со:

$$D(p||q) = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{q(x)} dx.$$

Дефиниција 7.5. Взаемна информација на две случајни променливи X и Y од апсолутно непрекинат тип, зададени со нивната заедничка густина $p(x, y)$, се дефинира со:

$$I(X; Y) = D(p(x, y)||p_X(x)p_Y(y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \log \frac{p(x, y)}{p_X(x)p_Y(y)} dx dy.$$

Со користење на дефиницијата на заемната информација и равенството $p_X(x|y) = \frac{p(x, y)}{p_Y(y)}$, се добива дека:

$$\begin{aligned} I(X; Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \log \frac{p(x, y)}{p_X(x)p_Y(y)} dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \log \frac{p_X(x|y)}{p_X(x)} dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) [\log p_X(x|y) - \log p_X(x)] dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \log p_X(x|y) dx dy - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \log p_X(x) dx dy \\ &= -h(X|Y) - \int_{-\infty}^{+\infty} \log p_X(x) \left(\int_{-\infty}^{+\infty} p(x, y) dy \right) dx \\ &= -h(X|Y) - \int_{-\infty}^{+\infty} p_X(x) \log p_X(x) dx \\ &= h(X) - h(X|Y). \end{aligned}$$

На сличен начин, со користење на дефиницијата на заемната информација и равенството $p_Y(y|x) = \frac{p(x,y)}{p_X(x)}$, се добива дека:

$$I(X; Y) = h(Y) - h(Y|X).$$

На крај, со користење на равенството (7.1), се добива:

$$\begin{aligned} I(X; Y) &= h(X) - h(X|Y) \\ &= h(X) - [h(X, Y) - h(Y)] \\ &= h(X) + h(Y) - h(X, Y). \end{aligned}$$

Својствата на релативната ентропија и заемната информација во апсолутно непрекинат случај се исти како и во дискретен случај. Со следната теорема и последица се покажува дека релативната ентропија и заемната информација на случајни променливи од апсолутно непрекинат тип се ненегативни величини.

Теорема 7.2. (Ненегативност на релативната ентропија) Релативната ентропија е ненегативна, т.е.

$$D(p||q) \geq 0.$$

Притоа, $D(p||q) = 0$ ако $p(x) = q(x)$.

Доказ: Слично како и во доказот на Теорема 2.4, ќе го искористиме равенството (2.6), т.е. (2.7), според кое $-\log x \geq 1 - x$. Притоа, равенство важи ако и само ако $x = 1$. Во овој случај, добиваме:

$$\begin{aligned} D(p||q) &= \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{q(x)} dx \\ &= - \int_{-\infty}^{+\infty} p(x) \log \frac{q(x)}{p(x)} dx \\ &\geq \int_{-\infty}^{+\infty} p(x) \left(1 - \frac{q(x)}{p(x)}\right) dx \\ &= \int_{-\infty}^{+\infty} p(x) dx - \int_{-\infty}^{+\infty} q(x) dx \\ &= 0. \end{aligned}$$

Притоа, равенство важи ако и само ако $\frac{p(x)}{q(x)} = 1$, за секој $x \in \mathbb{R}$, т.е. ако и само ако $p(x) = q(x)$, за секој $x \in \mathbb{R}$. Ова покажува дека релативната ентропија ќе биде 0 само ако p и q се еднакви распределби. \square

Последица 7.1. (Ненегативност на заемна информација) За произволни две случајни променливи X и Y важи

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

Притоа $I(X; Y) = 0$ ако X и Y се независни случајни променливи.

Доказ: Според Теорема 7.2, релативната ентропија е ненегативна, па

$$I(X; Y) = D(p(x, y) || p_X(x)p_Y(y)) \geq 0.$$

Притоа, равенство важи ако и само ако двете распределби се еднакви, т.е. $p(x, y) = p_X(x)p_Y(y)$, за секои $x, y \in \mathbb{R}$. Ова покажува дека равенство важи ако X и Y се независни случајни променливи. \square

Последица 7.2. (Условно редуцирање на ентропијата)

$$h(X|Y) \leq h(X),$$

при што равенство важи ако X и Y се независни случајни променливи.

Доказ: Доказот следува директно од Последица 7.1, т.е. од ненегативноста на заемната информација. Имено,

$$0 \leq I(X; Y) = h(X) - h(X|Y),$$

па

$$h(X|Y) \leq h(X).$$

Равенство важи ако и само ако $I(X; Y) = 0$, т.е. ако и само ако X и Y се независни случајни променливи. \square

Теорема 7.3. (Верижно правило за ентропијата) Нека векторот (X_1, X_2, \dots, X_n) има густина на распределба $p(x_1, x_2, \dots, x_n)$. Тогаш

$$h(X_1, X_2, \dots, X_n) = \sum_{i=1}^n h(X_i | X_{i-1}, \dots, X_1).$$

Доказ: Доказот се изведува со математичка индукција, слично како доказот на Теорема 2.2 за верижното правило на ентропијата за случајни променливи од дискретен тип. \square

Последица 7.3. За произволни случајни променливи X_1, X_2, \dots, X_n важи:

$$h(X_1, X_2, \dots, X_n) \leq \sum_{i=1}^n h(X_i),$$

при што равенство важи ако X_i се независни случајни променливи.

Доказ: Од Последицата 7.2 следува дека

$$h(X_i | X_{i-1}, \dots, X_1) \leq h(X_i),$$

за произволен $i = 1, 2, \dots, n$. Со користење на ова неравенство и верижното правило се добива следното:

$$\begin{aligned} h(X_1, X_2, \dots, X_n) &= \sum_{i=1}^n h(X_i | X_{i-1}, \dots, X_1) \\ &\leq \sum_{i=1}^n h(X_i). \end{aligned}$$

Притоа, равенство важи ако и само ако $h(X_i | X_{i-1}, \dots, X_1) = h(X_i)$, т.е. ако и само ако X_i е независна од X_{i-1}, \dots, X_1 , за секој $i = 1, 2, \dots, n$. Оттука, X_i , $i = 1, 2, \dots, n$ се независни случајни променливи. \square

Последната теорема од овој дел покажува дека од множеството од сите случајни променливи со еднаква дисперзија, случајната променлива со нормална распределба има најголема диференцијална ентропија.

Теорема 7.4. Нормалната распределба ја максимизира диференцијалната ентропија, т.е. од множеството од сите случајни променливи со еднаква дисперзија, случајната променлива со нормална распределба има најголема диференцијална ентропија.

Доказ: Нека $g(x)$ е густина на $N(a, \sigma^2)$ распределба, а $p(x)$ е произволна густина на распределба со иста дисперзија. Бидејќи диференцијалната ентропија е инваријантна на додавање на константа (Теорема 7.1), можеме да претпоставиме дека и математичкото очекување на распределбата со густина

$p(x)$ е исто со тоа на $g(x)$, т.е. е еднакво на a . За релативната ентропија помеѓу $p(x)$ и $g(x)$, се добива:

$$\begin{aligned}
0 \leq D(p||g) &= \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{g(x)} dx \\
&= \int_{-\infty}^{+\infty} p(x) [\log p(x) - \log g(x)] dx \\
&= \int_{-\infty}^{+\infty} p(x) \log p(x) dx - \int_{-\infty}^{+\infty} p(x) \log g(x) dx \\
&= -h(p) - \int_{-\infty}^{+\infty} p(x) \log g(x) dx. \tag{7.2}
\end{aligned}$$

Во последниот интеграл ќе искористиме дека $g(x)$ е густина на $N(a, \sigma^2)$ распределба, па

$$\begin{aligned}
\int_{-\infty}^{+\infty} p(x) \log g(x) dx &= \int_{-\infty}^{+\infty} p(x) \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}} \right) dx \\
&= \int_{-\infty}^{+\infty} p(x) \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{(x-a)^2}{2\sigma^2} \log e \right] dx \\
&= -\frac{1}{2} \log(2\pi\sigma^2) \int_{-\infty}^{+\infty} p(x) dx - \log e \int_{-\infty}^{+\infty} p(x) \frac{(x-a)^2}{2\sigma^2} dx \\
&= -\frac{1}{2} \log(2\pi\sigma^2) \int_{-\infty}^{+\infty} p(x) dx - \frac{\log e}{2\sigma^2} \int_{-\infty}^{+\infty} (x-a)^2 p(x) dx.
\end{aligned}$$

Во последниот израз, првиот интеграл е еднаков на 1, како интеграл од густина на случајна променлива од апсолутно непрекинат тип над $(-\infty, +\infty)$, а вториот интеграл е дисперзија на случајната променлива со густина $p(x)$ и

математичко очекување a , т.е. вториот интеграл е еднаков на σ^2 . Оттука,

$$\begin{aligned} \int_{-\infty}^{+\infty} p(x) \log g(x) dx &= -\frac{1}{2} \log(2\pi\sigma^2) - \frac{\log e}{2\sigma^2} \cdot \sigma^2 \\ &= -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log e \\ &= -\frac{1}{2} \log(2\pi e\sigma^2) \\ &= -h(g). \end{aligned}$$

Со замена на последниот израз за интегралот во (7.2), се добива:

$$0 \leq D(p||g) = -h(p) + h(g),$$

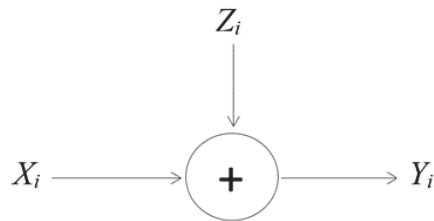
т.е.

$$h(g) \geq h(p),$$

со што тврдењето е покажано. \square

7.2. Гаусов канал

Најзначаен непрекинат канал за пренос на податоци е *Гаусовиот канал*. Тој е временски дискретен канал каде излезот во време i е Y_i и тој излез се добива како збир од влезот X_i и шумот Z_i (Слика 7.1). Шумот Z_i е составен од неза-



Слика 7.1: Гаусов канал

висни и еднакво распределени случајни променливи со нормална (Гаусова) $N(0, N)$ распределба и го нарекуваме *додаден бел Гаусов шум* (*Additive White Gaussian Noise, AWGN*), а каналот се нарекува Гаусов (AWGN) канал. Всушност,

$$Y_i = X_i + Z_i,$$

каде $Z_i \sim N(0, N)$. Претпоставуваме дека шумот Z_i е независен од сигналот X_i . Ова е добар модел со кој може да се опишат повеќе видови на комуникациски канали.

Ако дисперзијата на шумот е нула, тогаш приемникот го прима пренесениот сигнал коректно. Бидејќи X може да прими која било реална вредност, каналот може да пренесе произволен реален број без грешка.

Ако дисперзијата на шумот не е нула и нема ограничување на влезот, можеме да избереме бесконечно подмножество од произволни влезови (доволно оддалечени), така што добиените излези ќе се разликуваат помеѓу себе со мала веројатност на грешка. Таквиот канал ќе има бесконечен капацитет.

Најчесто ограничување на влезот е ограничување на моќноста. За кој било коден збор (x_1, x_2, \dots, x_n) пренесен преку каналот, се бара:

$$\frac{1}{n} \sum_{i=1}^n x_i^2 \leq P$$

каде што P е фиксен број.

Да претпоставиме дека сакаме да испратиме 1 бит преку каналот. За даденото ограничување на моќноста P , најдоброто што може да се направи е да се испрати сигнал чија моќност е со едно од двете аналогни нивоа $+\sqrt{P}$ или $-\sqrt{P}$.

Приемникот го гледа соодветниот примен сигнал Y и се обидува да одлучи на кое од двете нивоа е испратениот сигнал. Претпоставувајќи дека двете нивоа се еднакво веројатни (ова е случај ако сакаме да испратиме точно 1 бит информација), оптималното правило за декодирање е да одлучиме дека е пратен сигнал со моќност $+\sqrt{P}$ ако $Y > 0$, или сигнал со моќност $-\sqrt{P}$ ако $Y < 0$.

7.2.1. Капацитет на Гаусов канал

Во продолжение, ќе го определиме капацитетот на Гаусовиот канал.

Дефиниција 7.6. Информациски капацитет на Гаусов канал со ограничена моќност P се дефинира со:

$$C = \max_{p(x): EX^2 \leq P} I(X; Y),$$

т.е. информацискиот капацитет се пресметува како максимум од заемната информација меѓу влезот и излезот на сите распределби на влезот што го задоволува условот за ограничување на моќноста.

Взаемната информација $I(X; Y)$, со помош на диференцијалната ентропија, се запишува:

$$I(X; Y) = h(Y) - h(Y|X) = h(Y) - h(X + Z|X) = h(Y) - h(Z|X),$$

каде што $h(\cdot)$ ја означува диференцијалната ентропија на соодветната случајна променлива. Но, X и Z се независни случајни променливи, па според тоа:

$$I(X; Y) = h(Y) - h(Z). \quad (7.3)$$

Диференцијалната ентропија на случајната променлива $Z \sim N(0, N)$ е:

$$h(Z) = \frac{1}{2} \log(2\pi eN).$$

Бидејќи X и Z се независни и нивната сума ја дава Y , имаме:

$$EY^2 = E(X+Z)^2 = EX^2 + 2EXZ + EZ^2 = EX^2 + 2EXEZ + EZ^2 \leq P + N,$$

бидејќи $EZ = 0$. Оттука, и $DY \leq EY^2 \leq P + N$. Користејќи ја Теорема 7.4 дека од сите распределби со иста дисперзија, нормалната распределба има максимална ентропијата, добиваме дека

$$h(Y) \leq \frac{1}{2} \log 2\pi e(P + N).$$

Со замена на $h(Y)$ и $h(Z)$ во равенството (7.3), имаме:

$$I(X; Y) \leq \frac{1}{2} \log(2\pi e(P + N)) - \frac{1}{2} \log(2\pi eN) = \frac{1}{2} \log \left(1 + \frac{P}{N} \right).$$

Оттука, информацискиот капацитет на Гаусов канал е:

$$C = \max_{p(x): EX^2 \leq P} I(X; Y) = \frac{1}{2} \log \left(1 + \frac{P}{N} \right)$$

и тој максимум се постигнува кога $X \sim N(0, P)$ распределба. Со ова е покажана следната теорема.

Теорема 7.5. Капацитетот на Гаусов канал со ограничување на моќноста P и дисперзија на шумот N е:

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N} \right).$$

7.2.2. SNR

Во аналогните и дигиталните комуникации, односот сигнал–шум, означен со SNR (Signal-to-Noise Ratio), е мерка за јачината на саканиот сигнал во однос на бучавата во позадината (несаканиот сигнал). Всушност, SNR е мерка употребувана во науката и инженерството, што ги споредува нивото на посакуваниот сигнал со нивото на шумот во позадина. Се дефинира како однос на моќноста на сигналот спрема моќноста на шумот:

$$SNR = \frac{P_{signal}}{P_{noise}}.$$

Ако дисперзијата на сигналот и шумот се познати и сигналот има математичко очекување еднакво на 0, тогаш:

$$SNR = \frac{\sigma_{signal}^2}{\sigma_{noise}^2}.$$

Ако сигналот и шумот се мерат со исти единици, тогаш SNR може да се добие со пресметување на односот на квадратот на амплитудите:

$$SNR = \left(\frac{A_{signal}}{A_{noise}} \right)^2.$$

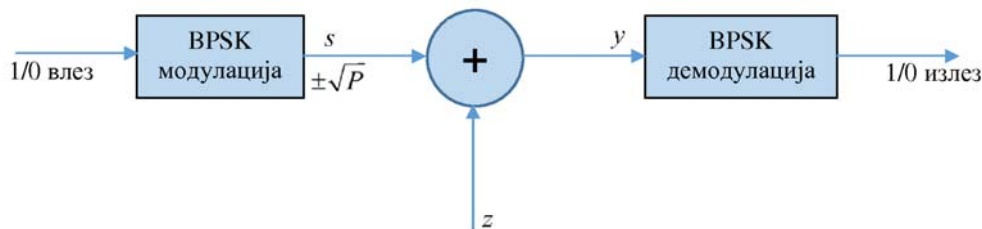
Бидејќи многу сигнали имаат многу широк динамичен опсег, SNR често се изразува со користење на логаритамска скала на децибели. Во децибели SNR се дефинира на следниов начин:

$$SNR_{db} = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right) = 20 \log_{10} \left(\frac{A_{signal}}{A_{noise}} \right).$$

7.2.3. Веројатност на бит грешка при BPSK модулација

Дигиталните модулации се користат при пренос на податоците во мобилните телефони, во научните и геомагнетните инструменти, итн. Која било

дигитална модулациска шема користи конечен број различни симболи, за да се претстават дигиталните податоци. Една од овие модулации е и PSK (Phase-shift keying). Таа користи конечен број на фази, секоја поврзана со единствен модел (патерн) на бинарни цифри. Обично, секоја фаза кодира еднаков број на битови. Секој модел на битови формира симбол, кој е претставен со одредена фаза. Демодулаторот кој е специјално дизајниран за множеството симболи користени од страна на модулаторот, ја определува фазата на примениот сигнал и го пресликува назад во симболите кои го претставуваат, со што се враќаме на оригиналните податоци. Најпрост облик на PSK модулаторот е $BPSK$ (Binary Phase-shift keying) кој користи само 2 фази. Со $BPSK$, бинарните цифри 1 и 0 можат да бидат претставени со аналогните нивоа $+\sqrt{P}$ и $-\sqrt{P}$, соодветно (слика 7.2).



Слика 7.2

Ќе ја пресметаме веројатноста на бит-грешка (Bit-Error Rate), или кратко BER со менување на вредноста на SNR (Signal-to-Noise Ratio). Низ каналот се пушта сигнал s . На него дејствува шум Z , кој има нормална $N(0, \sigma^2)$ распределба. Под дејство на шумот, сигналот s може да се промени. На излезот од каналот се јавува сигнал Y , кој се пресметува со следниве равенки:

- $Y = -\sqrt{P} + Z$, кога се пренесува битот 0,
- $Y = +\sqrt{P} + Z$, кога се пренесува битот 1.

За определување на распределбата на Y се користи следната теорема.

Теорема 7.6. Ако $Z \sim N(0, \sigma^2)$, а $Y = c + Z$ (каде што c е дадена константа), тогаш $Y \sim N(c, \sigma^2)$.

Доказ: Густината на распределба на Z е од облик:

$$p_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/(2\sigma^2)}.$$

За функцијата на распределба на случајната променлива Y се добива:

$$F_Y(y) = P\{Y < y\} = P\{c + Z < y\} = P\{Z < y - c\} = F_Z(y - c).$$

Оттука, за густината на Y имаме:

$$p_Y(y) = F'_Y(y) = F'_Z(y - c) = p_Z(y - c) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-c)^2/(2\sigma^2)}.$$

Значи, $Y \sim N(c, \sigma^2)$. □

Нека сигналот s_1 е модулиран со моќност $+\sqrt{P}$, а сигналот s_0 со моќност $-\sqrt{P}$. Ако е пратен сигналот s_1 , тогаш на излезот од каналот ќе се добие $Y = +\sqrt{P} + Z$. Според Теорема 7.6, условната распределба на Y кога е пратен сигнал s_1 е $N(+\sqrt{P}, N_0)$, па условната густина на Y е:

$$p_Y(y|s_1) = \frac{1}{\sqrt{2\pi N_0}} e^{-(y-\sqrt{P})^2/(2N_0)}.$$

Ако, пак, е пратен сигналот s_0 , тогаш на излезот од каналот ќе се добие $Y = -\sqrt{P} + Z$. Повторно, според Теорема 7.6, условната распределба на Y кога е пратен сигнал s_0 е $N(-\sqrt{P}, N_0)$, па условната густина на Y е:

$$p_Y(y|s_0) = \frac{1}{\sqrt{2\pi N_0}} e^{-(y+\sqrt{P})^2/(2N_0)}.$$

Ќе претпоставиме дека s_0 и s_1 се еднаквоверојатни, т.е. $P(s_0) = P(s_1) = 1/2$.

Демодулацијата се дефинира на следниот начин:

- ако примениот сигнал $y < 0$, тогаш претпоставуваме дека е испратен сигналот s_0 ;
- ако примениот сигнал $y \geq 0$, тогаш претпоставуваме дека е испратен сигналот s_1 .

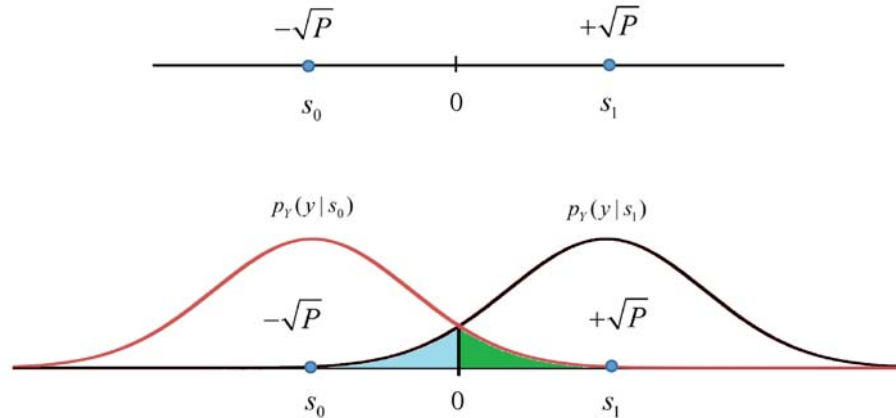
Ако на влезот на каналот е пратен сигнал s_1 , грешка ќе се појави ако, со додавањето на шумот Z , на излезот се добие $Y = +\sqrt{P} + Z < 0$. Веројатноста за тоа е

$$P(e|s_1) = \int_{-\infty}^0 p_Y(y|s_1) dy = \frac{1}{\sqrt{2\pi N_0}} \int_{-\infty}^0 e^{-(y-\sqrt{P})^2/(2N_0)} dy.$$

Во последниот интеграл се воведува смена $z = -\frac{y-\sqrt{P}}{\sqrt{2N_0}}$, па тој добива облик:

$$P(e|s_1) = -\frac{1}{\sqrt{\pi}} \int_{+\infty}^{\sqrt{\frac{P}{2N_0}}} e^{-z^2} dz = \frac{1}{\sqrt{\pi}} \int_{\sqrt{\frac{P}{2N_0}}}^{+\infty} e^{-z^2} dz = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{P}{2N_0}} \right),$$

каде $\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-t^2} dt$. Да воочиме дека $P(e|s_1)$ се совпаѓа со плоштината на делот обоен со сино на слика 7.3.



Слика 7.3

Од друга страна, ако е пратен сигнал s_0 , грешка ќе се појави ако, со додавањето на шумот Z , на излезот се добие $Y = -\sqrt{P} + Z > 0$. Веројатноста

за тоа (површината со зелена боја на слика 7.3) е:

$$P(e|s_0) = \int_0^{+\infty} p_Y(y|s_0) dy = \frac{1}{\sqrt{2\pi N_0}} \int_0^{+\infty} e^{-(y+\sqrt{P})^2/(2N_0)} dy.$$

Слично како и претходно, во овој интеграл воведуваме смена $z = \frac{y + \sqrt{P}}{\sqrt{2N_0}}$, па интегралот добива облик:

$$P(e|s_0) = \frac{1}{\sqrt{\pi}} \int_{\sqrt{\frac{P}{2N_0}}}^{+\infty} e^{-z^2} dz = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{P}{2N_0}} \right).$$

Тоталната веројатност на бит грешка ќе биде

$$P_b = P(s_1)P(e|s_1) + P(s_0)P(e|s_0).$$

Со замена на соодветните веројатности ја добиваме следната равенка за веројатноста на бит грешка:

$$P_b = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{P}{2N_0}} \right).$$

7.3. Решени задачи

Задача 7.3.1. Да се определи диференцијална ентропија на X , ако

- а) $X \sim \varepsilon(\lambda)$,
- б) $X \sim U(a, b)$.

Решение: а) Ако $X \sim \varepsilon(\lambda)$, тогаш $p(x) = \lambda e^{-\lambda x}$, $x \geq 0$. За диференцијалната ентропија на X се добива:

$$\begin{aligned} h(X) &= - \int_0^{+\infty} \lambda e^{-\lambda x} \ln(\lambda e^{-\lambda x}) dx = - \int_0^{+\infty} \lambda e^{-\lambda x} (\ln \lambda - \lambda x) dx \\ &= - \ln \lambda \int_0^{+\infty} \lambda e^{-\lambda x} dx + \lambda \int_0^{+\infty} x \lambda e^{-\lambda x} dx. \end{aligned}$$

Првиот интеграл во последниот израз е еднаков на 1, како интеграл од густина на $\varepsilon(\lambda)$ распределба. Вториот интеграл е математичко очекување на $\varepsilon(\lambda)$ распределба, па тој е еднаков на $1/\lambda$. Така, добиваме:

$$h(X) = -\ln \lambda \cdot 1 + \lambda \cdot \frac{1}{\lambda} = -\ln \lambda + 1 = \ln \frac{e}{\lambda} \text{ нити.}$$

б) Ако $X \sim U(a, b)$, тогаш $p(x) = \frac{1}{b-a}$, $a < x < b$. За диференцијалната ентропија на X се добива

$$\begin{aligned} h(X) &= - \int_a^b \frac{1}{b-a} \log \frac{1}{b-a} dx = \frac{1}{b-a} \log(b-a) \int_a^b dx \\ &= \frac{1}{b-a} \cdot \log(b-a) \cdot (b-a) = \log(b-a). \end{aligned}$$

□

Задача 7.3.2. Да се определи диференцијална ентропија за збирот на X_1 и X_2 , каде X_1 и X_2 се независни нормално распределени случајни променливи со математички очекувања μ_1 и μ_2 и дисперзии σ_1^2 и σ_2^2 , соодветно.

Решение: Ако $X \sim N(\mu, \sigma^2)$ распределба, диференцијалната ентропија е

$$h(X) = \frac{1}{2} \ln(2\pi e \sigma^2).$$

Бидејќи $X_1 \sim N(\mu_1, \sigma_1^2)$ и $X_2 \sim N(\mu_2, \sigma_2^2)$ нивниот збир $Y = X_1 + X_2$ има нормална распределба $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Оттука, за диференцијалната ентропија на збирот добиваме:

$$h(Y) = \frac{1}{2} \ln(2\pi e(\sigma_1^2 + \sigma_2^2)).$$

□

7.4. Задачи

Задача 7.4.1. Да се определи диференцијалната ентропија на случајна променлива X од апсолутно непрекинат тип со густина на распределба $p(x) = 2x$, за $x \in [0, 1)$.

Задача 7.4.2. Да се определи диференцијалната ентропија на случајна променлива X од апсолутно непрекинат тип со густина на распределба $p(x) = \frac{\lambda}{2}e^{-\lambda|x|}$, за $x \in (-\infty, \infty)$ (Лапласова распределба со математичко очекување 0).

Глава 8

Линеарни кодови

Идејата за развивање на кодови на каналот се појавува поради неизбежното постоење на грешки во кој било тип на комуникациски канал. На радиобрановите, електричните сигнали, па дури и на светлосните бранови преку каналите со оптички влакна делува шум на медиумот, што доведува до менување на сигналот. Кодовите на каналот може да се поделат во две големи групи: кодови за откривање на грешки и кодови за поправање на грешки. И едните и другите кодови на влезната порака додаваат одредени редунарни симболи. Тие редунарни симболи помагаат да се открие дали е настаната грешка при преносот или не. Кај кодовите кои откриваат грешки, ако декодерот открие дека настанала грешка при пренос, бара од праќачот пораката да ја прати уште еднаш. Кај кодовите кои поправаат грешки, пак, декодерот се обидува користејќи ги редунарните симболи, да ги поправи грешките при пренос и да ја врати оригиналната порака.

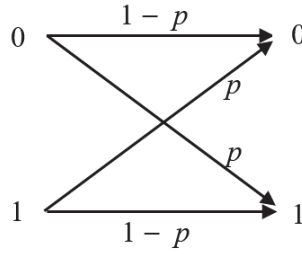
8.1. Концепт на кодирање и декодирање

Разгледуваме бинарен симетричен канал (BSC), каде што $R_X = R_Y = \{0, 1\}$. Шематскиот приказ на овој канал е даден на слика 8.1.

Матрицата на овој канал е:

$$\Pi = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}.$$

Избираме $0 < p < 1/2$. Ако $p > 1/2$, тогаш е поверојатно дека секој бит ќе биде погрешно пренесен. Во тој случај, декодерот би требало, најпрво, да



Слика 8.1

ги замени битовите, 0 со 1, и обратно, па потоа да се обиде да ја декодира примената порака.

Нека M е множеството од сите кодни зборови со должина n кои се јавуваат на влезот на каналот. Ќе претпоставиме дека распределбата на веројатностите на множеството $M \subset \{0, 1\}^n$ е рамномерна, т.е. за сите $\mathbf{x} = (a_1, a_2, \dots, a_n) \in M$ важи $P(\mathbf{x}) = 1/k$, каде $k = |M|$, т.е. k е бројот на сите можни влезни низи (кодни зборови). Ако на излез од каналот е добиен \mathbf{y} , прашањето е како да се декодира за да се добие најмала веројатност на грешка. Нека сега,

$$\max_{\mathbf{x} \in M} P(\mathbf{y}|\mathbf{x}) = P(\mathbf{y}|\mathbf{x}'),$$

па дефинираме $g(\mathbf{y}) = \mathbf{x}'$, за $\mathbf{y} \in \{0, 1\}^n$. Имено, \mathbf{y} се декодира со онаа порака \mathbf{x}' за која $P(\mathbf{y}|\mathbf{x}')$ е најголема од сите условни веројатности $P(\mathbf{y}|\mathbf{x})$. Со тоа е дефиниран алгоритам за декодирање со минимална веројатност за грешка.

Проблемот на овој алгоритам за декодирање е во тоа што за дадено \mathbf{y} , треба да се пресметаат сите веројатности $P(\mathbf{y}|\mathbf{x})$ и да се определи најголемата. Пресметувањето на ваквите условни веројатности не е секогаш едноставна работа. Затоа, се поставува прашањето, дали може да се најде некоја друга величина која ќе биде поедноставна за пресметување, а сепак ќе овозможи декодирање со најмала можна веројатност за грешка. За таа цел, дефинираме функција на растојание $d : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \mathbb{R}$ на следниот начин: за две низи $\mathbf{x} \in \{0, 1\}^n$ и $\mathbf{y} \in \{0, 1\}^n$ го дефинираме растојанието $d(\mathbf{x}, \mathbf{y})$ помеѓу нив како број на координати во кои \mathbf{x} и \mathbf{y} се разликуваат, т.е. ако $\mathbf{x} = (a_1, a_2, \dots, a_n)$ и $\mathbf{y} = (b_1, b_2, \dots, b_n)$, тогаш

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |a_i - b_i|. \quad (8.1)$$

На пример, ако $n = 5$, $\mathbf{x} = 00110$ и $\mathbf{y} = 10101$, тогаш $d(\mathbf{x}, \mathbf{y}) = 3$, бидејќи \mathbf{x} и \mathbf{y} се разликуваат во првата, четвртата и петтата координата.

Лесно се проверува дека дадената функција за растојание (8.1) е добро дефинирана, т.е. за произволни $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \{0, 1\}^n$ важат следните својства.

Својство 1. $d(\mathbf{x}, \mathbf{y}) \geq 0$.

Својство 2. $d(\mathbf{x}, \mathbf{y}) = 0$ ако $\mathbf{x} = \mathbf{y}$.

Својство 3. $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$.

Својство 4. $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$.

Доказ: Доказот на својствата 1, 2 и 3 е очигледен. Во продолжение, ќе го покажеме Својството 4. Нека $\mathbf{x} = (a_1, a_2, \dots, a_n)$, $\mathbf{y} = (b_1, b_2, \dots, b_n)$ и $\mathbf{z} = (c_1, c_2, \dots, c_n)$. Тогаш

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}) &= \sum_{i=1}^n |a_i - b_i| + \sum_{i=1}^n |b_i - c_i| \\ &= \sum_{i=1}^n (|a_i - b_i| + |b_i - c_i|) \\ &\geq \sum_{i=1}^n |a_i - b_i + b_i - c_i| \\ &= \sum_{i=1}^n |a_i - c_i| \\ &= d(\mathbf{x}, \mathbf{z}). \end{aligned}$$

Во изведувањето се користи неравенство на триаголник за $|a_i - b_i|$ и $|b_i - c_i|$, т.е.

$$|a_i - b_i| + |b_i - c_i| \geq |a_i - b_i + b_i - c_i|,$$

за секој $i = 1, 2, \dots, n$. □

Вака дефинираната функција d на множеството $S = \{0, 1\}^n$ се нарекува *Хамингово растојание*. Затоа, може да се зборува за Хамингово растојание на бинарните низи \mathbf{x} и \mathbf{y} . Бидејќи $\mathbf{x} \in M$, $\mathbf{y} \in S$, како влез и излез во BSC, се исто така елементи во $\{0, 1\}^n$, може да се мери оддалеченоста на излезната низа \mathbf{y} од влезната низа \mathbf{x} .

Следната теорема ја дава врската на растојанието помеѓу влезна и излезна низа од каналот со условната веројатност на излезната низа кога е дадена влезната.

Теорема 8.1. Нека е даден BSC со својата матрица

$$\Pi = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}, \quad 0 < p < \frac{1}{2}.$$

Тогаш важи

$$P(\mathbf{y}|\mathbf{x}_1) > P(\mathbf{y}|\mathbf{x}_2) \quad \text{ако} \quad d(\mathbf{x}_1, \mathbf{y}) < d(\mathbf{x}_2, \mathbf{y})$$

за секој $\mathbf{y} \in S$ и соодветните фиксни $\mathbf{x}_1, \mathbf{x}_2 \in M$.

Доказ: Нека $d(\mathbf{x}, \mathbf{y}) = m$ ($0 \leq m \leq n$). Тогаш, од тоа што бинарниот симетричен канал е дискретен канал без меморија, имаме:

$$P(\mathbf{y}|\mathbf{x}) = P((b_1, \dots, b_n)|(a_1, \dots, a_n)) = P(b_1|a_1) \dots P(b_n|a_n),$$

каде $a_j \in \{0, 1\}$ и $b_j \in \{0, 1\}$, $j = 1, 2, \dots, n$. Притоа, $P(b_j|a_j)$ е еднакво или на p (ако битот a_j не е точно пренесен) или на $(1-p)$ (ако битот a_j е точно пренесен). Бидејќи, $d(\mathbf{x}, \mathbf{y}) = m$ следува дека во горниот производ m множителите ќе бидат p , а останатите $(n-m)$ ќе бидат $(1-p)$. Оттука,

$$P(\mathbf{y}|\mathbf{x}) = p^m(1-p)^{n-m}, \quad \text{за } 0 \leq m \leq n.$$

Значи, ако $m_1 = d(\mathbf{x}_1, \mathbf{y})$ и $m_2 = d(\mathbf{x}_2, \mathbf{y})$, може да се напише:

$$P(\mathbf{y}|\mathbf{x}_1) = p^{m_1}(1-p)^{n-m_1},$$

$$P(\mathbf{y}|\mathbf{x}_2) = p^{m_2}(1-p)^{n-m_2}.$$

Сега,

$$\frac{P(\mathbf{y}|\mathbf{x}_1)}{P(\mathbf{y}|\mathbf{x}_2)} = \frac{p^{m_1}(1-p)^{n-m_1}}{p^{m_2}(1-p)^{n-m_2}} = \left(\frac{1-p}{p}\right)^{m_2-m_1}. \quad (8.2)$$

Бидејќи претпоставивме дека $0 < p < 1/2$, следува дека $1/2 < 1-p < 1$. Оттука, $1-p > p$, па

$$\frac{1-p}{p} > 1.$$

Нека претпоставиме дека $P(\mathbf{y}|\mathbf{x}_1) > P(\mathbf{y}|\mathbf{x}_2)$. Тогаш од (8.2) следува дека

$$1 < \frac{P(\mathbf{y}|\mathbf{x}_1)}{P(\mathbf{y}|\mathbf{x}_2)} = \left(\frac{1-p}{p}\right)^{m_2-m_1}.$$

Бидејќи $\frac{1-p}{p} > 1$, следува дека $m_2 - m_1 > 0$, т.е.

$$d(\mathbf{x}_1, \mathbf{y}) = m_1 < m_2 = d(\mathbf{x}_2, \mathbf{y}).$$

Обратно, нека $m_1 < m_2$.

$$1 < \left(\frac{1-p}{p}\right)^{m_2-m_1} = \frac{P(\mathbf{y}|\mathbf{x}_1)}{P(\mathbf{y}|\mathbf{x}_2)}.$$

т.е. $P(\mathbf{y}|\mathbf{x}_1) > P(\mathbf{y}|\mathbf{x}_2)$. □

Врз основа на претходното, јасно е како може да се конструира алгоритам за декодирање за дадено множество $M \subset \{0, 1\}^n$ од кодни зборови и даден бинарен симетричен канал со најмала веројатност на грешка. Имено, за секоја излезна низа $\mathbf{y} \in S$, треба да се најде онаа низа $\mathbf{x}' \in M$, за која Хаминговото растојание $d(\mathbf{x}', \mathbf{y})$ е најмало. Ако има повеќе такви низи, тогаш може да се земе која било од нив. Значи, за секој $\mathbf{y} \in S$, $g(\mathbf{y}) = \mathbf{x}'$, каде што \mathbf{x}' е кодниот збор кој е „најблиску“ до \mathbf{y} , т.е. кодниот збор чие Хамингово растојание до \mathbf{y} е најмало. Оттука, \mathbf{x}' е определен со:

$$\min_{\mathbf{x} \in M} d(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}', \mathbf{y}).$$

Од досегашните разгледувања, може да заклучиме дека за посигурен пренос на информации низ даден бинарен симетричен канал, мора да се изберат потребните k кодни зборови од $S = \{0, 1\}^n$, така што нивните меѓусебни Хамингови растојанија да се што е можно поголеми. Во тој случај, може да се појават и неколку грешки, а притоа, излезната низа да не се „оддалечи“ премногу од влезниот коден збор. Затоа, влезното кодирање на еднаквоверојатните пораки, кои доаѓаат предвид за пренос низ бинарниот симетричен канал, треба да се изврши така што елементите во множеството $M \subset \{0, 1\}^n$ да бидат што е можно пораздалечени во смисла на Хаминговото растојание.

Да претпоставиме дека при преносот на бинарен коден збор со должина n се допушта да настапат најмногу r ($0 \leq r \leq n$) грешки (промена на 1 во 0 или 0 во 1). При оваа претпоставка, главен проблем е следниот: ако бројот на можни пораки е k , како да се одреди минималната должина n на кодните зборови, така што со алгоритмот за декодирање и соодветно влезно кодирање ќе се обезбеди коректен пренос на пораките.

Најпрво да го воочиме следното: Нека $M = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ и нека n е такво што важи

$$d(\mathbf{x}_i, \mathbf{x}_j) \geq 2r + 1, \quad \text{за сите } i \neq j, \text{ каде } i, j \in \{1, \dots, k\},$$

т.е. кои било два различни елементи од M се на Хамингово растојание од барем $2r + 1$. Во тој случај, при r или помалку грешки при пренос на n -члена бинарна низа, со предложениот алгоритам за декодирање, влезната порака ќе се декодира без грешка.

Постапката за декодирање се сведува на следното: за секој $\mathbf{x}_i \in M$, се формира множество S_i од сите $y \in \{0, 1\}^n$, такви што $d(\mathbf{x}_i, \mathbf{y}) \leq r$, т.е.

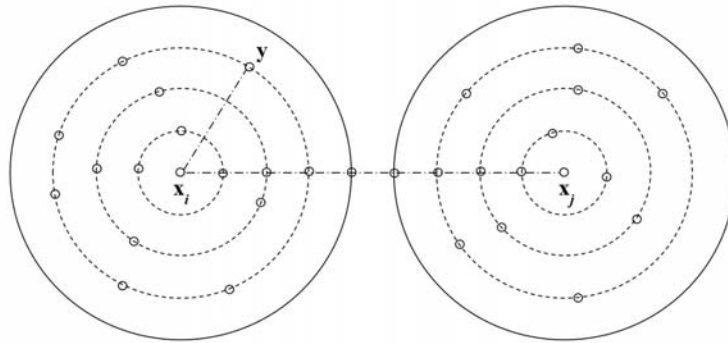
$$S_i = \{\mathbf{y} \in \{0, 1\}^n \mid d(\mathbf{x}_i, \mathbf{y}) \leq r\}, \quad i = 1, 2, \dots, k.$$

Сега, ако на излезот од каналот се добие $\mathbf{y} \in S_i$, тој се декодира со x_i , т.е. дефинираме

$$g(\mathbf{y}) = \mathbf{x}_i, \quad \text{за секој } \mathbf{y} \in S_i, \quad i = 1, 2, \dots, k.$$

Бидејќи, $d(\mathbf{x}_i, \mathbf{x}_j) \geq 2r + 1$, ќе следува дека $S_i \cap S_j = \emptyset$, за $i \neq j$. Според тоа, ако влезната порака е $\mathbf{x}_i \in M$, и ако настанат r или помалку грешки, тогаш излезната порака \mathbf{y} ќе припаѓа на множеството S_i и примената излезна низа ќе се декодира како \mathbf{x}_i , па имаме коректен прием, т.е. декодираната порака ќе се совпадне со влезната. Ако настанат повеќе од r грешки при пренос на влезната низа \mathbf{x}_i , тогаш излезната низа нема да припаѓа во множеството S_i , па таа ќе биде погрешно декодирана и ќе се појави грешка при декодирање.

Множеството S_i може геометриски да се интерпретира како n -димензионална топка со радиус r со центар во точката \mathbf{x}_i (во просторот $S = \{0, 1\}^n$ од сите n -члени бинарни низи со Хамингонова функција на растојание).



Слика 8.2

На слика 8.2, топките се претставени дводимензионално, затоа што не може да се нацрта топка во n -димензионален простор. На сликата, $r = 4$,

$d(\mathbf{x}_i, \mathbf{x}_j) = 9 = 2 \cdot r + 1$. Излезната порака \mathbf{y} се наоѓа во топката со центар \mathbf{x}_i (затоа што $d(\mathbf{x}_i, \mathbf{y}) = 3 < 4 = r$), па затоа \mathbf{y} се декодира со \mathbf{x}_i .

Следната теорема дава потребен услов за должината на кодните зборови, за да се овозможи декодирање без грешка на k еднаквоверојатни пораки, ако максималниот број на грешки при пренос на една n -члена порака е однапред познат.

Теорема 8.2. (Хамингов услов) Нека низ бинарен симетричен канал треба да се пренесуваат k еднаквоверојатни кодни зборови и притоа бројот на погрешно пренесени битови во една n -члена порака да е најмногу r . Тогаш, ќе може да се дефинира алгоритам кој ќе овозможи декодирање без грешка, ако должината n на кодните зборови го исполнува следниот услов:

$$\frac{2^n}{\sum_{i=0}^r \binom{n}{i}} \geq k.$$

Доказ: Најпрво, за фиксно i , да пресметаме колку елементи има во множеството S_i .

- Јасно е дека $\mathbf{x}_i \in S_i$, бидејќи \mathbf{x}_i е центар на топката S_i .
- На оддалеченост 1 од средината \mathbf{x}_i има n елементи. Тоа се сите бинарни n -торки кои се разликуваат од \mathbf{x}_i само во една координата.
- На оддалеченост 2 од \mathbf{x}_i има $\binom{n}{2}$ елементи. Тоа се сите бинарни n -торки кои се разликуваат од \mathbf{x}_i во точно две координати.
- ИТН.

Оттука, јасно е дека бројот на елементи во множеството S_i е

$$1 + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{r} = \sum_{i=0}^r \binom{n}{i}.$$

За да се овозможи декодирање без грешка, мора $S_i \cap S_j = \emptyset$, за $i \neq j$, $i, j \in \{1, 2, \dots, k\}$. Ова значи дека топките S_i , $i = 1, 2, \dots, k$, не смеат да

се пресечат. Затоа, за конструкција на k такви дисјунктни топки, потребни се барем

$$k \sum_{i=0}^r \binom{n}{i}$$

бинарни n -торки. Бидејќи, вкупниот број на бинарни n -торки е 2^n , мора да важи:

$$k \sum_{i=0}^r \binom{n}{i} \leq 2^n,$$

од каде што следува тврдењето. □

Во Табела 8.1, за $r = 2$ и за неколку вредности на $n \in \{5, 6, 7, 8, 9, 10\}$, дадени се вредностите на $2^n / \sum_{i=0}^r \binom{n}{i}$. Од табелата се гледа дека за конструкција на код кој овозможува декодирање без грешка, во случај кога бројот на влезни кодни збора е $k = 10$, и притоа максималниот број на грешки при пренос на еден блок со должина n е 2, тогаш мора за должината на кодните зборови да се земе најмалку $n = 9$. Значи, од можните $2^9 = 512$ деветорки, за вакво сигурносно кодирање на влезните зборови може да се користат не повеќе од 11. Кога не се прави сигурносно кодирање, тогаш за кодирање на 10 влезни пораки, доволно е да се земе $n = 4$, бидејќи постојат $2^4 = 16$ бинарни четворки.

$r = 2$	
n	$2^n / \sum_{i=0}^r \binom{n}{i}$
5	2
6	$32/11 \approx 2.9$
7	$128/29 \approx 4.4$
8	$256/37 \approx 6.9$
9	$256/23 \approx 11.1$
10	$128/7 \approx 18.3$

Табела 8.1

8.2. Линеарни блок кодови

Идејата за овие кодови ќе ја илустрираме на следниот пример.

Пример 8.1. Нека $M_0 = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ е множеството од кодни зборови за даден бинарен симетричен канал. Ако кодните зборови се испраќаат низ каналот такви како што се дадени (без додавање на редундантни битови), секоја грешка во преносот на некој бит, доведува до погрешен прием.

Затоа, на секој коден збор се допишува по 1 бит, така што збирот на битови да биде парен. Со тоа добиваме множество M на кодни зборови, каде $M = \{(0, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0)\}$.

На излезот од каналот може да се добие било која од $2^3 = 8$ бинарни тројки. Декодерот тестира дали во излезната порака бројот на битови е парен. Така, тој може да открие грешка во преносот, ако бројот на погрешно пренесени битови е непарен (1 или 3 погрешно пренесени бита). Ако бројот на погрешно пренесени битови е парен (2 погрешно пренесени бита), декодерот нема да открие грешка и пораката ќе ја прими како точно пренесена.

Ваквите кодови се нарекуваат *кодови кои откриваат грешки*. Кај овие кодови, кога ќе се открие грешка во преносот, се бара од испраќачот, пораката (или еден блок од неа) да ја испрати повторно. \square

Во продолжение, на овој пример ќе му дадеме формална (математичка) позадина, а потоа идејата ќе ја генерализираме. За таа цел, ќе дефинираме поле на Галоа (Galois).

Дефиниција 8.1. Множеството $\{0, 1\}$ со операции собирање и множење по модул 2, дефинирани со

$$\begin{array}{c|cc} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \qquad \begin{array}{c|cc} \cdot & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

се нарекува *поле на Галоа* од ред 2 и се означува со $GF(2)$.

Да воочиме дека во $GF(2)$, за секој елемент $x \in \{0, 1\}$, важи $x + x = 0$, т.е. $x = -x$. Во Пример 8.1, за секој $x = (a_1, a_2, a_3) \in M$, каде $M = \{(0, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0)\}$, важи:

$$a_1 + a_2 + a_3 = 0,$$

така што тестирањето на парноста се сведува на испитување дали равенството е задоволено во $GF(2)$.

Сега, идејата дадена во пример 8.1 ќе ја обопштиме. Нека $\mathbf{F} = [f_{ij}]_{m \times n}$, каде $n \geq m$, е матрица чии елементи се 0 или 1, така што се разгледува следниот систем линеарни равенки со коефициенти од $GF(2)$:

$$\begin{cases} f_{11}a_1 + f_{12}a_2 + \dots + f_{1n}a_n = 0 \\ f_{21}a_1 + f_{22}a_2 + \dots + f_{2n}a_n = 0 \\ \vdots \\ f_{m1}a_1 + f_{m2}a_2 + \dots + f_{mn}a_n = 0 \end{cases} \quad (8.3)$$

Секое решение на овој систем линеарни равенки во $GF(2)$ е некоја n -торка од битови, така што множеството M од сите решенија на овој систем е некое подмножество од множеството $\{0, 1\}^n$. Идејата е, елементите од множеството M да се искористат како кодни зборови, а за секоја излезна низа се тестира дали ги задоволува равенките од системот (8.3).

Матрицата $\mathbf{F} = [f_{ij}]_{m \times n}$ се нарекува *матрица на парност* или *контролна матрица*.

Во општ случај, системот (8.3) може да се запише во матрична форма како $\mathbf{F}\mathbf{x} = \mathbf{o}$, каде

$$\mathbf{x} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}_{n \times 1}, \quad \mathbf{o} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{m \times 1}.$$

Ако рангот на матрицата \mathbf{F} е t , т.е. $\text{rang}(\mathbf{F}) = t$, каде што $1 \leq t \leq m$, тогаш системот $\mathbf{F}\mathbf{x} = \mathbf{o}$ има $n - t = s$ ($0 \leq s \leq n - 1$) независни променливи за кои може да се изберат произволни вредности од полето на Галоа. За секој избор на вредности на тие s непознати, се добиваат одредени вредности за преостанатите $t = n - s$ непознати. Бидејќи за секоја независна променлива може да се избере една од две вредности (0 или 1), бројот на решенија на системот е 2^s . Секоја од тие 2^s бинарни n -торки може да се употреби како коден збор при пренос низ бинарен симетричен канал.

Со ова е добиен блок код (k, n) , каде што $k = 2^s$ ($s = n - r(\mathbf{F})$). Овој блок код се нарекува *линеарен код* со контролна матрица \mathbf{F} . Коефициентот на пренос или ратата на овој код изнесува

$$R = \frac{\log k}{n} = \frac{\log 2^s}{n} = \frac{s}{n}$$

и може да се интерпретира како просечна количина на информација (во битови) која ја „носи“ поединечен сигнал (бит) во n -члената низа од битови.

Пример 8.2. Нека контролната матрица е вектор редица:

$$\mathbf{F} = \underbrace{[1 \ 1 \ \dots \ 1]}_n.$$

Системот $\mathbf{F}\mathbf{x} = \mathbf{0}$ може да се запише во облик:

$$a_1 + a_2 + \dots + a_n = 0.$$

Оттука, $a_n = a_1 + a_2 + \dots + a_{n-1}$, па за секој избор на вредности на a_1, a_2, \dots, a_{n-1} може да се пресмета вредноста на a_n . Бројот на можни избори на вредности за a_1, a_2, \dots, a_{n-1} е 2^{n-1} , па добиениот линеарен код се состои од $k = 2^{n-1}$ кодни зборови.

Рангот на оваа матрица е $t = 1$ и $s = n - 1$, па коефициентот на пренос е

$$R = \frac{n-1}{n} = 1 - \frac{1}{n}.$$

Да воочиме дека со избор на голем природен број n , коефициентот на пренос може да се направи многу блиску до 1. Очигледно е дека добиениот линеарен код овозможува големи брзини на пренос, но можноста за отстранување на грешки е многу мала, затоа што Хамингоновото растојание помеѓу два кодни збора е само 2. \square

Пример 8.3. Да конструираме сега код со голема можност за отстранување на грешките при пренос на поединечни битови во n -члена бинарна низа. Нека матрицата на парност е од облик.

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}_{(n-1) \times n}$$

Системот $\mathbf{F}\mathbf{x} = \mathbf{0}$ во развиена форма, добива облик:

$$\begin{cases} a_1 + a_2 & = 0 \\ a_1 + a_3 & = 0 \\ a_1 + a_4 & = 0 \\ \vdots & \\ a_1 + a_n & = 0 \end{cases} \Leftrightarrow \begin{cases} a_2 = a_1 \\ a_3 = a_1 \\ a_4 = a_1 \\ \vdots \\ a_n = a_1 \end{cases}$$

Оттука, за секој избор на вредности на a_1 може да се пресметаат вредностите на a_2, \dots, a_n . Бројот на можни избори за a_1 е 2 (или 0 или 1), па бројот на решенија на системот е 2:

$$\mathbf{x}_1 = (0, 0, \dots, 0), \quad \mathbf{x}_2 = (1, 1, \dots, 1).$$

Овој код е таканаречен повторувачки код. За големо n , веројатноста дека декодерот ќе открие грешка во пренос е многу голема, практично блиску до 1.

Рангот на оваа матрица е $t = r(\mathbf{F}) = n - 1$ и $s = n - t = 1$. Тогаш $R = s/n = 1/n$, па може да се воочи дека добиениот линеарен код, за големо n , има многу мал коефициент на пренос. \square

Кодот со „дословна проверка на парност“ (пример 8.2) и кодот со „повеќекратно повторување на битови“ (пример 8.3) се два екстремни примера на линеарни кодови. За големо n , првиот овозможува голема брзина на пренос, но веројатноста за точен пренос е многу мала. Од друга страна, вториот овозможува голема веројатност да се открие грешка, но коефициентот на пренос е многу мал. Затоа, се поставува како една од главните задачи, конструирањето на таков линеарен код кој овозможува доволно голем коефициент на пренос, со доволно мала веројатност за грешка при преносот.

8.3. Векторски простор и потпростор

Во ова поглавје ќе ги дадеме дефинициите на векторски простор и потпростор кои ќе ни бидат потребни за понатамошни својства на кодовите.

Дефиниција 8.2. Множеството S е *векторски простор* над полето F (означуваме со $S(F)$), ако се исполнети следните својства:

A1. $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$, за секои $\mathbf{x}, \mathbf{y}, \mathbf{z} \in S$;

- A2. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$, за секои $\mathbf{x}, \mathbf{y} \in S$;
- A3. Постои $\mathbf{o} \in S$, така што $\mathbf{o} + \mathbf{x} = \mathbf{x} + \mathbf{o} = \mathbf{x}$, за секој $\mathbf{x} \in S$;
- A4. За секој $\mathbf{x} \in S$, постои $-\mathbf{x} \in S$, така што $\mathbf{x} + (-\mathbf{x}) = -\mathbf{x} + \mathbf{x} = \mathbf{o}$.
- B1. $\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$, за секои $\mathbf{x}, \mathbf{y} \in S$ и за секој $\lambda \in F$;
- B2. $(\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$, за секој $\mathbf{x} \in S$ и за секои $\lambda, \mu \in F$;
- B3. $\lambda(\mu\mathbf{x}) = (\lambda\mu)\mathbf{x}$, за секој $\mathbf{x} \in S$ и за секои $\lambda, \mu \in F$;
- B4. $1 \cdot \mathbf{x} = \mathbf{x}$, за секој $\mathbf{x} \in S$, каде 1 е единицата во F .

Дефиниција 8.3. Множеството M е *потпростор* од векторскиот простор $S(F)$, ако за секои $\mathbf{x}, \mathbf{y} \in M$ и за секои $\lambda, \mu \in F$, важи:

$$\lambda\mathbf{x} + \mu\mathbf{y} \in M.$$

Да го разгледаме сега множеството $S = \{0, 1\}^n$. Во S се дефинира собирање на вектори како собирање на подредени n -торки по модуло 2:

$$(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n),$$

и множење на вектор со скалар со:

$$\lambda(a_1, a_2, \dots, a_n) = (\lambda a_1, \lambda a_2, \dots, \lambda a_n),$$

каде што $\lambda \in \{0, 1\}$. Лесно се проверува дека множеството S ги задоволува својствата A1-A4, B1-B4, па е векторски простор над полето на Галоа $GF(2)$. Ова е основниот факт на кој ќе се темелат понатамошните разгледувања.

Да воочиме дека за секој $\mathbf{x} \in S$, $\mathbf{x} + \mathbf{x} = \mathbf{o}$, т.е. $\mathbf{x} = -\mathbf{x}$, така што во просторот S , одземањето се совпаѓа со собирањето.

8.4. Конструкција на контролна матрица за даден линеарен блок код

Во поглавјето 8.2 видовме дека решенијата на системот $\mathbf{F}\mathbf{x} = \mathbf{o}$ може да се користат како кодни зборови. Следната теорема покажува дека множеството од сите решенија на овој систем равенки е векторски потпростор од $S = \{0, 1\}^n$.

Теорема 8.3. Множеството $M = \{\mathbf{x} \in S \mid \mathbf{F}\mathbf{x} = \mathbf{o}\}$ од сите решенија на равенката $\mathbf{F}\mathbf{x} = \mathbf{o}$, е конечен потпростор од S .

Доказ: Согласно со дефиниција 8.3, треба да покажеме дека за произволни $\mathbf{x}, \mathbf{y} \in M$ и произволни $\lambda, \mu \in GF(2)$, $\lambda\mathbf{x} + \mu\mathbf{y} \in M$, т.е. $\lambda\mathbf{x} + \mu\mathbf{y}$ е исто така решение на системот $\mathbf{F}\mathbf{x} = \mathbf{o}$. Од тоа што $\mathbf{x}, \mathbf{y} \in M$ следува дека $\mathbf{F}\mathbf{x} = \mathbf{o}$ и $\mathbf{F}\mathbf{y} = \mathbf{o}$. Сега, добиваме:

$$\mathbf{F}(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda\mathbf{F}\mathbf{x} + \mu\mathbf{F}\mathbf{y} = \lambda \cdot \mathbf{o} + \mu \cdot \mathbf{o} = \mathbf{o}.$$

Оттука, $\lambda\mathbf{x} + \mu\mathbf{y} \in M$, па множеството M ($M \subset S$), од сите кодни зборови на линеарниот блок код $(2^s, n)$ е потпростор од векторскиот простор S над полето $GF(2)$. \square

Теоремата покажува дека ако е дадена контролна матрица и сите решенија на системот $\mathbf{F}\mathbf{x} = \mathbf{o}$ се земат за кодни зборови, тогаш множеството од тие кодни зборови е потпростор од $S(F)$, каде што $S = \{0, 1\}^n$. Следната теорема покажува дека важи и обратното. Имено, ако е даден потпростор M од кодни зборови, тогаш може да се најде контролна матрица A , така што сите решенија на системот $\mathbf{A}\mathbf{x} = \mathbf{o}$ се кодните зборови од M .

Теорема 8.4. За секој потпростор $M \subset S = \{0, 1\}^n$ постои матрица \mathbf{A} со n колони, таква што за секој $\mathbf{x} \in M$, важи $\mathbf{A}\mathbf{x} = \mathbf{o}$. Уште повеќе, секое решение на $\mathbf{A}\mathbf{x} = \mathbf{o}$ е елемент на множеството M .

Доказ: Најпрво, ќе покажеме дека постои природен број $s \leq n$, така што бројот на елементи во множеството M е 2^s .

За да се покаже тоа, се разгледува матрицата \mathbf{V} од ред $\mu \times n$, чии редици се сите вектори од M . Нека матрицата \mathbf{V} има ранг $s \leq n$. Тогаш секој вектор $\mathbf{x} \in M$ може да се претстави како линеарна комбинација на s линеарно независни вектори $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_s \in M$, така што важи:

$$\mathbf{x} = \lambda_1\mathbf{e}_1 + \lambda_2\mathbf{e}_2 + \dots + \lambda_s\mathbf{e}_s, \quad \lambda_i \in \{0, 1\}, \quad i = 1, \dots, s.$$

Коефициентите $\lambda_1, \lambda_2, \dots, \lambda_s$ може да се изберат на 2^s различни начина и со секој различен избор се добива различен елемент $\mathbf{x} \in M$. Затоа множеството M има 2^s елемента.

Понатаму, ќе ја формираме матрицата \mathbf{A} и ќе покажеме дека секој $\mathbf{x} \in M$ ја задоволува равенката $\mathbf{Ax} = \mathbf{o}$. Бидејќи $r(\mathbf{B}) = s$, помеѓу колоните на матрицата \mathbf{B} има s линеарно независни. Без губење на општоста, нека тоа се последните s , т.е. $(n-s+1)$ -та, $(n-s+2)$ -та, ..., $(n-1)$ -та и n -та колона. Тогаш секоја од првите $n-s$ колони на матрицата \mathbf{B} може да се претстави како линеарна комбинација на последните s колони. Тогаш, j -тата колона на \mathbf{B} може да се претстави како:

$$\begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{\mu j} \end{bmatrix} = \lambda_{j1} \begin{bmatrix} b_{1,n-s+1} \\ b_{2,n-s+1} \\ \vdots \\ b_{\mu,n-s+1} \end{bmatrix} + \lambda_{j2} \begin{bmatrix} b_{1,n-s+2} \\ b_{2,n-s+2} \\ \vdots \\ b_{\mu,n-s+2} \end{bmatrix} + \dots + \lambda_{j,s-1} \begin{bmatrix} b_{1,n-1} \\ b_{2,n-1} \\ \vdots \\ b_{\mu,n-1} \end{bmatrix} + \lambda_{js} \begin{bmatrix} b_{1n} \\ b_{2n} \\ \vdots \\ b_{\mu n} \end{bmatrix},$$

за $j = 1, 2, \dots, n-s$. Значи, за секој $j = 1, 2, \dots, n-s$, се добива по еден систем со μ линеарни равенки:

$$b_{ij} = \lambda_{j1}b_{i,n-s+1} + \lambda_{j2}b_{i,n-s+2} + \dots + \lambda_{j,s-1}b_{i,n-1} + \lambda_{js}b_{in}, \quad i = 1, \dots, \mu. \quad (8.4)$$

Но, секој од тие системи може да се редуцира на систем од s линеарно независни равенки за определување на коефициентите λ_{ji} ($j = 1, 2, \dots, n-s$, $i = 1, 2, \dots, s$). На тој начин, се добива матрицата

$$\mathbf{A} = [\mathbf{I}_{n-s}, -\mathbf{\Lambda}],$$

каде \mathbf{I}_{n-s} е единечна матрица од ред $n-s$, а $\mathbf{\Lambda} = [\lambda_{ij}]$ е матрица од претходно определените коефициенти.

Ако се земе $\mathbf{x}_i = (b_{i1}, b_{i2}, \dots, b_{in}) \in M$, се добива:

$$\begin{aligned} \mathbf{Ax}_i &= \begin{bmatrix} 1 & 0 & \dots & 0 & -\lambda_{11} & -\lambda_{12} & \dots & -\lambda_{1s} \\ 0 & 1 & \dots & 0 & -\lambda_{21} & -\lambda_{22} & \dots & -\lambda_{2s} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & -\lambda_{n-s,1} & -\lambda_{n-s,2} & \dots & -\lambda_{n-s,s} \end{bmatrix} \cdot \begin{bmatrix} b_{i1} \\ b_{i2} \\ \vdots \\ b_{in} \end{bmatrix} \\ &= \begin{bmatrix} b_{i1} - \lambda_{11}b_{i,n-s+1} - \lambda_{12}b_{i,n-s+2} - \dots - \lambda_{1s}b_{in} \\ b_{i2} - \lambda_{21}b_{i,n-s+1} - \lambda_{22}b_{i,n-s+2} - \dots - \lambda_{2s}b_{in} \\ \vdots \\ b_{i,n-s} - \lambda_{n-s,1}b_{i,n-s+1} - \lambda_{n-s,2}b_{i,n-s+2} - \dots - \lambda_{n-s,s}b_{in} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \end{aligned}$$

Последното равенство важи, затоа што λ_{ji} се решенија на системите равенки (8.4).

На крај, ќе покажеме дека секое решение на системот равенки $\mathbf{Ax} = \mathbf{0}$ е елемент на множеството M . Јасно е дека матрицата $\mathbf{A} = [\mathbf{I}_{n-s}, -\Lambda]$ има ранг $r(\mathbf{A}) = n - s$, така што општото решение на системот равенки $\mathbf{Ax} = \mathbf{0}$ содржи s произволни параметри за кои може да се изберат 2 вредности (0 или 1). Оттука, постојат 2^s решенија на системот $\mathbf{Ax} = \mathbf{0}$, а на почетокот докажавме дека множеството M содржи 2^s елемента. Значи, не постои решение на системот равенки $\mathbf{Ax} = \mathbf{0}$ кое не е елемент на M . \square

Да нагласиме уште еднаш дека последната теорема дава можност за дадени кодни зборови (кои формираат векторски потпростор од $S = \{0, 1\}^n$), да се најде контролна матрица со која декодерот ќе проверува дали се појавила грешка при пренос (која може да се открие) или не.

Пример 8.4. Нека $n = 6$ и нека потпросторот $M \subset \{0, 1\}^6$ ги содржи елементите:

$$\begin{aligned} \mathbf{b}_0 &= (0, 0, 0, 0, 0, 0), \\ \mathbf{b}_1 &= (1, 0, 1, 0, 0, 1), \\ \mathbf{b}_2 &= (1, 1, 0, 0, 1, 0), \\ \mathbf{b}_3 &= (0, 1, 1, 0, 1, 1), \\ \mathbf{b}_4 &= (1, 1, 1, 1, 0, 0), \\ \mathbf{b}_5 &= (0, 1, 0, 1, 0, 1), \\ \mathbf{b}_6 &= (0, 0, 1, 1, 1, 0), \\ \mathbf{b}_7 &= (1, 0, 0, 1, 1, 1). \end{aligned}$$

Матрицата чии редици се елементите од потпросторот M е

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

Да воочиме дека $r(\mathbf{B}) = 3$ и, на пример, векторите \mathbf{b}_1 , \mathbf{b}_2 и \mathbf{b}_4 се линеарно независни, а останатите вектори може да се изразат како линеарна комбинација

на овие три вектори. Навистина:

$$\begin{aligned}\mathbf{b}_3 &= \mathbf{b}_1 + \mathbf{b}_2 \\ \mathbf{b}_5 &= \mathbf{b}_1 + \mathbf{b}_4 \\ \mathbf{b}_6 &= \mathbf{b}_2 + \mathbf{b}_4 \\ \mathbf{b}_7 &= \mathbf{b}_1 + \mathbf{b}_2 + \mathbf{b}_4.\end{aligned}$$

Формираме нова матрица чии редици се овие три линеарно независни вектори:

$$\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

Оваа матрица се нарекува *генераторна матрица* за дадениот код. Името доаѓа од таму што секој коден збор од дадениот код може да се претстави како линеарна комбинација на редиците од матрицата со коефициенти од $GF(2)$.

Последните три колони на оваа матрица се линеарно независни, па секоја од првите три колони може да се претстави како линеарна комбинација на тие линеарно независни колони. Така, се добиваат три системи линеарни равенки. Првиот систем во векторска форма е:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \lambda_{11} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \lambda_{12} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \lambda_{13} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Оттука,

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda_{13} \\ \lambda_{12} \\ \lambda_{11} \end{bmatrix},$$

па решенијата на овој систем се $\lambda_{11} = \lambda_{12} = \lambda_{13} = 1$. Вториот систем линеарни равенки во векторска форма има облик:

$$\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \lambda_{21} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \lambda_{22} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \lambda_{23} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

и неговите решенија се: $\lambda_{21} = 1$, $\lambda_{22} = 1$, $\lambda_{23} = 0$. И третиот систем е

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \lambda_{31} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \lambda_{32} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \lambda_{33} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

со решенија $\lambda_{31} = 1$, $\lambda_{32} = 0$, $\lambda_{33} = 1$. Согласно доказот на Теорема 8.4, матрицата на парност е од облик:

$$\mathbf{A} = [\mathbf{I}_{n-s}, -\Lambda],$$

каде што $n = 6$, $s = r(\mathbf{B}) = 3$, па

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix},$$

така што $\mathbf{F} = \mathbf{A}$ е контролна матрица (матрица на парност) за блок кодот ($k = 8$, $n = 6$) со зададени кодни зборови $\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_7$. \square

Во алгоритмот кој го презентиравме во доказот на Теорема 8.4 земавме дека последните s колони на генераторната матрица се линеарно независни, па првите $n - s$ колони беа претставени како линеарна комбинација од нив. На тој начин, ги определивме коефициентите λ_{ij} и добивме дека контролната матрица е од облик

$$\mathbf{F} = [\mathbf{I}_{n-s}, -\Lambda].$$

Прашање е што ако последните s колони не се линеарно независни? Како тогаш ќе се конструира контролната матрица? Со постапка слична на претходно опишаната, се покажува дека при формирање на контролната матрица, коефициентите λ_{ij} се ставаат во матрицата F во оние колони чии редни броеви соодветствуваат на редните броеви на линеарно независните колони во генераторната матрица. Останатите колони на контролната матрица се пополнуваат со единечни вектори. На пример, ако во генераторната матрица, линеарно независни се втората и четвртата колона, тогаш во контролната матрица втората и четвртата колона ќе бидат формирани од коефициентите λ_{ij} , а останатите колони ќе се пополнат со единечни вектори. Ако, пак, линеарно независни колони во генераторната матрица се првите s , тогаш согласно претходната дискусија, контролната матрица е од облик

$$\mathbf{F} = [-\Lambda^T, \mathbf{I}_{n-s}].$$

Можеме да ги издвоиме и следните два специјални случаи кога линеарно независните колони формираат единечна матрица. Имено,

- ако генераторната матрица е од облик $\mathbf{G} = [g_{ij}]_{s \times n} = [\mathbf{I}_s, \mathbf{A}]$, тогаш контролната матрица \mathbf{F} е од облик

$$\mathbf{F} = [-\mathbf{A}^T, \mathbf{I}_{n-s}].$$

– ако генераторната матрица е од облик $\mathbf{G} = [g_{ij}]_{s \times n} = [\mathbf{A}, \mathbf{I}_s]$, тогаш контролната матрица \mathbf{F} е од облик

$$\mathbf{F} = [\mathbf{I}_{n-s}, -\mathbf{A}^T].$$

8.5. Алгоритам за корекција на грешки кај линеарен блок код

Досега, линеарните блок кодови ги разгледувавме како кодови кои откриваат грешки. Со користење на контролната матрица, декодерот проверува дали пораката што се добива на излезот од каналот е коден збор или не е. Ако не е, тогаш бара од испраќачот да ја прати таа порака (или дел од пораката) уште еднаш. Во овој дел, ќе воведеме алгоритам кој ќе овозможи декодерот да поправи одреден број грешки при пренос на пораките низ каналот.

Нека \mathbf{F} е матрица од ред $m \times n$ ($m \leq n$), $r(\mathbf{F}) = m$ и нека \mathbf{F} е контролна матрица за линеарен блок код чие множество кодни зборови е M . За секоја излезна низа $\mathbf{y} \in S = \{0, 1\}^n$, во општ случај, ќе важи $\mathbf{F}\mathbf{y} = \mathbf{c}$, каде $\mathbf{c} \in \{0, 1\}^m$ е одредена вектор-колона. Векторот \mathbf{c} се нарекува *синдром* (или *коректор*) на векторот \mathbf{y} . Сега, за секое $\mathbf{c} \in \{0, 1\}^m$, може да се разгледува множеството

$$S_{\mathbf{c}} = \{\mathbf{y} \in S \mid \mathbf{F}\mathbf{y} = \mathbf{c}\}.$$

Значи, во $S_{\mathbf{c}}$ се наоѓаат сите $\mathbf{y} \in S$ кои имаат заеднички синдром \mathbf{c} .

Од друга страна, ќе избереме $\mathbf{y}_0 \in S_{\mathbf{c}}$ и ќе го разгледаме множеството

$$\mathbf{y}_0 + M = \{\mathbf{y} \in S \mid \mathbf{y} = \mathbf{y}_0 + \mathbf{x}, \mathbf{x} \in M\} \subseteq S.$$

Ќе покажеме дека важи $S_{\mathbf{c}} = \mathbf{y}_0 + M$. Имено, ако $\mathbf{y} \in S_{\mathbf{c}}$, тогаш $\mathbf{F}\mathbf{y}_0 = \mathbf{F}\mathbf{y} = \mathbf{c}$. Оттука,

$$\mathbf{F}(\mathbf{y} - \mathbf{y}_0) = \mathbf{F}\mathbf{y} - \mathbf{F}\mathbf{y}_0 = \mathbf{o}.$$

Тоа значи дека $\mathbf{y} - \mathbf{y}_0$ е решение на системот равенки $\mathbf{F}\mathbf{x} = \mathbf{o}$, т.е. $\mathbf{y} - \mathbf{y}_0$ е коден збор и припаѓа на множеството кодни зборови M . Оттука, постои $\mathbf{x} \in M$, така што $\mathbf{x} = \mathbf{y} - \mathbf{y}_0$, т.е. $\mathbf{y} = \mathbf{y}_0 + \mathbf{x}$. Тоа значи дека $\mathbf{y} \in \mathbf{y}_0 + M$, т.е. $S_{\mathbf{c}} \subseteq \mathbf{y}_0 + M$.

Обратно, ако $\mathbf{y} \in \mathbf{y}_0 + M$, тогаш $\mathbf{y} = \mathbf{y}_0 + \mathbf{x}$, за некој $\mathbf{x} \in M$. Оттука,

$$\mathbf{F}\mathbf{y} = \mathbf{F}(\mathbf{y}_0 + \mathbf{x}) = \mathbf{F}\mathbf{y}_0 + \mathbf{F}\mathbf{x} = \mathbf{c} + \mathbf{o} = \mathbf{c},$$

т.е. $\mathbf{y} \in S_{\mathbf{c}}$, па $\mathbf{y}_0 + M \subseteq S_{\mathbf{c}}$. Значи, $S_{\mathbf{c}} = \mathbf{y}_0 + M$.

Пред да го конструираме алгоритмот за корекција на грешки кај линеарен блок код, за секој вектор $\mathbf{y} = (b_1, b_2, \dots, b_n) \in S$, може да се дефинира величината

$$T(\mathbf{y}) = \sum_{i=1}^n b_i, \quad (8.5)$$

која се нарекува *тежина* на бинарната низа \mathbf{y} и која покажува колку единици има во овој вектор. Да нагласиме дека збирот во (8.5) се пресметува декадно. Ако се примени дефиницијата на Хамингово растојание, може да се напише:

$$T(\mathbf{y}) = \sum_{i=1}^n b_i = \sum_{i=1}^n |b_i - 0| = d(\mathbf{0}, \mathbf{y}).$$

Од формулата е јасно дека тежината $T(\mathbf{y})$ може да се интерпретира како растојание од векторот \mathbf{y} до координатниот почеток.

Да воочиме дека Хаминговото растојание, дефинирано со

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |a_i - b_i|,$$

каде што $\mathbf{x} = (a_1, a_2, \dots, a_n)$ и $\mathbf{y} = (b_1, b_2, \dots, b_n)$, го има и следното својство.

Својство 8.1.

$$d(\mathbf{x} + \mathbf{z}, \mathbf{y} + \mathbf{z}) = d(\mathbf{x}, \mathbf{y}),$$

за секои $\mathbf{x}, \mathbf{y}, \mathbf{z} \in S$.

Доказ: Нека $\mathbf{x} = (a_1, a_2, \dots, a_n)$, $\mathbf{y} = (b_1, b_2, \dots, b_n)$ и $\mathbf{z} = (c_1, c_2, \dots, c_n)$. Тогаш имаме:

$$\mathbf{x} + \mathbf{z} = (a_1 + c_1, \dots, a_n + c_n), \quad \mathbf{y} + \mathbf{z} = (b_1 + c_1, \dots, b_n + c_n).$$

Сега,

$$\begin{aligned} d(\mathbf{x} + \mathbf{z}, \mathbf{y} + \mathbf{z}) &= \sum_{i=1}^n |(a_i + c_i) - (b_i + c_i)| \\ &= \sum_{i=1}^n |a_i - b_i| \\ &= d(\mathbf{x}, \mathbf{y}). \end{aligned}$$

□

Својствата на тежината, дефинирана со (8.5), се дадени со следната теорема.

Теорема 8.5.

- i.* $T(\mathbf{y}) \geq 0$, за секој $\mathbf{y} \in \mathbf{S}$;
- ii.* $T(\mathbf{y}) = 0$ ако $\mathbf{y} = \mathbf{0}$;
- iii.* $T(\mathbf{x} + \mathbf{y}) \leq T(\mathbf{x}) + T(\mathbf{y})$, за секои $\mathbf{x}, \mathbf{y} \in \mathbf{S}$.

Доказ: Доказот на *i.* и *ii.* е очигледен, па затоа ќе го покажеме само *iii.*
Добиваме:

$$\begin{aligned} T(\mathbf{x} + \mathbf{y}) &= d(\mathbf{0}, \mathbf{x} + \mathbf{y}) \leq d(\mathbf{0}, \mathbf{x}) + d(\mathbf{x}, \mathbf{x} + \mathbf{y}) \\ &= d(\mathbf{0}, \mathbf{x}) + d(\mathbf{0}, \mathbf{y}) \quad (\text{согласно со (8.1)}) \\ &= T(\mathbf{x}) + T(\mathbf{y}). \end{aligned}$$

□

Од Својство 8.1 на Хаминговото растојание произлегува следното:

$$d(\mathbf{y} - \mathbf{y}_0, \mathbf{y}) = d(\mathbf{y} - \mathbf{y}_0 - \mathbf{y}, \mathbf{y} - \mathbf{y}) = d(-\mathbf{y}_0, \mathbf{0}) = d(\mathbf{0}, \mathbf{y}_0) = T(\mathbf{y}_0).$$

Значи, за произволни $\mathbf{y}, \mathbf{y}_0 \in S$ важи:

$$d(\mathbf{y} - \mathbf{y}_0, \mathbf{y}) = T(\mathbf{y}_0).$$

Ако избереме таков $\mathbf{y}_1 \in S$, за кој важи $T(\mathbf{y}_1) \geq T(\mathbf{y}_0)$, тогаш од претходното равенство е јасно дека важи:

$$d(\mathbf{y} - \mathbf{y}_0, \mathbf{y}) \leq d(\mathbf{y} - \mathbf{y}_1, \mathbf{y}).$$

Сега, ако на излез од каналот се добие $\mathbf{y} \in S$, се поставува прашање како да се декодира? Според претходното, идеален алгоритам за декодирање на даден линеарен блок код со контролна матрица F и множество M од кодни зборови, се конструира така што на излезот $\mathbf{y} \in S$ се придружува оној влез $\mathbf{x}_0 = g(\mathbf{y}) \in M$, кој се добива на следниот начин:

- Прво, се наоѓа множеството $S_c = \{\mathbf{y}_1 \in S \mid F\mathbf{y}_1 = F\mathbf{y} = \mathbf{c}\}$.
- Потоа, се определува оној $\mathbf{y}_0 \in S_c$ кој има минимална тежина, т.е. $T(\mathbf{y}_0) \leq T(\mathbf{y}_1)$, за секој $\mathbf{y}_1 \in S_c$. Ако постојат повеќе вектори со минимална тежина, тогаш се избира еден од нив. Така избраниот \mathbf{y}_0 се нарекува *лидер*. Се зема дека всушност \mathbf{y}_0 е вектор на грешка.

– Затоа, \mathbf{y} се декодира со $\mathbf{x}_0 = \mathbf{y} - \mathbf{y}_0$.

Во тој случај, од $d(\mathbf{y} - \mathbf{y}_0, \mathbf{y}) \leq d(\mathbf{y} - \mathbf{y}_1, \mathbf{y})$, следува дека

$$d(\mathbf{x}_0, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{y}),$$

за секој $\mathbf{x} = \mathbf{y} - \mathbf{y}_1 \in M$. Последното неравенство важи за секој $\mathbf{x} \in M$, бидејќи покажавме дека $S_c = \mathbf{y}_0 + M$, па ако \mathbf{y}_1 ги прими сите вредности од S_c , тогаш $\mathbf{y} - \mathbf{y}_1$ ќе ги прими сите вредности од M .

Како резултат на претходната дискусија, може да се даде следниот алгоритам за декодирање на линеарен блок код. На излез од каналот се добива $\mathbf{y} \in S$. Се поставува прашање како тој да се декодира.

Алгоритам за декодирање на линеарен блок код

Влез: излезот од каналот $\mathbf{y} \in S$, Излез: декодираниот збор \mathbf{x}_0

Чекор 1. Се наоѓа множеството $S_c = \{\mathbf{y}_1 \in S \mid \mathbf{F}\mathbf{y}_1 = \mathbf{F}\mathbf{y} = \mathbf{c}\}$.

Чекор 2. Се определува оној $\mathbf{y}_0 \in S_c$ кој има минимална тежина.

Чекор 3. \mathbf{y} се декодира со $\mathbf{x}_0 = \mathbf{y} - \mathbf{y}_0$.

Пример 8.5. Нека контролната матрица на кодот е

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Рангот $r(\mathbf{F})$ на матрицата \mathbf{F} е $t = 4$, па имаме $s = n - t = 6 - 4 = 2$ информациски бита и 4 контролни бита. Оттука, матричната равенка $\mathbf{F}\mathbf{x} = \mathbf{0}$ има $2^s = 2^2 = 4$ решенија. Соодветниот систем линеарни равенки гласи:

$$\begin{cases} a_1 & & & + a_5 & + a_6 & = 0 \\ & a_2 & & + a_5 & & = 0 \\ & & a_3 & + a_5 & + a_6 & = 0 \\ & & & a_4 & & + a_6 & = 0 \end{cases}$$

Со елементарни алгебарски операции во полето на Галоа, овој систем се сведува на

$$\begin{cases} a_1 & = a_5 + a_6 \\ a_2 & = a_5 \\ a_3 & = a_5 + a_6 \\ a_4 & = a_6 \end{cases}.$$

Решенијата на овој системот се:

$$\mathbf{x}_1 = (0, 0, 0, 0, \mathbf{0}, \mathbf{0}), \quad \mathbf{x}_2 = (1, 1, 1, 0, \mathbf{1}, \mathbf{0}),$$

$$\mathbf{x}_3 = (1, 0, 1, 1, \mathbf{0}, \mathbf{1}), \quad \mathbf{x}_4 = (0, 1, 0, 1, \mathbf{1}, \mathbf{1}),$$

каде што со црвено се означени вредностите на независните променливи. Вака добиените \mathbf{x}_1 , \mathbf{x}_2 , \mathbf{x}_3 и \mathbf{x}_4 се кодни зборови.

Бидејќи $m = 4$, има вкупно $2^4 = 16$ различни синдроми, па за секој синдром $\mathbf{c} \in \{0, 1\}^4$ се формира множеството $S_{\mathbf{c}}$ од излезни низи, каде

$$S_{\mathbf{c}} = \{\mathbf{y} \in \{0, 1\}^6 \mid \mathbf{F}\mathbf{y} = \mathbf{c}\}.$$

На секој синдром одговара соодветно множество $S_{\mathbf{c}} = \mathbf{y}_0 + M$ кое содржи 4 шесточлени бинарни низи. Ако $\mathbf{c} = \mathbf{0}$, тогаш елементите на $S_{\mathbf{c}}$ се кодните зборови. Нека $\mathbf{c} = (1, 0, 0, 0)$. Тогаш векторската равенка $\mathbf{F}\mathbf{y} = \mathbf{c}$ може да се развие во следниот систем линеарни равенки:

$$\left\{ \begin{array}{cccccc} a_1 & & & + & a_5 & + & a_6 & = & 1 \\ & a_2 & & & + & a_5 & & = & 0 \\ & & a_3 & & + & a_5 & + & a_6 & = & 0 \\ & & & a_4 & & + & a_6 & = & 0 \end{array} \right\} \iff \left\{ \begin{array}{l} a_1 = a_5 + a_6 + 1 \\ a_2 = a_5 \\ a_3 = a_5 + a_6 \\ a_4 = a_6 \end{array} \right.$$

Решенијата на системот

$$\mathbf{y}_0 = (1, 0, 0, 0, \mathbf{0}, \mathbf{0}), \quad \mathbf{y}_1 = (0, 1, 1, 0, \mathbf{1}, \mathbf{0}),$$

$$\mathbf{y}_2 = (0, 0, 1, 1, \mathbf{0}, \mathbf{1}), \quad \mathbf{y}_3 = (1, 1, 0, 1, \mathbf{1}, \mathbf{1}),$$

се елементи на $S_{\mathbf{c}}$ за $\mathbf{c} = (1, 0, 0, 0)$, т.е.

$$S_{\mathbf{c}} = \{(1, 0, 0, 0, 0, 0), (0, 1, 1, 0, 1, 0), (0, 0, 1, 1, 0, 1), (1, 1, 0, 1, 1, 1)\}.$$

Првиот вектор $\mathbf{y}_0 = (1, 0, 0, 0, 0, 0)$ има најмала тежина, па тој се зема за лидер, т.е. вектор на грешка.

Векторите (низите) од множеството $S_{\mathbf{c}}$ се декодираат на следниот начин:

$$\mathbf{y}_0 - \mathbf{y}_0 = (1, 0, 0, 0, 0, 0) + (1, 0, 0, 0, 0, 0) = (0, 0, 0, 0, 0, 0) = \mathbf{x}_1 \in M,$$

$$\mathbf{y}_1 - \mathbf{y}_0 = (0, 1, 1, 0, 1, 0) + (1, 0, 0, 0, 0, 0) = (1, 1, 1, 0, 1, 0) = \mathbf{x}_2 \in M,$$

$$\mathbf{y}_2 - \mathbf{y}_0 = (0, 0, 1, 1, 0, 1) + (1, 0, 0, 0, 0, 0) = (1, 0, 1, 1, 0, 1) = \mathbf{x}_3 \in M,$$

$$\mathbf{y}_3 - \mathbf{y}_0 = (1, 1, 0, 1, 1, 1) + (1, 0, 0, 0, 0, 0) = (0, 1, 0, 1, 1, 1) = \mathbf{x}_4 \in M.$$

Да се потсетиме дека во полето на Галоа $x - y = x + y$, за секои $x, y \in \{0, 1\}$. Сега ја формираме табелата каде што во првиот ред се внесува синдромот $(0, 0, 0, 0)$ и добиените кодни зборови. Во следните редици се внесува прво синдромот, а потоа добиените излези од каналот кои се пишуваат под кодните зборови со кои соодветно се декодираат. Значи, за овој код првите две редици од табелата ќе бидат:

\mathbf{c}	$S_{\mathbf{c}} = \mathbf{y}_0 + M$			
0000	000000	111010	101101	010111
1000	100000	011010	001101	110111

За да се поедностави записот, во табелата векторите се внесени без загради и записки. Во засенчената колона се наоѓа еден од елементите во соодветното множество $S_{\mathbf{c}}$ кој има минимална тежина и кој се избира за лидер.

За илустрација, ќе видиме уште што се случува, и ако добиениот синдром е $\mathbf{c} = (0, 1, 0, 0)$. Тогаш векторската равенка $\mathbf{F}\mathbf{y} = \mathbf{c}$ може да се развие во следниот систем линеарни равенки:

$$\begin{cases} a_1 & & & + a_5 + a_6 = 0 \\ & a_2 & & + a_5 = 1 \\ & & a_3 & + a_5 + a_6 = 0 \\ & & & a_4 + a_6 = 0 \end{cases} \iff \begin{cases} a_1 = a_5 + a_6 \\ a_2 = a_5 + 1 \\ a_3 = a_5 + a_6 \\ a_4 = a_6 \end{cases}$$

Решенијата на системот

$$\mathbf{y}_0 = (0, 1, 0, 0, 0, 0), \quad \mathbf{y}_1 = (1, 0, 1, 0, 1, 0),$$

$$\mathbf{y}_2 = (1, 1, 1, 1, 0, 1), \quad \mathbf{y}_3 = (0, 0, 0, 1, 1, 1),$$

се елементи на $S_{\mathbf{c}}$ за $\mathbf{c} = (0, 1, 0, 0)$, т.е.

$$S_{\mathbf{c}} = \{(0, 1, 0, 0, 0, 0), (1, 0, 1, 0, 1, 0), (1, 1, 1, 1, 0, 1), (0, 0, 0, 1, 1, 1)\}.$$

Векторот $\mathbf{y}_0 = (0, 1, 0, 0, 0, 0)$ има најмала тежина, па тој се избира за лидер.

Се декодира на следниот начин:

$$\mathbf{y}_0 - \mathbf{y}_0 = (0, 1, 0, 0, 0, 0) + (0, 1, 0, 0, 0, 0) = (0, 0, 0, 0, 0, 0) = \mathbf{x}_1 \in M,$$

$$\mathbf{y}_1 - \mathbf{y}_0 = (1, 0, 1, 0, 1, 0) + (0, 1, 0, 0, 0, 0) = (1, 1, 1, 0, 1, 0) = \mathbf{x}_2 \in M,$$

$$\mathbf{y}_2 - \mathbf{y}_0 = (1, 1, 1, 1, 0, 1) + (0, 1, 0, 0, 0, 0) = (1, 0, 1, 1, 0, 1) = \mathbf{x}_3 \in M,$$

$$\mathbf{y}_3 - \mathbf{y}_0 = (0, 0, 0, 1, 1, 1) + (0, 1, 0, 0, 0, 0) = (0, 1, 0, 1, 1, 1) = \mathbf{x}_4 \in M.$$

Сега, претходната табела може да се дополни со уште една редица, соодветна на синдромот $\mathbf{c} = (0, 1, 0, 0)$.

\mathbf{c}	$S_{\mathbf{c}} = \mathbf{y}_0 + M$			
0000	000000	111010	101101	010111
1000	100000	011010	001101	110111
0100	010000	101010	111101	000111

Ако оваа постапка се повтори и за останатите 13 синдроми, табелата го добива следниот облик.

	\mathbf{c}	$S_{\mathbf{c}} = \mathbf{y}_0 + M$				$T(\mathbf{y}_0)$
K_0	0000	000000	111010	101101	010111	0
K_1	1000	100000	011010	001101	110111	1
	0100	010000	101010	111101	000111	1
	0010	001000	110010	100101	011111	1
	0001	000100	111110	101001	010011	1
	1110	000010	111000	101111	010101	1
	1011	000001	111011	101100	010110	1
K_2	1100	110000	001010	011101	100111	2
	1010	101000	010010	000101	111111	2
	1001	100100	011110	001001	110011	2
	0110	100010	011000	001111	110101	2
	0011	100001	011011	001100	110110	2
	0101	010100	101110	111001	000011	2
	1111	010001	101011	111100	000110	2
K_3	1101	110100	001110	011001	100011	3
	0111	110001	001011	011100	100110	3

Табела 8.2: Табела за декодирање

Во Табела 8.2, синдромите \mathbf{c} не се поредени лексикографски, туку прво се оние кои соодветствуваат на множество $S_{\mathbf{c}} = \mathbf{y}_0 + M$ со лидер \mathbf{y}_0 чија минимална тежина е $T(\mathbf{y}_0) = 0$, потоа синдромите на кои соодветствува множество $S_{\mathbf{c}}$ со лидер чија минимална тежина е $T(\mathbf{y}_0) = 1$, па $T(\mathbf{y}_0) = 2$ и на крај, $T(\mathbf{y}_0) = 3$. Да воочиме дека, во класата K_0 има само едно множество $S_{\mathbf{c}}$, и тоа се совпаѓа со множеството (потпросторот) од кодни зборови и во него $\mathbf{y}_0 = (0, 0, 0, 0, 0, 0)$ е лидерот со тежина $T(\mathbf{y}_0) = 0$. Во класата K_1 има 6 множества $S_{\mathbf{c}}$ и во секое има по еден елемент со минимална тежина 1. Во класата K_2 има 7 множества $S_{\mathbf{c}}$ и во секое има по барем два елемента со тежина 2,

која е минимална. Во K_3 има две множества S_c и сите елементи во нив имаат иста тежина 3.

Од Табела 8.2 се гледа дека со овој алгоритам за декодирање ќе имаме коректно декодирање на пораката, ако при пренос на кодниот збор се појави една грешка, но ќе бидат коригирани и некои двократни и трократни грешки. Значи, за $n = 6$ и $t = 4$ ($s = 2$), коригирани ќе бидат сите еднократни грешки, и уште некои двократни и трократни грешки. \square

Да нагласиме уште еднаш како оди поправањето на грешките настанати со пренос низ канал со шум. Нека на излезот од каналот се добие порака y . Прво, се проверува дали добиената порака y е коден збор. Доколку е коден збор, не се прави корекција. Ако y не е коден збор тогаш во идеалната шема за декодирање се бара дали постои u . Ако постои, тогаш u се декодира со кодниот збор кој е на почетокот на колоната во која се наоѓа u . Лидерот во редицата во која се наоѓа u е всушност, векторот на грешка. Тој покажува на кои позиции настанала грешката при пренос. Ако u не е во табелата, тоа значи дека при пренос се настанати повеќе грешки од предвидените или грешките не се на предвидените позиции, па корекција не може да се направи.

Во претходниот пример, нека на излезот од каналот е добиена порака $y = 011011$. Оваа порака не е коден збор. Во табелата наоѓаме дека се наоѓа во колоната соодветна на кодниот збор 111010, па y се декодира со 111010. Лидер на колоната во која се наоѓа y е 100001, што значи дека при пренос на кодниот збор настанале грешки кај првиот и шестиот бит.

8.6. Решени задачи

Задача 8.6.1. Нека C_1 и C_2 се два линеарни бинарни кодови со следните генераторни матрици:

$$\mathbf{G}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{G}_2 = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

- Да се определат кодните зборови на C_1 и C_2 и да се пресмета минималното растојание на секој код.
- Да се определат матриците на парност (контролните матрици) за секој од кодовите.

Решение:

а) Секој коден збор може да се претстави како линеарна комбинација на редиците на G со коефициенти од $GF(2)$. Множеството кодни зборови за првиот код C_1 е:

$$M_1 = \left\{ \begin{array}{cccc} 00000, & 11110, & 00111, & 11001 \\ x_1 & x_2 & x_3 & x_4 \end{array} \right\}.$$

Растојанијата меѓу кодните зборови се:

$$\begin{aligned} d(x_1, x_2) &= 4, & d(x_1, x_3) &= 3, & d(x_1, x_4) &= 3, \\ d(x_2, x_3) &= 3, & d(x_2, x_4) &= 3, & d(x_3, x_4) &= 4. \end{aligned}$$

Минималното растојание на овој код $d(M_1) = \min d(x_i, x_j) = 3$. Со оглед на тоа што низата од нули (во овој случај низата од 5 нули) е секогаш коден збор, минималното растојание на секој код е еднакво на минималната тежина на ненултите кодните зборови.

Множеството кодни зборови за вториот код C_2 е:

$$M_2 = \{0000000, 1001101, 0101011, 0010111, 1100110, 1011010, 0111100, 1110001\}$$

Минималната тежина на ненултите кодни зборови е 4, оттука минималното растојание на овој код $d(M_2) = 4$.

б) Последните две колони на \mathbf{G}_1 се линеарно независни, па за определување на контролната матрица за овој код првите четири колони се изразуваат како линеарна комбинација од нив:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \lambda_{11} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda_{12} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow \lambda_{11} = 1, \lambda_{12} = 1$$

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \lambda_{21} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda_{22} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow \lambda_{21} = 1, \lambda_{22} = 1$$

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \lambda_{31} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda_{32} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow \lambda_{31} = 1, \lambda_{32} = 0$$

Се добива следната контролна матрица:

$$\mathbf{F} = [\mathbf{I}_3, -\mathbf{A}] = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Генераторната матрица за вториот код е од облик $\mathbf{G}_2 = [\mathbf{I}_3, \mathbf{A}]$, па според претходната дискусија, за контролната матрица на овој код се добива:

$$\mathbf{F} = [-\mathbf{A}^T \mid \mathbf{I}_4] = \left[\begin{array}{ccc|cccc} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{array} \right].$$

□

Задача 8.6.2. Да се формира табела за идеална шема за декодирање за бинарен линеарен код чија контролна матрица е:

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Решение:

Кодните зборови се добиваат како решенија на системот равенки $\mathbf{F} \cdot \mathbf{x} = \mathbf{0}$. Во развиена форма, системот равенки добива облик:

$$\begin{cases} x_1 + x_2 + x_3 & = 0 \\ x_2 + x_4 & = 0 \\ x_1 + x_5 & = 0 \end{cases} \Rightarrow \begin{cases} x_1 = x_5 \\ x_2 = x_4 \\ x_3 = x_1 + x_2 = x_4 + x_5 \end{cases}$$

Множеството на кодни зборови (множеството од бинарните зборови $x_1x_2x_3x_4x_5$ кои ги задоволуваат овие равенства) е:

$$M = \{00000, 10101, 01110, 11011\}.$$

Да воочиме дека кодните зборови соодветствуваат на синдромот $\mathbf{c} = 000$. За останатите синдроми се добиваат следните равенства и бинарни зборови

кои ги задоволуваат (со y_0 е означен лидерот – зборот со најмала тежина):

$$\underline{\mathbf{c} = 100}$$

$$\begin{cases} x_1 + x_2 + x_3 & = 1 \\ x_2 + x_4 & = 0 \\ x_1 + x_5 & = 0 \end{cases} \Rightarrow \begin{cases} x_1 = x_5 \\ x_2 = x_4 \\ x_3 = 1 + x_4 + x_5 \end{cases} \Rightarrow \begin{array}{l} 00100 = y_0 \\ 10001 \\ 01010 \\ 11111 \end{array}$$

$$\underline{\mathbf{c} = 010}$$

$$\begin{cases} x_1 + x_2 + x_3 & = 0 \\ x_2 + x_4 & = 1 \\ x_1 + x_5 & = 0 \end{cases} \Rightarrow \begin{cases} x_1 = x_5 \\ x_2 = 1 + x_4 \\ x_3 = x_5 + x_4 + 1 \end{cases} \Rightarrow \begin{array}{l} 01100 \\ 11001 \\ 00010 = y_0 \\ 10111 \end{array}$$

$$\underline{\mathbf{c} = 001}$$

$$\begin{cases} x_1 + x_2 + x_3 & = 0 \\ x_2 + x_4 & = 0 \\ x_1 + x_5 & = 1 \end{cases} \Rightarrow \begin{cases} x_1 = x_5 + 1 \\ x_2 = x_4 \\ x_3 = x_5 + 1 + x_4 \end{cases} \Rightarrow \begin{array}{l} 10100 \\ 00001 = y_0 \\ 11010 \\ 01111 \end{array}$$

$$\underline{\mathbf{c} = 110}$$

$$\begin{cases} x_1 + x_2 + x_3 & = 1 \\ x_2 + x_4 & = 1 \\ x_1 + x_5 & = 0 \end{cases} \Rightarrow \begin{cases} x_1 = x_5 \\ x_2 = x_4 + 1 \\ x_3 = x_5 + x_4 \end{cases} \Rightarrow \begin{array}{l} 01000 = y_0 \\ 11101 \\ 00110 \\ 10011 \end{array}$$

$$\underline{\mathbf{c} = 101}$$

$$\begin{cases} x_1 + x_2 + x_3 & = 1 \\ x_2 + x_4 & = 0 \\ x_1 + x_5 & = 1 \end{cases} \Rightarrow \begin{cases} x_1 = x_5 + 1 \\ x_2 = x_4 \\ x_3 = x_5 + x_4 \end{cases} \Rightarrow \begin{array}{l} 10000 = y_0 \\ 00101 \\ 11110 \\ 01011 \end{array}$$

$$\underline{\mathbf{c} = 011}$$

$$\begin{cases} x_1 + x_2 + x_3 & = 0 \\ x_2 + x_4 & = 1 \\ x_1 + x_5 & = 1 \end{cases} \Rightarrow \begin{cases} x_1 = x_5 + 1 \\ x_2 = x_4 + 1 \\ x_3 = x_5 + x_4 \end{cases} \Rightarrow \begin{array}{l} 11000 \\ 01101 \\ 10110 \\ 00011 = y_0 \end{array}$$

$$\underline{\mathbf{c} = 111}$$

$$\begin{cases} x_1 + x_2 + x_3 & = 1 \\ x_2 + x_4 & = 1 \\ x_1 + x_5 & = 1 \end{cases} \Rightarrow \begin{cases} x_1 = x_5 + 1 \\ x_2 = x_4 + 1 \\ x_3 = 1 + x_5 + x_4 \end{cases} \Rightarrow \begin{array}{l} 11100 \\ 01001 \\ 10010 = y_0 \\ 00111 \end{array}$$

Идеалната шема за декодирање се формира, така што за секој синдром прво се внесува зборот y_0 со најмала тежина, а потоа останатите зборови за соодветниот синдром. Останатите зборови за даден синдром може да се добијат и како збир на y_0 (зборот со најмала тежина за тој синдром) и ненултните кодни зборови во M . Идеалната шема за декодирање за овој код е:

\mathbf{c}	$S_{\mathbf{c}} = \mathbf{y}_0 + M$			
000	00000	10101	01110	11011
100	00100	10001	01010	11111
010	00010	10111	01100	11001
001	00001	10100	01111	11010
110	01000	11101	00110	10011
101	10000	00101	11110	01011
011	00011	10110	01101	11000
111	10010	00111	11100	01001

Нека символите кои се кодираат со овој код се А, В, С и D и кодот е дефиниран со:

$$A \rightarrow 00000$$

$$B \rightarrow 10101$$

$$C \rightarrow 01110$$

$$D \rightarrow 11011$$

Ако влезната порака е: AABCACD, тогаш оваа порака ќе се кодира со:

00000 00000 10101 01110 00000 01110 11011

Нека по преносот низ канал со пречки се добива:

00001 00000 11101 01111 10010 01100 110007

Корекцијата на грешки се прави блок по блок. Првиот блок 00001 не е коден збор, но тој во табелата во која е претставена шемата за декодирање се наоѓа во колоната соодветна на кодниот збор 00000, па тој блок се декодира со 00000. Вториот блок 00000 е коден збор, па тука не се прави корекција. Третиот блок е 11101 и тој не е коден збор, но во табелата се наоѓа во колоната соодветна на 10101, па го декодираме со 10101. Со продолжување на постапката се добива следната низа на кодни зборови:

00000 00000 10101 01110 00000 01110 11011

Оваа низа зборови се декодира во: AABCACD. Декодираната порака е иста со пратена, што значи дека кодот ги поправил сите грешки настанати при преносот низ канал со пречки. \square

8.7. Задачи

Задача 8.7.1. Нека линеарен бинарен код е зададен со следната генераторна матрица:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

- Да се определи множеството M на кодните зборови за овој код.
- Да се најде контролната матрица за овој код.
- Да се формира табела за идеална шема за декодирање на овој код.

Задача 8.7.2. Нека линеарен бинарен код е зададен со следната генераторна матрица:

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

- а) Да се определи множеството M на кодните зборови за овој код.
- б) Да се најде контролната матрица за овој код.
- в) Да се формира табела за идеална шема за декодирање на овој код.

Задача 8.7.3. Нека линеарен бинарен код е зададен со следната контролна матрица:

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

- а) Да се определи множеството M на кодните зборови за овој код.
- б) Да се најде генераторната матрица за овој код.
- в) Да се формира табела за идеална шема за декодирање на овој код.

Задача 8.7.4. Нека линеарен бинарен код е зададен со следната генераторна матрица:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}.$$

- а) Да се определи множеството на кодните зборови за овој код.
- б) Да се пресмета минималното растојание на кодот.
- в) Да се најде контролната матрица за овој код.
- г) Да се формира табела за идеална шема за декодирање на овој код.

Литература

- [1] T. M. Cover, J. A. Thomas: *Elements of Information Theory*, John Wiley&Sons, Inc. (1991)
- [2] Zh. Paushe: *Uvod u teoriju informacije*, Školska knjiga, Zagreb (1980)
- [3] D. J. C. MacKay: *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press (2003)
- [4] R. Ash: *Information Theory*, Dover Publication, Inc. (1990)
- [5] D. A. Huffman: *A Method for the Construction of Minimum-Redundancy Codes*, Proceedings of the I.R.E., (1952), pp. 1098-1101
- [6] J. C. Bowman: *Coding Theory*, University of Alberta, Edmonton, Canada, 2003.
- [7] H. M. Taylor, S. Karlin: *An Introduction to Stochastic Modeling*, Academic Press (1998)
- [8] A. Papoulis: *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, Inc., New York (1965)
- [9] V. S. Pless and W. C. Huffman (editors): *Handbook of Coding Theory*, Elsevier Science B.V., Amsterdam, The Netherlands (1998)
- [10] M. Purser: *Introduction of Error-Correcting Codes*, Artech House, Boston, 1995.
- [11] В. Бакева: *Веројатност*, УКИМ, Скопје, <http://www.ukim.edu.mk/e-izdanija/FINKI/Verojatnost.pdf>

Ниту еден дел од оваа публикација не смее да биде репродуциран на било кој начин без претходна писмена согласност на авторот

Е-издание:

<https://www.ukim.edu.mk/e-izdanija/FINKI/>

Teorija-na-informacii-so-digitalni-komunikacii.pdf

